INFORMATION TO USERS

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. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.
- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.
- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., 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 as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

Order Number 8713763

Molecular cloning and expression in Escherichia coli of the human α-galactosidase A gene

Hantzopoulos, Petros A., Ph.D.

City University of New York, 1987

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

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 _____

_____

_____

MOLECULAR CLONING AND EXPRESSION IN ESCHERICHIA COLI OF THE HUMAN α-GALACTOSIDASE A GENE

By

PETROS A. HANTZOPOULOS

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

1987

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

25 MAR 87
Date

David Calhoun
David H. Calhoun, Ph.D.
Chairman of Examining Committee

26 March 1987
Date

Terry A. Krulwich, Ph.D.
Executive Officer

Dr. C. Bona, M. D., Ph. D.

Dr. M-T Hsu, Ph. D.

John Karkas
Dr. J. Karkas, Ph. D.

James G. Wetmur
Dr. J. Wetmur, Ph. D.

Supervisory Committee

The City University of New York

Abstract

MOLECULAR CLONING AND
EXPRESSION IN ESCHERICHIA COLI OF
THE HUMAN α-GALACTOSIDASE A GENE

by

Petros A. Hantzopoulos

Adviser: Dr. David H. Calhoun

Fabry disease is an X-linked inborn error of sphingolipid metabolism resulting from the deficient enzymatic activity of the lysosomal hydrolase α-galactosidase A. Previous clinical trials indicated that enzyme replacement therapy may be a feasible approach for the treatment of patients with Fabry disease. We have used recombinant DNA methods to isolate and characterize a human α-galactosidase A cDNA clone from a λgt11 human liver cDNA expression library. For the isolation of this clone, we used monospecific antibodies made against purified α-galactosidase A, as well as oligonucleotide mixtures corresponding to amino terminal and internal amino acid sequences. The authenticity of this clone was determined by nucleotide sequencing of both strands using the enzymatic and chemical methods. The cDNA insert contains a 1234 bp sequence with an open reading frame encoding 398 amino acids of the propeptide, as well as five amino acids of the leader peptide. For the expression of the α-galactosidase A propeptide in Escherichia coli, the 5'-terminus of the cDNA insert was engineered so that an ATG initiator codon precedes the first codon of the mature form of the enzyme. The engineered cDNA insert was cloned into the ClaI site of the prokaryotic expression vector ptrpL1, and expression was monitored in maxicells. Clones that

contained different sequences in the region separating the ribosome binding site and the ATG initiation codon were constructed and tested for efficiency of expression. Clones with α-galactosidase A specific cDNA encoding the proenzyme produce a protein of 45 kilodaltons (kDa), the size expected for the intact proenzyme. The 45 kDa protein is specifically precipitated by antibody to α-galactosidase A, and its expression is repressed by tryptophan and induced by 3-β-indoleacrylic acid as expected for this expression vector. The human enzyme is produced in E. coli in a catalytically active form at levels sufficient to support the growth of cells in minimal media using α-galactosides as sole sources of carbon and energy. In addition, bacterial colonies that produce the human enzyme turn blue in the presence of 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal). The entire α-galactosidase A cDNA, including the 5'-end untranslated sequence, and sequences encoding the complete prepropeptide was constructed when a genomic clone containing this segment was isolated. Primer extension experiments with poly(A)$^+$ RNA identified the probable CAP site for the α-galactosidase A transcript. The clones constructed in this study will be useful to evaluate the production of human recombinant α-galactosidase A for enzyme replacement therapy of Fabry disease using prokaryotic and eukaryotic systems.

ACKNOWLEDGMENTS

I would like to express my deep appreciation and gratitude to Dr. David Calhoun, who gave me the opportunity to learn and grow as a member of the scientific community. His continuous support helped me develop and evaluate my ideas in a fruitful way.

I would also like to acknowledge the help of my friends and colleagues, Merrigene Quinn, Harold Bernstein, David Bishop, Bob Desnick, Enzo Fidanza, without a preferred order, they were all very important to me.

Finally, I thank my mother, because it was her love, care, encouragment and confidence that inspired me to achieve this goal. This thesis is dedicated to my late father, for his continuous moral support throughout my studies, from the primary school to my highest degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

INTRODUCTION

Fabry disease is an inborn error of glycosphingolipid metabolism that results from the defective activity of the lysosomal hydrolase α-galactosidase A (α-D-galactosidase, α-D-galactoside galactohydrolase, EC 3.2.1.22). The disease is transmitted by an X-linked recessive gene. Deficiency of α-galactosidase A results in the progressive accumulation of the major glycosphingolipid substrate, galactosylgalactosylglucosylceramide (globotriaosyceramide), in many tissues of the body. This birefringent lipid is deposited in endothelial, perihelial and smooth-muscle cells of blood vessels in ganglion cells and perineural cells of the autonomous nervous system, in reticuloendothelial, myocardial, and connective tissue cells, and in epithelial cells of the cornea, kidney and other tissues. In hemizygous males progressive substrate deposition in the plasma membrane and vascular endothelium leads to skin lession (Desnick et al., 1983, Johnson et al., 1978). Clinical sequelae include ischemia, pain, and burning sensation in the extremeties, fever, and cardiac and renal failure. Early demise results from developing vascular disease of the heart, kidneys and/or brain (Desnick et al., 1983). Heterozygous females are generally asymptomatic (Desnick et al., 1973). The disease was first described by Fabry in Germany (Fabry, 1898) and Anderson in England (Anderson, 1898). Chemical analyses have confirmed the lipid nature of the storage material. In 1963, Sweeley and Klionsky (Sweeley et al., 1963) isolated and characterized two neutral glycosphingolipids, galactosylgalactosylglucosylceramide and digalactosyl ceramide, from a kidney of a Fabry hemizygote. On the basis of these findings,

they classified Fabry's disease as a sphingolipidosis. Brady and coworkers first demonstrated that the enzymatic defect in this inborn error of glycosphingolipid catabolism was the absence of α-galactosidase A enzymatic activity (Brady et al., 1967b).

The α-galactosidase A gene was originally assigned to the X chromosome by pedigree studies and by linkage to the X chromosome marker X_g^a of the blood group antigens (Opitz et al., 1965). The assignment of α-galactosidase A to the X chromosome was confirmed by somatic cell genetic studies (Grzeschik et al., 1972). Human-hamster hybrids selected in HAT medium (containing hypoxanthine, aminopterin, and thymidine), which requires the presence of the X-linked enzyme, hypoxanthine phosphoribosyltransferase for growth (Ricciuti and Ruddle, 1973), always had human α-galactosidase A activity. When the expression of α-galactosidase A in such hybrid clones was examined by gel electrophoresis, an enzymatic activity which migrated to a position between the human and chinese hamster α-galactosidase A isozymes was observed. The presence of this hybrid protein was consistent with earlier studies of the purified human enzyme (Rebourcet et al., 1974), which indicated that it is a homodimer. α-Galactosidase A was localized to Xq22-->q24 region of X chromosome (Fox et al., 1984) by obtaining different X/autosome translocations (Gerald et al., 1974, Pearson et al., 1975, Shows et al., 1978). These hybrids retained the X chromosome segment when grown in HAT media.

Therapeutic strategies for inherited metabolic diseases have until recently been necessarily directed at the symptom relief level. Our lack of understanding at the molecular level of both the primary biochemical defect and the pathologic processes, which lead to the clinical disease has been a major impediment to therapeutic progress.

Efforts to treat Fabry disease have been designed to deplete circulating substrate in order to prevent vascular deposition. The rationale for enzyme replacement in this lysosomal storage disease was first based on in vitro reversal of accumulation of substances in cultured skin fibroblasts derived from patients with other metabolic storage diseases (Hurler's and Hunter's syndromes). Some of the early observations along this line were the correction of metabolic defects in fibroblasts obtained from patients with mucopolysaccharidoses (Hurler's and Hunter's syndromes) by co-cultivation with normal cells, as well as those from patients with different mucopolysaccharidose storage disorders (Frantatoni et al., 1968). Also, partially purified enzymes were shown to reduce accumulated sphingolipids in cultured cells derived from patients with certain lysosomal storage disorders (Porter et al., 1971, Hickman et al., 1972). The first clinical trial in 1970 consisted of single infusions of fresh normal plasma with α-galactosidase A activity to three Fabry patients (Mapes et al., 1970). A 50% reduction of the circulating substrate was observed over a period of several days. When partially purified α-galactosidase A (Brady et al., 1967a) infused into two males with Fabry disorder a 50% decrease in the quantity of globotriaosylceramide in the circulation within a 45 min period of time was observed (Brady et al., 1973) with a return to the preinfusion levels by 48 hrs. Desnick and coworkers (Desnick et al., 1979, Desnick et al., 1980) in a similar study showed that α-galactosidase A preparation from human spleen and plasma reduced the globotriaosylceramide in the circulation upon multiple intravenous injections into Fabry patients. The half-life of the enzyme derived from plasma in the blood of the recipients was much longer than that of the splenic isozyme preparation and it was more effective than the

spleen enzyme in reducing globotriaosylceramide. In addition, no immune response was detected after successive administrations. After each dose of the splenic isozyme, the concentration of the accumulated circulating substrate, globotriaosylceramide, decreased maximally (about 50% of initial values) in 15 min and returned to preinfusion levels by 2-3 hrs. In marked contrast, injection of the plasma isozyme decreased the circulating substrate levels 50-70% by 2-6 hrs and its concentration gradually returned to preinfusion values by 36-72 hrs. It was realized after these clinical trials that the enzyme replacement therapy approach may be feasible and that a much larger quantity of enzyme would be required in order to obtain a sustained effect with a favorable clinical response. Appropriate intravenous injection of the missing enzyme may be followed by endocytosis and fusion with the lysosomes and eventual substrate hydrolysis.

In order to understand the nature of the nucleic acid abnormalities underlying Fabry disease, the normal gene coding for the α-galactosidase A must be isolated and characterized, and then compared to that of Fabry patients. The application of enzyme replacement therapy seems most likely to be successful with those lysosomal storage diseases that develop relatively slowly and do not seriously affect the central nervous system (e.g. Fabry disease, Gaucher disease). The effectiveness of enzyme replacement treatment remains to be established. Three main problems have been associated with this approach: (a) the inability to deliver the enzyme to the central nervous system due to the blood-brain barrier, (b) rapid enzyme clearance in the plasma and primary hepatic uptake and (c) the lack of sufficient quantities of the human enzyme for targeting to primary sites of pathology.

Like other similar lysosomal enzymes, α-galactosidase A is initially synthesized in the rough endoplasmic reticulum (Walter et al., 1984) as a precursor having higher molecular weight than the mature form found in the lysosomes (LeDonne et al., 1983). Studies of α-galactosidase A biosynthesis in Chang liver cells (LeDonne et al., 1983) and human fibroblasts (Bishop and coworkers, private communication) indicated that a glycosylated propeptide of molecular weight 55-58 kDa is processed to a mature subunit of molecular weight 49-50 kDa. The mature active enzyme purified from human tissues and plasma is a homodimer of molecular weight of approximately 101 kDa that contains one or more asparagine-linked oligosaccharide chains (Dean and Sweeley, 1979, Bishop and Desnick, 1981).

Lysosomal enzymes have been observed to undergo amino- and carboxyl-terminal processing (Erickson et al., 1983). A common processing step implies that the carboxyl-terminal peptides share a common information content. An intriguing possibility, proposed by Blobel (1980) is that the carboxyl-terminal peptide is a 'sorting' sequence characteristic of lysosomal but not secretory proteins. Such a sorting sequence could function in one of two ways. (i) It could be that the carboxyl-terminal peptide contains a sequence recognized by the phosphotransferase that transfers N-acetylglucosamine 1-phosphate to lysosomal enzymes or (ii) the transient carboxyl-terminal sequence might be a sorting sequence recognized by a presently unknown receptor that is distinct from the mannose 6-phosphate receptor (Sly and Fischer, 1982). If lysosomal enzymes react through a carboxyl-terminal peptide with this receptor whose number or turnover is limited, then some enzymes might fail to find a receptor and thus be secreted. This scheme would account for those lysosomal enzymes which are eventually

secreted. Recently it has been proposed (Fitting and Kabat, 1982, Lodish et al., 1983) that some sorting of secretory proteins occurs in the endoplasmic reticulum and that receptor-mediated transport explains the different rates observed for removal of certain secretory proteins from the endoplasmic reticulum. If receptor-mediated transport of secretory proteins initiates in the endoplasmic reticulum, then a similar mechanism might operate for lysosomal proteins.

With the exception of rodent β-glucuronidase (Hieber, 1982, Catterall and Leary, 1983) lysosomal hydrolases are present at relatively invariant levels in all tissues studied so far, 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 enrichment regimes 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, α-galactosidase A constitutes approximately 0.002% of total cellular protein in human lung. Furthermore, the fact that lysosomal hydrolases turn over relatively slowly in the tissues (Porter et al., 1971, Stahl et al., 1980) 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. The cDNA clone corresponding to another lysosomal hydrolase, α-fucosidase, has a mRNA abundance of 0.002% and a protein concentration of 0.01% of liver protein (Grantham et al., 1981).

Fabry disease represents an ideal model for the application of recombinant DNA technology for its treatment through enzyme replacement therapy. α-Galactosidase A is a homodimeric enzyme of molecular weight

49,000 daltons (Dean et al., 1979, Bishop et al., 1981), thus only a single gene is needed for cloning and expression in prokaryotic or eukaryotic systems. In contrast, some other lysosomal hydrolases, such as β-hexosaminidase A, are composed of nonidentical subunits (Mahouran et al., 1982). Now, improved and advanced techniques employed in recombinant DNA technology make the isolation and cloning of rare genes feasible. Current techniques employ the use of synthetic oligonucleotides (Szostak et al., 1979) for screening cDNA and genomic libraries, and specifically priming mRNA species. Alternative techniques include mRNA enrichment by polysome immunoprecipitation with antibodies to specific proteins (Wickens et al., 1978, Shapiro et al., 1981, Korman et al., 1982, Kraus et al., 1982), immunological screening of cDNA expression libraries using microbial systems (Huynh et al., 1984, Young and Davis, 1983). and positive-selection through hybridization assays (Parnes et al., 1981). Also, recent genetic and biochemical modifications may make it possible to produce neoglycoproteins that will enhance their endocytosis by specific target cells.

The molecular cloning of human α-galactosidase A gene would make it possible to obtain genomic clones containing the genetic regulatory sequences of the α-galactosidase A gene, characterize the structural organization of the gene, study the expression of the gene, and to elucidate the molecular defects of patients with Fabry disease. In addition, a cDNA clone containing the complete coding sequence for the processed, enzymatically active enzyme, as well as the signal peptide sequence would be very useful for the production of human α-galactosidase A in prokaryotic and eukaryotic expression systems. In addition, expression of α-galactosidase A in eukaryotic systems would

provide an insight in the nature of lysosomal enzyme biosynthesis as well as the recognition events that are involved in lysosomal targeting and trafficking. Since α-galactosidase A gene undergoes dosage compensation by random X-inactivation (Mohandas et al., 1984), studies with genomic clones may provide valuable information as to how the control of gene expression operates in X-linked housekeeping genes such as α-galactosidase A. Since the approach of enzyme replacement therapy may be feasible with Fabry disease (Desnick et al., 1979), the use of microbial and eukaryotic systems for the synthesis of recombinant human enzyme may overcome a major obstacle to this mode of therapy.

Recent advances in microbial molecular genetics offer the opportunity to isolate and clone specific genes of interest and to express their products in high yields, that may be used therapeutically in clinical situations. Bacteria are less fastidious than vertebrate cells in terms of growth requirements in culture and have shorter generation times. The biosynthetic machinery of prokaryotic cells is better defined than that of the eukaryotic cells. Several factors will influence the expression of human genes in microorganisms. Since the heterologous gene product may be toxic, it is advantageous to place the gene under the control of a regulatable promoter. In this way, the cloning steps can be conducted without high levels of expression, but the appropriate induction will permit synthesis prior to harvesting the cells (Halewell and Emtage, 1980, Jay et al., 1981, Yoakam et al., 1982, Janowitz et al., 1982, Rosenberg et al., 1983, Mott et al., 1985). Numerous vertebrate genes have been succesfully expressed in bacteria (Goeddel et al., 1980, Luck et al., 1986, Mandecki et al., 1985) and some, including human insulin (Goeddel et al., 1979a) and

human growth hormone (Goeddel et al., 1979b, Becker and Hsiung, 1986) have been approved for clinical trials.

Lysosomal enzymes, secreted proteins and integral membrane proteins of vertebrates contain signal peptides that are functionally homologous to prokaryotic signal peptides (Blobel et al., 1979). It has been shown,. for example, that rat preproinsulin synthesized in Escherichia coli is recognized and cleaved by the prokaryotic signal peptidase, which results in transport to the periplasmic space (Talmadge et al., 1980, Talmadge and Gilbert, 1982).

The differential patterns of codon utilization in prokaryotic and eukaryotic systems may influence the level of expression of human genes in bacteria (Gouy and Gautier, 1982, Grantham et al., 1981, Bennetzen et al., 1982). In Escherichia coli tRNA abundance correlates with the observed levels of protein synthesis, but some codons that are rare in Escherichia coli are commonly used in vertebrate genes (Gouy and Gautier, 1982).

Bacteria apparently lack the machinery required for protein glycosylation, so that this is one inherent limitation of prokaryotic systems. In this context, other hosts, including yeast, fungi, baculovirus and vertebrate cell culture, have been considered (Broach et al., 1983, Jabbar et al., 1985, Howley et al., 1983, Smith et al., 1985).

The objective of my thesis was to first participate in a group effort to isolate and characterize the normal gene coding for α-galactosidase A by screening cDNA and genomic libraries using immunological approaches (λgt11) and oligonucleotides derived from partial amino acid sequences obtained from several segments of the

purified human enzyme. Positive clones were analysed using restriction endonucleases, hybridization to additional oligonucleotides and DNA sequence analyses. Once a cDNA fragment of a clone was obtained that contained the coding region for α-galactosidase A, my specific task was to insert this cDNA downstream of a regulatable bacterial promoter. The vector selected contains the trp promoter, so that induction and repression of α-galactosidase A in Escherichia coli was achieved by supplying the inducer, 3-β-indoleacrylic acid or the repressor, tryptophan, respectively. Stability and immune reactivity of α-galactosidase A made in bacteria was determined in maxicells. Enzymatic activity of the α-galactosidase A was detected in vivo in the bacterial clones producing α-galactosidase A. Production of α-galactosidase A synthesized in bacteria will be available for further evaluation of enzyme replacement therapy.

MATERIALS AND METHODS

Bacterial strains, plasmids and bacteriophages

The bacterial strains used (Table 1) are all derivatives of Escherichia coli K12. The M13 phages used were constructed by Messing and coworkers (Messing and Vieira, 1982, Messing, 1983). Plasmids and phages constructed are in Table 1.

Materials

Restriction endonucleases, T4 DNA polymerase, T4 DNA ligase, AluI methylase, the 17-mer M13 sequencing primer (GTAAAACGACGGCCAGT, catalog number 1211) were from New England Biolabs. Phage T4 polynucleotide kinase and Staphylococcus aureus protein A was from Bethesda Research Laboratories. Nitrocellulose filters (type HATF) and Zetabind nylon transfer membranes were from Millipore and AMF Cuno, respectively. [α-^{32}P]dNTPs (3000 Ci/mmol), [^{35}S]dNTPs, [γ-^{32}P]ATP (3000 Ci/mmol), [^{35}S]methionine were from New England Nuclear. The human liver cDNA λgt11 library was provided by Dr. S.L.C. Woo. A human lung cDNA λgt11 library was obtained from Clontech (Palo Alto, CA). Rapid deletion subcloning kit and electrophoresis sequencing apparatus were purchased from International Biotechnologies, Inc. Cyclocerine, methicillin, carbenicillin, 3-β-indoleacrylic acid, and 1-O-methyl-α-D-galactopyranoside were from Sigma. Isopropyl-β-D-thiogalactopyranoside (IPTG),

TABLE 1

BACTERIAL STRAINS, BACTERIOPHAGES AND RECOMBINANT PLASMIDS USED IN THIS STUDY

Strain	Description	Reference
C600	F^-, thi-1, thr-1, leuB6, lacY1, tonA21, supE44, λ^-.	Appleyard, 1954
DH1	F-, recA1, endA1, gyrA96, thi-1, hsdR17, $r^-(K)$, $m^-(K)$, supE44, relA1, λ^-.	Hanahan, 1983
DH5a	F^-, endA1, hsdR17, $r^-(K)$, $m^-(K)$, supE44, thi-1, λ^-, recA1, gyrA96, relA1, Phi80dlacZΔM15.	Hanahan, 1983
JM103	Δ(lac-pro), thi, strA, supE, endA, sbcB, $hsdR^-$, F'traΔ36, proAB, $lacI^q$,lacZΔM15.	Messing et al., 1981
HB101	F^-, hsdS20, $r^-(B)$, $m^-(B)$, recA13, araP-14, proA2, lacY1, galK2, rpsL20 (str^R), xyl-5, mtl-1, supE44, λ^-.	Boyer and Rouland-Dussoix, 1969
CSR603	thr-1, leuB6, proA2, phr-1, recA1, argE3, thi-1, uvrA6, ara-14, lacY1, galK2, xyl-5, mtl-1, gyrA98 (nalA98), rpsL31, tsx-33, λ^-, F^-, supE44.	Sancar et al., 1979
X1776	F^-, tonA53, dapD8, minA1, gluV44, (supE42), Δ(gal-uvrB)40 λ^-, minB2, rfb-2, gyrA25, thyA142, oms-2, metC65 oms-1, (tte-1), Δ(bioH-asd)29, cycB2, hsdR2.	Curtiss et al., 1977
LE392	F^-, hsdR514, $r^-(K)$, $m^-(K)$, supE44, supF58, lacY1, galK2, galT22, metB1, trpR55, λ-	Murray et al., 1977

Plasmids	Description	Reference
pBR322	parental vector	Bolivar et al., 1977
pAG18	pBR322 with α-galactosidase A EcoRI cDNA insert (1.2 kb)	Calhoun et al., 1985
pUC9	pBR322 derivative, cloning vector	Vieira and Messing, 1982
pMS118	pUC9 with EcoRI fragment from pAG18	This work
pcDX	parental vector of the Okayama and Berg library	Okayama and Berg, 1983
ptrpL1	pBR322 derivative with a HindIII to ClaI fragment (about 275 bp) containing the trp promoter and operator of the trpL gene.	Hallewell and Emtage, 1980
pMS122	expression plasmid containing the cDNA for the α-galactosidase A propeptide with an ATG initiation codon 6 bp downstream of the ribosome binding site	This work
pMS132	same as pMS122, but 8 bp between the ribosome binding site and the ATG initiation codon.	This work
pAS1	expression vector containing the λcII gene S/D site, nutL and nutR sites, P_L, of λ phage on a 1850 bp restriction fragment. BamHI site at the initiation codon of the cII gene	Rosenberg et al., 1983.
pOTSNco12	derivative of pAS1. It contains a unique Nco12 restriction site instead of BamHI.	Devare et al., 1984
pMS137	expression plasmid containing a truncated form of the α-galactosidase A cDNA cloned into the NcoI site of pOTSNco12 vector.	This work
pKT218	pBR322 derivative, parental vector of the cDNA library obtain from Orkin. PstI cloning site.	Talmadge et al., 1980a
pPH15	contains 1.3 kb cDNA insert in PstI site of pKT218 that hybridized to oligonucleotide 1C	This work

Bacteriophages	Description	Reference
Charon 30	b1007, KH54, nin5, dup1(sbh2-3)	Rimm et al., 1980
λgt11	lac5, cI857, nin5, S100	Young and Davis, 1983
λAG18	λgt11 with EcoRI cDNA insert of α-galactosidase A.	Calhoun et al., 1985

M13 vectors	Description	Reference
M13mp18	cloning vectors	Messing, 1983
M13mp11	cloning vector	Messing, 1983
mAG25.0	M13mp18 with EcoRI fragment from pAG18 (message strand)	Bishop et al., 1986
mAG27.0	M13mp18 with EcoRI fragment from pAG18 (complementary strand)	Bishop et al., 1986.
mAG25.X	deletion derivatives of the parental vector mAG25.0 (X indicates decimal numbers shown in Fig. 5)	Bishop et al., 1986
mAG27.Y	deletion derivatives of the parental vector mAG27.0 (Y indicates decimal numbers shown in Fig. 5)	Bishop et al., 1986
mMQL	M13mp11 with 1.265 kb TaqI genomic insert of α-galactosidase A (complementary strand).	Quinn et al., 1987
M13.L21	mMQL deletion derivative in M13mp11 with 365 bp fragment (complementary strand) encompassing the first exon, part of the first intron and 5'-untranslated sequences of the TaqI genomic insert.	Quinn et al., 1987
M13.L21.1	M13mp11 containing the leader and propeptide coding sequences of α-galactosidase A.	This work

Klenow fragment of E. coli DNA polymerase I, 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal), 5-bromo-4-chloro-indolyl-β-D-galacto-pyranoside (X-Gal), and ClaI linkers were from Boehringer Mannheim. Deoxynucleotide and dideoxynucleotide triphosphates used for sequencing by the enzymatic method (Sanger et al., 1980) were from P-L Biochemicals. AMV (avian myoblastosis virus) reverse transcriptase was from Life Sciences, Inc. Plastic disposable columns (10 cm X 1 cm) were from Kontes, N. J.

Synthesis and purification of oligonucleotides

In the first 37 amino acids of the α-galactosidase A sequence, regions of low codon redundancy were identified. Mixed oligodeoxy-nucleotides corresponding to these regions were purchased from commercial sources (oligonucleotide 2A, 2B, Table 2, and Fig. 1), were synthesized by Dr. Peter Model, Rockefeller University (oligonucleotide 1C), or were synthesized in the Microbiology Department at Mount Sinai on a SAM-ONE synthesizer using phosphotriester or phosphoramidite chemistry as described by the manufacturer (oligonucleotide 1A, 1B, 3, 4, and 5, Table 2, and Fig. 1). When synthesis was completed the oligodeoxynucleotides were separated from the protecting groups and purified by gel electrophoresis on a 20% polyacrylamide gels containing 8 M urea. The oligonucleotides were visualised by UV shadowing (Maxam and Gilbert, 1980) and eluted from the gel using the crush and soak procedure of Maxam and Gilbert (1980). Following concentration of the oligonucleotides they were 5'-labelled with T4 polynucleotide kinase and used as hybridization probes.

TABLE 2

OLIGONUCLEOTIDES USED IN THIS STUDY

Oligonucleotide mixture	Number of species	Length	Source**	T_m(°C) Range	Sense*
1A	64	23	AT	61-69	+
1B	128	23	AT	60-68	+
1C	32	20	AT	57-62	+
2A	4	14	AT	34-48	+
2B	4	14	AT	36-40	+
3	96	17	CB	46-55	+
4	128	14	AT	38-48	-
5	1	47	TP	34-84	+
H	1	23	AT	65	-
J3	1	18	AT	54	+

(*)sense: (+):complementary to the mRNA sequence. (-):same polarity as mRNA. (**)Source of amino acid sequence is amino terminal (AT), Cyanogen bromide (CB), or tryptic peptides (TP).

Amino Acid Sequences:

Amino terminal:

```
    1       5        10        15        20        25        30        35
NH2-L-D-N-G-L-A-R-T-P-T-M-G-W-L-H-W-E-R-F-M-C-N-L-D-S-Q-E-R-P-D-S-X-I-X-S-
      Oligo 4          Oligos 1A/1B    Oligos 2A
                                1C          2B
                  <------Oligonucleotide 5------>

          1       5        10        15
TPEP/53B: -A-L-L-Q-D-K-D-V-I-A-I-N-Q-D-P-L-G-K-
                             Oligo 3
```

cDNA insert:

```
                      1                   5                  10                  14
                      L   D   N   G   L   A   R   T   P   T   M   G   W   L
CCT GGG GCT AGA GCA CTG GAC AAT GGA TTG GCA AGG ACG CCT ACC ATG GGC TGG CTG
              oligonucleotide H                 oligonucleotide J3
<-----leader------| |-----------α-galactosidase A propeptide----------->
```

Oligonucleotide Mixtures:

```
    11  12  13  14  15  16  17  18      19  20  21  22  23
1A: TAC CCN ACC AAC GTA ACC CTC TC      AAA TAC ACA TTA GA     :2A
                T   G       T G                 G   G

1B: TAC CCN ACC GAN GTA ACC CTC TC      AAG TAC ACA TTA GA     :2B
                    G       T G                 G   G

1C: TAC CCN ACC GAN GTG ACC CT
                    A

    10  11  12  13  14  15
                A
3:  CGN TAG TTA GTT CTA GG
        T   G   C   G

    2   3   4   5   6
4:  gat aat ggn ttn gc
      c   c     c

    8   9   10  11  12  13  14  15  16  17  18  19  20  21  22  23
5:  TGG GGG TGG TAC CCG ACC GAG GTG ACC CTT TCC AAG TAG ACG TTG GA

    -2  -1  1   2   3   4   5   6
H:  aga gca ctg gac aat gga ttc gc

     8   9   10  11  12  13
J3:  TGC GGA TGG TAC CCG ACC
```

Figure 1

Oligonucleotide mixtures derived from partial amino acid sequences of α-galactosidase A. Amino terminal, tryptic peptide and CNBr sequences of human lung α-galactosidase A were used to construct synthetic oligonucleotide mixtures 1A, 1B, 1C, 2A, 2B, 3, 4, 6, 7 and the unique oligonucleotides 5, H and J3. All oligonucleotides are written in the 3′ to 5′ direction except oligonucleotides 4 and H (lower case letters), which are written in the 5′ to 3′ direction. Abbreviations for amino acids and nucleotides are: A=Ala, C=Cys, D=Asp, E=Glu, F=Phe, G=Gly, H=His, I=Ile, K=Lys, L=Leu, M=Met, N=Asn, P=Pro, Q=Gln, R=Arg, S=Ser, T=Ser, T=Thr, V=Val, W=Trp, Y=Tyr, N=A, G, C, T, Oligo(s)= oligonucleotide(s).

Partial amino acid sequence from tryptic peptides and CNBr digests (Fig. 1) were also available and regions with low codon redundancy suitable for oligonucleotide synthesis were identified. All oligonucleotides except oligonucleotide mixture 4 and oligonucleotide H were designed to hybridize to mRNA (Table 2).

Oligonucleotide mixtures 1A and 1B are 23-mers (mixtures of 128 and 64 respectively) and correspond to amino acid residues 11 through 18 (Fig. 1). Oligonucleotides 1A and 1B were designed to differ in that they were specific for the leu codons UUA/G and CUN respectively.

Oligonucleotide mixture 1C is a shorter version of oligonucleotide mixture 1A and 1B (a mixture of 32 different 20-mers) and corresponds to amino acids 11 through 17 (Fig. 1). A guanine residue was selected to pair with the third base in the histidine codon (amino acid 15) in order to reduce the complexity of the mixture. Guanosine was selected rather than adenosine due to codon frequency (12% uracil vs 28% cytosine) and the G/T wobble (Beusage and Caruthers, 1981, Matteucci and Caruthers, 1981).

Oligonucleotide mixtures 2A and 2B consist of two sets of oligodeoxynucleotides, each composed of 4 different 14-mers corresponding to the amino acids 19 through 23 of the amino-terminal of the α-galactosidase A (Fig. 1). The complexity of these mixtures was reduced by selecting G for the first nucleotide of leucine based on its frequency (94%) of the human codon usage (Beusage and Caruthers, 1981).

Oligonucleotide mixture 3 is composed of 96 different 17-mers corresponding to an internal tryptic peptide (TPEP/53B), (Fig. 1).

Oligonucleotide mixture 4 is derived from amino acids 2 through 6 and is a mixture of 128 (all possible) different 14-mers (Fig. 1). This oligonucleotide was designed to hybridize to the anti-sense DNA

sequence of α-galactosidase A gene (Table 2).

Oligonucleotide 5 is a unique 47-mer derived from amino acids 8 through 23 and overlaps with oligonucleotides 1 and 2 (Fig. 1). For the design of this oligonucleotide, 14 nucleotides were selected at ambiguous positions based on the codon frequency and G/T wobble.

Oligonucleotide H is a unique 23-mer derived from amino acids -2 through 6 of the cDNA insert (Fig. 1). It was made to screen α-galactosidase A constructs containing partial leader peptide sequences and to hybridize to the antisense DNA sequence of α-galactosidase A (Table 2).

Oligonucleotide J3 is a unique 18-mer derived from amino acids 8 through 13 (Fig. 1). It was made to screen various expression constructs of α-galactosidase A containing plasmids and to hybridize to the antisense DNA sequence of α-galactosidase A (Table 2).

Labelling of oligonucleotides

An oligonucleotide mixture (50 pmoles) was added to 18 µl of distilled water, 5 µl of 10X kination buffer [0.5 M Tris-HCl (pH 7.6), 100 mM $MgCl_2$, 50 mM dithiothreitol (DTT), 1 mM EDTA, 1 mM spermidine], 25 µl of [γ-^{32}P]ATP (specific activity >5000 Ci/mmol) and 2 µl of T4 polynucleotide kinase (10 units/µl). Incubation was at 37°C for 30 minutes. The reaction was stopped with the addition of 50 µl of STE [0.1 M NaCl, 10 mM Tris-HCl (pH 7.6), 1 mM EDTA]. The labelled oligonucleotides was separated from unreacted ATP by spun-column chromatography using Sephadex G-50 (superfine) as described by Maniatis et al. (1982).

Amino acid sequence analysis of α-galactosidase A

The amino acid sequences of the amino terminal, tryptic and cyanogen bromide peptides were determined by automated gas-phase chromatography microsequencing and HPLC identification (Browne et al., 1982) of phenylthiohydantoin derivatives of the amino acids in the laboratory of Dr. Leroy Hood, Caltech, CA. The method used was as described by Hunkapiller and Hood, (1983).

Screening of cDNA and genomic libraries

Table 3 lists the libraries (cDNA, genomic, expression) available for screening with the synthetic oligonucleotides and antibody probes. The cDNA libraries of Orkin, Okayama and Berg, Woo (λgt11) and the genomic library of Wood were screened by me in an effort to identify an α-galactosidase A containing clone.

Screening of the cDNA libraries of Orkin and of Okayama and Berg: The colony hybridization method of Hanahan and Meselson (1983), was used to screen the human fetal and adult liver cDNA libraries of Orkin (Michelson et al., 1983, Proehownik et al., 1983) as well as the human fibroblast cDNA library of Okayama and Berg (Okayama and Berg, 1983) for clones that may contain sequences complementary to synthetic oligonucleotide probes. Bacteria (in a 200-400 µl volume) were spread on a nitrocellulose filter (137 mm, HATF, Millipore Co.) on a YT agar plate containing 20 µg/ml tetracycline (for Orkin libraries) or 30

TABLE 3

SUMMARY OF cDNA AND GENOMIC LIBRARIES AVAILABLE FOR SCREENING

Library*	Source	Type	Vector	Insert size (kb)	Number of recombinants
1.Orkin	Fetal liver	cDNA	pKT218	1.0**	1.4×10^5
2.Orkin	Adult liver	cDNA	pKT218	1.0**	2.3×10^5
3.Okayama/ Berg	Fibroblast	cDNA	pCD	0.3-7.0	1.0×10^6
4.Woo	Human liver	cDNA/ expression	λgt11	1.5	1.4×10^7
5.Wood	Lymphoblast	4X genomic	Charon30	12-17	1.7×10^6
6.Davies	Flow-sorted X	genomic	λgtWes.2B	4-14	5.0×10^4
7.Cooke	Flow-sorted X	genomic	λgtWes.2B	5-15	7.0×10^4
8.Siniscalco	Human X-mouse hybrid	genomic	λgt7	7-14	8.0×10^4

* These libraries were obtained from the individuals named above and are described in Michelson et al., (1983) (1), Proehownik et al., (1983) (2), Okayama and Berg, (1983) (3), personal communication (4), Wood et al. (1984) (5), Leder et al. (1977) (6), personal communication (7), and Siniscalco et al. (1982), (8).

** These numbers reflect the average size of the cDNA inserts in these libraries.

μg/ml ampicillin (for Okayama and Berg library) to yield approximately 50,000 colonies per plate. The plates were then incubated until colonies 0.1 mm in diameter appeared. Incubation at 30°-32°C facilitated control of colony size on the master (template) filters. The master filter was peeled off its agar plate and laid, colonies up, on a bed of sterile Whatman 3MM circles of 137 mm diameter. A wetted sterile nitrocellulose filter held between two flat bladed forceps was placed on the master filter. The sandwich was pressed firmly together with a velvet covered replica tool. The filters were keyed to each other by making an asymmetric set of holes in the sandwich with a large needle. The replica was peeled off the template and placed on a fresh agar plate containing the appropriate antibiotic. The filters were returned to plates and incubated until colonies were 0.5 mm in diameter. At this point, the master plates were stored at 4°C and served as colony source. The replica filters were grown until colony size reached 0.5 mm and then transfered to agar plates containing chloramphenicol (10 μg/ml) for an additional 12 hr period in order to amplify the plasmid copy number.

Replica filters to be probed were lysed in situ (colonies up) for 10 minutes on a sheet of Whatman 3MM paper saturated with denaturation buffer [1.5 M NaCl, 0.5 M NaOH] followed by neutralization with neutralization buffer [1.5 M NaCl, 0.5 M Tris-HCl (pH 8.0)]. After brief air drying, the filters were baked under vacuum at 80°C for 2 hrs.

The filters were placed in heat sealable freezer bags (Kapak pouches) with a maximum of 5 filters per bag. The hybridization solution containing 6X SSPE [0.9 M NaCl, 0.06 M sodium phosphate, (pH 7.4), 6 mM EDTA] 5x Denhardt's solution [0.1% Ficoll, (mol. weight =

400,000), 0.1% polyvinylpyrrolidone, (mol. weight = 360,000), 0.1% bovine serum albumin, 0.45 M NaCl, 0.045 M sodium citrate (pH 7.0), and 0.5% Sodium dodecyl sulfate, (SDS)] was added and the bag was sealed and then incubated for 2 hrs in a 68°C water bath followed by 2 hr incubation at the appropriate hybridization temperature (usually 10°C below the melting temperature of the probe). The hybridization buffer was emptied and 8-10 mls of fresh buffer containing [^{32}P]-labelled probe (1 x 10^6 cpm/ml) was added and bag was resealed. The bags were returned to the shaking bath at the appropriate hybridization temperature and incubated overnight.

Oligonucleotide probes were hybridized at temperatures that depend on the length and the nucleotide composition. The hybridization temperature selected for 14-mers to 17-mers was usually 10°C below the melting temperature (T_m) calculated by the following formula (Hanahan and Meselson, 1983):

$$T_m = \{2^\circ \text{ X [AT base pairs]} + 4^\circ \text{ X [GC base pairs]}\}$$

The formula shown in Figure 2 was used for longer oligonucleotides. The oligonucleotides were used at a final concentration of 1 nM.

The filters were then washed in 6X SSC [0.9 M NaCl, 0.09M sodium citrate] at room temperature for 15 min and then for an additional 20 min at the hybridization temperature. After air drying the filters were exposed to X-ray film. Keying back from an autoradiogram to the master plate localized the positive signals. A few colonies were removed from the region(s) of hybridization and dispersed in medium, and an appropriate dilution was spread on a fresh nitrocellulose filter to give 100-200 colonies per plate. This enriched population was

$$T_m = \frac{372}{\{1 - (372/4000N_{tot})\}\ \{\ln[(1.5 \times 10^{-3})\ (C_o)\ (1.81^{GC-AT})]\}}$$

where: $372 = 273^{\circ}K + 99^{\circ}$,

($99^{\circ}C = T_m$ for E. coli DNA assuming 50% G+C content).

4000, 1.5×10^{-3}, 1.81 are constants.

N_{tot} : length of oligonucleotide.

C_o : Molar concentration of oligonucleotide.

GC-AT: number of GC less number AT pairs.

T_m is given in $^{\circ}C$.

Figure 2

Hybridization formula used to determine the precise melting temperatures (Tm) of oligonucleotides in hybridization experiments (J. Wetmur, personal communication).

replicated and probed, allowing isolation of pure colonies of positive clones.

Screeening of the genomic library of Wood et al. (1984): The 49, XXXXY human lymphoblast genomic library of Wood et al. (1984) was screened twice. This library contains 1.7 x 10^6 independent recombinants with inserts 12-17 kb long which have been cloned into Charon 30 vector. Each screening was performed on 10 YT plates (150 mm diameter) at a density of 20,000 plaque forming units per plate. Bottom YT plates were prepared and overlaid with 8 mls of top agar containing 0.4 mls of an overnight culture of E. coli strain LE392 (Table 1) previously adsorbed with phage at a multiplicity of infection of 100. Plaques were transferred to nitrocellulose filters, denatured in NaOH, neutralized, dried and hybridized as described for colony hybridization above. Positive plaques were picked and rescreened as described above. This purification procedure was repeated until well isolated, hybridization positive plaques were obtained.

Screening of the λgt11 library: Antibody screening of a human liver λgt11 library (provided by Dr. S. L. C. Woo) was initiated in D. Bishop's laboratory (Department of Human Genetics, Mount Sinai) and the subcloning and initial nucleotide sequence analysis was concluded in Dr. Calhoun's laboratory. Oligonucleotide screening of the λgt11 cDNA library was performed twice at a 1 x 10^6 recombinants each time using the method of Benton and Davis (1977). The human liver cDNA λgt11 library, which contains approximately 1.4 x 10^7 independent clones (Table 3), was plated at a density of 1 x 10^5 phage per 150 mm petri dish and screened as described above. After 4 hr growth at 42°C, the

plaques were overlaid for 2 hrs at 37°C with dry 137 mm nitrocellulose filters which had been soaked previously in 10 mM IPTG. Plaques containing α-galactosidase A determinants were detected following overnight incubation of each filter with 10 ml of a 1:500 dilution of pre-absorbed anti-α-galactosidase A antibodies, followed by a 2 hr incubation with 10 ml of a 1:1000 dilution of peroxidase-conjugated goat anti-rabbit IgG. Antibody positive clones were subjected to competition studies with α-galactosidase A absorbed antiserum to demonstrate binding specificity (Calhoun et al., 1985a).

Analysis of cDNA and genomic clones

Restriction endonuclease analysis of positive clones: To isolate plasmids present in positive cDNA clones, the alkaline rapid extraction procedure of Birnboim and Doly (1979) was used. For the Orkin library cDNA clones, PstI restriction endonuclease is the enzyme of choice, since it liberates the cDNA inserts cloned at the PstI site of vector pKT218. Use of other restriction enzymes revealed more information about the cDNA inserts and a minimun restriction map of the positive clones was constructed. Horizontal agarose gels (0.7% to 1.7%) and vertical polyacrylamide gels (5% or 8%) were used for the analysis of restriction enzyme fragments of various sizes (Maniatis et al., 1982).

For the Okayama and Berg library cDNA clones, BamHI treatment of the hybrid plasmids made it possible to visualize the vector and insert since this enzyme cleaves just outside of the unique cloning site.

DNA from clones isolated from the λgt11 library of Woo (personal

communication) that were antibody-positive clones was isolated and hybridized with the various oligonucleotides probes in order to test for binding of the oligonucleotides to denatured DNA. In addition, DNA from antibody positive clones was digested with EcoRI in combination with HaeIII, HinfI, MspI, AluI, Sau3AI and electrophoresed in agarose gels.

Southern blot analysis of positive clones: Positive cDNA and genomic clones were digested with several restriction endonucleases to generate insert fragments of varying length. Electrophoresis, blotting to nitrocellulose paper as described by Southern (Southern, 1975), and probing with [^{32}P]-labelled synthetic oligonucleotides identified restriction fragments positive to one or more oligonucleotides.

DNA sequence analysis of probe positive clones: Definitive identification of both cDNA and genomic clones relied upon direct DNA sequence analysis and comparison to the regions of known amino acid sequence in α-galactosidase A. The chemical method of Maxam and Gilbert (Maxam and Gilbert, 1980) as well as the enzymatic method of Sanger et al., (1981) were used (see below).

Subcloning of α-galactosidase A sequences into pBR322, pUC9 and M13 vectors

DNA (0.1 μg) from clone λAG18 digested with EcoRI was ligated with 0.1 μg of EcoRI digested pBR322 DNA, at 4°C for 16 hrs. The ligation mixture was used to transform C600 competent cells. Tetracycline

resistant colonies were selected on YT plates supplemented with tetracycline. A set of resistant colonies was subjected to alkaline rapid screen (Birnboim and Doly, 1979) and the DNA was digested with EcoRI and electrophoresed on a 1% agarose gel in order to detect the presence of a 1.2 kb insert. A recombinant clone containing the α-galactosidase A cDNA insert was chosen and the plasmid was designated pAG18.

The α-galactosidase A EcoRI insert was also subcloned into the EcoRI site of pUC9 (by Quinn and Hantzopoulos) and the plasmid was designated pMS118. The same α-galactosidase A cDNA insert was subcloned into the EcoRI site of M13mp18 (D. Bishop's laboratory) in both orientations.

DNA manipulations and cloning procedures

T4 DNA polymerase reactions. (i) Exonuclease step: At a ratio of 1.25 units of T4 DNA polymerase/μg of DNA, 20 nucleotides/min are excised from each 3'-end (Maniatis et al., 1982). Incubation varied from 3 to 15 min depending on the number of nucleotides desired to be excised. The reaction was carried out in a 20 μl volume containing 1 to 5 μg of DNA, TA buffer [33 mM Tris-Acetate (pH 7.9), 66 mM potassium-acetate, 10 mM magnesium-acetate, and 0.5 mM DTT] and 1.25 units of T4 DNA polymerase/μg.

(ii) Repair reaction: To the above reaction 1 μl containing the four deoxynucleotides at a concentration of 2 mM each was added and the reaction was incubated for an additional 15-30 min at 37°C. The reaction was stopped by the addition of 2 μl of 0.5 M EDTA and heating

to 68°C for 5 min.

In vitro methylation of DNA: AluI methylase was incubated with 10 µg of plasmid pMS118 DNA in 50 µl of buffer [50mM Tris-HCl, (pH 7.5), 10mM EDTA, 5 mM β-mercaptoethanol and 80 µM S-adenosylmethionine] at 37°C for 16 hrs. The reaction was inactivated at 65°C for 20 min. The extent of methylation was determined by addition of buffer [10mM Tris-HCl, (pH 7.4), 20mM $MgCl_2$, 50mM NaCl, 6mM β-mercaptoethanol] and 50 units of AluI restriction endonuclease. Incubation at 37°C for 3 hrs was followed by analysis on an agarose gel (Maniatis et al., 1982).

Dephosphorylation of DNA: Removal of the 5'-, or 3'- phosphate groups from DNA vectors or restriction fragments was done as described by Maniatis et al., (Maniatis et al., 1982).

T4 DNA ligation reaction: For sticky-end ligations, reactions were performed at a DNA concentration (vector plus insert) of 5 µg/ml. A molar ratio of vector-to-insert of 2:1 was used when vector was dephosphorylated, and a ratio of 1:5 was used when the vector was intact (phosphorylated). T4 DNA ligase was added (1 unit/ml, final concentration) and reaction was incubated at 16°C for 16 hrs.

For blunt end ligations, vector DNA was ligated to insert DNA at a molar ratio of 1:5. DNA concentration was 50 µg/ml and 100 units of T4 DNA ligase/ml was added. The reaction was incubated at room temperature for 2-16 hrs.

Bacterial transformations: Bacterial transformations were performed using the RbCl method as described by Hanahan (Hanahan, 1983)

or the $CaCl_2$ method as described by Maniatis et al. (Maniatis et al., 1982). Transformation was carried out by adding the DNA solution (ligation mixture or intact plasmid DNA) in a volume of <20 µl into a 200 µl of competent cell suspension, and swirling to mix, followed by incubation on ice for 30 min. Then the cells were heat shocked by placing them in a 42°C water bath for 90 sec, then placed on ice for 2 min to cool. An aliquot of 800 µl of growth medium was added followed by incubation at 37°C with agitation (100 rpm) for 1 hr (2 hrs for CSR603 competent cells). An aliquot of transformed cells was spread on YT agar plates containing appropriate antibiotics (ampicillin at 50 µg/ml, or tetracycline at 15 µg/ml) to select transformants.

DNA sequencing determination of α-galactosidase A

Definitive identification of α-galactosidase A cDNA clones relied upon direct DNA sequence analysis and comparison to the region of known amino acid sequence in the peptide. The chemical method of Maxam and Gilbert (Maxam and Gilbert, 1980) was used to sequence the 5'-end and an internal EcoRI-to-MspI fragment of pAG18 cDNA clone.

M13mp18 deletion subclones were sequenced by the Sanger method (Sanger et al., 1980) using the 17-mer M13 sequencing primer (5'GTAAAACGACGGCCAGT 3') that initiates 4 nucleotides from the polylinker. Overlapping nucleotide sequences were aligned in a microcomputer using the MicroGenie program (Beckman). Protein structural analysis and DNA and amino acid homology searches of the National Institutes of Health GenBank and the National Biomedical Research Foundation protein data base were performed in February 1986.

The method of supercoil sequencing of Chen and Seeburg (1985) was also used in some experiments. The 5' sequences of ptrpL1 recombinant expression constructs were confirmed by this method: Double-stranded DNA (2 µg, obtained by the alkaline rapid screen method) containing α-galactosidase A inserts was mixed with 7.5 pmoles of 5'-end labelled synthetic oligonucleotide J3 (Table 2), in annealing buffer [10 mM Tris-HCl, (pH 7.5), 0.1 mM EDTA, 20 mM NaCl]. This mixture was boiled for 3 min and quick cooled in ice-water bath for at least 10 min. Following primer annealing, 1.5 µl of Klenow buffer [10 mM Tris-HCl, (pH 7.5), 5 mM $MgCl_2$, 0.1 mM EDTA], 1.5 µl of 1 mM DTT, and 1 µl of E. coli DNA polymerase I, large fragment (1 unit/µl) were added and 3 µl samples were dispensed to each of the four reaction tubes containing 2 µl of the appropriate dideoxy- and deoxynucleotide triphosphates (Sanger et al., 1977) (ratio of dideoxy- to deoxynucleotide 1:2). Following an incubation of 15 min at 30°C, 2 µl of chase mixture (Sanger et al., 1977) was added and reaction was further incubated for 15 min. Reactions were stopped by the addition of 10 µl 98% formamide, 0.2% xylene cyanol, and 3 µl was applied to a 5%, 8%, or 20% polyacrylamide sequencing gel (Maxam and Gilbert, 1980).

Construction of ptrpL1 expression plasmids encoding the mature α-galactosidase A sequences

Plasmid pMS118 contains the EcoRI cDNA insert for α-galactosidase A cloned into pUC9 vector (Messing and Vieira, 1982). The HgiAI site at base pair 150 was methylated in vitro using AluI methylase as described above. Digestion with EcoRI and HgiAI released a 1214 bp

fragment from the vector which was isolated by low-melting agarose electrophoresis and elutip-d (Schleicher and Schuell) affinity chromatography (Schmitt et al., 1983). The 1214 bp fragment was made blunt-ended by T4 DNA polymerase in the presence of all four deoxynucleotides, as described above. The blunt-ended fragment was ligated to ClaI linkers, CATCGATG followed by ClaI digestion to cleave the excess linkers off the DNA fragment. To remove the excess linkers, the DNA was purified by gel chromatography through a Sephadex G-100 or a Sepharose CL4 B made in a disposable plastic column of 10 cm x 1 cm using 2 mM Tris-HCl, pH 7.5, 0.01 mM EDTA as a buffer. The excluded DNA was pooled (usually 2-3 fractions of 4 drops each), lyophilized to reduce volume and then cloned into the unique ClaI site of dephosphorylated DNA of the ptrpL1 expression vector. The ligation products were used to transform strain HB101 and positive clones were identified by filter hybridization (Hanahan and Meselson, 1983) using oligonucleotide probe J3. A set of positive clones was examined using alkaline rapid screen DNA (Birnboim and Doly, 1979) and digestion with HinfI and ClaI restriction endonuclease for orientation, and insert analysis, respectively. Digestion with EcoRI was used to confirm the end-filling of the 5'-protruding end and the subsequent ligation of the ClaI linker at the 3'-end of the α-galactosidase A insert. Several positive clones were obtained, including plasmid pMS122 that lacks the ClaI site at the 3'-end. In order to optimize the distance between the Shine-Dalgarno (Shine and Dalgarno, 1975) and ATG added to α-galactosidase A, plasmid pMS122 DNA was cut with ClaI. The ClaI staggered ends at the 5'-end of α-galactosidase A insert were converted to blunt-ends using the Klenow fragment of DNA polymerase to end-fill the two protruding bases (Maniatis et al., 1982). Blunt end self

ligation of this DNA generates an NruI restriction site. The nucleotide sequence of the junction region around the NruI site of plasmid pMS132 was confirmed by the double stranded nucleotide sequencing protocol of Chen and Seeburg (1985) using oligonucleotide J3 as primer. The resultant plasmid pMS132 was used for expression of α-galactosidase A in maxicells.

Maxicell experiments for detection of α-galactosidase A

After transformation of the E. coli strain CSR603 with the appropriate recombinant plasmid, a single ampicillin resistant transformant was isolated and purified. The protocol was as previously described (Sancar et al., 1979, Calhoun and Gray, 1981, Gray et al., 1982), except that after irradiation, methicillin (10 μg/ml) and carbenicillin (10 μg/ml) were included along with cycloserine (20 μg/ml) during overnight incubation in minimum salt media (Miller, 1972). In order to test in maxicells for control of the expression of the human α-galactosidase A, tryptophan (200 μg/ml) or 3-β-indoleacrylic acid (20 μg/ml) were present during the starvation period (1 hour) and labelling period (1 hour). In order to determine the stability of α-galactosidase A in maxicells, [^{35}S]methionine was added, and after a 1 minute incorporation period, unlabelled methionine (final concentration of 100 μg/ml) was added and samples were removed at intervals of 0, 10, 20, 30 and 60 min.

Immunoprecipitation of α-galactosidase A made in maxicells

Irradiated maxicells (10 ml) containing no plasmid, plasmid ptrpL1 without an insert, or plasmid ptrpL1 with the α-galactosidase A cDNA, were centrifuged and resuspended in 0.75 ml of 0.01 M sodium phosphate buffer, (pH 6.5), and sonicated on ice 3 times in pulses of 30 seconds. Cell debris and cell membranes were removed by sequential centrifugations at 5°C for 5 min at 14,000 x g (microfuge), followed by a 20 min centrifugation at 436,000 x g (Beckman TL100). Pre-immune rabbit serum (1 µl) was added to 400 µl of the supernatant and incubated on ice for 3 hrs. Then a 100 µl aliquot of Staphylococcus aureus protein A was added, incubated on ice for 1 hr and centrifuged in the cold 10 min at 14,000 x g. The pH of the supernatant was adjusted to 7.5 using 2 µl aliquots of NaOH (0.1 M). Immune or pre-immune serum (1 µl) was added and incubated 18 hr at 4°C. Sheep anti-rabbit immunoglobulin (50 µl) was added, incubated on ice for 1 hr, and centrifuged at 5°C for 5 min at 14,000 x g. The pellet was washed first with 500 µl of 50 mM Tris-HCl, (pH 7.5), 1.2 M KCl, 1.2% Triton X-100, and then with 50 mM Tris-HCl, (pH 7.5), 0.1 M NaCl. The pellet was resuspended in 30 µl of cracking buffer [50mM Tris-HCl, (pH 6.5), 1% SDS, 2mM EDTA, 1% β-mercaptoethanol, and 10% (v/v) glycerol] and 7.5 µl aliquots were subjected to electrophoresis and autoradiography as previously described (Calhoun and Gray, 1981).

Primer extension of total mRNA preparation

Hybridization primer (0.5 pmoles of the HgiAI 131 bp fragment purified by polyacrylamide gel electrophoresis) and 2 µg of mRNA sample (from D. Bishop) were resuspended in 30 µl of hybridization mixture [80% formamide, 400 mM NaCl, 40 mM Pipes (pH 6.8), 0.5 mM EDTA] (Broome and Gilbert, 1985, Benoist and Chambon, 1981) denatured at 70°C for 10 min and incubated at 52°C for 16 hrs. The sample was then ethanol precipitated twice, dried and resuspended in 15 µl of 10 mM Tris-HCl (pH 8.3), 0.1 mM EDTA. An equal volume of reaction mixture [50 mM Tris-HCl (pH 8.3), 20 mM $MgCl_2$, 60 mM β-mercaptoethanol, 2 mM of each dGTP, dCTP, dTTP and [α-^{32}P]dATP) and RNasin at 1 unit/µl were added. Elongation was carried out for 60 min at 41°C using 1000 units/ml of avian myoblastosis virus reverse transcriptase. This reaction mixture was ethanol precipitated in the presence of 2M ammonium acetate and then extracted twice with phenol and chloroform. The mRNA was hydrolysed with 0.6 N NaOH for 1 hr at 45°C. After neutralization with 1 N HCl and ethanol precipitation, a small aliquot was tested on a 5% polyacrylamide gel containing 7M urea to visualize the primer extension products.

Cloning of leader sequences of α-galactosidase A

Plasmid pMS137 contains a NcoI-to-BamHI restriction fragment of α-galactosidase A, whose 29 bp (10 amino acids) from the 5'-end are deleted. Plasmid pMS137 was double digested with NcoI and BamHI to release the truncated α-galactosidase A fragment (1191 bp).

Clone M13.L21 (obtained from M. Quinn) contains 5'- untranslated sequences, the first exon, and part of the first intron cloned as a 365 nucleotides insert in the EcoRI and SacI sites of M13mp11. Plasmid M13.L21 was digested with NcoI and BamHI to release a 185 bp fragment containing part of the exon sequences, the intron sequences and a 12 bp polylinker fragment. Following dephosphorylation of the M13.L21 DNA with calf intestine phosphatase (Maniatis et al., 1982), it was mixed with the digested pMS137 DNA in equimolar ratio, and ligation mixtures were used to transform E. coli strain JM103 (Maniatis et al., 1982). Several white plaques were obtained and screened for recombinant phages. DNA from these phages was digested with EcoRI and BamHI to release a 1.4 kb fragment that contained the α-galactosidase A leader sequences cloned upstream from the α-galactosidase A propeptide sequences (Fig. 19).

RESULTS

Purification and amino acid composition of human α-galactosidase A

Human lung was the source for extracting sufficient homogeneous α-galactosidase A preparations (Calhoun et al., 1985). Cyanogen bromide and tryptic peptides were generated and enzyme preparations were used for microsequencing in Dr. L. Hood's laboratory (Hunkapiller and Hood, 1983). Microsequencing of the intact mature polypeptide provided us with 37 amino acid residues derived from the amino terminal portion of α-galactosidase A. In addition, cyanogen bromide and tryptic peptides were sequenced, providing a total number of 101 residues (Bishop et al., 1986).

Oligonucleotide synthesis

Previous studies demonstrated the feasibility of using oligonucleotides to detect specific DNA or RNA sequences, either directly as probes (Wallace et al., 1979), or as primers (Noyes et al., 1979). Synthetic oligonucleotides have now become an obvious choice as tools for cloning rare DNA or mRNA sequences for which standard message enrichment techniques are not applicable (Itakura et al., 1984). Initially we identified a suitable region of low redundancy within the

first 37 amino acids of α-galactosidase A. Seven oligonucleotide mixtures were synthesized to include the possible codon combination predicted from adjacent and/or overlapping regions of α-galactosidase A (i.e. oligonucleotides 1A, 1B, 1C, 2A, 2B, 4 and 5, Fig. 1). Amino acid sequence information derived from cyanogen bromide or tryptic peptides (Fig. 1) was used to construct additional oligonucleotides (e.g. oligonucleotide 3, Fig. 1). Several oligonucleotides were synthesized as mixtures to include all possible combination of codons for the corresponding amino acids. In several cases complexity of oligonucleotides was reduced based on the frequency of the human codon usage (i.e. oligonucleotides 2A and 2B) (Grantham et al., 1981), or G/T wobble (Matteucci and Caruthers, 1981) (i.e. oligonucleotide 5).

Initial screening of cDNA and genomic libraries

The human adult and fetal cDNA libraries described by Orkin and his collegues (Michelson et al., 1983, Proehownik et al., 1983) cloned into pKT218 E. coli vector, (Table 3), were screened 3 times (250,000 clones per screening) using several oligonucleotides (2A, 2B, 1C and 5) available (Table 2). The goal was to identify clones that bound two or more oligonucleotides at high hybridization stringency. An inherent limitation of this approach is that a discrepancy in our tentative amino acid sequence could result in the absence of binding by the authentic α-galactosidase A cDNA by an oligonucleotide mixture, or in a partial mismatch with one or more oligonucleotide mixtures. No clones were identified that convincingly bound two or more oligonucleotides at stringencies expected for a perfect match. However, considering the limitations of this approach, the most promising clones were

characterized further. Four putative positive clones were selected by hybridization with oligonucleotides 2A, 2B and 1C (Fig. 1). These cDNA clones were selected on the basis of hybridization at moderate stringency conditions for each oligonucleotide mixture. One clone designated pPH15, containing a 1.3 kb insert, bound to a 20-mer oligonucleotide (e.g oligonucleotide 1C) at high hybridization stringency. DNA sequence analysis revealed that clone pPH15 had a 17-base (15 contiguous bases) match to the 20-mer oligonucleotide mixture, of which 15 bases formed a block of perfect matches with the oligonucleotide and two bases interspersed in the remaining 5 bases of the oligonucleotide. Examination of the sequence flanking the oligonucleotide binding site showed no homology with the known amino acid sequence (not shown). Oligonucleotide 5, a unique 47-mer in which 14 nucleotides were selected at ambiguous positions based on the codon frequencies for human genes and the G/T wobble pairing, was used in all previous screenings. However, no hybridization signals were detected at the stringencies expected to obtain positive signals of significant hybridization (see Discussion).

In addition, the 49, XXXXY human fibroblast genomic library of Wood et al. (Wood et al., 1984) containing inserts 12-17 kb long in Charon 30 (Table 3), was screened using oligonucleotides 1A, 1B, 2A, 2B, 3, and 5. Several clones that exhibited binding to more than one oligonucleotides (e.g. oligonucleotides 1A/1B and 3) at moderate or high stringencies were selected. However, further analysis was discontinued after the identification of a positive cDNA clone from the expression λgt11 cDNA library (Woo library) after screening with both antibody and oligonucleotide probes (see below).

Identification of a cDNA clone coding for the α-galactosidase A

Multiple screenings were initiated in our and D. Bishop's (Department of Human Genetics) laboratories using a human liver cDNA λgt11 expression library. This library contained 1.4×10^7 independent clones (Table 3) and was screened for α-galactosidase A by using oligonucleotide probes as well as antibody detection methods (Young and Davis, 1983) as modified by deWet et al. (de Wet et al., 1984). Because the site of insertion for foreign DNA in λgt11 is within the structural gene for β-galactosidase, foreign cDNA sequences in this vector have the potential to be expressed as fusion proteins with β-galactosidase. An antibody positive clone, designated λAG18, was identified, isolated and plaque purified (Calhoun et al., 1985). The authenticity of this clone was confirmed by demonstrating competition of antibody binding with highly purified α-galactosidase A. In addition, the 1.25 kb insert from Southern blots hybridized strongly with three different synthetic oligonucleotide mixtures (e.g. 1A/1B, 2A/2B and 3). Two of the oligonucleotide probes (1A/1B and 2A/2B) corresponded to adjacent amino-terminal nucleotide sequences, separated by only one nucleotide (Fig. 1). Also, a 142 bp AluI restriction fragment showed stringent hybridization with these two amino-terminal probes (Calhoun et al., 1985). In addition, probe 3 corresponding to an internal tryptic peptide sequence (Fig. 1) hybridized to the EcoRI insert of α-galactosidase A cDNA. Colinearity between known amino acid sequence and predicted amino acid sequence obtained from partial sequencing of the cDNA (see below) further confirmed the identity of

this cDNA clone. The size of the cDNA insert of λAG18 was estimated to be 1250 bp, a length that could include the entire α-galactosidase A coding sequence.

Subcloning of the α-galactosidase A cDNA sequences into pBR322

DNA from clone λAG18 was digested with EcoRI and the 1250 bp insert fragment was ligated into the EcoRI site of pBR322. A clone, designated pAG18, contained the α-galactosidase A insert in a relative orientation that positions the amino-terminal of α-galactosidase A next to the HindIII site of pBR322. DNA from clone pAG18 was used to obtain preliminary nucleotide sequence from the 5'-end portion of the insert. This sequence, when translated from the appropriate open reading frame, gave an exact correspondence between the predicted amino acid sequence and the known amino acid sequence. A minimal restriction map of clone pAG18 was deduced using a battery of various restriction enzymes (Fig. 3) in order to facilitate a sequencing strategy using the base specific chemical cleavage method (Maxam and Gilbert, 1980).

Nucleotide sequence of α-galactosidase A cDNA

Dephosphorylated EcoRI α-galactosidase A insert DNA was labelled with [^{32}P]ATP and the 1250 bp fragment was digested with PvuII restriction endonuclease, which created two new fragments, of approximately 750 bp and 500 bp. The larger fragment was subjected to chemical sequencing (Maxam and Gilbert, 1980) and about 250 nucleotides

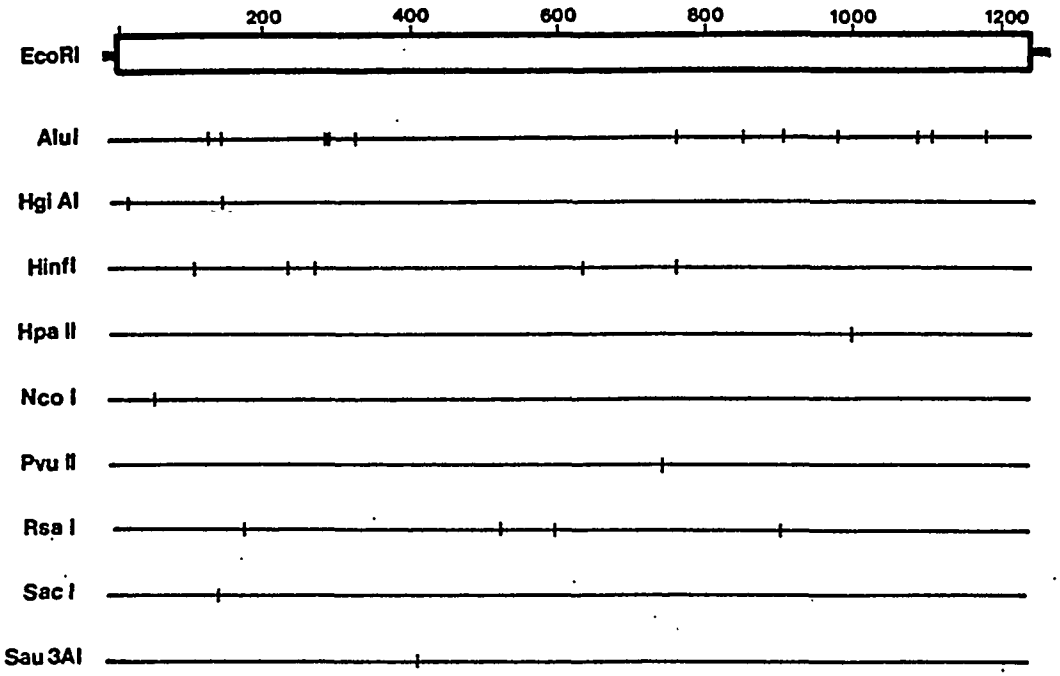

Figure 3

Restriction enzyme map of α-galactosidase A cDNA insert of clone pAG18. Enzymes represented are EcoRI, AluI, HgiAI, HinfI, HpaII, NcoI, PvuII, RsaI, SacI, Sau3AI. EcoRI indicates the EcoRI cloning site of pBR322.

were obtained from the 5'-end. It was evident from the nucleotide sequence that the cDNA insert of clone pAG18 included a 5'- coding region with an amino acid sequence that corresponded precisely to the 37 known amino terminal residues (Fig. 4). In addition, the extra 150 nucleotides provided sequence information beyond the already known amino terminal sequence.

Plasmid pAG18 DNA was digested with MspI restriction endonuclease, [^{32}P]-end-labeled and digested with EcoRI. The 950 bp MspI-to-EcoRI fragment (Fig. 3) was isolated by gel electrophoresis and subjected into chemical sequencing. A total of about 350 nucleotides were obtained from this internal MspI site (Fig. 5). Oligonucleotide 3 binding to the MspI-to-EcoRI fragment was detected using dot blot hybridization. Oligonucleotide 3 was derived from an internal tryptic peptide fragment (T-53B, Fig. 1) and its correspondence with the known amino acid sequence of tryptic peptide T-53B. The 1250 bp EcoRI insert of pAG18 was ligated in both orientations into the EcoRI site of M13mp18 (Bishop et al., 1986). The two subclones, designated mAG18.27 and mAG18.25 (Fig. 5) for the two orientations, were used to construct a series of overlapping deletion sublones according to previously published procedure (Dale et al., 1985). Single-stranded DNA was isolated from the sublones and their nucleotide sequence was determined during an intensive group effort using the enzymatic method of Sanger (Sanger, 1981). M13 clones 25.17, 25.21, 27.19 and 27.18 (Fig. 5) were sequenced by me using the dideoxy sequencing method of Sanger (1981). In addition two internal synthetic oligonucleotide primers were used to sequence previously ambiguous areas of both the message and complementary strands (Bishop et al., 1986), (Fig. 5, closed squares).

Sequence files were made from each clone and analysed in an IBM

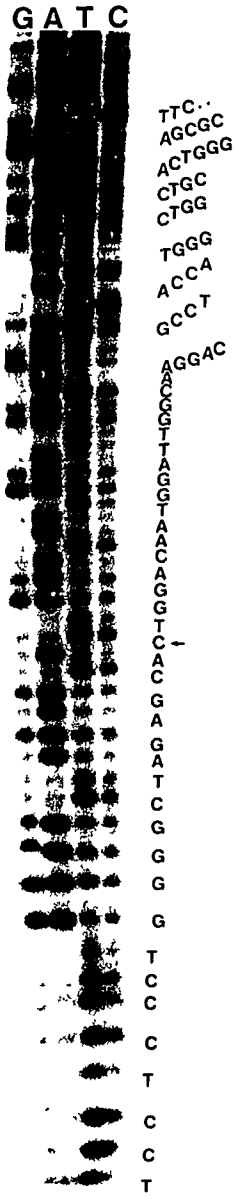

Figure 4

Autoradiography of the 5'-end nucleotide sequence of the cDNA insert of clone pAG18. EcoRI-PvuII fragment that was labelled at 5'-end was subjected to base specific chemical cleavage (Maxam and Gilbert, 1980). Arrow shows the first nucleotide of the mature α-galactosidase A sequence. The lanes labelled G, A, T, and C correspond to cleavage at G, G+A, T+C, and C>T, respectively.

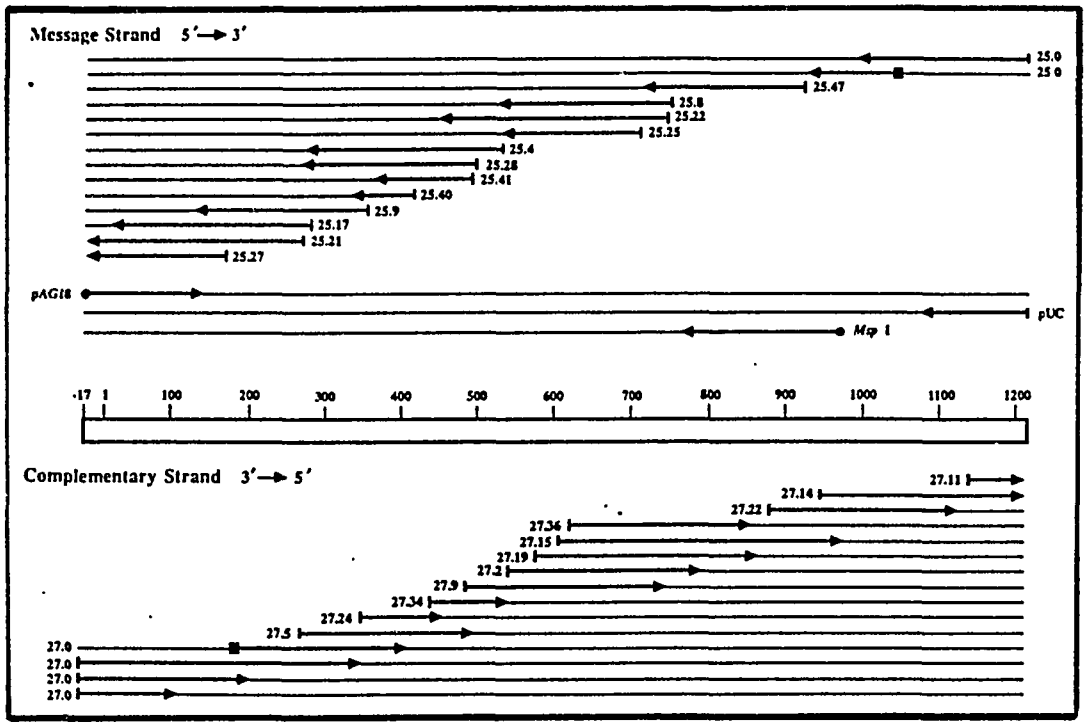

Figure 5

Strategy for sequencing the human α-galactosidase A cDNA clones. M13 clones mAG25.0 and mAG27.0 package the entire message and complementary strand, respectively, from the λAG18 EcoRI insert and were used to generate deletion subclones (e.g. 25.4, 27.2, etc.). 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 (Sanger et al., 1980) 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. Arrows originating from closed circles indicate sequences determined by the method of Maxam and Gilbert (Maxam and Gilbert, 1980), (taken from Bishop et al., 1986).

PC-XT microcomputer, using the Microgenie Sequence Software (Beckman) for DNA/RNA/Protein analysis. The overall strategy for sequencing of the human α-galactosidase A cDNA is shown in Fig. 5. The complete nucleotide sequence of clone pAG18, confirmed in its entirety from both strands is depicted in Fig. 6.

Predicted amino acid sequence of the mature form of α-galactosidase A

The entire insert cDNA sequence of 1234 bp corresponds to an open reading frame, that encodes for 406 amino acids. Molecular weight estimates of the mature α-galactosidase A, before and after N-glycanase treatment, indicated that the mature glycoprotein contained approximately 370 amino acid residues and about 16% carbohydrates (Calhoun et al., 1985). Therefore, clone pAG18 is of sufficient length to include the entire coding sequence of the mature enzyme. Sequence analysis showed that the 5'-end of the insert encodes 6 amino acids of the prepeptide sequence plus two amino acids coded by the EcoRI linker at the 5'-end (nucleotides 1 to 23). The first amino acid of the mature, processed form of α-galactosidase A begins with the CTG codon for leucine at nucleotide 24 in the cDNA insert. The stop codon UAA is at nucleotide 1218 followed by poly(A) tail. Attached to the poly(A) tail is another set of EcoRI linkers (Fig. 6). Interestingly, the two A residues in the UAA stop codon form part of the poly(A) tail. The most common polyadenylation signal AATAAA is not found near the 3'-end of the sequence. Instead, a less common signal, ATTAAA occurs 17 bases before the poly(A) tail (Fig. 6). In addition, another hexanucleotide

```
      10         20         30         40         50         60         70
AATTCCTCCC TGGGGCTAGA GCACTGGACA ATGGATTGGC AAGGACGCCT ACCATGGGCT GGCTGCACTG
   GGAGGG ACCCCGATCT CGTGACCTGT TACCTAACCG TTCCTGCGGA TGGTACCCGA CCGACGTGAC

      80         90        100        110        120        130        140
GGAGCGCTTC ATGTGCAACC TTGACTGCCA GGAAGAGCCA GATTCCTGCA TCAGTGAGAA GCTCTTCATG
CCTCGCGAAG TACACGTTGG AACTGACGGT CCTTCTCGGT CTAAGGACGT AGTCACTCTT CGAGAAGTAC

     150        160        170        180        190        200        210
GAGATGGCAG AGCTCATGGT CTCAGAAGGC TGGAAGGATG CAGGTTATGA GTACCTCTGC ATTGATGACT
CTCTACCGTC TCGAGTACCA GAGTCTTCCG ACCTTCCTAC GTCCAATACT CATGGAGACG TAACTACTGA

     220        230        240        250        260        270        280
GTTGGATGGC TCCCCAAAGA GATTCAGAAG GCAGACTTCA GGCAGACCCT CAGCGCTTTC CTCATGGGAT
CAACCTACCG AGGGGTTTCT CTAAGTCTTC CGTCTGAAGT CCGTCTGGGA GTCGCGAAAG GAGTACCCTA

     290        300        310        320        330        340        350
TCGCCAGCTA GCTAATTATG TTCACAGCAA AGGACTGAAG CTAGGGATTT ATGCAGATGT TGGAAATAAA
AGCGGTCGAT CGATTAATAC AAGTGTCGTT TCCTGACTTC GATCCCTAAA TACGTCTACA ACCTTTATTT

     360        370        380        390        400        410        420
ACCTGCGCAG GCTTCCCTGG GAGTTTTGGA TACTACGACA TTGATGCCCA GACCTTTGCT GACTGGGGAG
TGGACGCGTC CGAAGGGACC CTCAAAACCT ATGATGCTGT AACTACGGGT CTGGAAACGA CTGACCCCTC

     430        440        450        460        470        480        490
TAGATCTGCT AAAATTTGAT GGTTGTTACT GTGACAGTTT GGAAAATTTG GCAGATGGTT ATAAGCACAT
ATCTAGACGA TTTTAAACTA CCAACAATGA CACTGTCAAA CCTTTTAAAC CGTCTACCAA TATTCGTGTA

     500        510        520        530        540        550        560
GTCCTTGGCC CTGAATAGGA CTGGCAGAAG CATTGTGTAC TCCTGTGAGT GGCCTCTTTA TATGTGGCCC
CAGGAACCGG GACTTATCCT GACCGTCTTC GTAACACATG AGGACACTCA CCGGAGAAAT ATACACCGGG

     570        580        590        600        610        620        630
TTTCAAAAGC CCAATTATAC AGAAATCCGA CAGTACTGCA ATCACTGGCG AAATTTTGCT GACATTGATG
AAAGTTTTCG GGTTAATATG TCTTTAGGCT GTCATGACGT TAGTGACCGC TTTAAAACGA CTGTAACTAC

     640        650        660        670        680        690        700
ATTCCTGGAA AAGTATAAAG AGTATCTTGG ACTGGACATC TTTTAACCAG GAGAGAATTG TTGATGTTGC
TAAGGACCTT TTCATATTTC TCATAGAACC TGACCTGTAG AAAATTGGTC CTCTCTTAAC AACTACAACG

     710        720        730        740        750        760        770
TGGACCAGGG GGTTGGAATG ACCCAGATAT GTTAGTGATT GGCAACTTTG GCCTCAGCTG GAATCAGCAA
ACCTGGTCCC CCAACCTTAC TGGGTCTATA CAATCACTAA CCGTTGAAAC CGGAGTCGAC CTTAGTCGTT

     780        790        800        810        820        830        840
GTAACTCAGA TGGCCCTCTG GGCTATCATG GCTGCTCCTT TATTCATGTC TAATGACCTC CGACACATCA
CATTGAGTCT ACCGGGGAGAC CCGATAGTAC CGACGAGGAA ATAAGTACAG ATTACTGGAG GCTGTGTAGT

     850        860        870        880        890        900        910
GCCCTCAAGC CAAAGCTCTC CTTCAGGATA AGGACGTAAT TGCCATCAAT CAGGACCCCT TGGGCAAGCA
CGGGAGTTCG GTTTCGAGAG GAAGTCCTAT TCCTGCATTA ACGGTAGTTA GTCCTGGGGA ACCCGTTCGT

     920        930        940        950        960        970        980
AGGGTACCAG CTTAGACAGG GAGACAACTT TGAAGTGTGG GAACGACCTC TCTCAGGCTT AGCCTGGGCT
TCCCATGGTC GAATCTGTCC CTCTGTTGAA ACTTCACACC CTTGCTGGAG AGAGTCCGAA TCGGACCCGA

     990       1000       1010       1020       1030       1040       1050
GTAGCTATGA TAAACCGGCA GGAGATTGGT GGACCTCGCT CTTATACCAT CGCAGTTGCT TCCCTGGGTA
CATCGATACT ATTTGGCCGT CCTCTAACCA CCTGGAGCGA GAATATGGTA GCGTCAACGA AGGGACCCAT

    1060       1070       1080       1090       1100       1110       1120
AAGGAGTGGC CTGTAATCCT GCCTGCTTCA TCACACAGCT CCTCCCTGTG AAAAGGAAGC TAGGGTTCTA
TTCCTCACCG GACATTAGGA CGGACGAAGT AGTGTGTCGA GGAGGGACAC TTTTCCTTCG ATCCCAAGAT

    1130       1140       1150       1160       1170       1180       1190
TGAATGGACT TCAAGGTTAA GAAGTCACAT AAATCCCACA GGCACTGTTT TGCTTCAGCT AGAAAATACA
ACTTACCTGA AGTTCCAATT CTTCAGTGTA TTTAGGGTGT CCGTGACAAA ACGAAGTCGA TCTTTTATGT

    1200       1210       1220       1230
ATGCAGATGT CATTAAAAGA CTTACTTTAA AAAAAAAAAA AAGG
TACGTCTACA GTAATTTTCT GAATGAAATT TTTTTTTTTT TTCCTTAA
```

Figure 6

Complete double-stranded nucleotide sequence of the pAG18 clone cDNA insert encoding the human mature α-galactosidase A subunit. The entire cDNA insert is 1234 bp long (EcoRI-to-EcoRI). Nucleotides 1 to 23 encode five amino acids of the leader peptide plus two amino acids due to the EcoRI linker, at the 5′ end. Polyadenylation signals AATACA and ATTAAA are overlined at nucleotides 1185 and 1202, respectively. In this cDNA insert, poly(A) tail consists of only 14 adenine residues followed by EcoRI linkers at the 3′ end of the insert.

poly(A) signal, AATACA, occurs 28 nucleotides prior to the UAA stop codon. The ATTAAA signal is present in about 12% of vertebrate messages, whereas AATACA occurs in only about 2% (Wickens et al., 1984, Brinstiel et al., 1985).

The complete amino acid sequence of the mature form of α-galactosidase A derived from the cDNA sequence is shown in Fig. 7. The cDNA sequence predicts that the enzyme has a molecular weight of 45,356 daltons (Table 4) with four putative glycosylation sites of the type Asn-X-Ser/Thr (Fig. 7). The predicted amino acid composition of the mature α-galactosidase A protein is shown in Table 4. Table 5 shows the distribution of the codons used in the open reading frame of the mature α-galactosidase A cDNA sequence.

Local secondary structure was predicted by the algorithm of Garnier et al. (1978) to adapt the method of Chou and Fassman (Chou and Fassman, 1978) for better computer computation. Regions of α-helical structure of 10 or more contiguous amino acids were located at residues 31-50, 146-155, 253-267, 275-284, and 385-398 (Fig. 8). β-sheet regions involving approximately 20-40 amino acids occured at residues 87-128, 220-240, and 325-365 (Fig. 8). The four possible N-glycosylation sites were located in β-turns within hydrophilic regions of the α-galactosidase A, as predicted by their probable surface localization (Aubert et al., 1976). There are no statistically significant repeated sequences or inverted repeats within the α-galactosidase A cDNA sequence.

Computer-generated analysis by the algorithm of Kyte and Doolittle (Kyte and Doolittle, 1982) with a span setting of 6, revealed hydrophilic regions of at least 20 amino acids dispersed along the

```
1                                                                   30                                                                  60
CTG GAC AAT GGA TTG GCA AGG ACG CCT ACC ATG GGC TGG CTG CAC TGG GAG CGC TTC ATG
Leu Asp Asn Gly Leu Ala Arg Thr Pro Thr Met Gly Trp Leu His Trp Glu Arg Phe Met

                                                                    90                                                                 120
TGC AAC CTT GAC TGC CAG GAA GAG CCA GAT TCC TGC ATC AGT GAG AAG CTC TTC ATG GAG
Cys Asn Leu Asp Cys Gln Glu Glu Pro Asp Ser Cys Ile Ser Glu Lys Leu Phe Met Glu

                                                                   150                                                                 180
ATG GCA GAG CTC ATG GTC TCA GAA GGC TGG AAG GAT GCA GGT TAT GAG TAC CTC TGC ATT
Met Ala Glu Leu Met Val Ser Glu Gly Trp Lys Asp Ala Gly Tyr Glu Tyr Leu Cys Ile

                                                                   210                                                                 240
GAT GAC TGT TGG ATG GCT CCC CAA AGA GAT TCA GAA GGC AGA CTT CAG GCA GAC CCT CAG
Asp Asp Cys Trp Met Ala Pro Gln Arg Asp Ser Glu Gly Arg Leu Gln Ala Asp Pro Gln

                                                                   270                                                                 300
CGC TTT CCT CAT GGG ATT CGC CAG CTA GCT AAT TAT GTT CAC AGC AAA GGA CTG AAG CTA
Arg Phe Pro His Gly Ile Arg Gln Leu Ala Asn Tyr Val His Ser Lys Gly Leu Lys Leu

                                                                   330                                                                 360
GGG ATT TAT GCA GAT GTT GGA AAT AAA ACC TGC GCA GGC TTC CCT GGG AGT TTT GGA TAC
Gly Ile Tyr Ala Asp Val Gly Asn Lys Thr Cys Ala Gly Phe Pro Gly Ser Phe Gly Tyr
                                   CHO--------

                                                                   390                                                                 420
TAC GAC ATT GAT GCC CAG ACC TTT GCT GAC TGG GGA GTA GAT CTG CTA AAA TTT GAT GGT
Tyr Asp Ile Asp Ala Gln Thr Phe Ala Asp Trp Gly Val Asp Leu Leu Lys Phe Asp Gly

                                                                   450                                                                 480
TGT TAC TGT GAC AGT TTG GAA AAT TTG GCA GAT GGT TAT AAG CAC ATG TCC TTG GCC CTG
Cys Tyr Cys Asp Ser Leu Glu Asn Leu Ala Asp Gly Tyr Lys His Met Ser Leu Ala Leu

                                                                   510                                                                 540
AAT AGG ACT GGC AGA AGC ATT GTG TAC TCC TGT GAG TGG CCT CTT TAT ATG TGG CCC TTT
Asn Arg Thr Gly Arg Ser Ile Val Tyr Ser Cys Glu Trp Pro Leu Tyr Met Trp Pro Phe
CHO--------

                                                                   570                                                                 600
CAA AAG CCC AAT TAT ACA GAA ATC CGA CAG TAC TGC AAT CAC TGG CGA AAT TTT GCT GAC
Gln Lys Pro Asn Tyr Thr Glu Ile Arg Gln Tyr Cys Asn His Trp Arg Asn Phe Ala Asp
           CHO--------

                                                                   630                                                                 660
ATT GAT GAT TCC TGG AAA AGT ATA AAG AGT ATC TTG GAC TGG ACA TCT TTT AAC CAG GAG
Ile Asp Asp Ser Trp Lys Ser Ile Lys Ser Ile Leu Asp Trp Thr Ser Phe Asn Gln Glu

                                                                   690                                                                 720
AGA ATT GTT GAT GTT GCT GGA CCA GGG GGT TGG AAT GAC CCA GAT ATG TTA GTG ATT GGC
Arg Ile Val Asp Val Ala Gly Pro Gly Gly Trp Asn Asp Pro Asp Met Leu Val Ile Gly

                                                                   750                                                                 780
AAC TTT GGC CTC AGC TGG AAT CAG CAA GTA ACT CAG ATG GCC CTC TGG GCT ATC ATG GCT
Asn Phe Gly Leu Ser Trp Asn Gln Gln Val Thr Gln Met Ala Leu Trp Ala Ile Met Ala

                                                                   810                                                                 840
GCT CCT TTA TTC ATG TCT AAT GAC CTC CGA CAC ATC AGC CCT CAA GCC AAA GCT CTC CTT
Ala Pro Leu Phe Met Ser Asn Asp Leu Arg His Ile Ser Pro Gln Ala Lys Ala Leu Leu

                                                                   870                                                                 900
CAG GAT AAG GAC GTA ATT GCC ATC AAT CAG GAC CCC TTG GGC AAG CAA GGG TAC CAG CTT
Gln Asp Lys Asp Val Ile Ala Ile Asn Gln Asp Pro Leu Gly Lys Gln Gly Tyr Gln Leu

                                                                   930                                                                 960
AGA CAG GGA GAC AAC TTT GAA GTG TGG GAA CGA CCT CTC TCA GGC TTA GCC TGG GCT GTA
Arg Gln Gly Asp Asn Phe Glu Val Trp Glu Arg Pro Leu Ser Gly Leu Ala Trp Ala Val

                                                                   990                                                                1020
GCT ATG ATA AAC CGG CAG GAG ATT GGT GGA CCT CGC TCT TAT ACC ATC GCA GTT GCT TCC
Ala Met Ile Asn Arg Gln Glu Ile Gly Gly Pro Arg Ser Tyr Thr Ile Ala Val Ala Ser

                                                                  1050                                                                1080
CTG GGT AAA GGA GTG GCC TGT AAT CCT GCC TGC TTC ATC ACA CAG CTC CTC CCT GTG AAA
Leu Gly Lys Gly Val Ala Cys Asn Pro Ala Cys Phe Ile Thr Gln Leu Leu Pro Val Lys

                                                                  1110                                                                1140
AGG AAG CTA GGG TTC TAT GAA TGG ACT TCA AGG TTA AGA AGT CAC ATA AAT CCC ACA GGC
Arg Lys Leu Gly Phe Tyr Glu Trp Thr Ser Arg Leu Arg Ser His Ile Asn Pro Thr Gly
                                                                                                                    CHO--------

                                                                  1170
ACT GTT TTG CTT CAG CTA GAA AAT ACA ATG CAG ATG TCA TTA AAA GAC TTA CTT TAA
Thr Val Leu Leu Gln Leu Glu Asn Thr Met Gln Met Ser Leu Lys Asp Leu Leu TER
```

Figure 7

The predicted amino acid sequence of the human mature α-galactosidase A subunit written below the open reading frame of the cDNA sequence. Nucleotide 24 of the cDNA insert of pAG18 clone is nucleotide 1 of the mature α-galactosidase A sequence. CHO indicate potential sites of N-glycosylation (Asn-Xaa-Thr/Ser).

TABLE 4

AMINO ACID COMPOSITION OF THE MATURE FORM OF HUMAN α-GALACTOSIDASE A AS PREDICTED FROM THE cDNA SEQUENCE

(Numbers in parentheses indicate percentages for each amino acid)

Ala	28	(7.0)	Leu	41	(10.3)
Arg	19	(4.8)	Lys	17	(4.3)
Asn	20	(5.0)	Met	15	(3.8)
Asp	29	(7.3)	Phe	15	(3.8)
Cys	12	(3.0)	Pro	19	(4.8)
Gln	22	(5.5)	Ser	23	(5.8)
Glu	18	(4.5)	Thr	14	(3.5)
Gly	31	(7.8)	Trp	16	(4.0)
His	7	(1.8)	Tyr	15	(3.8)
Ile	21	(5.3)	Val	16	(4.0)

Acidic	(Asp + Glu)	47	(11.8)
Basic	(Arg + Lys)	36	(9.0)
Aromatic	(Phe + Trp + Tyr)	46	(11.5)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	139	(34.8)

Molecular Weight = 45356.

TABLE 5

THE DISTRIBUTION OF CODONS IN THE OPEN READING FRAME
OF THE MATURE α-GALACTOSIDASE A
AS PREDICTED FROM THE cDNA NUCLEOTIDE SEQUENSE
(Number in parentheses indicate percentages for each codon)

TTT Phe	9 (2.3)	TCT Ser	3 (0.8)	TAT Tyr	8 (2.0)	TGT Cys	5 (1.3)
TTC Phe	6 (1.5)	TCC Ser	5 (1.3)	TAC Tyr	7 (1.8)	TGC Cys	7 (1.8)
TTA Leu	6 (1.5)	TCA Ser	5 (1.3)	TAA End	1 (0.3)	TGA End	0 (0.0)
TTG Leu	7 (1.8)	TCG Ser	0 (0.0)	TAG End	0 (0.0)	TGG Trp	16 (4.0)
CTT Leu	7 (1.8)	CCT Pro	11 (2.8)	CAT His	1 (0.3)	CGT Arg	0 (0.0)
CTC Leu	10 (2.5)	CCC Pro	5 (1.3)	CAC His	6 (1.5)	CGC Arg	4 (1.0)
CTA Leu	5 (1.3)	CCA Pro	3 (0.8)	CAA Gln	5 (1.3)	CGA Arg	4 (1.0)
CTG Leu	6 (1.5)	CCG Pro	0 (0.0)	CAG Gln	17 (4.3)	CGG Arg	1 (0.3)
ATT Ile	10 (2.5)	ACT Thr	4 (1.0)	AAT Asn	15 (3.8)	AGT Ser	6 (1.5)
ATC Ile	8 (2.0)	ACC Thr	4 (1.0)	AAC Asn	5 (1.3)	AGC Ser	4 (1.0)
ATA Ile	3 (0.8)	ACA Thr	5 (1.3)	AAA Lys	8 (2.0)	AGA Arg	6 (1.5)
ATG Met	15 (3.8)	ACG Thr	1 (0.3)	AAG Lys	9 (2.3)	AGG Arg	4 (1.0)
GTT Val	6 (1.5)	GCT Ala	12 (3.0)	GAT Asp	14 (3.5)	GGT Gly	6 (1.5)
GTC Val	1 (0.3)	GCC Ala	8 (2.0)	GAC Asp	15 (3.8)	GGC Gly	10 (2.5)
GTA Val	4 (1.0)	GCA Ala	8 (2.0)	GAA Glu	9 (2.3)	GGA Gly	9 (2.3)
GTG Val	5 (1.3)	GCG Ala	0 (0.0)	GAG Glu	9 (2.3)	GGG Gly	6 (1.5)

```
                                                  10                                                  20
Leu Asp Asn Gly Leu Ala Arg Thr Pro Thr Met Gly Trp Leu His Trp Glu Arg Phe Met
     T  T                          T       T           A  A  A  T  T  T

                                                  30                                                  40
Cys Asn Leu Asp Cys Gln Glu Glu Pro Asp Ser Cys Ile Ser Glu Lys Leu Phe Met Glu
 T  T       T  T  T  A  A  T  T  A  A  A  A  A  A  A  A  A  A

                                                  50                                                  60
Met Ala Glu Leu Met Val Ser Glu Gly Trp Lys Asp Ala Gly Tyr Glu Tyr Leu Cys Ile
 A  A  A  A  A  A  A  A  A  A           T  T  T  T  B  B  B  T

                                                  70                                                  80
Asp Asp Cys Trp Met Ala Pro Gln Arg Asp Ser Glu Gly Arg Leu Gln Ala Asp Pro Gln
 T  T  T  T  B           T       T       T  T  T       T           T  T

                                                  90                                                 100
Arg Phe Pro His Gly Ile Arg Gln Leu Ala Asn Tyr Val His Ser Lys Gly Leu Lys Leu
        T  T           B  T  B  B  B  B  B  B  T  T  T  T  B  B

                                                 110                                                 120
Gly Ile Tyr Ala Asp Val Gly Asn Lys Thr Cys Ala Gly Phe Pro Gly Ser Phe Gly Tyr
 B  B  B  B  T  T  T  T  T  T  B  B  B  B        T  T  T  T  T

                                                 130                                                 140
Tyr Asp Ile Asp Ala Gln Thr Phe Ala Asp Trp Gly Val Asp Leu Leu Lys Phe Asp Gly
 T  T  B  B  B  B  B  B       T  A  A  A  A  B  T  T  T  T  T

                                                 150                                                 160
Cys Tyr Cys Asp Ser Leu Glu Asn Leu Ala Asp Gly Tyr Lys His Met Ser Leu Ala Leu
 T  T  T  T  T  A  A  A  A  A  A  A  A  A  A  B  B  B  A  A

                                                 170                                                 180
Asn Arg Thr Gly Arg Ser Ile Val Tyr Ser Cys Glu Trp Pro Leu Tyr Met Trp Pro Phe
     T  T  T  T  B  B  B  B  T  T  T           T  T  T           T

                                                 190                                                 200
Gln Lys Pro Asn Tyr Thr Glu Ile Arg Gln Tyr Cys Asn His Trp Arg Asn Phe Ala Asp
           T  T  T  B  B  B  T  T  T  T  T  T  T  T  T  T  T  T

                                                 210                                                 220
Ile Asp Asp Ser Trp Lys Ser Ile Lys Ser Ile Leu Asp Trp Thr Ser Phe Asn Gln Glu
 T  T  T  A  A  A        T  B  B  B  B  T  T                          T

                                                 230                                                 240
Arg Ile Val Asp Val Ala Gly Pro Gly Gly Trp Asn Asp Pro Asp Met Leu Val Ile Gly
 T  B  B  B  B  B        T  T  T                 B  B  B  B  B  B  B

                                                 250                                                 260
Asn Phe Gly Leu Ser Trp Asn Gln Gln Val Thr Gln Met Ala Leu Trp Ala Ile Met Ala
 T                                    B  B  A  A  A  A  A  A  A  A

                                                 270                                                 280
Ala Pro Leu Phe Met Ser Asn Asp Leu Arg His Ile Ser Pro Gln Ala Lys Ala Leu Leu
 A  A  A  A  A  A  A  T       T  T                 A  A  A  A  A  A
                                                 290                                                 300
Gln Asp Lys Asp Val Ile Ala Ile Asn Gln Asp Pro Leu Gly Lys Gln Gly Tyr Gln Leu
 A  A  A  A  B  B  B  B  B                    T  T  T  T  T  B  B  B

                                                 310                                                 320
Arg Gln Gly Asp Asn Phe Glu Val Trp Glu Arg Pro Leu Ser Gly Leu Ala Trp Ala Val
 B  T                    B  B  A  A           T           A  A  A  A  A

                                                 330                                                 340
Ala Met Ile Asn Arg Gln Glu Ile Gly Gly Pro Arg Ser Tyr Thr Ile Ala Val Ala Ser
 A  A  A        T  T  T  T                    T       B  B  B  B  B  B  B

                                                 350                                                 360
Leu Gly Lys Gly Val Ala Cys Asn Pro Ala Cys Phe Ile Thr Gln Leu Leu Pro Val Lys
 T  T  T  T  T  B  B  B  T  T  B  B  B  B  B  B  B        T  T

                                                 370                                                 380
Arg Lys Leu Gly Phe Tyr Glu Trp Thr Ser Arg Leu Arg Ser His Ile Asn Pro Thr Gly
 B  B           B        T  T  A  T  T  T  T        T                    T

                                                 390
Thr Val Leu Leu Gln Leu Glu Asn Thr Met Gln Met Ser Leu Lys Asp Leu Leu
 B  B  B  B  A  A  A  A  A  A  A  A  A  A  A  A  A  A  A  A
```

Figure 8

Secondary structure of the mature α-galactosidase A polypeptide written below its amino acid sequence. The secondary structure was predicted by computer analysis (Microgenie, Beckman), using the algorithm of Garnier et al. (1978). A=α helix, B=β sheet, T=turn, blank=random coil

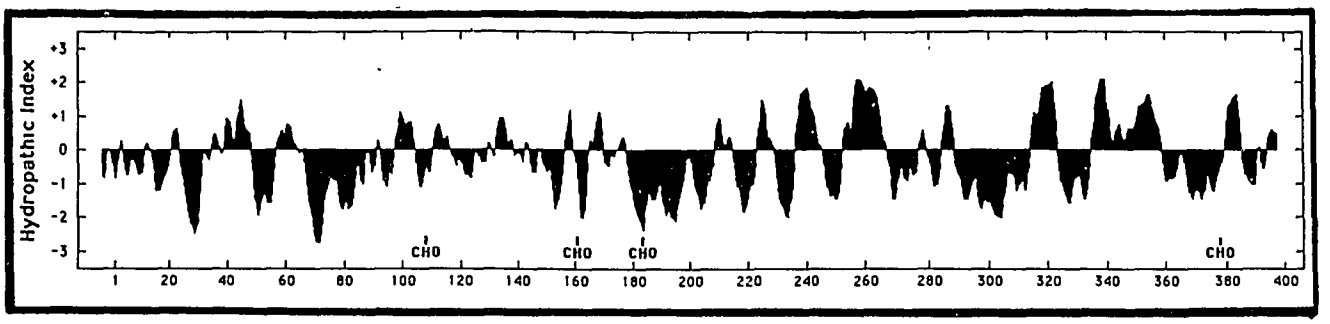

Figure 9

Hydropathy profile of the mature α-galactosidase A amino acid sequence predicted from clone pAG18. Hydropathy profile was deduced by computer-assisted analysis (Microgenie, Beckman) using the algorithm of Kyte and Doolittle (Kyte and Doolittle, 1982). For hydrophilic regions a span of 6 amino acids was set. For hydrophobic regions a span of 10 amino acids was set. Positive and negative numbers correspond to hydrophobic and hydrophilic regions, respectively. CHO indicates possible N-glycosylation sites (see Fig. 6). (Figure taken and modified from Bishop et al., 1986).

entire sequence of α-galactosidase A, occuring at residues 66-86, 179-208, 289-313, and 359-379 (Fig. 9). In contrast, hydrophobic regions of 10 amino acids or more were primarily located in the carboxy-terminal region of the mature polypeptide, occuring at residues 37-47, 252-266, 314-323, and 335-358 (Fig. 9).

Amino acid homology of α-galactosidase A with other proteins

Computer searches revealed little, if any, amino acid homology with that predicted for α-galactosidase A. The highest degree of nucleotide sequence homology was with human factor VIII (Wood et al., 1984). When two gaps totaling 23 residues were introduced, nucleotides 5489-5595 of factor VIII had 47% homology with α-galactosidase A nucleotides 495-623. Possible N-glycosylation sites were located within 6 bases of the 5'-ends of both aligned sequences. Comparison of available data for lysosomal enzymes, including human cathepsin D (Faust et al., 1985), α-fucosidase (Fukushima et al., 1985), β-glucosidase (Sorge et al., 1985), β-glucuronidase (Guise et al., 1985), β-hexosaminidase A α subunit (Myerowitz et al., 1985), and β subunit (O'Dowd et al., 1985), as well as rat cathepsin B (Segundo et al., 1985), revealed little nucleotide or amino acid sequence similarity, with the possible exeption of α-fucosidase, of which nucleotides 31-178 were 52% homologous to α-galactosidase A nucleotides 34-168, when four gaps totaling 18 bases were introduced.

Choice of the expression vector ptrpL1

There are many reports of succesful high-level expression of proteins with the trp promoter (deBoer et al., 1983, Emtage et al., 1980, Goeddel et al., 1980, Itoh et al., 1984). Early plasmid vectors designed for gene expression with the trp promoter, included the trp leader plus portions of trpD or trpE genes (Halewell and Emtage, 1980). Plasmid ptrpL1 has the trp operator-promoter region and ribosome binding site for the trp leader peptide. Expression of the five genes of the trp operon of E. coli is regulated by repression by tryptophan at the trp operator and attenuation within the trp leader (Yanofsky et al., 1981). Together, these mechanisms allow operon expression to vary over an approximately 500-fold range.

The trp promoter gene constructions on plasmid ptrpL1 has a measurable rate of constitutive activity even in the presence of excess tryptophan. This is due to the capacity of the trp operator, when present on a multicopy plasmid to titrate partially the trpR product (Trp repressor) produced by a single chromosomal copy of the trpR. Plasmid ptrpL1 and can be used as an expression vector by inserting heterologous segments (with an ATG initiation codon) immediately downstream of the promoter and ribosome binding site of trpL into the unique ClaI site (Edman et al., 1981) (Fig. 10).

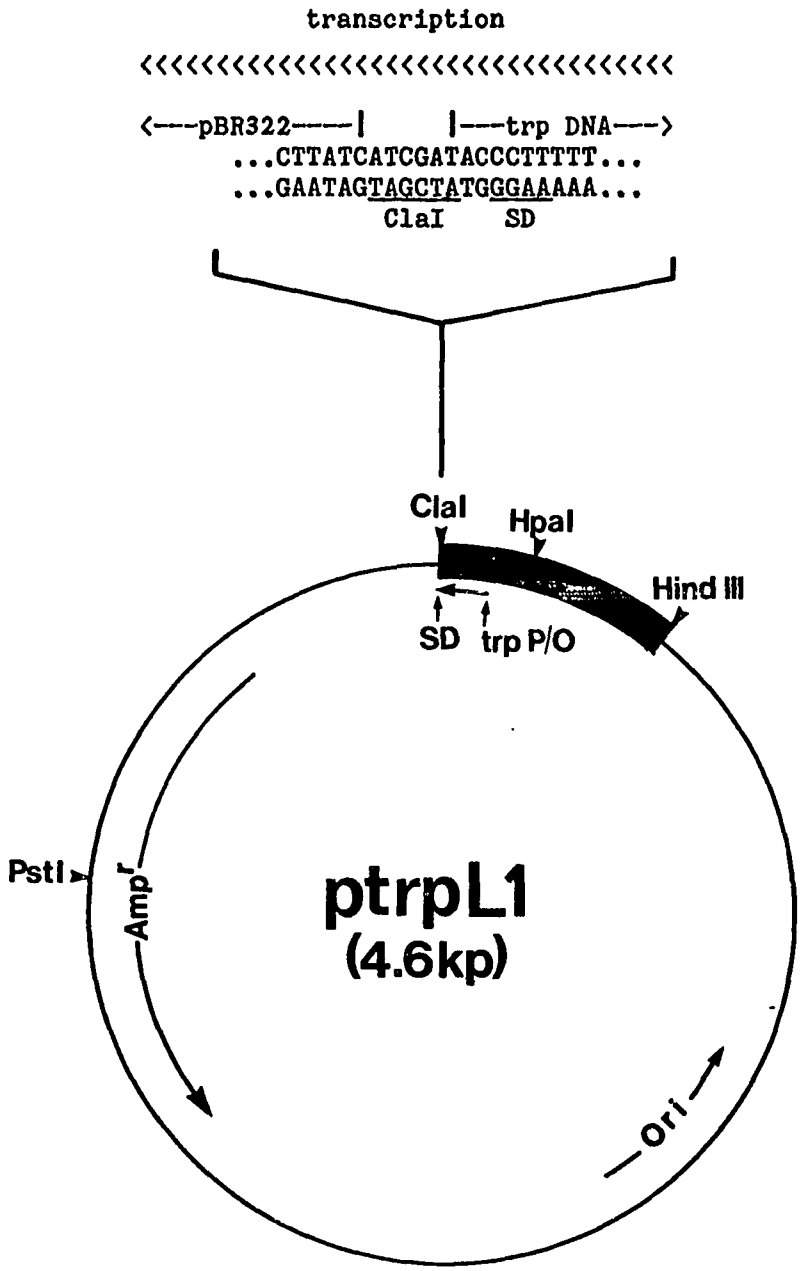

Figure 10

Diagram showing the expression vector ptrpL1. Top of the figure shows the DNA sequence around the ClaI site of ptrpL1 and the location of trp Shine-Dalgarno (SD) sequence. The direction of transcription is also shown by the leftward arrowheads. The trp promoter/ operator (trp P/O above) site is also shown relative to cloning site. The ClaI-to-HindIII fragment containing the regulatory sequences of trpL gene is approximately 275 bp.

Construction of plasmid ptrpL1 derivatives

Fig. 11 describes the steps leading to the construction of derivatives of plasmid ptrpL1 that express the mature form of the human α-galactosidase A from the regulatable trp promoter. Plasmid pMS118 contains the α-galactosidase A cDNA on a 1234 bp EcoRI fragment subcloned into pUC9 (Messing and Viera, 1982). This cDNA clone contains 17 bp upstream of the CTG leucine codon that encodes the amino terminal residue of the mature polypeptide. Plasmid pMS118 was first treated with AluI methylase to block cleavage at the internal HgiAI site in the cDNA insert, but not at the HgiAI site spanning the amino terminal leucine codon (Brown et al., 1980). Then the plasmid was digested with HgiAI and EcoRI resulting in a 1217 bp fragment. This restriction fragment was then purified and treated with T4 DNA polymerase to generate a flush ended molecule with the desired CTG codon at the 5'-end. The addition of ClaI linkers provided an ATG initiation codon at the 5'-end and ClaI specific cohesive ends at both ends of the molecule for cloning to the unique ClaI site of plasmid ptrpL1 (Fig. 10). This ClaI site is immediately downstream of the trp promoter and ribosome binding site for the trp leader peptide (Fig. 10, and Table 6). One plasmid derivative, pMS122 (Fig. 11) that fortuitously received ClaI linkers only at the 5'-end of the cDNA was used to generate derivatives with alterations in the sequence between the ribosome binding site and ATG codon (Table 6). Plasmid pMS122 was treated with ClaI, end filled with T4 polymerase, and self ligated to generate plasmid pMS132 (Fig. 11) that has a CG insertion relative to plasmid pMS122 (Fig. 11, and Table 6). Plasmid pMS132 was examined by nucleotide sequence analysis of the proximal region of the

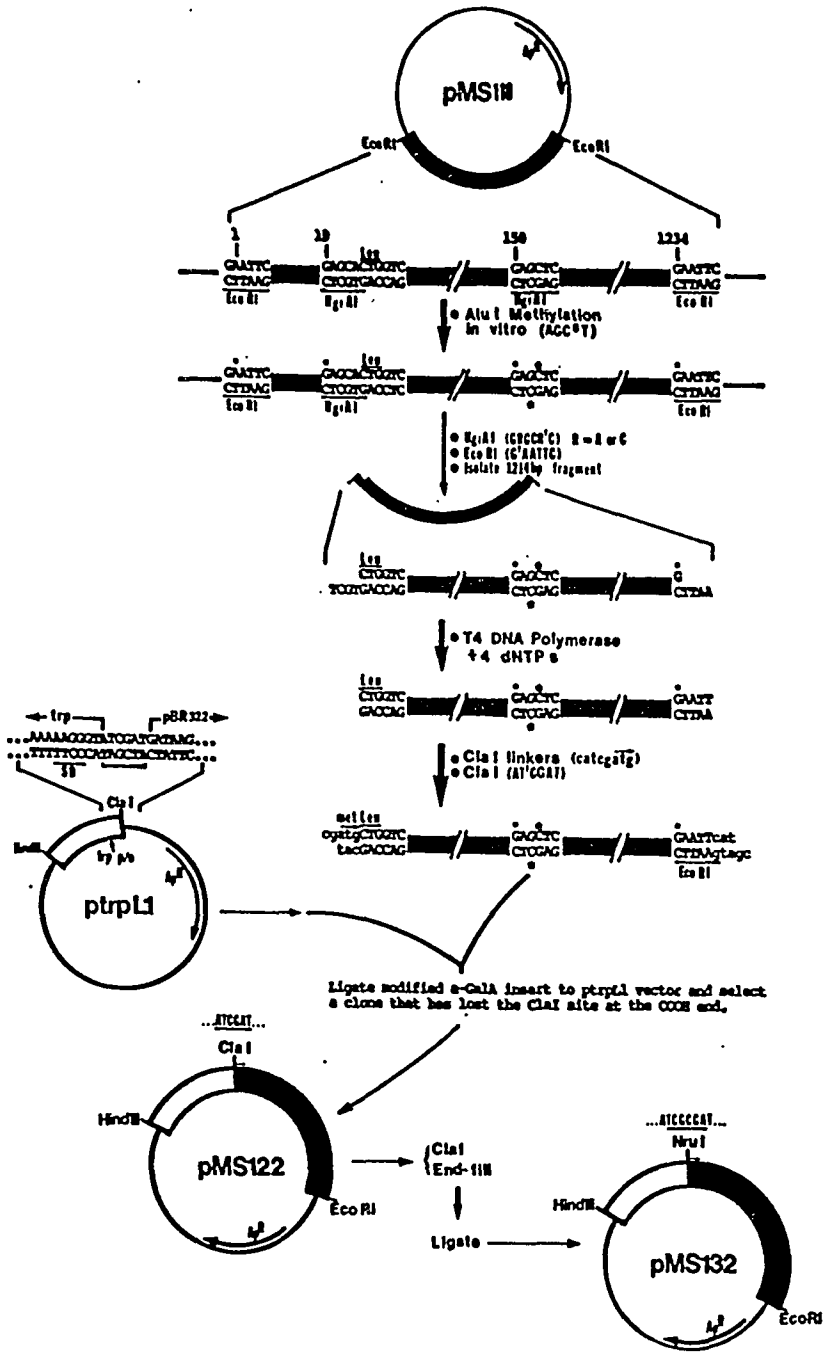

Figure 11

Construction of plasmid ptrpL1 derivatives pMS122 and pMS132 that express the mature form of the human α-galactosidase A in E. coli K-12. An EcoRI segment derived from our λgt11 cDNA clone (Calhoun et al., 1985, Bishop et al., 1986) was treated as indicated to obtain plasmids which conditionally express the human enzyme in response to tryptophan levels within the cell. See text for details.

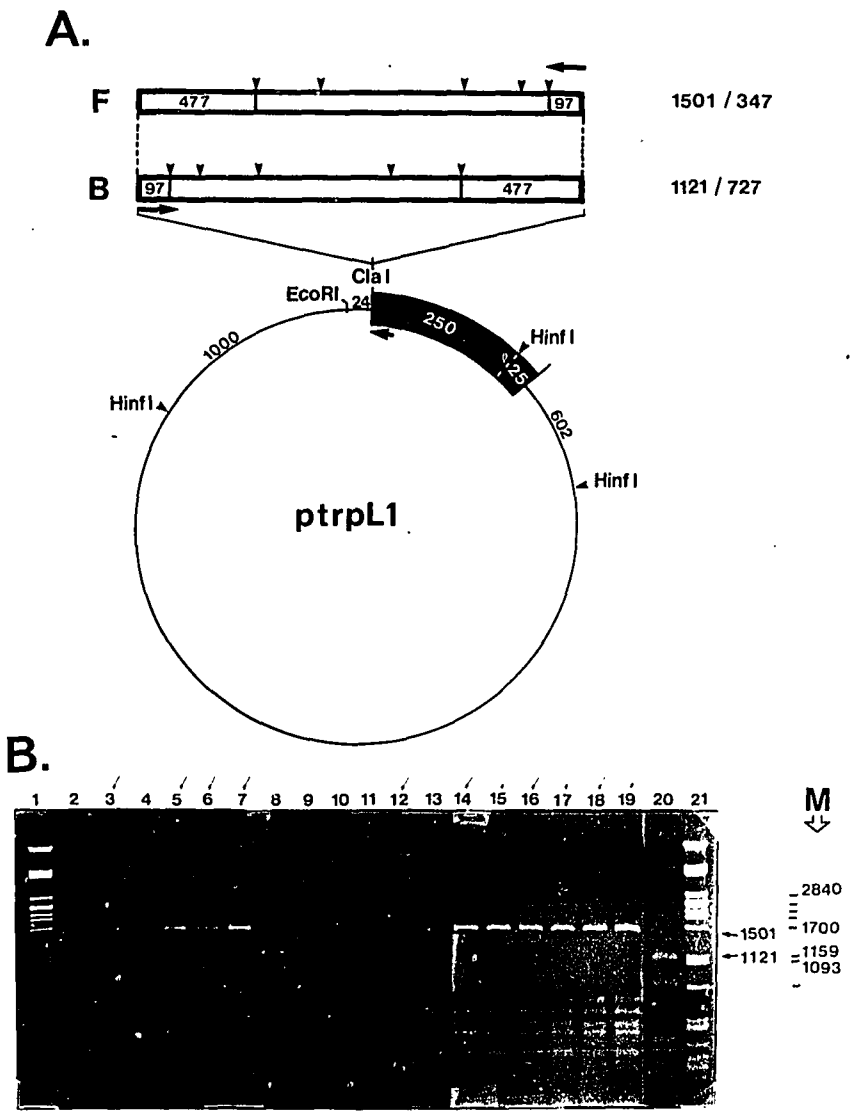

Figure 12

Orientation analysis of ptrpL1 recombinant clones. Panel A: α-galactosidase A insert cloned in vector ptrpL1 in the possible orientation is shown in the top. F:forward orientation, B:backward orientation relative to the trp promoter. Small arrows indicate HinfI sites in ptrpL1 (only two sites are shown) and α-galactosidase A insert. The size of junction fragments are indicated by numbers in bp (e.g. 477 and 97 bp). For each orientation a characteristic fragment size is expected (e.g. 1501 for F orientation and 1121 bp for B orientation). Large arrows indicate direction of transcription. Panel B: 1% agarose gel showing HinfI digestion of recombinant ptrpL1 clones based on the above orientation analysis (Panel A). Lanes 1 and 21: PstI digested λ DNA. Lanes 4, 10, 13: no DNA. Thin arrows above numbered lanes indicate positive clones (i.e. F orientation).

TABLE 6

NUCLEOTIDE SEQUENCE FLANKING THE RIBOSOME BINDING SITE IN ptrpL1 DERIVATIVES

Plasmid	Sequence	Distance[c]
trpL1	.....A A A G G G T A T C G A C A A T G A A A	9bp
pMS122	.....A A A G G G T A T - - - C G A T G C T G	6bp
pMS132	.....A A A G G G T A T C G - C G A T G C T G	8bp

[c] The number of nucleotides between the AAGG ribosome binding site (overlined) and the ATG initiation codon (underlined) is indicated.

α-galactosidase A gene. The examined nucleotide sequence corresponded exactly to the expected sequence for this region. The above procedure increased the separation of the ATG and ribosome binding site from 6 to 8 bp, and it also generated an NruI site. The full-length of α-galactosidase A gene could be excised from this plasmid by NruI and EcoRI double cleavage.

Synthesis of α-galactosidase A in Escherichia coli

The trp promoter present in plasmid ptrpL1 is expressed at a low level in the presence of exogenous tryptophan, at intermediate levels in minimal medium without tryptophan, and at high levels when 3-β-indoleacrylic acid is present to interfere with the action of the trpR coded tryptophan repressor protein (Rose and Yanofsky, 1974). A protein of approximately 45 kDa is produced in maxicells containing plasmid pMS132 (Fig. 13, lanes 1, 3, and 5) but not with the vector alone, (Fig. 13, lanes 2, 4, and 6). The production of this 45 kDa protein responds as expected in the presence of 3-β-indoleacrylic acid (Fig. 13, lane 1), tryptophan (Fig. 13, lane 3), and unsupplemented minimal medium (Fig. 13, lane 5). Identification of this 45 kDa protein as the α-galactosidase A product is further confirmed by its dissappearance in the plasmids containing no α-galactosidase A insert. The 45 kDa band seen in lanes 3 and 5 of Fig. 13 correspond to α-galactosidase A protein made at a measurable constitutive levels even in the presence of excess tryptophan (200 μg/ml). The other vector coded polypeptide, the bla gene product (β-lactamase), is seen in all maxicell preparations as a protein of approximately 31 kDa. An

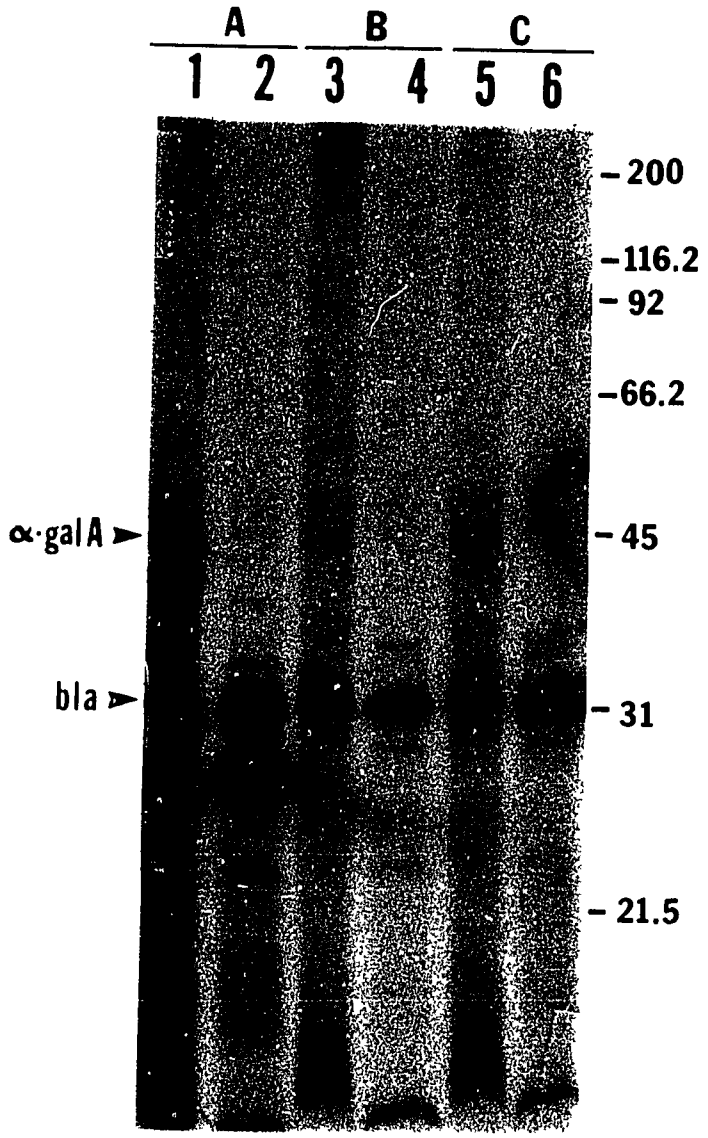

Figure 13

Regulated expression of the human α-galactosidase A in maxicells containing plasmid pMS132 (lanes 1, 3, and 5) compared to the ptrpL1 control (lanes 2, 4, and 6). The cells were treated with 3-β-indoleacrylic acid to derepress the trp promoter (lanes 1 and 2), tryptophan to repress expression (lane 3 and 4) or no treatment, a situation which results in intermediate levels of promoter activity (lanes 5 and 6). The positions of the human α-galactosidase A (α-galactosidase A), the vector coded β-lactamase (bla) and the molecular weight markers (size in kDa) are indicated.

additional polypeptide of approximately 42 kDa is produced by plasmid pMS132 but not by the vector (Fig. 13) and it is apparently related to the 45 kDa polypeptide since both are selectively precipitated by specific antibodies to the human α-galactosidase A (see below).

The levels of the bla gene product are controlled by the low level, constitutive, bla promoter and by the regulatable trp promoter located just upstream. The majority of the bla gene product in plasmid ptrpL1 is directed by the more active trp promoter, (Edman et al., 1981, J. C. Edman, personal communication, Hantzopoulos and Calhoun, unpublished observations). Therefore, the bla gene product seen in Fig. 13 primarily reflects the translation of bi-cistronic mRNA molecules that were initiated at trpP and encode the human α-galactosidase A at the 5'-end and the bacterial β-lactamase at the 3'-end. The relative ratios of human α-galactosidase A to bacterial β-lactamase were variable (Fig. 13-16) but in this and other experiments (unpublished) the levels of β-lactamase (which has 10 methionine residues) were usually higher than that of α-galactosidase A (which has 15 methionine residues). This difference could be due to reduced stability of the 5'-end of the transcript, inefficient translation initiation or elongation for the human α-galactosidase A, or proteolytic degradation of the human α-galactosidase A polypeptide (see below).

Our rationale for constructing plasmid pMS132 from the initial construct, pMS122, was to determine if the distance separating the ribosome binding site and ATG might affect the efficiency of α-galactosidase A expression. We noted that the level of α-galactosidase A produced in maxicells containing plasmid pMS132 was about twice that of plasmid pMS122 (Fig. 14).

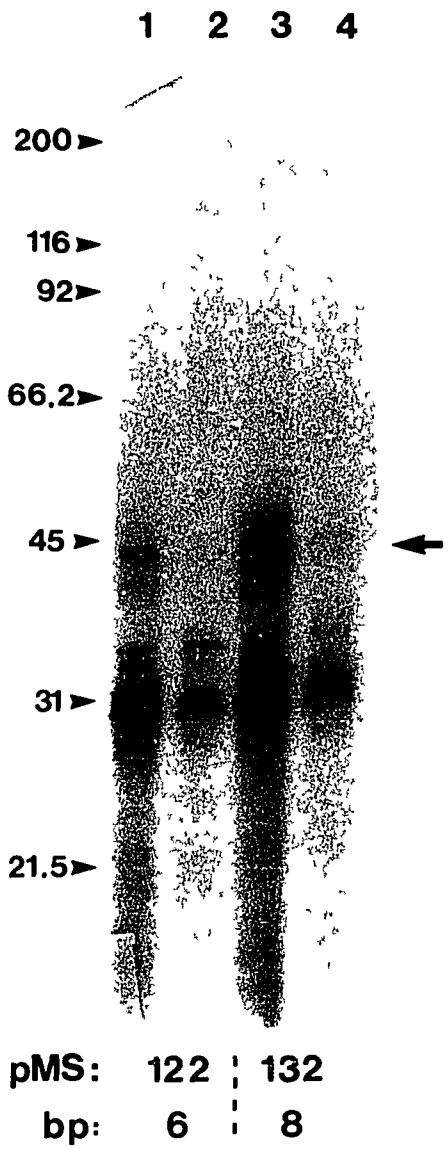

Figure 14

Comparison of human α-galactosidase A levels in plasmids pMS122 and pMS132. Plasmids pMS122 (lanes 1 and 2, and pMS132 (lanes 3 and 4), were examined in maxicells treated with tryptophan (lanes 2 and 4) or with 3-β-indoleacrylic acid (lanes 1 and 3).

Stability of α-galactosidase A in Escherichia coli

In order to monitor the stability of the human α-galactosidase A in E. coli, irradiated maxicells were incubated with [^{35}S]methionine for 1 min, excess unlabelled methionine was added, and samples were taken at intervals of 0, 10, 20, 30, 60 min. It can be seen (Fig. 15) that the human enzyme shows very little degradation during the 60 min interval and its stability is comparable to that of the vector coded β-lactamase polypeptide. We find no evidence for a precursor to product relationship between the 45 kDa and 42 kDa polypeptides.

Authenticity of α-galactosidase A produced in Escherichia coli

In order to confirm the authenticity of the human α-galactosidase A synthesized in E. coli, we used specific polyclonal antiserum raised against the purified enzyme. The maxicell sample was first cleared by treatment with normal rabbit serum to eliminate non-specific precipitations. The precipitates obtained following addition to the cleared supernatant of normal rabbit serum or immune serum were then analyzed by polyacrylamide gel electrophoresis. The results reveal the specific precipitation of the 45 kDa polypeptide by the immune serum (Fig. 16, lane 3) but not by normal rabbit serum (Fig. 16, lane 4). It can also be seen (Fig. 16, lane 3) that additional polypeptides of a range of sizes between approximately 45 kDa and 42 kDa are present and selectively precipitated by immune serum. The origin of polydisperse forms of intermediate size seen in the immunoprecipitates (Fig. 16,

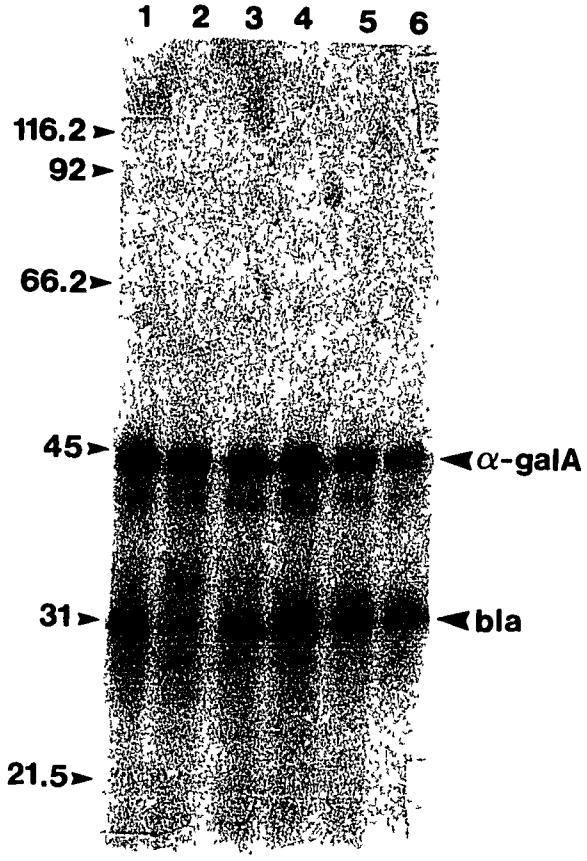

Figure 15

Stability of the human α-galactosidase A in E. coli. Radioactive methionine was added and incubated for one minute. Then an excess of non-radioactive methionine was added and samples were taken at 0, 10, 20, 30, and 60 min (lanes 2-6, respectively). The sample in lane 1 was incubated for 60 min with labelled methionine and the chase with unlabelled methionine was omitted.

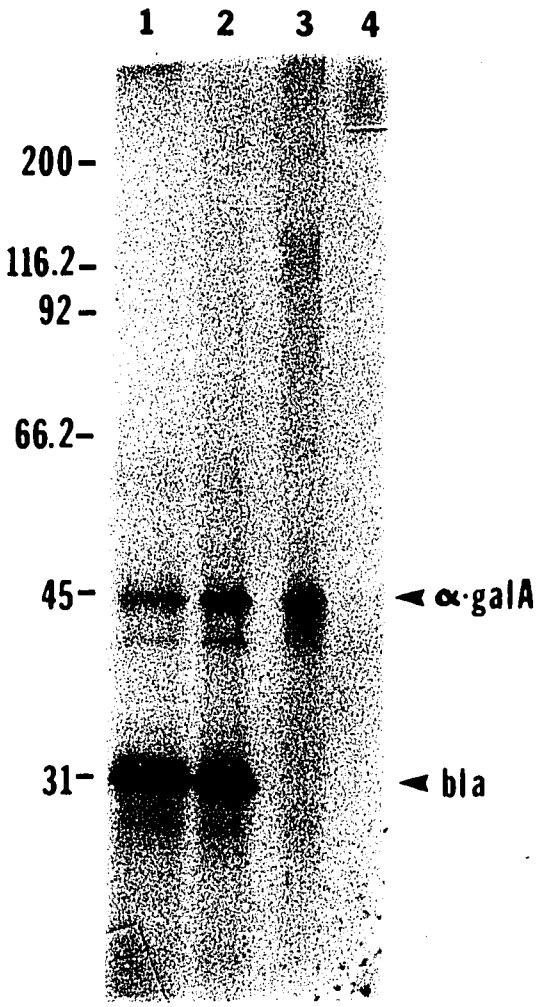

Figure 16

Immune precipitation of human α-galactosidase A synthesized in E. coli. The samples in lane 1 were loaded directly and those in lanes 2-4 were first treated with normal rabbit serum and centrifuged to remove non-specific precipitates. This supernatant was then loaded without the additional treatment (lane 2) or following precipitation using specific antibody to α-galactosidase A (lane 3) or normal rabbit serum (lane 4).

lane 3) is not certain. It is possible that the 45 kDa form, which corresponds in the size to the 45,356 dalton product predicted from the cDNA sequence, is subject to degradation during the immunoprecipitation procedures. It may or may not be fortuitous that the 42 kDa polypeptide that is routinely observed (Fig. 13-16) is approximately the same size as the 41,800 dalton mature enzyme (Calhoun et al., 1985) obtained by removal of the carboxy-terminal propeptide (Quinn et al., unpublished observations). The 42 kDa polypeptide could result from proteolytic degradation in E. coli of the 45 kDa enzyme, or it could result from site(s) of internal initiation or termination transcription, or translation.

Catalytic activity of the α-galactosidase A produced in E. coli

A major goal of these studies is to produce large quantities of the human α-galactosidase A in bacterial cells in order to facilitate studies aimed at testing the feasibility of enzyme replacement therapy in the treatment of patients with Fabry disease. We were obviously interested, therefore, in determining whether the α-galactosidase A produced in E. coli possesses catalytic activity. We also wished to develop a simple and reliable assay to identify or select bacterial colonies that may express the human enzyme in vivo in a catalytically active form to assist in the construction and modifications of various plasmid constructs. The human α-galactosidase A is known to be active with synthetic substrates (Table 7), including O-nitrophenyl-α-D-galactopyranoside (Kint, 1970) and 4-methylumbelliferyl-α-D-

TABLE 7

SUBSTRATES FOR HUMAN α-GALACTOSIDASE A

SUBSTRATE	COMMENTS
gal-α↓gal-β-glu-ceramide	natural substrate
O-nitrophenyl-α↓D-galactopyranoside	in vitro assay
gal-α↓CH_3	methyl-α-D-galactopyranoside
gal-α↓umbilliferose	fluorescent substrate
gal-α↓glu	melibiose
gal-α↓H	α-D-galactose
X-α↓Gal	5-Br-4-Cl-3-indolyl-α-D-galactopyranoside

KEY:

gal=galactose

glu=glucose

α=α bond

β=β bond

Br=bromo

Cl=Chloro

arrow indicates putative cleavage sites within substrate.

galactopyranoside (Kint, 1970), in addition to the natural substrate, globotriaosylceramide (Desnick and Sweeley, 1983). Therefore, we tested several α-galactosides as potential sources of carbon and energy in minimal salts medium in an effort to make the growth of the cells dependent upon the catalytic activity of the human enzyme. This would also provide a sensitive and convenient method to identify and select derivatives that may express higher levels of the human enzyme by monitoring for growth capacity and colony size. At the same time, we sought the custom synthesis of 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal), which is analogous to the widely used 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). When X-Gal is present in petri plates, colonies that produce the bacterial β-galactosidase turn blue, which provides a very sensitive assay for even very low levels of enzyme activity.

When methyl-α-D-galactopyranoside (Table 7), was present at 0.5% in minimal salts medium as the sole source of carbon and energy, we observed rapid growth (normal colony size in 1-2 days at 37°C) in strain CSR603 derivatives containing plasmids identified above that express the human α-galactosidase A, but no growth was present, even after prolonged incubation, using cells lacking plasmids or using cells with the ptrpL1 vector alone (Fig. 17). In addition, we have now identified a commercial source of X-α-Gal, and we observe a similar specific formation of blue colonies only in cells that express the human α-galactosidase A. We were not able to detect growth using other α-galactosides as potential carbon and energy sources, including the disaccharide, α-D-melibiose, or the trisaccharide, stachyose (Table 7). This result could be due to either permeability barriers that prevent the uptake of these compounds into cells, or the lack of

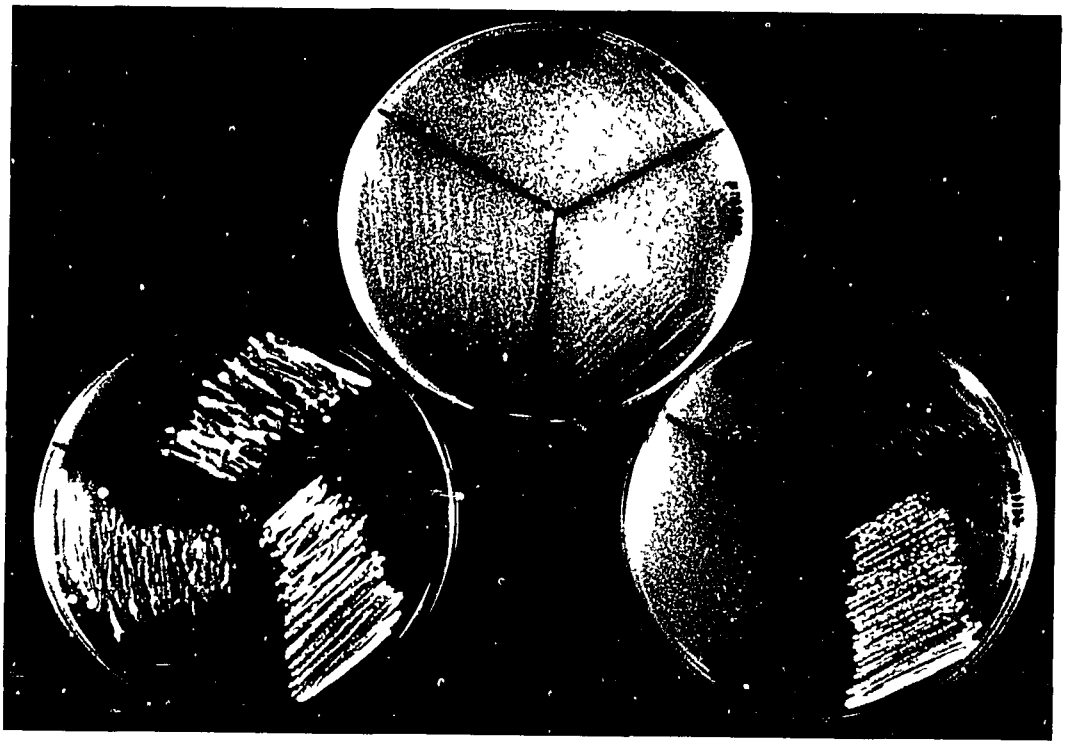

Figure 17

Catalytic activity of the human α-galactosidase A expressed in vivo in growing cells of E. coli K12. The three petri plates shown in the photo contain minimal salts medium with no carbon source (top), glucose (lower left), or α-methyl- galactoside (lower right) as carbon source. Each plate was divided into three sectors and inoculated on the top, right, and left sectors with strain CSR603 derivatives containing no plasmid, pMS122, and ptrpL1 with no insert, respectively.

activity of the human enzyme produced in E. coli with these compounds.

We have not yet quantitated the activity of α-galactosidase A in crude cell extracts and we have not yet attempted purification of this enzyme. In fact, in our preliminary experiments we detected little or no activity using O-nitrophenyl-α-D-galactopyranoside as a substrate in cells sonicated and tested in buffers that are optimal for the mature enzyme obtained from human sources. The straightforward evidence for enzyme activity in vivo (Fig. 17), however, suggests that some modifications of the protocol, e.g. cell disruption techniques or buffers, will permit stabilization of catalytic activity in cell extracts and enzyme purification.

Detection of the 5'-end sequences of α-galactosidase A mRNA by primer extension

To analyze the structure of the 5'-end of the α-galactosidase A mRNA, we hybridized a 131-nucleotide primer, complementary to sequences encoding the open reading frame of the amino terminal portion of the α-galactosidase A (pre)peptide to mRNA preparation. Elongation towards the 5'-end of mRNA would generate a product that corresponds to a species containing both the α-galactosidase A leader and 5'-untranslated sequences of the mRNA. Primer extension generates a background pattern of products with a major band of approximately 277 nucleotides long and shorter products presumambly due to pauses associated with RNA structure (Fig. 18). This indicates the presence of approximately 146 nucleotides in the mRNA transcript upstream of the

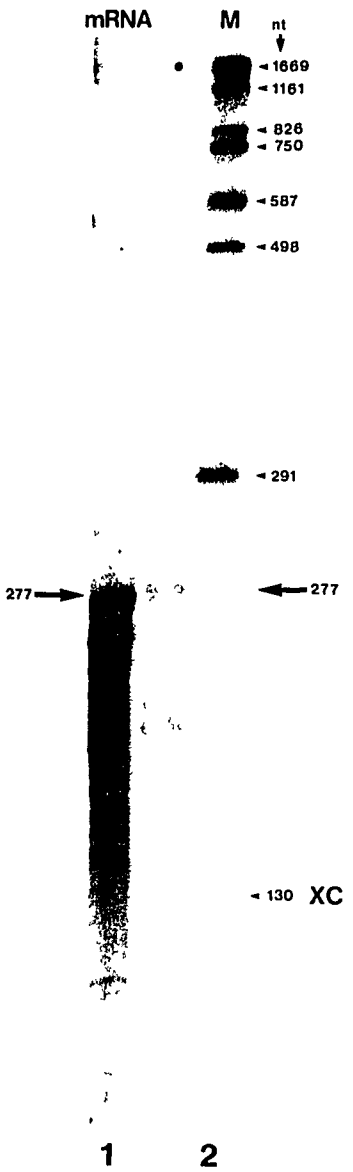

Figure 18

Identification of the 5'-end of α-galactosidase A mRNA by primer extension. Lane 1: mRNA sample was hybridized with a 131 nucleotides HgiAI-to-HgiAI DNA at 52°C and the primer was extended using reverse transcriptase. The extension products were examined on a 5% polyacrylamide/7M urea gel. Large arrow indicates the major run-off product corresponding to a nucleotide size of about 277. Lane 2: (M) size markers, XC: xylene cyanol (130 bp).

restriction fragment primer, which agrees with the analysis of a genomic clone containing this region (M. Quinn, unpublished).

Cloning of the leader sequences of the α-galactosidase A

Clone M13.L21, obtained from M. Quinn, (Fig. 19) packages the complementary strand of a about 365-nucleotide fragment, which contains 57 bp of 5'- untranslated sequences, the first exon (194 bp) and part of the first intron (104 bp). This clone is a deletion construct made for nucleotide sequencing and contains nucleotides 886 to 1240 of the parental genomic clone pMQ1 (Quinn et al., 1987). M13.L21 contains a unique NcoI site in the exon sequences of α-galactosidase A and a BamHI site in the polylinker region to the right of intron sequences (Fig. 19). Digestion of M13.L21 DNA with NcoI and BamHI restriction enzymes excised a 185 nucleotide fragment that contained 69 nucleotides of the exon sequences, 104 nucleotides intron sequences and a 12 nucleotides polylinker fragment. To clone the rest of the propeptide downstream of the leader coding sequences, plasmid pMS137 was used. Plasmid pMS137 contains the α-galactosidase A propeptide sequences shorter by 29 nucleotides (i.e. 10 amino acids) at the amino terminal. At the 3'-end of α-galactosidase A there is a set of BamHI linker added to α-galactosidase A 3'-end for the construction of plasmid pMS137. Double digestion of pMS137 with NcoI and BamHI liberates a 1191 bp fragment containing all the propeptide sequences sequences except the first 10 amino acids.

The two digested plasmids, M13.L21 and pMS137, were ligated together and the resultant recombinant clone M13.L21.1 contained an

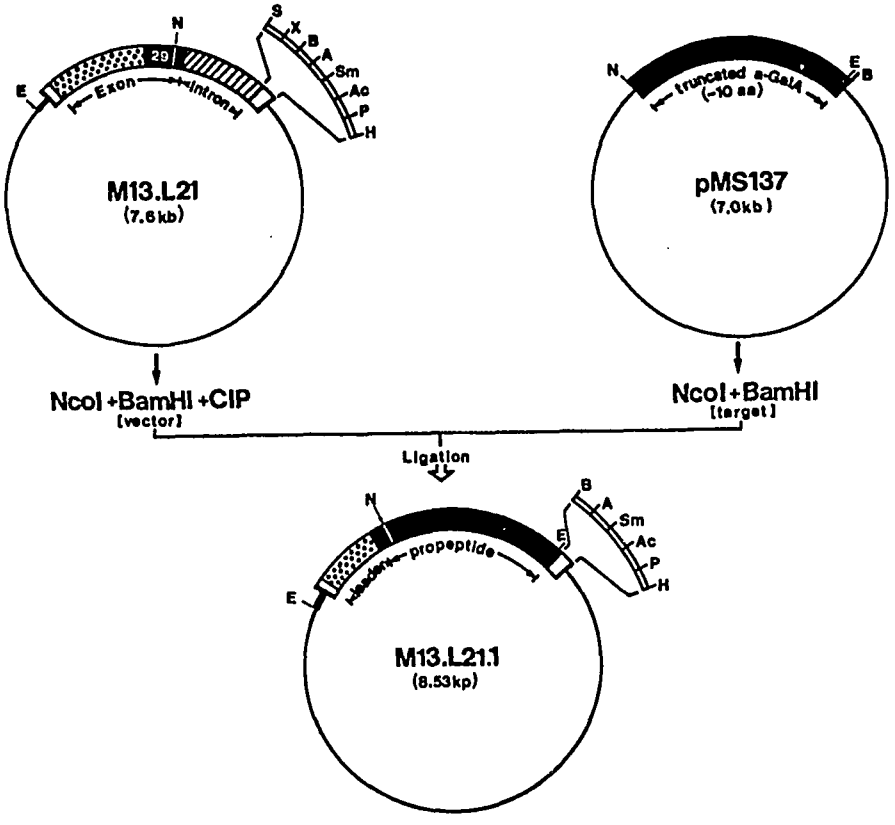

Figure 19

Strategy for cloning the leader peptide sequences upstream from the human α-galactosidase A propeptide in M13mp8. Dotted area represents coding sequences for leader peptide. Black areas represent sequences of α-galactosidase A propeptide. Hatched areas represent intron sequences. Empty box to the left of leader sequences represent 5' untranslated sequences. Empty box to the right of intron and propeptide represents a polylinker fragment containing various restriction enzyme sites (A:AvaI, Ac:AccI, B:BamHI, E:EcoRI, H:HindIII, N:NcoI, P:PstI, S:SacI, Sm:SmaI, X:XbaI).

insert fragment of about 1.4 kb, as judged by double digestion with EcoRI and BamHI restriction enzymes (Fig. 19). The 1.4 kb fragment contains the leader sequences cloned upstream of the propeptide coding sequences. Clone M13.L21.1 was stored for further analysis and expression of the α-galactosidase A with the leader peptide.

Fig. 20 shows the entire amino acid sequence of the human α-galactosidase A protein including the leader peptide. The entire length of the pre-propeptide is 429 amino acids. The length of the leader peptide is 31 amino acids and has an amino-terminal component, n (amino acids 1-4), a hydrophobic region, h (amino acids 6-26) and a carboxy-terminal region, c (amino acids 27-31) as it was predicted by the von Heijne rules (von Heinje, 1985).

```
ATG CAG CTG AGG AAC CCA GAA CTA CAT CTG GGC TGC GCG CTT GCG CTT CGC TTC CTG GCC   60
Met Gln Leu Arg Asn Pro Glu Leu His Leu Gly Cys Ala Leu Ala Leu Arg Phe Leu Ala   20

CTC GTT TCC TGG GAC ATC CCT GGG GCT AGA GCA CTG GAC AAT GGA TTG GCA AGG ACG CCT  120
Leu Val Ser Trp Asp Ile Pro Gly Ala Arg Ala Leu Asp Asn Gly Leu Ala Arg Thr Pro   40

ACC ATG GGC TGG CTG CAC TGG GAG CGC TTC ATG TGC AAC CTT GAC TGC CAG GAA GAG CCA  180
Thr Met Gly Trp Leu His Trp Glu Arg Phe Met Cys Asn Leu Asp Cys Gln Glu Glu Pro   60

GAT TCC TGC ATC AGT GAG AAG CTC TTC ATG GAG ATG GCA GAG CTC ATG GTC TCA GAA GGC  240
Asp Ser Cys Ile Ser Glu Lys Leu Phe Met Glu Met Ala Glu Leu Met Val Ser Glu Gly   80

TGG AAG GAT GCA GGT TAT GAG TAC CTC TGC ATT GAT GAC TGT TGG ATG GCT CCC CAA AGA  300
Trp Lys Asp Ala Gly Tyr Glu Tyr Leu Cys Ile Asp Asp Cys Trp Met Ala Pro Gln Arg  100

GAT TCA GAA GGC AGA CTT CAG GCA GAC CCT CAG CGC TTT CCT CAT GGG ATT CGC CAG CTA  360
Asp Ser Glu Gly Arg Leu Gln Ala Asp Pro Gln Arg Phe Pro His Gly Ile Arg Gln Leu  120

GCT AAT TAT GTT CAC AGC AAA GGA CTG AAG CTA GGG ATT TAT GCA GAT GTT GGA AAT AAA  420
Ala Asn Tyr Val His Ser Lys Gly Leu Lys Leu Gly Ile Tyr Ala Asp Val Gly Asn Lys  140

ACC TGC GCA GGC TTC CCT GGG AGT TTT GGA TAC TAC GAC ATT GAT GCC CAG ACC TTT GCT  480
Thr Cys Ala Gly Phe Pro Gly Ser Phe Gly Tyr Tyr Asp Ile Asp Ala Gln Thr Phe Ala  160

GAC TGG GGA GTA GAT CTG CTA AAA TTT GAT GGT TGT TAC TGT GAC AGT TTG GAA AAT TTG  540
Asp Trp Gly Val Asp Leu Leu Lys Phe Asp Gly Cys Tyr Cys Asp Ser Leu Glu Asn Leu  180

GCA GAT GGT TAT AAG CAC ATG TCC TTG GCC CTG AAT AGG ACT GGC AGA AGC ATT GTG TAC  600
Ala Asp Gly Tyr Lys His Met Ser Leu Ala Leu Asn Arg Thr Gly Arg Ser Ile Val Tyr  200

TCC TGT GAG TGG CCT CTT TAT ATG TGG CCC TTT CAA AAG CCC AAT TAT ACA GAA ATC CGA  660
Ser Cys Glu Trp Pro Leu Tyr Met Trp Pro Phe Gln Lys Pro Asn Tyr Thr Glu Ile Arg  220

CAG TAC TGC AAT CAC TGG CGA AAT TTT GCT GAC ATT GAT GAT TCC TGG AAA AGT ATA AAG  720
Gln Tyr Cys Asn His Trp Arg Asn Phe Ala Asp Ile Asp Asp Ser Trp Lys Ser Ile Lys  240

AGT ATC TTG GAC TGG ACA TCT TTT AAC CAG GAG AGA ATT GTT GAT GTT GCT GGA CCA GGG  780
Ser Ile Leu Asp Trp Thr Ser Phe Asn Gln Glu Arg Ile Val Asp Val Ala Gly Pro Gly  260

GGT TGG AAT GAC CCA GAT ATG TTA GTG ATT GGC AAC TTT GGC CTC AGC TGG AAT CAG CAA  840
Gly Trp Asn Asp Pro Asp Met Leu Val Ile Gly Asn Phe Gly Leu Ser Trp Asn Gln Gln  280

GTA ACT CAG ATG GCC CTC TGG GCT ATC ATG GCT GCT CCT TTA TTC ATG TCT AAT GAC CTC  900
Val Thr Gln Met Ala Leu Trp Ala Ile Met Ala Ala Pro Leu Phe Met Ser Asn Asp Leu  300

CGA CAC ATC AGC CCT CAA GCC AAA GCT CTC CTT CAG GAT AAG GAC GTA ATT GCC ATC AAT  960
Arg His Ile Ser Pro Gln Ala Lys Ala Leu Leu Gln Asp Lys Asp Val Ile Ala Ile Asn  320

CAG GAC CCC TTG GGC AAG CAA GGG TAC CAG CTT AGA CAG GGA GAC AAC TTT GAA GTG TGG 1020
Gln Asp Pro Leu Gly Lys Gln Gly Tyr Gln Leu Arg Gln Gly Asp Asn Phe Glu Val Trp  340

GAA CGA CCT CTC TCA GGC TTA GCC TGG GCT GTA GCT ATG ATA AAC CGG CAG GAG ATT GGT 1080
Glu Arg Pro Leu Ser Gly Leu Ala Trp Ala Val Ala Met Ile Asn Arg Gln Glu Ile Gly  360

GGA CCT CGC TCT TAT ACC ATC GCA GTT GCT TCC CTG GGT AAA GGA GTG GCC TGT AAT CCT 1140
Gly Pro Arg Ser Tyr Thr Ile Ala Val Ala Ser Leu Gly Lys Gly Val Ala Cys Asn Pro  380

GCC TGC TTC ATC ACA CAG CTC CTC CCT GTG AAA AGG AAG CTA GGG TTC TAT GAA TGG ACT 1200
Ala Cys Phe Ile Thr Gln Leu Leu Pro Val Lys Arg Lys Leu Gly Phe Tyr Glu Trp Thr  400

TCA AGG TTA AGA AGT CAC ATA AAT CCC ACA GGC ACT GTT TTG CTT CAG CTA GAA AAT ACA 1260
Ser Arg Leu Arg Ser His Ile Asn Pro Thr Gly Thr Val Leu Leu Gln Leu Glu Asn Thr  420

ATG CAG ATG TCA TTA AAA GAC TTA CTT TAA  1290
Met Gln Met Ser Leu Lys Asp Leu Leu +++   429
```

Figure 20

The entire amino acid sequence of the human α-galactosidase A protein including the leader sequences, written below the open reading frame of the cDNA sequence. Leader nucleotide sequence was kindly provided by M. Quinn (Quinn and Calhoun, 1987), in order to reconstruct the α-galactosidase A sequences with leader peptide.

DISCUSSION

A highly purified α-galactosidase A preparation was used to obtain accurate amino acid sequence information for the amino terminal region, as well as for internal cyanogen bromide and tryptic peptides. Knowledge of partial amino acid sequence assisted us to design oligonucleotide mixtures corresponding to known amino acid sequence. We initially screened three independent human cDNA libraries (Orkin fetal and adult liver libraries and Okayama and Berg fibroblast cDNA library) constructed in pBR322 derivative vectors (pKT218 and pcD vectors respectively, Table 3) with probes 2A, 2B, 1C and 5 (Fig. 1) by the colony hybridization method (Hanahan and Meselson, 1983). Several clones that bound to these probes were isolated and characterized. One clone, pPH15, was subjected to nucleotide sequence analysis, but this sequence was not compatible with the amino acid sequence of α-galactosidase A. However, this clone did contain nucleotide sequences corresponding to the synthetic probe 1C in 17 of a total of 20 nucleotide positions.

The clone containing α-galactosidase A cDNA was first detected as an antibody positive clone (Calhoun et al., 1985) after multiple screenings of the λgt11 expression cDNA library. After the initial screening and selection of antibody-positive clones, authenticity of α-galactosidase A cDNA clone was established by competition experiments with antiserum that was absorbed with an excess of homogeneous α-galactosidase A (D. Bishop's laboratory). These experiments demonstrated the α-galactosidase A specificity of the fusion protein

expressed by the positive clone λAG18. In addition, we further characterized the antibody positive clone λAG18 with oligonucleotide mixtures corresponding to both amino-terminal and internal peptide amino acid sequences. Purified insert DNA on Southern blots strongly hybridized to three different synthetic oligonucleotide mixtures. In addition, the predicted amino acid sequences were in agreement with the known amino acid sequence.

Electrophoretic analysis of the mature α-galactosidase A before and after glycanase treatment (Calhoun et al., 1985), indicated that it contains approximately 370 amino acid residues and approximately 15% carbohydrate. We concluded that clone λAG18 was of sufficient length to include the entire coding sequence of the mature enzyme. The subsequent subcloning and nucleotide sequence determination of the α-galactosidase A cDNA insert, revealed an open reading frame of 1209 nucleotides encoding the entire mature form of this lysosomal enzyme (Fig. 6 and 7). The demonstration of colinearity between the nucleotide sequence and the 86 nonoverlapping amino acid residues determined by microsequencing in Dr. L. Hood's laboratory identified that the amino-terminal codon for the mature enzyme was consistent with the reading frame throughout, and confirmed the authenticity of this clone. Poly(A)$^+$ RNA hybridization experiments revealed that the α-galactosidase A enzyme may be coded by a message of approximately 1.45 kb (Bishop et al., 1986). In previous immunologic studies of human α-galactosidase A biosynthesis, it had been shown that a glycosylated propeptide of a molecular weight of approximately 55-58 kDa underwent a proteolytic cleavage to the mature lysosomal form of approximately 50 kDa (LeDonne et al., 1983). The 1234 bp of α-galactosidase A cDNA insert encodes 403 amino acids (i.e. 398 of the

mature subunit and the last 5 of the leader peptide). Therefore the additional 225 nucleotides in the 1.45 kbp α-galactosidase A mRNA could code for the additional signal peptide as well as sequences for 5'-untranslated region and poly(A) tail. Examination of both the nucleotide sequence of the cDNA insert for α-galactosidase A (pAG18 clone) (Fig. 1) revealed that oligonucleotides 1C (20-mer), and 2B (14-mer) had a perfect match, and oligonucleotide 5 (47-mer) had a 43 out of 47 nucleotides match (assuming G/T wobble) and these probes do hybridize to to the α-galactosidase A cDNA at the temperature range used for library screening (Table 2). The reasons for our failure to detect an α-galactosidase A clone in our screenings of cDNA libraries using oligonucleotides are not known. Possibly, the given aliquot of these libraries did not contain a full length cDNA for α-galactosidase A. Indeed, α-galactosidase A constitutes approximately 0.002% of total cellular protein in human lung and it turns over very slowly in the tissues.

Examination of the predicted amino acid sequence revealed that there are four possible N-glycosylation sites (Fig. 7). This observation was consistent with previous studies of purified α-galactosidase A from human plasma and spleen (Bishop and Desnick, 1981), and immunoprecipitated enzyme from Chang liver cells (LeDonne et al., 1983), which indicated the presence of tri- and tetraantennary complex and high mannose-type oligosaccharides.

The α-galactosidase A cDNA sequence (Fig. 6), has two consensus sequences, ATTAAA and AATACA, for cleavage at the poly(A) site, which are 11 and 28 nucleotides upstream from the UAA stop codon, respectively. The ATTAAA signal is present in approximately 12% of

vertebrate messages, whereas the AATACA occurs in only 2% (Wickens and Stephenson, 1984). An unusual feature of the α-galactosidase A mRNA is the absence of a nucleotide segment that separates the UAA stop codon from the poly(A) tail. The lack of 3'- untranslated nucleotides has been observed previously only in human and bovine mitochondrial mRNAs (Anderson et al., 1981, Ojala et al., 1981, Anderson et al., 1982). A search of the data bases identified 256 genes in which the locations of the stop codon and poly(A) addition site are known. None of these had contiguous or overlapping bases of the stop codon and poly(A) segment. In the α-galactosidase A mRNA, at least one, or possibly both of the adenosine residues of the UAA termination codon could be coded by the template or added post-transcriptionally by poly(A) polymerase. In order to verify the absence of a 3'- untranslated sequence in α-galactosidase A mRNA, a second α-galactosidase A cDNA clone isolated from an independently constructed library was sequenced and found to contain an identical 3'- sequence, including the absence of a 3'- untranslated region between the termination codon and a 120 nucleotide poly(A) tail (Bishop et al., 1986). One could not eliminate the remote possibility that an identical cloning error may occur either during reverse transcription or during replication in E. coli. The human α-galactosidase A mRNA was the first example of a nuclear encoded mRNA species that lacks a 3'- untranslated region. More recently, another report of a transcript for the mouse thymidylate synthase lacking a 3'- untranslated region has appeared (Jehn et al., 1986). It appears, therefore, that a 3'- untranslated region is not essential for the accumulation or translation of these mRNAs.

Searches of amino acid and nucleotide sequence data bases identified only a few sequences of limited homology to α-galactosidase

A. Of the sequences available for mammalian lysosomal enzyme cDNAs, only α-L-fucosidase (Fukushima et al., 1985) had any significant homology.

We have used a well characterized expression system (Edman et al., 1981, Halewell et al., 1980) that utilizes the trp promoter cloned into the expression vector ptrpL1. For the expression in E. coli of the cDNA clones specific for the human α-galactosidase A propeptide we engineered the 5'-terminus of the cDNA insert, in order to insert an ATG initiation codon preceding the first codon of the mature form of the enzyme. Clones with α-galactosidase A specific cDNA insert encoding the propeptide were constructed (Fig. 11) and produced a protein of 45 kDa exactly the size expected for the intact proenzyme at measurable levels (Fig. 13). We demonstrated the authenticity of α-galactosidase A polypeptide made in bacteria by immunoprecipitation with polyclonal antibodies raised against the native mature glycosylated α-galactosidase A human protein (Fig. 16). At least a two-fold increase in expression of α-galactosidase A was observed when the distance between the Shine-Dalgarno sequence and initiator codon ATG was increased to 8 bp (Fig. 14). This observation was consistent with the 8 to 9 bp optimal spacing required between the Shine-Dalgarno and initiator codon for efficient expression (Itoh et al., 1984).

The human enzyme is produced in a catalytically active form at levels that can support the growth of cells using α-galactosides as the only source of carbon and energy. In addition to natural and synthetic substrates we identified a chemical substrate, 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal) (Table 7), which turns blue as

an indicator by cells harboring α-galactosidase A recombinant plasmids. It is not likely that the absence of glycosylation of the enzyme produced in E. coli significantly affects the catalytic activity, since enzymatic de-glycosylation in vitro of the enzyme obtained from humans does not reduce its enzyme activity (D. Bishop, unpublished observations).

Some exogenous proteins form intracellular granules composed of denatured polypeptides with aberrant disulfide linkages, and human proteins with multiple disulfide bonds are less likely to fold properly in E. coli. In those circumstances, denaturation and gentle renaturation has sometimes been shown to restore enzyme activity (Hager and Burgess, 1980). We have not yet tested for the presence of such granules in bacterial cells expressing α-galactosidase A, and we have not yet attempted to test for a possible increased activity using a renaturation protocol.

It is not likely that the vectors described in this study produce the human α-galactosidase A at levels that constitute an inordinately high fraction (e.g. 10% to 50%) of the total cell protein as has sometimes been reported in other systems (e.g. Rosenberg et al., 1983). For example, in log phase cells grown with or without indoleacrylic acid we do not detect a significant accumulation of polypeptides at 45 kDa or 42 kDa in Coomassie blue stained gels or in autoradiograms. However, owing to the availability of facile methods cited above for the identification and selection of high level expressors, straightforward methods exist to construct and identify suitable derivatives of ptrpL1 or other expression vectors. We have constructed derivatives of pUC9 (Messing and Vieira, 1982), pAS1 (Rosenberg et al., 1983, Gross et al., 1985) and pOTSNco12 (Devare et al., 1984) that will

be tested for α-galactosidase A expression. In addition we have begun the construction of derivatives of the yeast vector, YEp52 (Broach et al., 1983) and of baculovirus (Smith et al., 1983, Smith et al., 1985; Kuroda et al., 1986).

We have considered several factors that may influence the selection of the appropriate expression system for the human α-galactosidase A. The enzyme is glycosylated (LeDonne et al., 1983) and four possible sites (Asn-Xaa-Ser/Thr) for asparagine linked oligosaccharide addition were identified in our cDNA clone (Bishop et al., 1986). The enzyme produced in E. coli will not be glycosylated in vivo, and this could affect molecular folding, catalytic activity, stability, or specificity of uptake at the mammalian cell surface. For these reasons, a eukaryotic system, such as yeast or baculovirus, would offer advantages, even though the level of expression and ease of scale up would not be expected to match that obtainable in prokaryotic systems.

Another major consideration is the absence of a signal peptide in the cDNA clones tested here. Our original cDNA clone (Calhoun et al., 1985) used for the constructions described here contained sequences coding for only five residues of what we now know to be (Quinn et al., 1987) a 31 amino acid signal peptide. It is possible that a specific folding pathway exists during co-translational cleavage of the signal peptide and entry of the nascent polypeptide in a vectorial fashion into the lumen of the rough endoplasmid reticulum. The folding pathway in E. coli followed by the clones described here may differ from the naturally occuring pathway, and this could affect enzyme stability, activity, or even translational efficiency. We have now subcloned the coding region for the complete signal peptide to our

expression vectors to constuct clones that encode the complete prepeptide. It remains to be determined, how this affects expression in E. coli, and it is not known if, in yeast and baculovirus, the signal peptide will route significant fractions of the human α-galactosidase A to the lysosome or secretory pathways.

Finally, we can envision several potential difficulties associated with the use of this experimental approach for the development of enzyme replacement therapy in Fabry disease, including specificity of uptake, unanticipated toxicities, immune responses, and others. And yet, this pursuit seems justified in light of the absence of an adequate therapy for this inborn error of metabolism.

A variety of experimental approaches directed at the application of molecular techniques to Fabry disease depended upon the isolation of specific genomic and/or cDNA clones specific for α-galactosidase A. Accordingly, the efforts of Hantzopoulos, and Quinn in Dr. Calhoun's laboratory were initially combined with those of H. Bernstein and D. Bishop in Dr. Desnick's laboratory in order to isolate such a clone. The succesful results of these combined efforts are described in jointly authored manuscripts (Calhoun et al., 1985, Bishop et al., 1986). Subsequently, independent projects were undertaken specifically by Hantzopoulos, Quinn and Bernstein that focused on expression of α-galactosidase A in bacteria, isolation of genomic clones containing the promoter, and analysis of the molecular defects in Fabry patient fibroblast cell lines, respectively. This thesis attempts to discriminate between jointly undertaken projects in the first phase of the study, and individual projects undertaken in the second phase of

the study.

BIBLIOGRAPHY

Anderson, S., Bankier, 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., Scheier, P. H., Smith, A. J. H., Staden, K. and Young, I. G. 1981. Sequence and organization of the human mitochondrial genome. Nature (London) 290: 457-465.

Anderson, S., DeBruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F., and Young, I. G. 1982. Complete sequence of bovine mitochondrial DNA. J. Mol. Biol. 156: 683-717.

Anderson, W. 1898. A case of angiokeratoma. Brit. J. Derm. 10: 113.

Appleyard, R. K. 1954. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from Escherichia coli K12. Genetics 39: 440-450.

Aubert, J-P., Biserte, G., and Loucheus-Lefebvre, M-H. 1976. Carbohydrate-peptide linkage in glycoproteins. Arch. Biochem. Biophys. 175: 410-418.

Becker, G. W., and Hsiung, H. M. 1986. Expression, secretion and folding of human growth hormone in Escherichia coli. FEBS 204: 145-150.

Bennetzen, J. L. and Hall, B. D. 1982. Codon selection in yeast. J. Biol. Chem. 257, 3026-3031.

Benoist, C. and Chambon, P. 1981. In vivo sequence requirements of the SV40 early promoter region. Nature 290: 304-310.

Benton, W. D., and Davis, R. W. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196: 180-182.

Beucage, S. L., and Caruthers, M. H. 1981. Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. 22: 1859-1862.

Birnboim, H. C., and Doly, J. 1979. A rapid extraction procedure for screening recombinant plasmid DNA. Nucl. Acid Res. 7:1513-1520.

Birnstiel, M. L., Busslinger, M. and Strub, K. 1985. Transcription termination and 3' processing: the end is in site. Cell 41: 349-359.

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 clone encoding the mature enzyme. Proc. Natl. Acad. Sci. USA 83: 4859-4863.

Bishop, D. F., and Desnick, R. J. 1981. Affinity purification of

α-galactosidase A from human spleen, placenta, and plasma with elimination of pyrogen contamination. J. Biol. Chem. 256: 1307-1316.

Blobel, G., Walter, P., Chang, C. N., Erickson, A. H., and Lingappa, V. R. 1979. Symp. Soc. Exp. Biol. 33: 9-36.

Blobel, G. 1980. Intacellular protein topogenesis. Proc. Natl. Acad. Sci. (USA) 77: 1496-1500.

Bolivar, F., Rondriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2: 95-119.

Boyer, H. W., and Roulland-Dussoix, D. 1969. A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. 41: 459-471.

Brady, R. O., Gal, A. E., Bradley, R. M., and Martensson, E. 1967. The Metabolism of Ceramide Trihexosides. I. Purification and properties of an enzyme that cleaves the terminal galactose molecule of galactosylgalactosylglucosylceramide. J. Biol. Chem. 242: 1021-1026.

Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. L., and Laster, L. 1967. Enzymatic defects in Fabry's disease: ceramide trihexosidase deficiency. New Eng. J. Med. 276: 1163.

Brady, R. O., Tallman, J. F., Johnson, W. G., Gal, A. E., Leahy, W. E., Quirk, J. M., and Dekaban, A. S. 1973. Replacement therapy for inherited enzyme deficiency. N. Eng. J. Med. 289: 9-14.

Broach, J. R., Li, Y., -Y., Wu Chen L., -C., and Jayaram, M. 1983. Vectors for high level inducible expression of cloned genes in yeast. In 'Experimental Manipulation of Gene Expression', Inouye M., (Ed.), Acad. Press, New York, pp. 84-116.

Broome, S. and Gilbert, W. 1985. Rous sarcoma virus encodes a transcriptional activator. Cell 40: 537-546.

Brown, N. L., McClelland, M., and Whitehead, R. P. 1980. HgiAI: A restriction endonuclease from Herpetosiphon giganeus HP1023. Gene 9: 49-68.

Browne, C. A., Bennett, H. P. J., and Solomon, S. 1982. The isolation of peptides by high-performance liquid chromatography using predicted elution positions. Anal. Biochem. 124: 201-208.

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. Proc. Natl. Acad. Sci. USA 82: 7364-7368.

Calhoun, D. H. and Gray, J. E. 1981. Detection of proteins coded by cloned DNA segments. Focus (BRL) 3: 1.

Catteral, J. F., and Leary, S. L. 1983. Detection of early changes in adrogen renal β-glucuronidase messenger ribonucleic acid using cloned complementary deoxyribonucleic acid. Biochemistry 22: 6049-6053.

Chen, E., Y., and Seeburg, P. H. 1985. Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4: 165-170.

Chou, P. Y. and Fasman, G. D. 1978. Empirical predictions of protein conformation. Ann. Rev. Biochem. 47: 251-276.

Curtiss, R. III, Inoue, M., Pereira, D., Hsu J. C., Alexander L., and Rock L. 1977. Construction in use of safer bacterial host strains for recombinant DNA research. In 'Molecular cloning of recombinant DNA' (ed. Scott W. A. and Werner, R.), pp. 248-272, Academic Press, New York.

Dale, R. M. K., McClure, B. A., and Houchins, J. P. 1985. A Rapid Single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing for corn mitochondrial 18S DNA. Plasmid 13: 31-40.

Davies, K. E., Young, B. D., Elles, R. G., Hill, M. E., and Williamson, R. 1981. Cloning of a representative genomic library of the human X chromosome after sorting by flow cytometry. Nature 293: 374-379.

de Boer, H. A., Comstock, L. J., and Vasser, M. 1983. The tac promoter: A functional hybrid derived from the trp and lac promoters. Proc. Natl. Acad. Sci. USA 80: 21-25.

Dean, K. J., and Sweeley, C. C. 1979. Studies on human liver α-galactosidase. I. Purification of α-galactosidase A and its enzymatic properties with glycolipid and oligosaccharide substrate J. Biol. Chem. 254: 9994-10000.

Desnick, R., J., Dean, K., J., Grabowski, G., A., Bishop, D., F., and Sweeley, C., C. 1980. In 'Enzyme therapy in genetic diseases: 2', (R. J. Desnick, ed.), p. 393, Alan R. Liss, New York.

Desnick, R. J., Dean, K. J., Grabowski, G. A., Bishop, D. F. and Sweeley, C. C. 1979. Enzyme therapy in Fabry disease: differential in vivo plasma clearance and metabolic effectiveness of plasma and splenic α-galactosidase A isozymes. Proc. Natl. Acad. Sci. USA 76: 5326-5330.

Desnick, R. G., Allen, K. Y., Desnick, S. J., Raman, M. K., Bernlohr, R. W., and Krivit, W. 1973. Fabry's disease: enzymatic diagnosis of hemizygotes and heterogygotes. J. Lab. Clin. Med.:81, 157-171.

Desnick, R., J., and Sweeley, C., C. 1983. Fabry's disease: α-galactosidase A deficiency. In: 'The metabolic basis of inherited diseases', eds. Stanbury, J. B., Wyngaarden, J., B., Fredrickson, D. S., Goldstein, J., L., and Brown, M., S.

(MacGraw-Hill, New York), (5th Ed.), pp. 906-944.

Devare, S. G., Shatzman, A., Robbins, K. C., Rosenberg, M., and Aaronson, S. A. 1984. Expression of the PDGF-related transforming protein of simian sarcoma virus in Escherichia coli. Cell 36: 43-49.

de Wet, J. R., Fukushima, H., Dewiji, N. N., Wilcox, E., O'Brien, J. S., and Helinski, D. R. 1984. Chromogenic immunodetection of human serum albumin and α-L-fucosidase clones in a human hepatoma cDNA expression library. DNA 3: 437-447.

Edman, J. C., Hallewell, R. A., Valenzuela, P., Goodman, H. M. and Rutter, W. J. 1981. Synthesis of hepatitis B surface and core antigens in Escherichia coli. Nature 291: 503-506.

Emtage, J. S., Tacon, W. C. A., Catlin, G. H., Jenkins, B., Porter, A. G., and Carey, N. H. 1980. Influenza antigenic determinants are expressed from haemagglutinin genes cloned in E. coli. Nature 283:, 171-174

Erickson, A. H., and Blobel, G. 1983. Carboxyl-terminal proteolytic processing during biosynthesis of the lysosomal enzymes β-glucuronidase and cathepsin D. Biochem. 22: 5201-5205.

Fabry, J. 1898. Ein beitrag zur kenntnis der purpure haemorrhagica nodulae. (Purpure papulosa haemorrhagica herbae). Arch. Derm. Syph. 43: 183.

Faust, P. L., Kornfeld, S. and Chirgwin, J. M. 1985. Cloning and sequence analysis of cDNA for human cathepsin D. Proc. Natl. Acad. Sci. USA 82: 4910-4914.

Fitting, T., and Kabat, D. 1982. Evidence for a glycoprotein 'signal' involved in transport between subcellular organelles. J. Biol. Chem. 257: 14011-14017.

Fox, M. F., Ditoit, D. L., Warnick, L., and Retief, A. E. 1984. Regional localization of α-galactosidase (GLA) to Xpter-->q22, hexosaminidase B (HEXB) to 5q13-->qter, and asylsulfatase B (ARSB) to 5pter-->q13. Cytog. Cell Genet. 38:45-49.

Frantatoni, J. C., Hall, C. W., and Neufeld, E. F. 1968. Hurler and Hunter syndromes: mutual correction of the defect in cultured fibroblasts. Science 162: 570-573.

Fukushima, H., deWet, J., R., and O'Brien, J., S. 1985. Molecular cloning of a cDNA for human α-L-fucosidase. Proc. Natl. Acad. Sci. USA 82: 1262-1265.

Garnier J., Osguthorpe, D. J., and Robson, B. 1983. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97-120.

Gerald, P. S. and Brown, J. A. 1974. Report of the committee on the

genetic constitution of the X chromosome. Cytogenet. Cell Genet. 13: 29-34.

Ginns, E. I., Choudary, P. V., Martin, B. M., Winfield, S., Stubblefield, B., Mayor, J., Merkle-Lehman, D., Murray, G. J., Bowers, L. A., and Barranger, J. A. 1984. Isolation of cDNA clones for human β-glucocerebrosidase using the lambdagt11 expression system. Biochem. Biophys. Res. Comm. 123: 574-580.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A. Itakuta, K., and Riggs, A. D. 1979a. Expression in Escherichia coli of chemically synthesized genes for human insulin. Proc. Natl. Acad. Sci. (USA) 76: 106-110.

Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D. G., Ross, M. J., Miozzari, G., Crea, R., and Seeburg, P. H. 1979b. Nature 281: 544-548.

Goeddel, D. V., Shepard, H. M., Velverton, E., Leung, D., and Crea, R. 1980. Synthesis of human fibroblast interferon by E. coli. Nucl. Acids Res. 8: 4057-4074.

Gouy, M. and Gautier, C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucl. Acids Res. 10: 7055-7074.

Gray J. E., Wallen, J. W., and Calhoun D. H. 1982. Identification of a protein of 15,000 daltons related to isoleucine-valine biosynthesis in Esherichia coli K-12. J. Bacteriol. 151: 127-134.

Grantham, R., Gantier, C., Gouy, M., Jacobzone, M., and Mercier, R. 1981. Nucl. Acid Res. 9: r43-r73.

Gross M., Sweet, R. W., Sathe, G., Yokoyama, S., Fasano, O., Goldfarb, M., Wigler, M., and Rosenberg, M. 1985. Purification and characterization of human H-ras protein expressed in E. coli. Mol. Cell Biol. 5: 1015-1024.

Grzeschik, K., H., van Someren, H., Meera Khan, P., Weaterveld, A., and Bootsma, D. 1972. X-linkage of human α-galactosidase. Nature New Biol. 240: 48-51.

Guise, K. S., Korneluk, R. G., Waye, J., Lambonwah, A-M., Quan, F., Palmer, R., Ganschow, R. E., Sly, W. S., and Gravel, R. A. 1985. Isolation and expression in Escherichia coli of a cDNA clone encoding human β-glucuronidase. Gene 34: 105-110.

Hager, D. A., and Burgess, R. R. 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of Escherichia coli RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. Anal. Biochem. 109: 76-86.

Hallewell, R. A. and Emtage, S. 1980. Plasmid vectors containing the tryptophan operon promoter suitable for efficient regulated expression of foreign genes. Gene 9: 27-47.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166: 557-580.

Hanahan, D., and Meselson, M. 1983. Plasmid screening at high colony density. Methods Enzymol. 100: 333-342.

Hickman, S. and Neufeld, E., F. 1972. A hypothesis for I-cell disease: defective hydrolases that do not enter lysosomes. Biochem. Biophys. Res. Com. 49: 992-999.

Hieber, V. C. 1982. Cloning of a cDNA complementary to rat preputial gland β-glucuronidase mRNA. Biochem. Biophys. Res. Comm. 104: 1271-1278.

Howley, P. M., Sarver, N., and Law, M-F. 1983. Eukaryotic cloning vectors derived from bovine papillomavirus DNA. Methods Enzymol. 101: 387-402.

Hunkapiller, M. W., and Hood, L. E. 1983. Protein sequence analysis: automated microsequencing. Science 219: 650-659.

Huynh, T. V., Young, R. A., and Davis, R. W. 1984. Constructing and screening cDNA libraries in λgt10 and λgt11. In: DNA Cloning: A Practical Approach (D. Clover, ed.), IRL Press, Oxford.

Inouye, S., Hsu, C. P., Itakura, K. and Inouye, M. 1983. Requirements for signal peptide cleavage of Escherichia coli prolipoprotein. Science 221: 59-61.

Itakura, K., Rossi, J. J., and Wallace, R. B. 1984. Synthesis and use of synthetic oligonucleotides. Ann. Rev. Biochem. 53: 323-326.

Itoh, S., Mizukami, T., Matsumoto, T., Nishi, T., Saito, A., Oka, T., Furuya, A., Takaoka, C., and Tanigushi, T., 1984. Efficient expression in Escherichia coli of a mature and a modified human interferon β1. DNA 3: 157-165.

Jabbar, M., A., Sivasubramanian, N., and Nyak, D., P. 1985. Influenza viral (A/WSN/33) hemagglutinin is expressed and glycosylated in the yeast Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 82: 2019-2023.

Janowitz, Z. A., Henning, U., and Hollenberg, C. P. 1982. Synthesis of Escherichia coli outer membrane ompA protein in yeasts. Gene 20: 347-358.

Jay, G., Khoury, G., Seth, A. K., and Jay. E. 1981. Construction of a general vector for efficient expression of mammalian proteins in bacteria: Use of a synthetic ribosome binding site. Proc. Natl. Acad. Sci. USA 78: 5543-5548.

Jehn, G-H., Deng, T., Li, D., DeWille, J., and Johnson, L., F. 1986. Mouse thymidylate synthase messenger RNA lacks a 3' untranslated region. Proc. Natl. Acad. Sci. USA. 83: 8482-8486.

Johnson, D., L., and Desnick, R., J. 1978. Molecular pathology of Fabry's disease. Physical and kinetic properties of α-galactosidase in cultured human endothelial cells. Biochem. Biophys. Acta. 538: 195-204.

Kint, J. A. 1970. Fabry's Disease: α-galactosidase deficiency Science 167: 1268-1269.

Korman, A. J., Knudsen, P. J., Kaufmen, J. F., and Strominger, J. L. 1982. cDNA clones for the heavy chain HLA-DR antigens obtained after immunopurification of polysomes by monoclonal antibody. Proc. Natl. Acad. Sci. USA. 79: 1844-1848.

Kraus, J. P., and Rosenberg, L. E. 1982. Purification of low abundance messenger RNA from rat liver by polysome immunoadsorption. Proc. Natl. Acad. Sci. USA. 79: 4015-4019.

Kuroda, K., Hauser, C., Rott, R., Klenk, H-D., and Doerfler, W. 1986. Expression of the influenza virus haemagglutinin in insect cells by a baculovirus vector. EMBO Journal 5: 1359-1365.

Kyte, J., and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.

Leder, P., Tiemeier, D., and Enquist, L. 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: The λgtWES system. Science 196: 175-177.

LeDonne, N. C., Fairley, J. L., and Sweeley, C. C. 1983. Biosynthesis of α-galactosidase A in cultured Chang liver cells. Arch. Biochem. Biophys. 224: 186-195.

Lodish, H. F., Kong, N., Snider, M., and Strous, G. J. 1983. Hepatoma secretory proteins migrate from rough endoplasmic reticulum to Golgi at characteristic rates. Nature 304: 80-83.

Luck, D. N., Ngsee, J. K., Rottman, F. M., and Smith, M. 1986. Synthesis of bovine prolactin in E. coli. DNA 5: 21-28.

Mahuran, D. J., Tsui, F., Gravel, R., and Lowden, J. A. 1982. Evidence for two dissimilar polypeptide chains in the β2 subunit of hexosaminidase. Proc. Natl. Acad. Sci. USA. 79: 1602-1605.

Mandecki, W., Mollison, K. W., Bolling, T. J., Powell, B. S., Carter, G. W., and Fox, J. L. 1985. Chemical synthesis of a gene encoding the human complement fragment C5a and its expression in Escherichia coli. Proc. Natl. Acad. Sci. USA. 82: 3543-3547.

Maniatis, T., Fritch, E., F., and Sambrook, J. 1982. Molecular cloning: A laboratory manual, (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Mapes, C. A., Anderson, R. L., Sweeley, C. C., Desnick, R. J., and

Krivit, W. 1970. Enzyme replacement in Fabry's disease and inborn errors of metabolism. Science 169: 987-989.

Matteucci, M. D., and Caruthers, M. H. 1981. Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. 103: 3185-3190.

Maxam, A. M. and Gilbert, W. 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. 65: 499-560.

Messing, J., Crea, R., and Seeburg, P. H. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9: 309-325.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101: 20-78.

Messing, J. and Vieira, J. 1982. A new pair of M13 vector for selecting either DNA strand of double-digest restriction fragment. Gene 19: 269-276.

Michelson, A. M., Markhan, A. F., and Orkin, S. H. 1983. Isolation and DNA sequence of a full-length cDNA clone for human X chromosome encoded phosphoglycerate kinase. Proc. Natl. Acad. Sci. USA. 80: 472-476.

Miller, J. H. 1972. Experiments in molecular biology. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mohandas, T., Sparkes, R. S., Bishop, D. F., Desnick, R. J., and Shapiro, L. S. 1984. Frequency of reactivation and variability in expression of X-linked enzyme loci. Am. J. Hum. Genet. 36: 916-925.

Mott, J. E., Grant, R. A., Ho, Y-S., and Platt, T. 1985. Maximizing gene expression from plasmid vectors containing the lambda P_L promoter: strategies for overproducing transcription termination factor rho. Proc. Natl. Acad. Sci. USA. 82: 88-92.

Murray, N. E., Brammer, W. J., and Murray, K. 1977. Lamboid phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. 150: 53-75.

Myerowitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B., and Suzuki, K. 1985. Human β-hexosaminidase α-chain: coding sequence and homology with the β chain. Proc. Natl. Acad. Sci. USA. 82: 7830-7834.

Myerowitz, R., and Proia, R. 1984. cDNA clone for the α-chain of human β-hexosaminidase: deficiency of α-chain mRNA in Ashkenazi Tay-Sachs fibroblasts. Proc. Natl. Acad. Sci. USA. 81: 5394-5398.

Noyes, B. E., Mevarech, M., Stein, R., and Agarwal, K. 1979. Detection and partial sequence analysis of gastrin mRNA by using an oligonucleotide probe. Proc. Natl. Acad. Sci. USA 76: 1770-1774.

O'Dowd, B. F., Quan, F., Willard, H. F., Lamhonwah, A.-M., Korneluk, R. G., Lowden, J. A., Gravel, R. A., and Muhurna, D. J. 1985. Isolation of cDNA clones coding for the β subunit of human β-hexosaminidase. Proc. Natl. Acad. Sci. USA 82: 1184-1188.

Ojala, D., Montoya, J., and Attardi, G. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature (London) 290: 470-474.

Okayama, H., and Berg, P. 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell Biol. 3: 280-289.

Opitz. J. M., Stiles, F. C., Wise, D., Race, R. R., Sander, R., Von Gemmingen, G. R., Kierland, R. R., Cross, E. G., and de Groot, W. P. 1965. The genetics of angiokeratoma corporis diffusum (Fabry's disease) and its linkage relations with the Xg locus. Am. J. Hum. Genet. 17: 325-331.

Parnes, J. R., Velan, B., Felsenfeld, A., Ramanathan, L., Apella, E., and Seidman, J. G. 1981. Mouse β2 microglobulin cDNA clones: A screening procedure for cDNA clones corresponding to rare mRNAs. Proc. Natl. Acad. Sci. USA 78: 2253-2257.

Pearson, P. L., Sanger, R., and Brown, J. A. 1975. Report of the committee on the genetic constitution of the X-chromosome. Cytogenet. Cell Genetic. 14: 190-195.

Porter, M. T., Fluharty, A. L., and Kihara, H. 1971. Correction of abnormal cerebroside sulfate metabolism in cultured metachromatic leukodystrophy fibroblasts. Science 172: 1263-1265.

Proehownik, E. V., Markham, A. F., and Orkin, S. H. 1983. Isolation of a cDNA clone for human antithrombin III. J. Biol. Chem. 258: 8389-8394.

Quinn, M., Hantzopoulos, P., Fidanza, V., and Calhoun, D. H. 1987. Isolation of a genomic clone containing the promoter of a lysosomal enzyme gene (submitted).

Rebourcet, R., Weil, D., Van Gong, N., and Frezal, J. 1974. C. R. Acad. Sci. (Paris) 278: 3379-3396.

Ricciuti, R., and Ruddle, F. H. 1973. Assignement of three gene loci (PGK, HGPRT, G6PD) to the long arm of the human X chromosome by somatic cell genetics. Genetics 74: 661-678.

Rimm, D. L., Horness, D., Kucera, J., and Blattner, F. R. 1980. Construction of coliphage lambda charon vectors with BamHI cloning sites. Gene 12: 301-311.

Rose, J. K., and Yanofsky, C. 1974. Interaction of the operator of the tryptophan operon with repressor. Proc. Natl. Acad. Sci. USA 71: 3134-3138.

Rosenberg, M., Ho, Y.-S, and Shatzman, A. 1983. The use of pKC30 and its derivative for controlled expression of genes. Methods Enzymol. 101: 123-138.

Sancar, A., Hack, A. M., and Rupp, W. D. 1979. Simple method for identification of plasmid-coded protein. J. Bacteriol. 137: 692-693.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143: 161-178.

Sanger, F. 1981. Determination of nucleotide sequences in DNA. Science 214: 1205-1210.

Schmitt, J. J., and Cohen, B. N. 1983. Quantitative isolation of DNA restriction fragments from low-melting agarose by elutip-d affinity chromatography. Anal. Biochem. 133: 462-464.

Segundo, B. S., Chan, S. J., and Steiner, D. F. 1985. Identification of cDNA clones encoding a precursor of rat liver cathepsin B. Proc. Natl. Acad. Sci. 82: 2320-2324.

Shapiro, S. Z., and Young, J. R. 1981. An immunochemical method for mRNA purification. J. Biol. Chem. 256:1495-1498.

Shine, J., and Dalgarno, L. 1975. Detarminant of cistron specificity in bacterial ribosomes. Nature 254: 34-38.

Shows, T. B., Brown, J. A., Haley, L. L., Goggin, A. P., Eddy, R. L., and Byers, M. G. 1978. Assignement of α-galactosidase (α-Gal) to the q22—>pter region of the X chromosome in man. Cytogenet. Cell Genet. 22: 541-544.

Siniscalco, M., Szabo. P., Fillippi, G., and Rinaldi, A. 1982. Combination of old and new strategies for the molecular mapping of the human X-chromosome. In: Progress in Clinical and Biological Research. Alan R. Liss, Inc., New York, 103A: 103-124.

Sly, W. S., and Fischer, H. D. 1982. The phosphomannosyl recognition system for intracellular and intercellular transport of lysosomal enzymes. J. Cell. Biochem. 18: 67-85.

Smith, G. E., Ju, G., Ericson, B. L., Moschera, J., Lahm, H., Chizzonite, R., and Summers, M. 1985. Modification and secretion of human interleukin 2 produced in insect cells by a baculovirus expression vector. Proc. Natl. Acad. Sci. USA 82: 8404-8408.

Smith, G. E., Summers, M. D., and Fraser, M. J. 1983. Production of human β interferon in insect cells infected with a baculovirus expression vector. Mol. Cell. Biol. 3: 2156-2165.

Sorge, J., West, C., Westwood, B., and Beutler, E. 1985. Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA. Proc. Natl. Acad. Sci. USA 82: 7289-7293.

Southern, E. M. 1975. Detection of specific sequences among DNA fragnents separated by gel electrophoresis. J. Mol. Biol. 98: 503-515.

Stahl, P., Schlesinger, P. H., Sigardson, E., Rodman, J. S., and Lee, Y. C. 1980. Receptor-mediated pinocytosis of mannose glycoconjugates by macrophages: characterization and evidence for receptor recycling. Cell 19: 207-215.

Sweeley, C. C. and Klionsky, B. 1963. Fabry's disease: Classification as a sphingolipidosis and partial characterization of a novel glycolipid. J. Biol. Chem. 238: 3148-3150.

Szostak, J. W., Stiles, J. I., Tye, B.-K., Chiu, P., Sherman, F., and Wu, R. 1979. Hybridization with synthetic oligonucleotides. Methods Enzymol. 68: 419-428.

Talmadge, K., Stahl, S., and Gilbert, W. 1980a. Eukaryotic signal sequence transports insulin antigen in Escherichia coli. Proc. Natl. Acad. Sci. (USA) 77: 3369-3374.

Talmadge, K., Kaufman, J., and Gilbert, W. 1980b. Bacteria mature preproinsulin to proinsulin. Proc. Natl. Acad. Sci. USA 77:3988-3992.

Talmadge, K., and Gilbert, W. 1982. Cellular location affects protein stability in E. coli. Proc. Natl. Acad. Sci. USA 79: 1830-1833.

Vieira, J., and Messing, J. 1982. The pUC plasmids and M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259-268.

von Heijne, G. 1985. Signal sequences. The limits of variation. J. Mol. Biol. 184: 99-105.

Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. 1979. Hybridization of synthetic oligonucleotides to phiX174 DNA: the effect of sigle base pair mismatch. Nucl. Acid. Res. 6: 3543-3557.

Walter, P., Gilmore, R., and Blobel, G. 1984. Protein translocation across endoplasmic reticulum. Cell 38, 5-8.

Wickens, M. P., Puell, G. N., and Schimke, R. T. 1978. Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid, and ovalbumin mRNAs. J. Biol. Chem. 253: 2483-2495.

Wickens, M. and Stephenson, P. 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. Science 226: 1045-1051.

Wood, W. I., Capon, D. J., Simmonsen, C. C., Eaton, D. L., Gitschier,

J., Keyt, N., Seeburg, P. H., Smith, D. H., Hollingshead, P., Wion, K. L., Delward, E., Tuddenbaum, E. G. D., Vehar, G. A., and Lawn R. M. 1984. Expression of active human factor VIII from recombinant DNA clones. Nature 312: 330-337.

Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., vanCleemput, M., and Wu, A. M. 1981. The complete nucleotide sequence of the tryptophan operon of E. coli. Nucl. Acid. Res. 9: 6647-6668.

Yoakam, G. H., Yeung, A., Mattes, W., and Grossman, L. 1982. Amplification of the uvrA gene product of E. coli to 7% of cellular protein by linkage to the P_L Promoter of pKC30. Proc. Natl. Acad. Sci. USA 79: 1766-1770.

Young, R. A., and Davis, R. W. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science 222: 778-782.