Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 7101–7105, October 1985
Medical Sciences

Gene mapping and leader polypeptide sequence of human glucocerebrosidase: Implications for Gaucher disease

(lysosomal/chromosome 1q21/*in situ*/locus/heteromorphism)

EDWARD I. GINNS*, PRABHAKARA V. CHOUDARY*, SHOJI TSUJI*, BRIAN MARTIN*, BARBARA STUBBLEFIELD*, JEFFREY SAWYER†, JOHN HOZIER†, AND JOHN A. BARRANGER*

*Molecular and Medical Genetics Section, Developmental and Metabolic Neurology Branch, Intramural Research Program, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, MD 20205; and †Medical Genetics Laboratory, Florida Institute of Technology, Melbourne, FL 32901

Communicated by J. Herbert Taylor, July 1, 1985

ABSTRACT **Analysis of immunologic cross-reacting material in Chinese hamster–human somatic cell hybrids allowed assignment of the structural gene for glucocerebrosidase (glucosylceramidase; β-D-glucosyl-*N*-acylsphingosine glucohydrolase, EC 3.2.1.45) to chromosome 1 bands q21–q32. *In situ* hybridization of a radiolabeled human glucocerebrosidase cDNA to high resolution human chromosomes demonstrated that a single locus encoding glucocerebrosidase is on 1q21, adjacent to a region of chromosome 1 (1qh) abundant in structural heteromorphisms. We also have identified a hydrophobic leader polypeptide encoded by this locus, permitting a more complete description of the biosynthesis of the enzyme. These results suggest that the type-specific protein polymorphisms in Gaucher disease result from mutations at this single locus, whose segregation might be followed by linkage to visible chromosomal heteromorphisms.**

The membrane-associated lysosomal enzyme glucocerebrosidase (glucosylceramidase; β-D-glucosyl-*N*-acylsphingosine glucohydrolase, EC 3.2.1.45) is deficient in Gaucher disease, the most common Jewish genetic disorder. This reticuloendothelial storage disorder presents clinically as three major phenotypes: type 1 (chronic, nonneuronopathic), type 2 (acute, neuronopathic), and type 3 (chronic, neuronopathic). Although type 2 Gaucher disease is a clinically stereotypic disorder, types 1 and 3 are quite heterogeneous in symptoms and severity (1). The absence of functional complementation between phenotypes in somatic cell hybridization studies (2) and the ethnic predilection of only type 1 Gaucher disease (Ashkenazi carrier frequency as high as 1 in 13) (3) suggest that the different types of this disorder are a result of multiple allelic mutations of the same structural gene. This hypothesis is supported by the polymorphism of the immunologically cross-reacting material (CRM) to human placental glucocerebrosidase present in electrophoretic patterns of tissue extracts from each of the phenotypes (4, 5). Recent data also suggest that the mutations causing Gaucher disease result in the synthesis of catalytically defective glucocerebrosidases that are abnormally processed or improperly compartmentalized, or both (6).

Although previous immunological characterization of the synthesis of normal and mutant glucocerebrosidase has proven informative, the molecular and cytogenetic analysis presented here by using a cloned cDNA probe (13) should lead to a more rigorous explanation of the clinical heterogeneity within this disorder and resolve the present controversies concerning the localization of the structural gene for glucocerebrosidase (7). The unique mapping location that we establish in this study suggests a possible means of identifying homozygotes and heterozygotes for Gaucher disease.

MATERIALS AND METHODS

Cell Lines and Culture. Chinese hamster (designated Cl_3) and Chinese hamster–human (designated $X/1P_aD_1$ and $X/1P_aC_2$) cell lines were obtained from Arnold Reuser (Erasmus University, Rotterdam, Netherlands) (8). Line $X/1P_aD_1$ contained human 1q31–qter and line $X/1P_aC_2$ contained human 1pter–q31 in 38% and 90% of metaphases, respectively. GM1260, a type 2 Gaucher fibroblast line was obtained from the Human Genetic Mutant Cell Repository (Camden, NJ). Cell lines were maintained in McCoy's 5A medium containing 10% fetal calf serum and antibiotics.

Tissue Extraction for Immunologic Studies. Cells were grown to confluency, trypsinized, pelleted, and washed with phosphate-buffered saline, extracted with 60 mM potassium phosphate buffer (pH 6.6) containing 0.1% Triton X-100 and sonicated twice at 50 W for 10 sec each. After centrifugation at 48,000 × *g* for 20 min, the supernatant proteins were separated on a 12.5% polyacrylamide gel and transferred to nitrocellulose membranes. The CRM to glucocerebrosidase was detected as previously described (4).

Hybridization of Probes to Glucocerebrosidase cDNA Clones. Plasmids were grown in *Escherichia coli* C600, and DNA was isolated as described by Maniatis *et al.* (9). The Okayama–Berg Library (10) was generously provided by H. Okayama. After restriction endonuclease digestion, DNA was electrophoresed on a 1.4% agarose gel and transferred to nitrocellulose. Southern analysis (11) was performed with a ^{32}P-5′-end-labeled 17-mer oligonucleotide mixture. After prehybridization in 50% formamide containing 5× NaCl/Cit (1× = 0.15 M NaCl/0.015 M sodium citrate, pH 7), 5× Denhardt's solution (1× = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), 20 mM $NaPO_4$ (pH 6.5), 1% glycine, 0.01% *E. coli* DNA, and 0.001% poly(A) for 4 hr at 42°C, hybridization was performed in 50% formamide containing 5× NaCl/Cit, 1× Denhardt's solution, 20 mM $NaPO_4$ (pH 6.5) with 10% dextran sulfate, 0.005% *E. coli* DNA, and 0.001% poly(A) at 26°C overnight. Filters were washed 10 min at 23°C in 6× NaCl/Cit, 20 min at 23°C in 2× NaCl/Cit containing 0.1% $NaDodSO_4$, and 10 min at 23°C in 0.1× NaCl/Cit containing 0.1% $NaDodSO_4$. Autoradiographs were made from exposures of Kodak XAR-5 at −70°C for ≈1 wk.

Southern Analysis of Genomic DNA. Total cellular DNA was digested with restriction endonuclease *Msp* I (Bethesda Research Laboratories and New England Biolabs), electrophoresed on a 1.4% agarose gel, and analyzed by the method

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: CRM, cross-reacting material.

of Southern (11). Prehybridization and hybridization at 42°C were performed as described for oligonucleotide probes but included an additional final wash for 40 min at 55°C in 2× NaCl/Cit containing 0.1% $NaDodSO_4$.

***In Situ* Chromosome Localization.** High-resolution early metaphase and prometaphase chromosome spreads were prepared from methotrexate-synchronized whole blood cultures (12). The slides were allowed to age at least 1 wk before hybridization. Hybridization of a nick-translated ^3H-labeled, *Eco*RI-excised inset of a pBR322 subclone (13) (specific activity, 2×10^7 cpm/μg) was carried out at 80 ng/ml for 18 hr by the procedure of Harper and Saunders (14). Autoradiography was carried out with Kodak NTB-2 photographic emulsion (diluted 1:1). Autoradiographs were from 18-day exposures and were developed with Kodak Dektol developer for 2 min. Chromosomes were G-banded with Wright's stain in phosphate buffer for 4–6 min. The slides were destained in a series of graded ethanol (70–95%) solutions and then restained for two to three cycles to improve chromosome banding quality.

Other Analyses. Amino acid sequences were determined by using a Beckman 890M sequencer. Nucleic acids were sequenced by both the Maxam–Gilbert and dideoxynucleotide chain-termination procedures using both single and double-stranded templates (15–17). Protein was determined by the method of Bradford (18).

RESULTS AND DISCUSSION

Mapping of the Gene for Human Glucocerebrosidase. Barneveld *et al.* (8) assigned the structural gene for glucocerebrosidase using Chinese hamster–human cell hybrids to chromosome region 1q21–1q31. Another laboratory independently assigned the gene to region 1q42–1qter using different somatic cell hybrids (19). In an attempt to resolve this controversy, we have identified CRM to glucocerebrosidase in cell extracts from a panel of Chinese hamster–human cell lines (Fig. 1). The presence of CRM in the extracts of the hybrid cell line $X/1P_aC_2$ (containing human chromosome region 1q21–1q31) and its absence in the Chinese hamster–human hybrid cell line $X/1P_aD_1$ (containing the human chromosome region 1q31–1qter) support the localization of the glucocerebrosidase gene to the proximal rather than distal portion of the long arm of chromosome 1.

Using a cDNA library in the expression vector λgt11, Ginns *et al.* (13) have recently reported the isolation of two identical cDNA clones (λGC-1 and λGC-2) for human glucocerebrosidase. Recombinant bacteriophage that express the cloned DNA as part of a β-galactosidase fusion protein were identified by reactivity with antibodies to human glucocerebrosidase. The hybridization to the pBR322 subclones (pGC-1 and pGC-2) of a synthetic oligonucleotide probe derived from the amino acid sequence of homogeneous human placental glucocerebrosidase (Fig. 2) and the oligonucleotide-primed sequencing showing a 12-amino acid region of perfect match (13) confirm the identity of these clones. A comparison of additional amino acid sequence derived from the nucleotide sequence of the cDNA insert of pGC-1 subsequently revealed colinearity to >450 amino acid residues of human placental glucocerebrosidase obtained independently by chemical methods.

A more precise localization of the structural gene for glucocerebrosidase was obtained using the *Eco*RI-excised insert of pGC-1 as an *in situ* molecular hybridization probe to normal human chromosomes. The DNA was labeled by nick-translation and hybridized to high-resolution, early-metaphase, and prometaphase chromosome preparations from phytohemagglutinin-stimulated human lymphocyte cultures. After autoradiography, the chromosomes were G-banded. This *in situ* hybridization technique allows the simultaneous visualization of the autoradiographic grains from the ^3H-labeled DNA probe and the Giemsa-banding patterns of human chromosomes (14).

Analysis of 50 early metaphase and prometaphase spreads showed a high proportion of labeling on chromosome 1. Of a total of 165 grains appearing on all chromosomes, 52 (32%) were present on chromosome 1. Of the 52 grains hybridized to chromosome 1, 34 (65%) were clustered at band 1q21. Thus, examination of these 50 high-resolution spreads indicates that the structural gene for glucocerebrosidase is

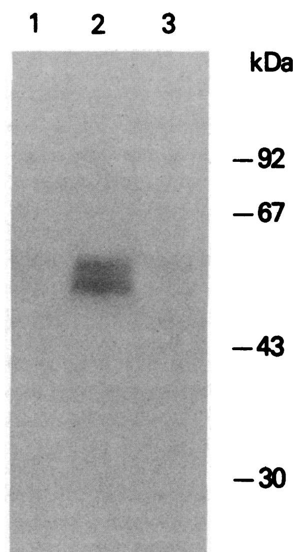

FIG. 1. Detection of CRM to human glucocerebrosidase in Chinese hamster and Chinese hamster–human hybrid cell lines using rabbit antibodies. Lanes: 1, Cl_3 cell line extract (120 μg of protein); 2, $X/1P_aC_2$ cell line extract (210 μg of protein); 3, $X/1P_aD_1$ cell extract (90 μg of protein).

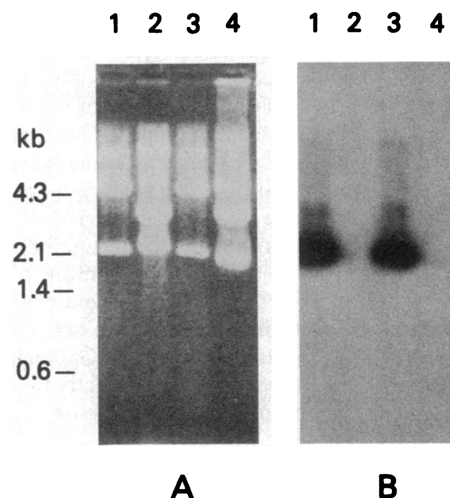

FIG. 2. (*A*) Ethidium bromide-stained agarose gel. Lanes: 1 and 3, pBR322 subclones (pGC1 and pGC2) containing cDNA inserts for glucocerebrosidase (13); 2 and 4, random clones from the Okayama–Berg library. Plasmids were grown in *E. coli* C600, and the DNA was isolated as described by Maniatis *et al.* (9). After digestion with either *Eco*RI (lanes 1 and 3) or *Bam*HI (lanes 2 and 4), DNA (10 μg per lane) was electrophoresed on a 1.4% agarose gel and transferred to nitrocellulose. (*B*) Southern blot analysis was performed with a ^{32}P-labeled oligonucleotide 17-mer probe derived from the amino acid sequence (Trp-Thr-Asp-Trp-Asn-Leu) from a tryptic peptide from human placental glucocerebrosidase.

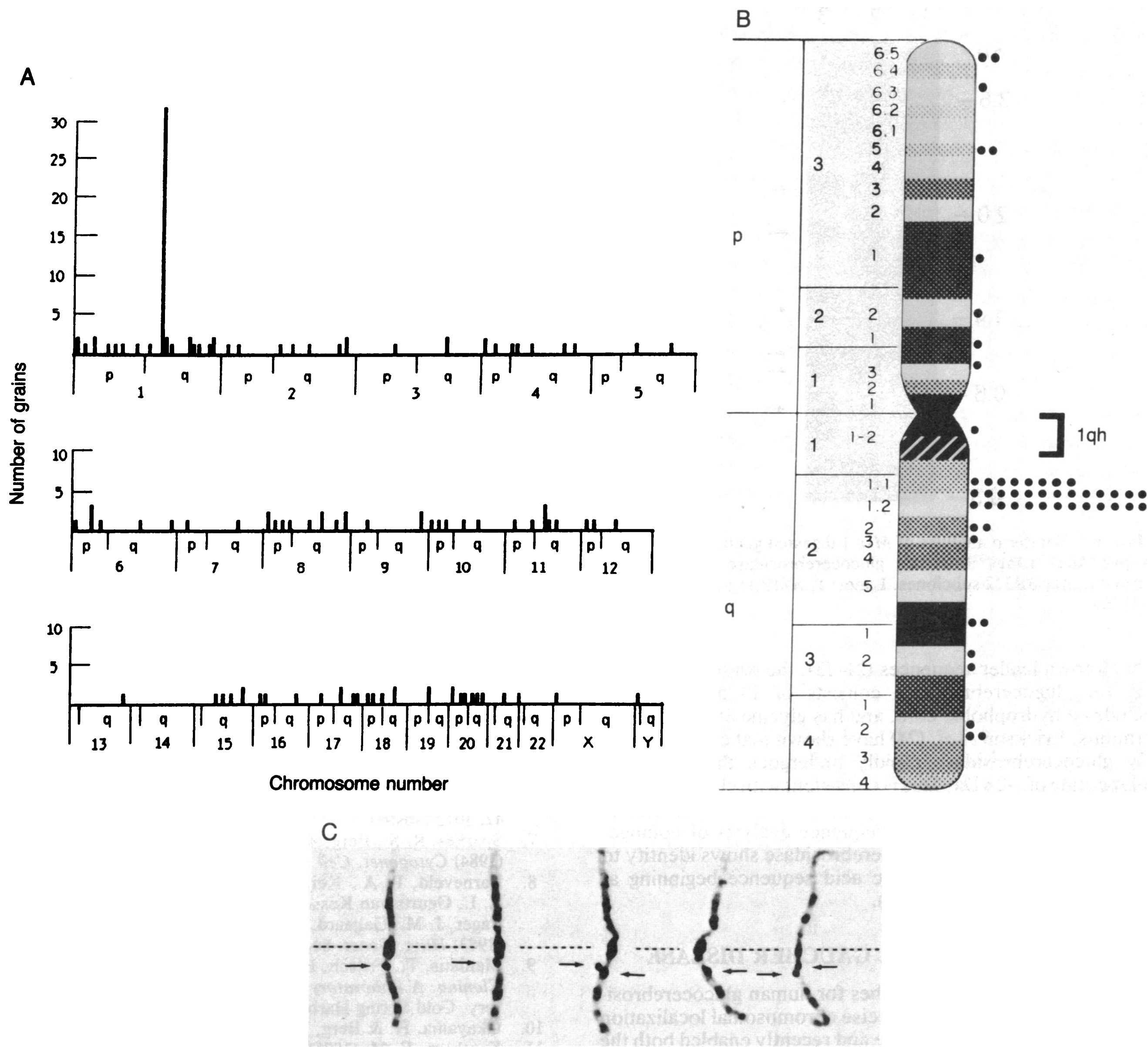

FIG. 3. Localization of the glucocerebrosidase gene to chromosome 1 by *in situ* hybridization. (*A*) Histogram illustrating the distribution of labeled sites over 50 metaphase spreads from two normal males. Significant grain accumulation is seen only over band 1q21. (*B*) Distribution of labeled sites on 25 labeled chromosomes 1 after *in situ* hybridization of the probe. Brackets indicate the adjacent highly heteromorphic 1qh region. (*C*) Examples of G-banded chromosomes 1 showing grains (arrows) localized to band 1q21. Dotted line indicates position of centromere.

located at 1q21 (Fig. 3) directly adjacent to bands 1q11–1q12, the 1qh region, rich in banding heteromorphisms. There are no other areas on chromosome 1 or any other position in the karyotype to which there is significant hybridization of the glucocerebrosidase probe.

Further conformation of the locus proximal to the centromeric region of chromosome 1 was obtained by preparing *Msp* I-digested genomic DNA from cell lines $X/1P_aC_2$, $X/1P_aD_1$, and GM1260 (a type 2 Gaucher fibroblast cell line). Southern transfers of these DNA patterns were hybridized to the ^{32}P-labeled *Eco*RI-excised insert of pGC-1 (Fig. 4). Genomic DNA fragments from cell line $X/1P_aC_2$ but not $X/1P_aD_1$ showed hybridization with labeled cDNA that was consistent with the presence of human DNA fragments encoding glucocerebrosidase.

Taken together, these observations indicate a single locus encoding glucocerebrosidase at 1q21, suggesting that the protein polymorphism seen among the types of Gaucher disease is probably a result of sets of mutations affecting only this locus. Although the clinical heterogeneity could be explained similarly, one must be careful in ascribing all of the phenotypic variation to different mutations at the structural gene locus. Stabilizer or activator mutations, as well as the presence of other genetic influences that can modify expression, may result in the production of different phenotypes (20).

Characterization of the Leader Polypeptide. The complex co- and post-translational processing events that eventually result in the formation of lysosomal glucocerebrosidase may be altered in Gaucher disease by mutations at the structural locus for this enzyme. These mutations could affect nascent peptide synthesis, the structure of the leader polypeptide, glycosylation, proteolytic processing, and/or formation of the active site environment. The isolation of a nearly full-length cDNA clone containing the entire structural gene for this membrane-associated enzyme has permitted the identification of the complete leader sequence of a human lysosomal enzyme (Fig. 5) Consistent with the properties of

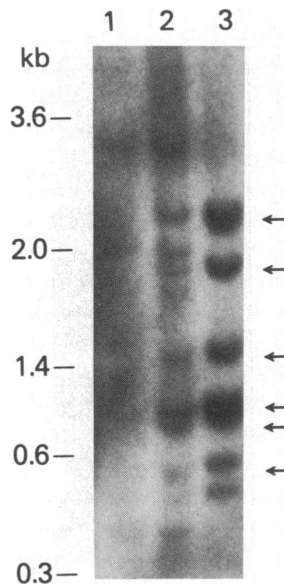

FIG. 4. Southern analysis of *Msp* I-digested genomic DNAs (10 μg per lane) using ^{32}P-labeled glucocerebrosidase insert cDNA derived from pBR322 subclones. Lanes: 1, $X/1P_aD_1$; 2, $X/1P_aC_2$; 3, GM1260.

other known leader sequences (21–23), the leader polypeptide for glucocerebrosidase consists of 19 amino acids, includes a hydrophobic core, and has glycine at its carboxyl terminus. Erickson *et al.* (24) have shown that cotranslationally glucocerebrosidase rapidly undergoes the loss of a polypeptide of ≈2 kDa. This is consistent with cleavage of the 19-amino-acid leader sequence from the amino terminus of the proenzyme. Amino acid sequence analysis of homogeneous human placental glucocerebrosidase shows identity to that derived from the nucleic acid sequence beginning at amino acid residue 20 (Fig. 5).

IMPLICATIONS FOR GAUCHER DISEASE

The availability of cDNA probes for human glucocerebrosidase has permitted a more precise chromosomal localization of the glucocerebrosidase gene and recently enabled both the isolation of genomic DNA for glucocerebrosidase and the expression of the cDNA in *E. coli* (25, 26). Expression of glucocerebrosidase cDNA in mammalian cells resulted in enzymatic activity (unpublished data). Elucidation of the mutations that result in the synthesis of abnormal glucocerebrosidase also will be facilitated by the availability of DNA probes for the entire gene, including that encoding the leader polypeptide. As further nucleic acid and amino acid sequences are obtained, more information regarding the primary structural and conformational differences between the normal and mutant glucocerebrosidases should be provided. Using the cDNA probe described here (13) restriction fragment length polymorphisms (RFLPs) involving the glucocerebrosidase gene have been detected in normal and Gaucher DNA digested with endonucleases *Pvu* II, *Bgl* I, and *Msp* I. In addition, the proximal location of the gene to the highly heteromorphic 1qh region should allow observation of segregation of mutant alleles by linkage to these heteromorphisms. By using C-, Giemsa-11, and lateral asymmetry banding, it is estimated that the segregation patterns of most chromosomes 1 (27, 28) can be unambiguously followed. If the chromosome 1 heteromorphic region and the structural gene for glucocerebrosidase are closely linked, it should be possible to use this technique to reliably identify both homozygotes and heterozygotes for Gaucher disease. This analysis of the gene mutations should provide explanations for the clinically significant heterogeneity seen within the phenotypes of this lysosomal disorder.

We thank Dr. Mary Harper for her advice and sharing of protocols for *in situ* hybridization and gratefully acknowledge the hybrid cell lines and suggestions given by Drs. Joseph Tager and Arnold Reuser and the technical assistance of Ms. June Mayor, Ms. Suzanne Winfeld, Ms. Susan Sorrell, Ms. Denise Merkle-Lehman, Mr. George Mook, and Mr. Carl Lauter. We thank Ms. Laura Metz for typing the manuscript and the National Gaucher Foundation for its support.

1. Brady, R. O. & Barranger, J. A. (1983) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 842–856.
2. Gravel, R. A. & Leung, A. (1983) *Hum. Genet.* **65**, 112–116.
3. Kolodny, E. H., Ullman, M. D., Mankin, H. J., Raghavan, S. S., Topol, J. & Sullivan, J. L. (1982) in *Gaucher Disease: A Century of Delineation and Research*, eds. Desnick, R. J., Gatt, S. & Grabowski, G. A. (Liss, New York), pp. 33–65.
4. Ginns, E. I., Brady, R. O., Pirruccello, S., Moore, C., Sorrell, S., Furbish, F. S., Murray, G. J., Tager, J. M. & Barranger, J. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5607–5610.
5. Ginns, E. I., Tegelaers, F. P. W., Barneveld, R., Brady, R. O., Tager, J. M., Galjaard, H., Reuser, A. J. J. & Barranger, J. A. (1983) *Clin. Chim. Acta* **131**, 283–287.
6. Ginns, E. I., Laszlow, D. J., Moore, C., Brady, R. O. & Barranger, J. A. (1983) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **42**, 1012 (abstr.).
7. Sparkes, R. S., Berg, K., Evans, H. J. & Klinger, H. P., eds. (1984) *Cytogenet. Cell Genet.* **37**, 1–4.
8. Barneveld, R. A., Keijzer, W., Tegelaers, F. P. W., Ginns, E. I., Geurts van Kessel, A., Brady, R. O., Barranger, J. A., Tager, J. M., Galjaard, H., Westerveld, A. & Reuser, A. J. J. (1983) *Hum. Genet.* **64**, 227–231.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
10. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289.
11. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
12. Yunis, J. J., Sawyer, J. R. & Ball, D. W. (1978) *Chromosoma* **67**, 293–307.
13. Ginns, E. I., Choudary P. V., Martin, B. M., Winfield, S., Stubblefield, B., Mayor, J., Merkle-Lehman, D., Murray, G. J., Bowers, L. A. & Barranger, J. A. (1984) *Biochem. Biophys. Res. Commun.* **123**, 574–580.
14. Harper, M. E. & Saunders, G. F. (1981) *Chromosoma* **83**, 431–439.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Wallace, R. B., Johnson, M. J., Saggs, S. V., Miyoshi, K., Bhatt, R. & Itakura, K. (1981) *Gene* **16**, 21–26.
18. Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254.
19. Devine, E. A., Smith, M., Arredondo, F. X., Shafit-Zagardo, B. & Desnick, R. J. (1982) *Cytogenet. Cell Genet.* **33**, 340–344.
20. Ginns, E. I. (1985) *BioEssays* **2**, 118–122.
21. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164.
22. Walter, P., Gilmore, R. & Blobel, G. (1984) *Cell* **38**, 5–8.

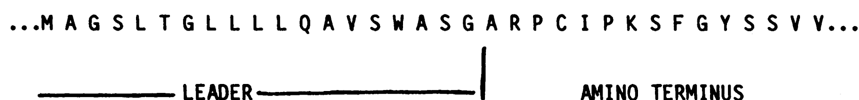

FIG. 5. Leader polypeptide and amino terminus sequence of human glucocerebrosidase. The standard single-letter amino acid code is used.

23. Rosenfeld, M. G., Kreibach, G., Popov, D., Kato, K. & Sabatini, D. (1982) *J. Cell Biol.* **93,** 135–143.
24. Erickson, A. H., Ginns, E. I. & Barranger, J. A. (1985) *J. Biol. Chem*, in press.
25. Choudary, P. V., Ginns, E. I. & Barranger, J. A. (1985) *DNA* **4,** 74.
26. Choudary, P. V., Ginns, E. I., Mayor, J. & Barranger J. A. (1985) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **44,** 7029 (abstr.).
27. Craig-Holmes, A. P. & Shaw, M. W. (1971) *Science* **174,** 702–704.
28. Magenis, R. E., Donlon, T. A. & Wyandt, H. E. (1978) *Science* **202,** 64–65.