Vol. 164, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 16, 1989 Pages 439-445

HUMAN α-L-FUCOSIDASE: COMPLETE CODING SEQUENCE FROM cDNA CLONES

Teresa Occhiodoro, Kerri R. Beckmann, C. Phillip Morris,* and John J. Hopwood

Lysosomal Diseases Research Unit, Department of Chemical Pathology, Adelaide Medical Centre for Women and Children, 72 King William Road, North Adelaide, South Australia 5006, Australia

Received August 22, 1989

The human lysosomal storage disorder fucosidosis results from the deficiency of α-L-fucosidase, a lysosomal enzyme essential for the catabolism of oligosaccharides containing α-L-fucosides. cDNA clones coding for human α-L-fucosidase have been isolated from λgt10 and λgt11 cDNA libraries derived from human liver, placenta and colon. Compilation of cDNA sequences results in a nucleotide sequence of 2053 base pairs encoding α-L-fucosidase. The sequence contains an open reading frame of 461 amino acids beginning with the first in-frame methionine and includes 439 amino acids which comprise the mature protein in addition to a hydrophobic signal peptide sequence of 22 amino acids.

α-L-Fucosidase (α-L-fucoside fucohydrolase E.C. 3.2.1.51) is a lysosomal enzyme which is involved in the degradation of fucose containing glycoproteins and glycolipids. It acts by hydrolysing glycosidically linked α-L-fucose moieties located at the non-reducing terminus of oligosaccharides (1). Deficiency of α-L-fucosidase activity in humans is the basic defect in fucosidosis, an autosomal recessive lysosomal storage disorder. This fatal disorder is characterised by the lysosomal accumulation of partially catabolised fucoglycoconjugates in the cells of most organs, particularly in the central nervous system (2).

Patients with fucosidosis display a considerable degree of clinical heterogeneity which has led to the delineation of two major subtypes of the disease (3). The more severe form (type I) is characterised by rapid neurological deterioration leading to death within the first decade of life, whereas the type II form presents itself as a disease of slower progressive neurological deterioration, allowing survival of the patient into adulthood.

* To whom correspondence should be addressed.

Abbreviations: bp, base pairs; NBRF, National Biomedical Research Foundation.

0006-291X/89 $1.50

In order to fully understand the molecular basis of the defect involved in fucosidosis and to investigate the usefulness of somatic cell gene therapy for the disease, cDNA clones encoding full length human α-L-fucosidase have been isolated.

To date, published nucleotide sequence for α-L-fucosidase has been limited to cDNA clones encoding portions of the α-L-fucosidase protein. The first report described a clone terminating at an internal EcoRI site and missing both the amino and carboxyl-termini of the protein (4), while a latter report described longer cDNA clones encoding the carboxyl terminus of the protein and 3′ untranslated sequence (5). However, although the compiled sequence of these clones is reported to cover approximately 90% of the mature enzyme, sequence information encoding the leader peptide sequence and the amino-terminal region of the protein is lacking.

In this communication we report the isolation of several α-L-fucosidase cDNA clones which, in combination, contain the complete coding sequence for both the signal peptide and the mature enzyme for human α-L-fucosidase.

MATERIALS and METHODS

Isolation of cDNA clones: A λgt10 human liver cDNA library kindly provided by Dr. G. Howlett (University of Melbourne) was screened using a 27-mer oligodeoxynucleotide (5′ACCCCAGTCTTGACAAGGACAGTGGTA 3′), synthesised using nucleotide sequence from a partial cDNA clone for α-L-fucosidase (4). T4 polynucleotide kinase and [γ-^{32}P]ATP were used to 5′ end-label the 27-mer for screening of the cDNA library as described (6). Positive clones were purified and subcloned into pUC19 for further analysis.

The position of restriction enzyme sites in the longest positive clone were mapped and the most amino-terminal fragment was isolated. The insert from the clone and this fragment were radioactively labelled using [α-^{32}P]dCTP with random oligonucleotide primers (Amersham Multiprime Labelling kit)(7) and used to screen a λgt11 human placenta cDNA library and a λgt10 human colon cDNA library (Clontech).

cDNA Sequencing: The cDNA inserts and restriction fragments were subcloned into M13mp18 or M13mp19 for nucleotide sequence analysis by the dideoxy chain termination method (8) using the Klenow fragment of DNA polymerase I at 37°C. GC-rich areas of the clones which proved difficult to sequence were resolved using the United States Biochemical Co. Sequenase Kit and dITP (9).

Sequence Analysis: The nucleotide sequence was screened against the GenBank database and the encoded protein sequence against the NBRF protein database. General sequence analysis and comparisons were performed using a package of programmes supplied by A. Reisner (10).

RESULTS and DISCUSSION

A 27-mer oligodeoxynucleotide (described in Materials and Methods) was synthesised complementary to a portion of the 1058 bp cDNA clone previously reported by Fukushima *et al* (4), specifically to bases 570-596. Using this

oligodeoxynucleotide as a probe, 2 x 10^5 recombinants of the liver cDNA library were screened resulting in five positive clones. The longest clone (λHF05), which contained an insert of 1254 bp, was analysed by sequencing of the 5′- and 3′-ends (Figure 1). In addition, a 244 bp EcoRI-RsaI restriction fragment from the 5′-end of the λHF05 insert was sequenced in both directions. The sequence revealed that this clone extended the known sequence by 209 bp in the 5′-direction and included the full sequence of mature α-L-fucosidase and 19 amino acids of the signal peptide, but did not contain the initiating methionine. The λHF05 insert was used as a probe to screen 2.4 x 10^5 recombinant clones from a random primed human colon cDNA library. The restriction enzyme sites in the resulting 8 positive clones were mapped relative to λHF05. One clone (λHF12) was found to contain sequence covering the full 3′-untranslated region of the mRNA, however none of these clones extended the sequence beyond the 5′-end of the λHF05.

Using the EcoRI-RsaI restriction fragment derived from the 5′-end of the λHF05 insert from nucleotide position 23 to 267 (see Figure 2), further screening was carried out on 9.3 x 10^5 recombinants from a λgt11 human placenta cDNA library and 1.3 x 10^6 recombinants from the previously screened randomly primed λgt10 human colon cDNA library. A total of 25 positive clones were isolated and rescreened with a 70 bp EcoRI-Sau96I restriction fragment from nucleotide position 23 to 93 (see Figure 2), used to eliminate any clones that did not extend to near the 5′-end of λHF05, resulting in 3 positive clones from the colon library and 16 from the placental library.

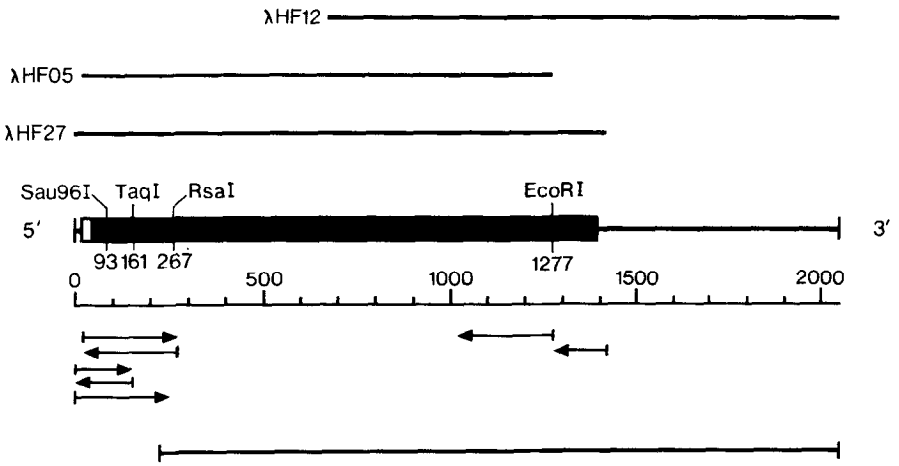

Figure 1. Diagrammatic representation of α-L-fucosidase cDNA clones relative to the compiled total α-L-fucosidase sequence. The open box indicates the sequence coding for the signal peptide; the solid box represents the sequence of the mature enzyme. Sequenced areas and the direction of sequencing are denoted by the arrows. Relevant restriction enzyme sites are indicated. The extent of the previously published cDNA sequence data is shown beneath the sequencing strategy as a bar.

```
                 Met Arg Ser Arg Pro Ala Gly Pro Ala Leu Leu Leu Leu Leu Leu Phe Leu Gly Ala Ala Glu Ser↓Val Arg
GAATTCCGGGCTCCGGGGATGAGGTCGCGGCCGGCGGGTCCCGCGCTGTTGCTGCTGCTGCTCTTCCTCGGAGCGGCCGAGTCGGTGCGT
        10        20        30        40        50        60        70        80        90

Arg Ala Gln Pro Pro Arg Arg Tyr Thr Pro Asp Trp Pro Ser Leu Asp Ser Arg Pro Leu Pro Ala Trp Phe Asp Glu Ala Lys Phe Gly
CGGGCCCAGCCTCCGCGCCGCTACACCCCAGACTGGCCGAGCCTGGATTCTCGGCCGCTGCCGGCCTGGTTCGACGAAGCCAAGTTCGGG
       100       110       120       130       140       150       160       170       180

Val Phe Ile His Trp Gly Val Phe Ser Val Pro Ala Trp Gly Ser Glu Trp Phe Trp Trp His Trp Gln Gly Glu Gly Arg Pro Gln Tyr
GTGTTCATCCACTGGGGCGTGTTCTCGGTGCCCGCCTGGGGCAGCGAGTGGTTCTGGTGGCACTGGCAGGGCGAGGGGCGGCCGCAGTAC
       190       200       210       220       230       240       250       260       270

Gln Arg Phe Met Arg Asp Asn Tyr Pro Pro Gly Phe Ser Tyr Ala Asp Phe Gly Pro Gln Phe Thr Ala Arg Phe Phe His Pro Glu Glu
CAGCGCTTCATGCGCGACAACTACCCGCCCGGCTTCAGCTACGCCGACTTCGGACCGCAGTTCACTGCGCGCTTCTTCCACCCGGAGGAG
       280       290       300       310       320       330       340       350       360

Trp Ala Asp Leu Phe Gln Ala Ala Gly Ala Lys Tyr Val Val Leu Thr Thr Lys His His Glu Gly Phe Thr Asn Trp Pro Ser Pro Val
TGGGCCGACCTCTTCCAGGCCGCGGGCGCCAAGTATGTAGTTTTGACGACAAAGCATCACGAAGGCTTCACAAACTGGCCGAGTCCTGTG
       370       380       390       400       410       420       430       440       450

Ser Trp Asn Trp Asn Ser Lys Asp Val Gly Pro His Arg Asp Leu Val Gly Glu Leu Gly Thr Ala Leu Arg Lys Arg Asn Ile Arg Tyr
TCTTGGAACTGGAACTCCAAAGACGTGGGGCCTCATCGGGATTTGGTTGGTGAATTGGGAACAGCTCTCCGGAAGAGGAACATCCGCTAT
       460       470       480       490       500       510       520       530       540

Gly Leu Tyr His Ser Leu Leu Glu Trp Phe His Pro Leu Tyr Leu Leu Asp Lys Lys Asn Gly Phe Lys Thr Gln His Phe Val Ser Ala
GGACTATACCACTCACTCTTAGAGTGGTTCCATCCACTCTATCTACTTGATAAGAAAAATGGCTTCAAAACACAGCATTTTGTCAGTGCA
       550       560       570       580       590       600       610       620       630

Lys Thr Met Pro Glu Leu Tyr Asp Leu Val Asn Ser Tyr Lys Pro Asp Leu Ile Trp Ser Asp Gly Glu Trp Glu Cys Pro Asp Thr Tyr
AAAACAATGCCAGAGCTGTACGACCTTGTTAACAGCTATAAACCTGATCTGATCTGGTCTGATGGGGAGTGGGAATGTCCTGATACTTAC
       640       650       660       670       680       690       700       710       720

Trp Asn Ser Thr Asn Phe Leu Ser Trp Leu Tyr [Asn] Asp Ser Pro Val Lys Asp Glu Val Val Val Asn Asp Arg Trp Gly Gln [Asn] Ser
TGGAACTCCACAAATTTTCTTTCATGGCTCTACAATGACAGCCCTGTCAAGGATGAGGTGGTAGTAAATGACCGATGGGGTCAGAACTCT
       730       740       750       760       770       780       790       800       810

Ser Cys His His Gly Gly Tyr Tyr Asn Cys Glu Asp Lys Phe Lys Pro Gln Ser Leu Pro Asp His Lys Trp Glu Met Cys Thr Ser Ile
TCCTGTCACCATGGAGGATACTATAACTGTGAAGATAAATTCAAGCCACAGAGCTTGCCAGATCACAAGTGGGAGATGTGCACCAGCATT
       820       830       840       850       860       870       880       890       900

Asp Lys Phe Ser Trp Gly Tyr Arg Arg Asp Met Ala Leu Ser Asp Val Thr Glu Glu Ser Glu Ile Ile Ser Glu Leu Val Gln Thr Val
GACAAGTTTTCCTGGGGCTATCGTCGTGACATGGCATTGTCTGATGTTACAGAAGAATCTGAAATCATTTCGGAACTGGTTCAGACAGTA
       910       920       930       940       950       960       970       980       990

Ser Leu Gly Gly Asn Tyr Leu Leu Asn Ile Gly Pro Thr Lys Asp Gly Leu Ile Val Pro Ile Phe Gln Glu Arg Leu Leu Ala Val Gly
AGTTTGGGAGGCAACTATCTTCTGAACATTGGACCAACTAAAGATGGACTGATTGTTCCCATCTTCCAAGAAAGGCTTCTTGCTGTTGGG
      1000      1010      1020      1030      1040      1050      1060      1070      1080

Lys Trp Leu Ser Ile Asn Gly Glu Ala Ile Tyr Ala Ser Lys Pro Trp Arg Val Gln Trp Glu Lys [Asn] Thr Thr Ser Val Trp Tyr Thr
AAATGGCTGAGCATCAATGGGGAGGCTATCTATGCCTCCAAACCATGGCGGGTGCAATGGGAAAAGAACACAACATCTGTATGGTATACC
      1090      1100      1110      1120      1130      1140      1150      1160      1170

Ser Lys Gly Ser Ala Val Tyr Ala Ile Phe Leu His Trp Pro Glu Asn Gly Val Leu Asn Leu Glu Ser Pro Ile Thr Thr Ser Thr Thr
TCAAAGGGATCGGCTGTTTATGCCATTTTTCTGCACTGGCCAGAAAATGGAGTCTTAAACCTTGAATCCCCCATAACTACCTCAACTACA
      1180      1190      1200      1210      1220      1230      1240      1250      1260

Lys Ile Thr Met Leu Gly Ile Gln Gly Asp Leu Lys Trp Ser Thr Asp Pro Asp Lys Gly Leu Phe Ile Ser Leu Pro Gln Leu Pro Pro
AAGATAACAATGCTGGGAATTCAAGGAGATCTGAAGTGGTCCACAGATCCAGATAAAGGTCTCTTCATCTCTCTACCCCAGTTGCCACCC
      1270      1280      1290      1300      1310      1320      1330      1340      1350

Ser Ala Val Pro Ala Glu Phe Ala Trp Thr Ile Lys Leu Thr Gly Val Lys  *
TCTGCTGTCCCCGCAGAGTTTGCTTGGACTATAAAGCTGACAGGAGTGAAGTAATCATTTGAGTGCAAGAAGAAAGAGGCGCTGCTCACT
      1360      1370      1380      1390      1400      1410      1420      1430      1440

GTTTTCCTGCTTCAGTTTTTCTCTTATAGTACCATCACTATAATCAACGAACTTCTCTTCTCCACCCAGAGATGGCTTTTCCAACACATT
      1450      1460      1470      1480      1490      1500      1510      1520      1530

TTAATTAAAGGAACTGAGTACATTACCCTGATGTCTAAATGGACCAAAGATCTGAGATCCATTGTGATTATATCTGTATCAGGTCAGCAG
      1540      1550      1560      1570      1580      1590      1600      1610      1620

AAGAAGGAACTGAGCAGTTGAACTCTGAGTTCATCAATTCTAATATTTGGAAATTATCTACAATGGAATCTTCCCTCTGTTCTCTGATAA
      1630      1640      1650      1660      1670      1680      1690      1700      1710

CCTACTTGCTTACTCAATGCCTTTAAGCCAAGTCACCCTGTTGCCTATGGGAGGAGGTGGAAGGATTTGGCAAGCTCAACCACATGCTAT
      1720      1730      1740      1750      1760      1770      1780      1790      1800

TTAGTTAGCATCAGTTGTCACCAACAGTCTTTCTGCAAAGGGCAGGAGAGCTTTGGGGGAAAGGAAAAGGCTTACCAGGCTGCTATGGTC
      1810      1820      1830      1840      1850      1860      1870      1880      1890

AACTCTTCAGAAATTTTCAGAGCAATCTAAAAGCGCCAAAATTCGCTATGTTTACAGTGATACTATTAAGAAAATGAATGTGATTCTGCT
      1900      1910      1920      1930      1940      1950      1960      1970      1980

CTGTCTTTTTAAGTATGATCA[AATAAA]AAATTTGTACATCACAATCATTTCTACCAAAAAAAAAAAAAAAAA
      1990      2000      2010      2020      2030      2040      2050
```

Figure 2. Compiled nucleotide sequence of the α-L-fucosidase cDNA clones and the deduced amino acid sequence of the encoded protein. Previously published sequence extends from base 230. The predicted signal peptidase cleavage site is indicated by an arrow, and the polyadenylation sequence and potential glycosylation sites are boxed. The stop codon is denoted by an asterisk. The nucleotide at position 809 was originally reported to be a G and formed part of a cysteine codon (4) however, it was subsequently reported to be a C changing it to a serine codon (5).

Inserts from the positive colon library clones were analysed by sequencing. One clone, λHF27, was shown by sequencing (see Figure 1) to extend the sequence of λHF05 in the 5'-direction by 23 base pairs and revealed an in-frame methionine beyond the amino-terminus of the λHF05 open reading frame. The various cDNA clones and the relevant restriction sites used for the sequencing of the human α-L-fucosidase cDNA are shown in Figure 1, relative to the previously published clones (4,5). The nucleotide sequence encoding the entire mature α-L-fucosidase enzyme along with its deduced amino acid sequence is shown in Figure 2. Previously published sequence extends from base 230 and areas which we have sequenced in this report are shown in the sequencing strategy presented in Figure 1. In the regions where our clones overlapped the previously reported clones, our sequence corresponded exactly to the published data (5).

The additional nucleotide sequence presented here extends the α-L-fucosidase sequence beyond the presumptive initiating methionine codon at position 19 which, together with the previously published deduced amino acid sequence (4,5), results in an open reading frame encoding an unprocessed protein of 461 amino acids, including 22 amino acids of signal peptide and 439 amino acids encoding mature human α-L-fucosidase. The nucleotide sequence upstream of the ATG codon is composed largely of C and G residues and is deficient in T residues, a feature commonly observed in sequences flanking eukaryotic initiator codons (11). This upstream sequence also conforms favourably with the consensus sequence for initiation of translation in vertebrates (11).

The amino terminus of the deduced amino acid sequence consists of a stretch of 22 amino acids with several features consistent with those of signal peptide sequences (12). This area of the protein contains a number of aliphatic and neutral amino acids making it one of the most hydrophobic domains in the protein, as indicated by the hydrophobicity plot shown in Figure 3. The sequence also conforms well to the commonly observed features surrounding the signal peptidase cleavage site (12), including the occurrence of small uncharged amino acid residues flanking the -2 position preceding the cleavage site. No sequence homology was found when the full sequence of the α-L-fucosidase cDNA clone and the encoded protein were compared with the GenBank nucleic acid database (release 59.0) and the NBRF protein database (release 20.0).

The availability of a dog colony affected with the canine form of fucosidosis is currently providing invaluable information on the use of bone marrow transplantation as a means of therapy for this disorder (13). Recent experiments have shown that transplantation of allogeneic bone marrow cells

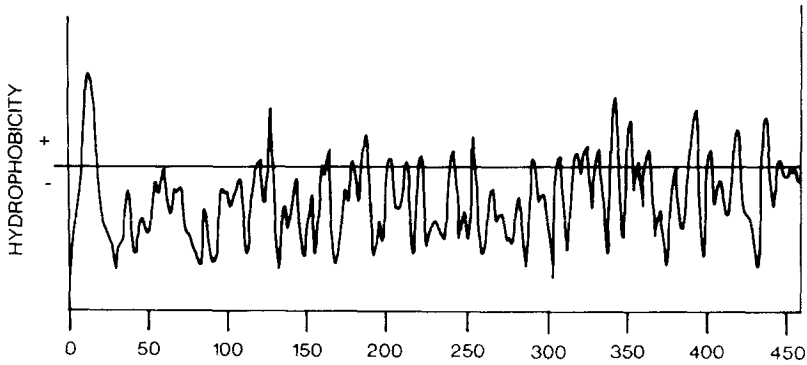

Figure 3. The hydrophobicity plot of the α-L-fucosidase protein was generated using the parameters of Kyte and Doolittle (15) using the "CHOU" programme described by Novotny and Auffray (16).

at an early stage of the disease reconstitutes deficient α-L-fucosidase activity in affected dogs (14). This raises the prospect of somatic cell gene therapy by introduction of bone marrow cells containing the α-L-fucosidase gene into affected dogs. This work may now proceed by the use of the human α-L-fucosidase cDNA clone, allowing the necessary preclinical evaluation in an animal model before gene therapy can be applied to humans.

ACKNOWLEDGMENTS

We would like to thank Daniel Robertson, of this Department, for helpful discussions and Dr. Geoffrey Howlett, Department of Biochemistry, University of Melbourne for providing the human liver cDNA library.

This work was supported by the Adelaide Children's Hospital Research Foundation.

REFERENCES

1. Flowers, H.M. (1981) Adv. Carbohydr. Chem. Biochem. **39**, 279-345.
2. Durand, P., Gatti, R. and Borrone, C. (1982) In Genetic errors of glycoprotein metabolism (eds. Durand, P. and O'Brien, J.S.), pp 49-81, Springer-Verlag, New York.
3. Kousseff, B.G., Beratis, N.G., Danesino, C. and Hirschhorn, K. (1973) Lancet **2**, 1387-1388.
4. Fukushima, H., de Wet, J.R. and O'Brien, J.S. (1985) Proc. Natl. Acad. Sci. USA **82**, 1262-1265.
5. O'Brien, J.S., Willems, P.J., Fukushima, H., de Wet, J.R., Darby, J.K., Di Cioccio, R., Fowler, M.L. and Shows, T.B. (1987) Enzyme **38**, 45-53.
6. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in DNA Cloning (ed. D.M. Glover), Vol. 1, pp 49-78. IRL Press Limited, Oxford.
7. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6-13.
8. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
9. Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771.

10. Reisner, A.H. and Bucholtz, C. (1985) Nature **314**, 310.
11. Kozak, M. (1987) Nucl. Acids Res. **15**, 8125-8131.
12. von Heijne, G. (1986) Nucl. Acids Res. **14**, 4683-4690.
13. Taylor, R.M., Stewart, G.J., Farrow, B.R.H. and Healy, P.J. (1986) Lancet **2**, 772-774.
14. Taylor, R.M., Farrow, B.R.H., Stewart, G.J., Healy, P.J. and Tiver, K. (1989) Transp. Proc. **20**, 89-93.
15. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. **157**, 105-132.
16. Novotny, J. and Auffray, C. (1984) Nuc. Acids Res. **12**, 243-255.