Vol. 205, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1994 Pages 1-5

COMPLETE CDNA SEQUENCE OF HUMAN LYSOSOME-ASSOCIATED MEMBRANE PROTEIN-2

David S. Konecki[1], Kay Foetisch, Magdalena Schlotter, and Uta Lichter-Konecki[2]

Molecular Genetic Research Group, University Childrens Hospital, Im Neuenheimer Feld 150, 69120 Heidelberg, FRG

Received September 24, 1994

SUMMARY: The isolation and sequencing of 15 independent human lysosome-associated membrane protein-2 (h-lamp-2) recombinants from a primary human liver cDNA library has resulted in the determination of a transcript sequence significantly longer than previously reported and reveals the utilization of each of the four potential polyadenylation signals (AATAAA) present in the 3' untranslated region. The most 5' extending cDNA clone initiates upstream of the proposed transcription initiation site. A number of differences with published sequences for the h-lamp-2 transcript were observed, some of which result in amino acid changes in the predicted primary structure of the h-lamp-2 protein, and two of which give rise to restriction fragment length polymorphisms. The knowledge of these sequence alterations and polymorphisms is an important consideration for the further analysis of the h-lamp-2 locus with regard to the delineation of function and association with human inherited disorders.

The human lysosome-associated membrane protein-2 (h-lamp-2) is one of the most abundant constituents of the human lysosomal membrane (1). The structure of this protein contains a luminal domain consisting of two homologous domains with four identically spaced cysteins linked by two disulfide bonds, a single 20 amino acid transmembrane spanning region, and a short cytoplasmic tail containing the lysosomal membrane targeting signal. This protein is heavily glycosylated by N-glycans, including polylactosamino glycans. Suggested functions for h-lamp-2 include protection of the lysosomal membrane from autodigestion, maintenance of the acidic environment of the lysosome, adhesion when expressed on the cell surface (plasma membrane), and inter- and intracellular signal transduction. Its expression has been recently shown to be uniquely regulated (2), with a significant increase at the cell surface in patients with the autoimmune disorder scleroderma. To date all sequence information concerning the h-lamp-2 gene, located at chromosome Xq24-25 (3), and its transcript have relied upon the initial cDNA recombinant and its sequence of 1327 nucleotides (4). As a first step for the study of association

[1] Current address: Institute for Virus Research, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, FRG.

[2] Corresponding author: Dr. U. Lichter-Konecki. fax: 49-6221-563324.

0006-291X/94 $5.00

between this gene and human disorders resulting from potential defects in the lysosomal membrane, we isolated a number of full-length h-lamp-2 cDNA recombinants, determined their sequence, and compared them to previously reported sequences (2, 4, 5).

MATERIALS AND METHODS

Complementary DNA Library Construction and Screening. The details of the construction of the human liver cDNA library, designated λ-1956, have been previously reported (6). This clone library, constructed in the vector λgt11,consisted of about 2.0 x 10^8 independent recombinants. Approximately 2.0 x 10^6 recombinant bacteriophage were initially screened by filter hybridization (7), with the radiolabelled (8) insert of a mouse lgp-110 recombinant generously provided by Drs. A. Helenius and W. Granger, Yale University. Hybridization positive recombinants were plaque-purified, the phage DNA isolated (9) and digested with *Not I* restriction endonuclease. Following resolution by low melting-point gel electrophoresis, the cDNA inserts of λ recombinants were isolated by the method of Higuchi et al. (10). The purified inserts were subsequently ligated into the plasmid vector pBlueScript II - SK(-) (Stratagene) for DNA sequence analysis.

Reverse Transcription / Polymerase Chain Reaction Studies. First strand cDNA synthesis utilized the reaction components of the First Strand cDNA Synthesis Kit obtained from Phamacia/LKB, total placental RNA (1.0 μg) isolated as described (11), and 220 ng of the oligonucleotide primer L2A-1 (5'-GAACAAGTTTGTCTCCAGGACC-3', which is complementary to the sense strand from base 1593 to base 1572 of the sequence shown in Figure 1). The synthesis reaction was performed for 1 hour at 42°C in a volume of 10 μl. The resulting products of this reaction were used as template for a polymerase chain reaction (PCR) using the oligonulceotide primers L2A-2 (5'-TGCAGCTGTTGTTGTACCG-3', complementary to bases 82-100 of the antisense strand, Figure 1) and L2A-3 (5'-GACTGATCTCAAATGCTGGG-3', complementary to the sense strand shown in Figure 1 from base1513 to base 1493). The PCR volume of 100 μl contained 120 ng of each primer, AmpliTaq polymerase (Perkin Elmer), and AmpliTaq reaction components. An Autogene II (Grant) thermocycler was used to perform the 50 cycle PCR consisting of 0.6 minutes at 96°C (denaturation), 0.6 minutes at 60°C (annealing), and 2.0 minutes at 72°C (extension) for each cycle; with a final extension step at 72°C for 7.0 minutes. Following PCR, the reaction products were resolved on 0.8% low-melting agarose (FMC) gels, with the expected amplified band of 1432 base pairs being excised and purified (10). Amplified DNA was subcloned into a PCR product cloning vector (T-vector) prepared from pBlueScript II - SK(-) (Stratagene) by the method of Holton and Graham (12).

DNA Sequence Analysis. Sequencing was performed by the dideoxy chain termination method (13) using the T7 Polymerase Sequencing Kit (Pharmacia/LKB) and implementing the modifications of Biggin et al. (14) and Del Sal et al. (15). The 5'- and 3'-ends of all isolated h-lamp-2 cDNA inserts were determined through the use of universal and reverse M13 sequencing primers (Promega) with the inserst of selected recombinants being sequenced completely. Regions in which differences were found with previous reports (2, 4, 5) were verified for all h-lamp-2 cDNA inserts through the use of specific primer-directed sequencing. Computer analysis of DNA sequence data was performed at the Deutsches Krebsforschungszentrum (Heidelberg, Germany) using the University of Wisconsin Genetics Computer Group (UWGCG) DNA sequence analysis package, version 7.0.

RESULTS AND DISCUSSION

Screening of the unamplified human liver cDNA library (5) λ-1956 with the insert of a mouse lgp-110 recombinant resulted in the isolation of the full-length h-lamp-2 recombinant

designated λ1956-72E3B. The insert of this recombinant was used to rescreen the above mentioned human liver library. Of the more than 60 clones identified by hybridization, 15 were isolated and subsequently sequenced. The sizes of the cDNA inserts ranged from 997 to 1821 base pairs, with the average length being 1555 base pairs. Each of the h-lamp-2 recombinants was found to be independent at the level of DNA sequence in that no two recombinants possessed the same initiation and termination of sequence. Compared to the previously published h-lamp-2 cDNA sequence (4), an additional 48 nucleotides of 5' untranslated region (UTR) and 588 nucleotides of 3' UTR were identified; as well as six additional nucleotides of sequence between position 1100 and 1114. Figure 1 shows our cDNA derived sequence of the h-lamp-2 transcript initiating 12 nucleotides 5' of the proposed transcription initiation (CAP) site (2) and terminating nearly 600 nucleotides downstream of the end of the only h-lamp-2 cDNA previously reported (4). A total of twelve nucleotide differences in the protein coding region of the h-lamp-2 transcript sequence were observed. To ascertain whether these nucleotide alterations were the result of sequence polymorphisms residing within this region, reverse transcription / polymerase chain reaction (RT/PCR) studies were conducted with total RNA isolated from two independent placental tissue samples, with the amplification products subcloned into the pBluscript vector prior to DNA sequence analysis. By this means of investigation only the nucleotide substitutions at nucleotide positions 293 (T/A) and 1064 (T/C) were found to be polymorphic. Both of these substitutions occur at the third positions of their respective codons and represent silent mutations, resulting in no amino acid changes in the protein. The presence of a T at position 293 abolishes a *Rsa I* site, while a T at position 1064 results in the loss of a *Nla IV* site. These two polymorphisms represent the foundation necessary for the establishment of a haplotype analysis system. None of the remaining sequence differences in the protein coding region at nucleotide positions 469 (amino acid 83), 795 (amino acid 192), 838-839 (amino acid 206) 1101 (amino acid 294), and 1110 to 1114 (amino acids 297 - 298) were determined to be polymorphic. Since the translated amino acids at these positions using our sequence are identical (amino acid 83, isoleucine; amino acid 206, leucine; 294 - 298, aspartate, alanine, proline, leucine, glycine) or represent a conservative change (amino acid 192, alanine versus valine in rodents) as compared to the predicted amino acids sequences of mouse (16) and rat (17), the sequence of Sawada et al. (2) at these nucleotide positions may represent sequencing errors.

In the previous report by Sawada et al. (2), sequencing of a genomic recombinant containing the 3'-end of the h-lamp-2 transcript provided a sequence of the 3' UTR which differs from the sequence presented at nine positions (Figure 1). Also, since most of the sequence from the previous report (2) of the 3' UTR was determined from genomic DNA and not cDNA, the functionality of the four potential polyadenylation signals in the 3' UTR could only be speculated upon. Among the 15 h-lamp-2 cDNA recombinants presented in this study, at least one recombinant was determined to have resulted from the utilization of each of the polyadenylation signals shown in Figure 1, although the most frequently used site was the second (nucleotide positions 1702-1707). This polyadenylation signal was found to have given rise to five (33%) of the h-lamp-2 recombinants investigated.

Due to the results of this investigation of h-lamp2 cDNA recombinants it is possible to speculate that the true transcription initiation site for this gene may lie somewhat further 5' than

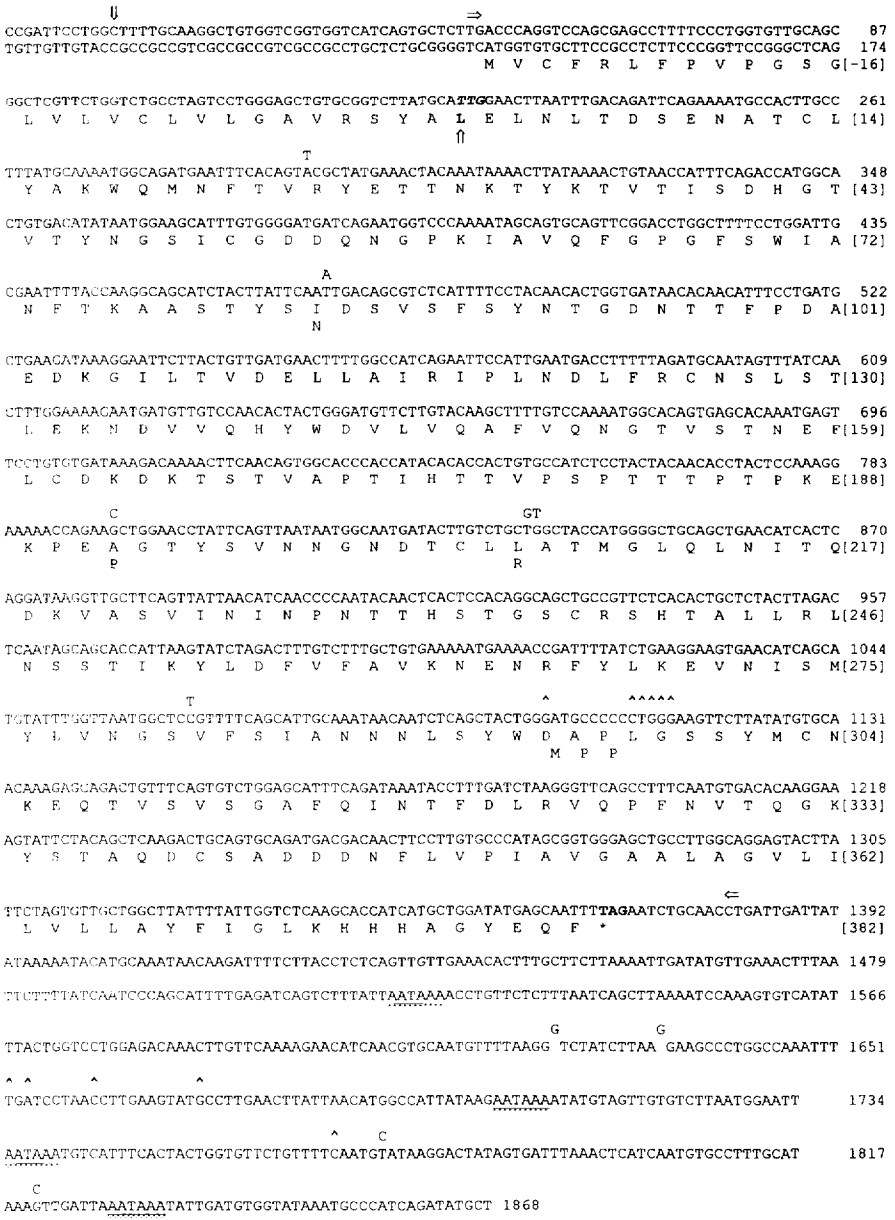

Figure 1. H-lamp-2 nucleotide sequence derived from cDNA recombinants isolated from the primary human liver cDNA library λ-1956 (EMBL Nucleotide Sequence Database Accession Number X77196 HSLAMP2). The position of the proposed transcription initiation site (2, indicated by ⇓), the beginning (⇒),and end (⇐) of the only h-lamp-2 cDNA previously reported (4) are shown. Numbers to the right of each line of sequence represent the position of the last base on the line. Below the sequence line, beginning at the initiator methionine and ending at the termination codon TAG, the postulated amino acid sequence of h-lamp-2 is shown. The numbers in brackets to the right of the single letter abbreviations for amino acids correspond to the amino acid number of the last residue on the line, with numbering begininning from the first residue of the mature protein (leucine, shown in bold and indicated by ⇑). Above the sequence line are those bases from the previous publication (4) which differ from our sequence. The symbol ^ is used to indicate nucleotides missing in the initial cDNA sequence (4) or the genomically derived 3' UTR sequence (2). In those instances in which alternative nucleotides within the coding region exist, the amino acid reported by Fukuda et al. (4) is shown below the respective amino acid derived from our sequence.The four polyadenylation signals in the 3' UTR are double underlined.

previously reported (2), or that multiple transcription initaition sites exist. The detection of two sequence polymorphisms within the protein coding region of the gene provide a first step towards the development of a haplotype analysis system for this locus, which will aid studies concerning the association of this gene with X-linked inherited disorders mapped to this region (Xq24-25).

ACKNOWLEDGMENT

This work was supported by the Deutsche Forschungsgemeinschaft grant Li 375/5-1 to U. L.-K.

REFERENCES

1. Kornfeld, S. and Mellman, I. (1989) *Annu. Rev. Cell Biol.* **5**, 483 - 525.
2. Sawada, R., Jardine, K.A., and Fukuda, M. (1993) *J. Biol. Chem.* **268**, 9014 - 9022.
3. Mattei, M.-G., Matterson, J., Chen, J.W., Williams, M.A., Fukuda, M. (1990) *J. Biol. Chem.* **265**, 7548 - 7551.
4. Fukuda, M., Viitala, J., Matteson, J., and Carlsson, S.R. (1988) *J. Biol. Chem.* **263**, 18920 - 18928.
5. Sawada, R., Jardine, K.A., and Fukuda, M. (1993) *J. Biol. Chem.* **268**, 13010.
6. Konecki, D.S., Wang, Y., Trefz, F.K., Lichter-Konecki, U., and Woo, S.L.C. (1992) *Biochemistry* **31**, 8363 - 8368.
7. Herrmann, B., Bucan, M., Mains, P.E., Frischauf, A.-M., Silover, L.M., and Lehrach, H. (1986) *Cell* **44**, 469 - 476.
8. Bucan M., Zimmer, M., Whaley, W.L., Poustka, A., Youngman, S., Allitto, A, Ormondroyd, E., Smith, B., Pohl, T.M., MacDonald, M., Bates, G.P., Richards, J., Volina, S., Gilliam, T.C., Sedlacek, Z., Collins, F.S., Wasmuth, J.J., Shaw, D.J., Gusella, J.F., Frischauf, A.-M., and Lehrach, H. (1990) *Genomics* **6**, 1-15.
9. Pohl, T.M., Zimmer, M., MacDonald, M.E, Smith, B.,Bucan M., Poustka, A., Volina, S., Searle, S., Zehetner, Wasmuth, J.J., Gusella, J. Lehrach, H., and Frischauf, A.-M. (1988) *Nucleic Acids Res.* **16**, 9185 - 9198.
10. Higuchi, R, Stang H.D., Browne, J.K., Martin, M.O., Huot, M., Lipeles, J., and Salser, W. (1981) *Gene* **15**, 177 - 186.
11. Chromczynski, P., and Sacchi, N. (1987) *Analytical Biochemistry* **162**, 156 -159.
12. Holton, T.A., and Graham, M.W. (1991) *Nucleic Acids Res.* **19**, 1156.
13. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 - 5467.
14. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* ***80,3963 - 3965.***
15. Del Sal, G., Manfioletti, G., and Schneider, C. (1989) *Biotechniques* **7**, 514 - 519.
16. Cha, Y., Holland, S.M., and August, J.T. (1990) *J. Biol. Chem.* **265**, 5008 - 5013.
17. Noguchi, Y., Himeno, M., Sasaki, H., Tanaka, Y., Kono, A., Sakaki, Y., and Kato, K. (1989) *Biochem. Biophys. Res. Commun.* **164**, 1113 - 1120.