3

Contract N00014-85-K-0703
COMMUNICATION

Isolation and Sequencing of a cDNA Clone Encoding Lysosomal Membrane Glycoprotein mLAMP-1: Sequence Similarity to Proteins Bearing Onco-Differentiation Antigens*

Jeff W. Chen‡, Ying Cha, K. Umit Yukselt, Robert W. Gracyt, and J. Thomas August

From the Department of Pharmacology and Molecular Sciences, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, and the tDepartment of Biochemistry, North Texas State University and Texas College of Osteopathic Medicine, Denton, Texas 76203

*This work was supported by grants from the National Institutes of Health (5 RO1 GM31168, AG01274 and AM14638), the Office of Naval Research (N00014-85-K-0703), the Robert A. Welch Foundation (B-502), and the Texas Advanced Technology and Research Program (Applied Enzymology). J.W.C. was the recipient of Medical Scientist Training Program Grant 5 T32 GM07309.

‡Current address: Department of Neurosurgery, UCSD Medical Center, San Diego, California 92103

Proofs should be sent to: Dr. J. Thomas August, Department of Pharmacology and Molecular Sciences, The Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205 (301-955-3985)

Running Title: cDNA Cloning of Mouse mLAMP-1

SUMMARY

We have isolated and sequenced a cDNA clone encoding the mouse mLAMP-1 major lysosomal membrane glycoprotein. The deduced protein sequence, which included the NH_2-terminal portion of the mLAMP-1 molecule, consisted of 382 amino acids (M_r 41,509). The predicted structure of this protein included an NH_2-terminal intralumenal domain consisting of two homology units of approximately 160 residues each, separated by a proline-rich hinge region. Each homology unit contained four cysteine residues, with two intercysteine intervals of 36 to 38 residues and one of 68 or 76 residues. The molecule also contained 20 Asn-linked glycosylation sites within residues 1 to 287, a membrane-spanning region from residues 347 to 370, and a carboxyl-terminal cytoplasmic domain of 12 residues. The biochemical properties and amino acid sequence of mLAMP-1 were highly similar to those of two other molecules that have been studied as cell surface onco-differentiation antigens: a highly sialylated, polylactosaminoglycan-containing glycoprotein isolated from human chronic myelogenous leukemia cells (Viitala, J., Carlsson, S.R., Siebert, P.D. and Fukuda, M., manuscript submitted), and the mouse gp130 (P2B) glycoprotein, in which an increase in β1-6 branching of Asn-linked oligosaccharides has been correlated with metastatic potential in certain tumor cells (Dennis, J. W., Laferte, S., Waghorne, C., Breitman, M.L., and Kerbel, R.S. [1987] Science 236, 582-585).

Keywords: gene mapping, (K7)

Glycoproteins localized primarily in the limiting membrane of lysosomes have recently been identified in mouse (1-4), rat (5-7), chicken (8), and human[1] cells. These major constituents of the lysosomal membrane represent a significant fraction of the total cell membrane glycoprotein and are notable for their extensive glycosylation. A number of the molecules consist of a ~40-kDa core polypeptide substituted with up to 20 Asn-linked complex-type oligosaccharides. The composition of these oligosaccharides differs markedly in various cells, as indicated by a broad range in apparent M_r of the mature glycoproteins (100 to 150 kDa).

We report here the isolation and sequencing of a cDNA clone encoding the mouse LAMP-1 (mLAMP-1) lysosomal membrane glycoprotein and the predicted primary structure of the molecule. The amino acid sequence of this protein is compared to those of mouse and human tumor cell glycoproteins containing oligosaccharides whose unique structures have been found to be associated with cell differentation and oncogenesis (9,10).

EXPERIMENTAL PROCEDURES

Analysis of Amino Acid Sequences of mLAMP-1 - Mouse mLAMP-1 was purified by monoclonal antibody affinity chromatography from 3T3 and P388 cells as described previously (11), with the following modifications: Prior to elution of the protein from the antibody column, the column was washed with borate buffer (100 mM Na borate, pH 8.5, with 1 M NaCl) containing 0.5% octylglucoside (Calbiochem) instead of Triton X-100. The protein was eluted with 100 mM diethylamine (pH 11.5) containing 0.5% octylglucoside into a neutralizing solution containing 0.1 volume of 0.5 M NaH_2PO_4 (pH 7.4.). Fractions containing the purified antigen were pooled and concentrated by negative pressure dialysis against 20 mM $NaPO_4$ (pH 7.5) containing 0.25% octylglucoside and fractionated on a TSK 3000 size

separation high pressure liquid chromatography (HPLC)[2] column in the same buffer. Fractions containing pure antigen were pooled and concentrated to 50 ul in 0.1% SDS using a Centricon microcentrator (Amicon). The NH_2-terminal amino acid sequence of pure mLAMP-1 from 3T3 cells or P388 cells (300 pmoles from each) was determined in duplicate by use of an automated gas-phase sequencer (Applied Biosystems) (12,13). Sequence analysis was also performed on peptides of the affinity-purified protein prepared by digestion with trypsin after reduction with dithiothreitol and alkylation with iodoacetamide (14). The peptide fragments were separated by reversed phase HPLC on a Vydac C-4 column.

A mouse embryo BALB/c 3T3 cDNA expression library in λ gt11 (Clontech Laboratories) was screened with polyclonal antisera raised against purified mLAMP-1 glycoprotein, as described (15,16). Approximately 3 X 10^5 phages were plated in NZY agarose and incubated at 42^0 C for 3 h. The plates were then overlaid with nitrocellulose filters impregnated with 10 mM isopropylthiogalactoside and incubated for 2.5 h at 37^0 C. Filters were blocked with fetal bovine serum and incubated with a 1:50 dilution of the polyclonal antiserum that had previously been absorbed twice using λ gt11 phage without inserts (17). An avidin/biotin/alkaline phosphatase detection system was used to detect antibody-binding clones (18). Putative positive clones were selected and plaque purified (19). DNA was isolated and digested with EcoRI, and the cDNA insert (1800 bp) was subcloned into M13 mp9 (20,21). Deletion clones were prepared (22) and overlapping subclones were sequenced by the dideoxy method (23) in duplicate: once with DNA polymerase I, Klenow fragment, and once with T7 DNA polymerase. Areas of overlap between the clones usually included 20-40 bp. Portions of the sequence were derived independently from both DNA strands.

RESULTS

mLAMP-1 was purified from both mouse embryo 3T3 and mouse P388 histiocytic lymphoma cell lines by antibody affinity chromatography (11) and TSK 3000 size separation HPLC. The purified fractions from each cell type contained a single protein of >90% purity, as assessed by polyacrylamide gel electrophoresis and silver staining or by autoradiography of ^{125}I-labeled protein (2). The NH_2-terminal amino acid sequences of the purified proteins were determined by automated Edman degradation and analysis of the phenylthiohydantoin-amino acid products. Although mLAMP-1 molecules from 3T3 and P388 cells were markedly heterogeneous and differed significantly in apparent M_r, the provisional NH_2-terminal amino acid sequences of the two molecules were identical through residue 33: Leu Phe Glu Val Lys Asn X Gly Thr Thr X Ile Met Ala Ser Phe Ser Ala X Phe Leu Thr Thr Tyr Glu Thr Ala X Gly Ser Gln Ile Val. The predicted amino acids for the unidentified residues, Asn at the potential N-glycosylation sites at positions 7 and 28 and Cys or Ser at positions 11 and 19, were confirmed by nucleotide sequence, as shown below. Sequence analysis was also performed on three peptides obtained by reversed phase HPLC of the products of tryptic digestion.

A mouse embryo 3T3 cDNA expression library in λgt11 was screened with polyclonal antiserum recognizing the purified LAMP-1 glycoprotein. A positive clone containing a cDNA insert of 1800 bp was purified, and the cDNA insert was subcloned in M13 vectors and sequenced (Fig. 2). The amino acid sequences deduced from the cDNA sequence were identical to those obtained by protein sequencing of residues 3 to 33 of the NH_2-terminus and of three tryptic peptides (residues 67 to 76, 80 to 87, and 105 to 116) of the purified protein, thus confirming the identity of the cDNA. The cDNA lacked the sequence corresponding to the two NH_2-terminal amino acid

residues identified by protein sequencing, possibly due to an artifact of library construction. The cDNA also lacked a 5'-untranslated sequence, initiator methionine, and signal sequence. A termination sequence and a 3'-untranslated sequence of 711 nucleotides followed nucleotide 1155. A putative transcription termination signal, AATAAA (25), was present at nucleotide residues 1860 to 1865.

The polypeptide deduced from the cDNA sequence contained 382 amino acids, with an M_r of 41,509 (Fig. 1). The NH_2-terminal 90% of the sequence formed two homology units comprising residues 1 to 165 and 188 to 344. These two sequences showed 37 identities out of 154 possible matches, with an ALIGN score (25) of 9.34 SD (Fig. 2). The homology units were separated by a potential hinge region that was rich in prolines, was strongly hydrophilic, and had no α-helical component. These NH_2-terminal domains also contained a total of 20 potential Asn-linked (Asn-X-Ser or Asn-X-Thr) glycosylation sites and eight cysteine residues, four of which were uniformly spaced at intervals of 36 to 38 residues and two at intervals of 68 and 76 residues. A strong hydrophobic region of 24 amino acids, residues 347 to 370, occurred near the carboxyl-terminus (Fig. 3); the hydrophobicity and absence of glycosylation sites and of positively charged residues make this carboxyl domain a strong candidate for the membrane-spanning region. The remaining short carboxyl-terminal sequence, residues 371 to 382, would then constitute the cytoplasmic domain of the molecule. The secondary structure characteristics of the protein, predicted by the algorithms of Garnier et al. (26) and Kyte and Doolittle (27), included 37% β-sheet, 21% β-turns, 23% random coil and 19% α-helix.

We have found that mLAMP-1 is closely similar in biochemical properties and amino acid sequence to two leukemia cell glycoproteins that have been studied as possible onco-differentiation antigens. The first of these

molecules is gp130 (P2B), a highly glycosylated glycoprotein with a core polypeptide of about 33 kDa that has been isolated from the mouse metastatic tumor cell line MDAY-D2[3]. The P2B molecule is of interest because changes in its concentration of β1-6-branched Asn-linked oligosaccharides have been shown to correlate with the metastatic potential of the cells in which it is expressed (9). The provisional NH_2-terminal amino acid sequence of 20 residues from P2B is 90% identical to that of mLAMP-1, and our monoclonal antibody against mLAMP-1 cross-reacts with P2B.[3]

mLAMP-1 is also closely similar to a highly sialylated, polylactosaminoglycan-containing glycoprotein that has been studied as a differentiation marker in hematopoietic cells (10) and has been purified from human chronic myelogenous leukemia cells.[4] In an exchange of reagents with Dr. M. Fukuda, La Jolla Cancer Research Foundation, it was found that this purified human leukemia cell glycoprotein was immunoprecipitated by our monoclonal antibody directed against hLAMP-1, the human analogue of mLAMP-1.[1] A cDNA clone encoding the leukemia cell glycoprotein has been isolated and sequenced,[4] and the protein predicted from these data contains an NH_2-terminal sequence (residues 14 to 37) which is identical to that of hLAMP-1, as determined by direct protein sequence analysis. The human leukemia cell glycoprotein is also very similar to mLAMP-1 (Fig. 4). These two proteins have an identical sequence in 66% of their residues (252 of 382 residues), all cysteines are conserved, and 13 of the 20 potential Asn-linked glycosylation sites on the mouse protein are conserved in the human protein. This leukemia cell protein and mLAMP-1 also share 20 conservative substitutions (I to V, D to E, or R to K). Moreover, the carboxyl-termini of the two molecules are almost totally conserved, with a stretch of 35 identical amino acids broken only by two neutral I to V substitutions.

A search of The National Biomedical Foundation Protein Databank did not

reveal significant amino acid sequence similarity of mLAMP-1 to any additional proteins.

DISCUSSION

The amino acid sequence predicted from the cDNA clone of mLAMP-1 further elucidates the structure of the LAMP-1 class of lysosomal membrane glycoproteins. The 41,509-Da mLAMP-1 molecule contained 20 potential Asn-linked glycosylation sites (consistent with previous predictions of 18 to 20 Asn-linked oligosaccharides [4,5]), all occurring in the NH_2-terminal 70% of the molecule. A single highly hydrophobic and putative membrane-spanning region occurred at residues 347 to 370 near the carboxyl terminus, suggesting that the majority of the mLAMP-1 molecule resides in the lumen of the lysosome and only a short (12 amino acid) domain extends into the cytoplasm. An internal homology was found between the two halves of the intralumenal domain, providing evidence that these sequences were derived by gene duplication. The principal basis for the high homology score obtained was the alignment of Cys and neighboring residues in the homology units. The regular spacing of these residues suggested the occurrence of a pair of immunoglobulin-like (V-related) loops generated by disulfide bonds linking Cys 50 to 125 and 235 to 303. The interval between Cys residues (approximately 70 amino acids) was appropriate for a V-related sequence, and the carboxyl-terminal region of each homology unit contained the Y-X-C group common to immunoglobulin superfamily sequences (28). However, none of the other sequence patterns characteristic of the immunoglobulin superfamily were found in the remaining portion of the two mLAMP-1 domains. Furthermore, ALIGN scores indicated that there was no significant sequence identity between the 20 residues preceding and following the Cys residues of the two LAMP domains and those of immunoglobulin superfamily V or V-related domains. More information concerning the disulfide bonding and 3-dimensional structure of the LAMP-1 molecule must be obtained before the possibility of a relationship between LAMP and Ig-related molecules can be

determined. Such a relationship would be of particular interest because it would suggest a potential receptor function for the LAMP molecule (28).

The interspecies sequence similarity (66% identity) of mLAMP-1 and the human leukemia cell glycoprotein corresponding to human hLAMP-1 also extends to the chicken LEP100 molecule (29) which has approximately 44% identity to mLAMP-1. Overall, 159 residues are common to the three proteins, including each of the eight cysteine residues. Moreover, the three molecules are virtually identical in the sequence of their hydrophobic domains (residues 347 to 370) and their short carboxyl-terminal cytoplasmic segments, implying that these carboxyl-terminal domains may have functional roles. The chicken cDNA sequence has in addition a 5'-untranslated region of 466 bases and an initiator methionine and signal sequence of 18 residues that are lacking in the mLAMP-1 cDNA. An initiator methionine is also present in the human leukemia cell cDNA sequence, but an apparent signal sequence of eight residues is followed by a gap that includes 12 amino acids of the NH_2-terminus of the protein. The close similarity between the mouse, chicken, and human proteins suggests that these are homologous molecules of the different species or that they are closely related proteins that have arisen by gene duplication.

The close similarity of the mouse P2B and the human leukemia cell glycoproteins to the LAMP-1 molecules has identified these onco-differentiation antigens as lysosomal membrane glycoproteins[3,4]. The role of these glycoproteins as cell-surface antigens implies that movement of the molecules from the lysosome to the plasma membrane must occur; this phenomenon has already been demonstrated for the chicken LEP-100 molecule (8). A small fraction of the LAMP-1 molecules (<10%) has been found on the surface of the mouse macrophage-like cell line P388 (1) and the human U937 and HL-60 myelomonocytic leukemia cells.[1] This expression of LAMP-1 on the

cell surface may be related to the state of differentiation of the cell and may reflect an alteration in the oligosaccharide composition of the molecules as well as selective movement of a fraction of the LAMP-1 molecules to the plasma membrane. Such a model is consistent with the marked difference in the concentration of LAMP-1 observed on the surface of various human blood cells (with the highest levels found on U937 and HL-60 myelomonocytic leukemia cell lines and none on normal peripheral blood monocytes) and the reduction in the cell-surface expression of this molecule that occurs when the leukemia cells are induced to differentiate to cells resembling mature macrophages or granulocytes.[1] Alterations in N-linked oligosaccharide composition of particular glycoproteins (including incomplete processing, increased sialic acid and polylactosamine content, and increased branching of the trimannosyl core) have been observed in a number of differentiating or oncogenically transformed cells (30-33). These alterations have been shown to affect a variety of processes, including substrate adhesion and tumor progression (9, 34-36). In particular, experiments have linked an increase in sialic acid and polylactosamine content to an increase in metastatic potential of certain tumor cells (9), and the P2B homologue of LAMP-1 has been identified as the major cell component in which such an increase occurs.[3] Furthermore, it has been reported that purified P2B binds to immobilized collagen and that removal of sialic acid, polylactosamine or complete Asn-linked chains from the molecule enhances its binding to extracellular matrix proteins, including collagens, laminin, and fibronectin.[5] These finding suggest that <u>in vivo</u> alterations in the oligosaccharide composition of LAMP-1 and related glycoproteins, in combination with the expression of the modified molecules on the cell surface, could have significant consequences for processes such as adhesion, metastasis, and immune recognition.

ACKNOWLEDGMENTS

We thank Drs. D. Warder and R. Reed for helpful suggestions, Drs. M. Amzel, A. Williams, and M. Guarnieri for aid in analyzing protein sequence data, Drs. J. Dennis, M. Fukuda and D. Fambrough for supplying copies of their manuscripts prior to publication, Dr. M. Strand for critical reading of the manuscript, and Dr. D. McClellan for editorial assistance.

FIGURE LEGENDS

Fig. 1 cDNA and deduced amino acid sequence of mLAMP-1 lysosomal membrane glycoprotein. Nucleotides and amino acids are numbered at the left of each line. The predicted protein sequence from amino acid residue 3 to 382 is shown below the DNA sequence. ++, the NH_2-terminal amino acids (Leu and Phe) identified by protein sequencing; _____, regions for which direct amino acid sequence of LAMP-1 was obtained; #, Cys residues; *, potential Asn-linked glycosylation sites; ===, proposed membrane-spanning region; ---, stop codon; _____, putative transcription termination sequence (24).

Fig. 2 Internal alignment of mLAMP-1 sequences. The mLAMP-1 sequences of residues 1 to 165 and 188 to 344 were compared by the ALIGN program (25), with a bias of 6 and penalty of 6, against 150 random runs. The alignment score = 9.34 SD. #, Cys residue.

Fig. 3 Hydropathy plot of mLAMP-1. Hydropathy values were determined using the algorithm of Kyte and Doolittle (27), with a window of n=11. Each point represents a single amino acid. Potential Asn-linked glycosylation sites (arrows) of type Asn-X-Ser or Asn-X-Thr and the predicted membrane-spanning region (box) are indicated on the stick diagram below the hydropathy plot.

Fig. 4 Amino acid sequence similarity between mLAMP-1 and a human leukemia cell glycoprotein.[4] The mouse sequence (top) is represented in the standard one-letter code. For the human sequence only amino acids different from the mouse protein are shown. Alignment was carried out using the accepted point mutation values (PAM) of Dayhoff et al. (25). Identical

residues are indicated by (*), gaps by (.). The human sequence is that reported by Viitala et al.[4]

FOOTNOTES

1. Mane, S.M., Marzella, L., Bainton, D.F., Holt, V.K., Cha, Y., Hildreth, J.E.K., and August, J.T. Purification and characterization of human lysosomal membrane glycoproteins and their presence in the plasma membrane of myelomonocytic leukemia cells (manuscript in preparation).

2. The abbreviation used is: HPLC, high pressure liquid chromatography.

3. Laferte, S., and Dennis, J.W. Purification of two glycoproteins expressing β1-6 branched Asn-linked oligosaccharides commonly associated with the malignant phenotype (manuscript submitted).

4. Viitala, J., Carlsson, S.R., Siebert, P.D. and Fukuda, M. A human lysosomal membrane glycoprotein consists of two heavily N-glycosylated domains separated by a hinge region (manuscript submitted).

5. Laferte, S., and Dennis, J.W. Collagen-binding activities of two membrane glycoproteins are modulated by their glycosylation (manuscript submitted).

REFERENCES

1. Chen, J.W., Murphy, T.L., Willingham, M.C., Pastan, I., and August, J.T. (1985) J. Cell Biol. 101, 85-95
2. Chen, J.W., Pan, W., D'Souza, M.P., and August, J.T. (1985) Arch. Biochem. Biophys. 239, 574-586
3. Chen, J.W., Chen, G.L., D'Souza, M.P., Murphy, T.L., and August, J.T. (1986) Biochem. Soc. Symp. 51, 97-112
4. D'Souza, M.P., and August, J.T. (1986) Arch. Biochem. Biophys. 249, 522-532
5. Lewis, V., Green, S.A., Marsh, M., Vihko, P., Helenius, A., and Mellman, I. (1985) J. Cell Biol. 100, 1839-1847
6. Barriocanal, J.G., Bonifacino, J.S., Yuan, L., and Sandoval, I.V. (1986) J. Biol. Chem. 261, 16755-16763
7. Reggio, H., Bainton, D., Harms, R., and Louvard, D. (1984) J. Cell Biol. 99, 1511-1526
8. Lippincott-Schwartz, J., and Fambrough, D.M. (1986) J. Cell Biol. 102, 1593-1605
9. Dennis, J.W., Laferte, S., Waghorne, C., Breitman, M.L., and Kerbel, R.S. (1987) Science 236, 582-585
10. Fukuda, M. (1985) Biochem. Biophys. Acta 780, 119-150
11. Hughes, E.N., and August, J.T. (1982) J. Biol. Chem. 257, 3970-3977
12. Hewick, R.M., Hunkapiller, M.W., Hood, L.E., and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990-7997
13. Hunkapillar, M.W., and Hood, L.E. (1983) Science 219, 650-659
14. Speicher, D.W., Davis, G., and Marchesi, V.T. (1983) J. Biol. Chem. 258, 14938-14947
15. Young, R.A., and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194-1198

16. Huyng, T.V., Young, R.A., and Davis, R.W. (1985) DNA Cloning: A Practical Approach, Vol. 1, pp. 49-78, IRL Press, Oxford

17. Helfman, D.M., Feramisco, J.R., Fiddes, J.C., Thomas, G.P., and Hughes, S.H. (1984) BRL Focus 6, 1-5

18. DeWet, J.R., Fukushima, H., Dewji, N.N., Wilcox, E., O'Brien, J.S., and Helinski, D.R. (1984) DNA 3, 437-447

19. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, New York

20. Messing, J. (1983) Methods Enzymol. 101, 20-78

21. Messing, J., and Vieira, J. (1982) Gene 19, 269-276

22. Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985) Plasmid 13, 31-40

23. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Rose, B. (1980) J. Mol. Biol. 143, 161-178

24. Birnstiel, M.L., Busslinger, M., and Strub, K. (1985) Cell 41, 349-359

25. Dayhoff, M.O., Barker, W.C., and Hunt. L.T. (1983) Methods Enzymol. 91, 524-545

26. Garnier, J., Osguthorpe, D.J., and Robson, B. (1978) J. Mol. Biol. 120, 97-120

27. Kyte, J., and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132

28. Williams, A.F. (1987) Immunol. Today 8, 298-303

29. Fambrough, D.M., Takeyasu, K., Lippincott-Schwartz, J., and Siegel, N.R., with the technical assistance of Somerville, D. (1988) J. Cell Biol., in press

30. Feizi, T. (1985) Nature 314, 53-57

31. Cossu, G., and Warren, L. (1983) J. Biol. Chem. 258, 5603-5607

32. Reading, C.L., and Hutchins, J.T. (1985) Cancer Metastasis Rev. 4, 221-260

33. Hubbard, S.C. (1987) J. Biol. Chem. **262**, 16403-16411

34. Zhu, B.C.R., and Laine, R.A. (1985) J. Biol. Chem. **260**, 4041-4045

35. Humphries, M.J., Matsumoto, K., White, S.L., and Olden, K. (1986) Proc. Natl. Acad. Sci. USA **83**, 1752-1756

36. Mercurio, A.M. (1986) Proc. Natl. Acad. Sci. USA **83**, 2609-2613

FIGURE 1

```
   1 TAC CGC GAA ATT CCG GAG GTG AAA AAC AAT GGC ACA ACG TGT ATA ATG
                 Leu Phe Glu Val Lys Asn Asn Gly Thr Thr Cys Ile Met
                  +   +                *                 #
  49 GCC AGC TTC TCT GCC TCC TTT CTG ACC ACC TAC GAG ACT GCG AAT GGT
  14 Ala Ser Phe Ser Ala Ser Phe Leu Thr Thr Tyr Glu Thr Ala Asn Gly
                                                      *

  97 TCT CAG ATC GTG AAC ATT TCC CTG CCA GCC TCT GCA GAA GTA CTG AAA
  30 Ser Gln Ile Val Asn Ile Ser Leu Pro Ala Ser Ala Glu Val Leu Lys
                    *

 145 AAT GGC AGT TCT TGT GGT AAA GAA AAT GTT TCT GAC CCC AGC CTC ACA
  46 Asn Gly Ser Ser Cys Gly Lys Glu Asn Val Ser Asp Pro Ser Leu Thr
   *                #                 *

 193 ATT ACT TTT GGA AGA GGA TAT TTA CTG ACA CTC AAC TTC ACA AAA AAT
  62 Ile Thr Phe Gly Arg Gly Tyr Leu Leu Thr Leu Asn Phe Thr Lys Asn
                                                *                 *

 241 ACA ACA CGT TAC AGT GTC CAG CAT ATG TAT TTT ACA TAT AAC TTG TCA
  78 Thr Thr Arg Tyr Ser Val Gln His Met Tyr Phe Thr Tyr Asn Leu Ser
                                                          *

 289 GAT ACA GAA CAT TTT CCC AAT GCC ATC AGC AAA GAG ATC TAC ACC ATG
  94 Asp Thr Glu His Phe Pro Asn Ala Ile Ser Lys Glu Ile Tyr Thr Met

 337 GAT TCC ACA ACT GAC ATC AAG GCA GAC ATC AAC AAA GCA TAC CGG TGT
 110 Asp Ser Thr Thr Asp Ile Lys Ala Asp Ile Asn Lys Ala Tyr Arg Cys
                                                                #

 385 GTC AGT GAT ATC CGG GTC TAC ATG AAG AAT GTG ACC GTT GTG CTC CGG
 126 Val Ser Asp Ile Arg Val Tyr Met Lys Asn Val Thr Val Val Leu Arg
                                  *

 433 GAT GCC ACT ATC CAG GCC TAC CTG TCG AGT GGC AAC TTC AGC AAG GAA
 142 Asp Ala Thr Ile Gln Ala Tyr Leu Ser Ser Gly Asn Phe Ser Lys Glu
                                              *

 481 GAG ACA CAC TGC ACA CAG GAT GGA CCT TCC CCA ACC ACT GGG CCA CCC
 158 Glu Thr His Cys Thr Gln Asp Gly Pro Ser Pro Thr Thr Gly Pro Pro
                #

 529 AGC CCC TCA CCA CCA CTT GTG CCC ACA AAC CCC ACT GTA TCC AAG TAC
 174 Ser Pro Ser Pro Pro Leu Val Pro Thr Asn Pro Thr Val Ser Lys Tyr
                                      *

 577 AAT GTT ACT GGT AAC AAC GGA ACC TGC CTG CTG GCC TCT ATG GCA CTG
 190 Asn Val Thr Gly Asn Asn Gly Thr Cys Leu Leu Ala Ser Met Ala Leu
   *                              #

 625 CAA CTG AAT ATC ACC TAC CTG AAA AAG GAC AAC AAG ACG GTG ACC AGA
 206 Gln Leu Asn Ile Thr Tyr Leu Lys Lys Asp Asn Lys Thr Val Thr Arg
           *                          *

 673 GCG TTC AAC ATC AGC CCA AAT GAC ACA TCT AGT GGG AGT TGC GGT ATC
 222 Ala Phe Asn Ile Ser Pro Asn Asp Thr Ser Ser Gly Ser Cys Gly Ile
           *                  *                          #

 721 AAC TTG GTG ACC CTG AAA GTG GAG AAC AAG AAC AGA GCC CTG GAA TTG
 238 Asn Leu Val Thr Leu Lys Val Glu Asn Lys Asn Arg Ala Leu Glu Leu

 769 CAG TTT GGG ATG AAT GCC AGC TCT AGC CTG TTT TTC TTG CAA GGA GTG
 254 Gln Phe Gly Met Asn Ala Ser Ser Ser Leu Phe Phe Leu Gln Gly Val
                    *

 817 CGC TTG AAT ATG ACT CTT CCT GAT GCC CTA GTG CCC ACA TTC AGC ATC
 270 Arg Leu Asn Met Thr Leu Pro Asp Ala Leu Val Pro Thr Phe Ser Ile
           *

 865 TCC AAC CAT TCA CTG AAA GCT CTT CAG GCC ACT GTG GGA AAC TCA TAC
 286 Ser Asn His Ser Leu Lys Ala Leu Gln Ala Thr Val Gly Asn Ser Tyr
           *

 913 AAG TGC AAC ACT GAG GAA CAC ATC TTT GTC AGC AAG ATG CTC TCC CTC
 302 Lys Cys Asn Thr Glu Glu His Ile Phe Val Ser Lys Met Leu Ser Leu
        #

 961 AAT GTC TTC AGT GTG CAG GTC CAG GCT TTC AAG GTG GAC AGT GAC AGG
 318 Asn Val Phe Ser Val Gln Val Gln Ala Phe Lys Val Asp Ser Asp Arg

1009 TTT GGG TCT GTG GAA GAG TGT GTT CAG GAT GGT AAC AAC ATG TTG ATC
 334 Phe Gly Ser Val Glu Glu Cys Val Gln Asp Gly Asn Asn Met Leu Ile
                              #                              ----------

1057 CCC ATT GCT GTG GGC GGT GCC CTG GCA GGG CTG ATC CTC ATC GTC CTC
 350 Pro Ile Ala Val Gly Gly Ala Leu Ala Gly Leu Ile Leu Ile Val Leu
     ------------------------------------------------------------

1105 ATT GCC TAC CTC ATT GGC AGG AAG AGG AGT CAC GCC GGC TAT CAG ACC
 366 Ile Ala Tyr Leu Ile Gly Arg Lys Arg Ser His Ala Gly Tyr Gln Thr
     ==================

1153 ATC TAG CCTGGTGGGCAGGTGCACCAGAGATGCACAGGGGCCTGTTCTCACATCCCCAAGC
 382 Ile ---

1214 TTAGATAGGTGTGGAAGGGAGGCACACTTTCTGGCAAACTGTTTTAAAATCTGCTTTATCAAA

1277 TGTGAAGTTCATCTTGCAACATTTACTATGCACAAAGGAATAACTATTGAAATGACGGTGTTA

1340 ATTTTGCTAACTGGGTTAAATATTGATGAGAAGGCTCCACTGATTTGACTTTTAAGACTTGGT

1403 GTTTGGTTCTTCATTCTTTTACTCAGATTTAAGCCTATCAAAGGGATACTCTGGTCCAGACCT

1466 TGGCCTGGCAAGGGTGGCTGATGGTTAGGCTGCACACACTTAAGAAGCAACGGGAGCAGGGAA

1529 GGCTTGCACACAGGCACGCACAGGGTCAACCTCTGGACACTTGGCTTGGGCTACCTGGCCTTG

1592 GGGGGGCTGAACTCTGGCATCTGGCTGGGTACACACCCCCCCAATTTCTGTGCTCTGCCACCC

1655 GTGAGCTGCCACTTTCCTAAATAGAAAATGGCATTATTTTTATTTACTTTTTTGTAAAGTGAT

1718 TTCCAGTCTTGTGTTGGCGTTCAGGGTGGCCCTGTCTCTGCACTGTGTACAATAATAGATTCA

1781 CACTGCTGACGTGTCTTGCAGCGTAGGTGGGTTGTACACTGGGCATCAGCTCACGTAATGCAT

1844 TGCCTGTAACGATGCTAATAAAAA
```

FIGURE 2

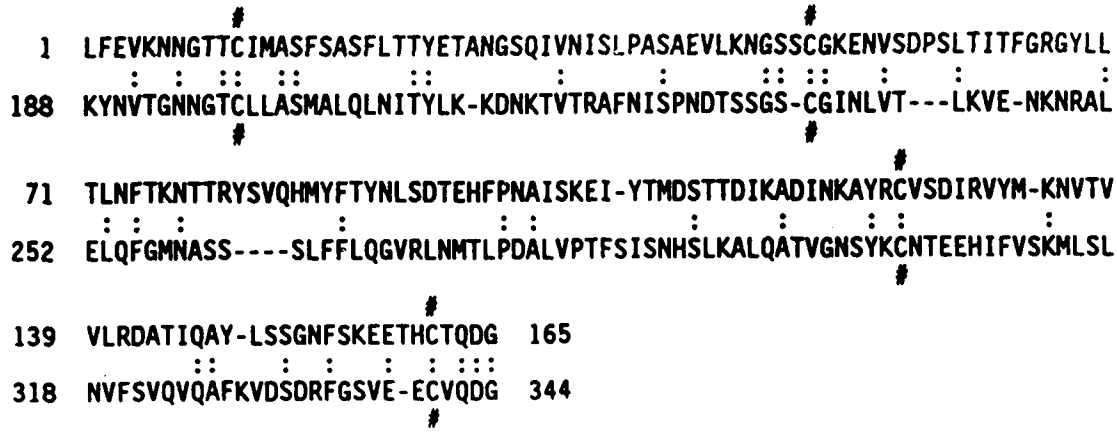

1 LFEVKNNGTTCIMASFSASFLTTYETANGSQIVNISLPASAEVLKNGSSCGKENVSDPSLTITFGRGYLL
188 KYNVTGNNGTCLLASMALQLNITYLK-KDNKTVTRAFNISPNDTSSGS-CGINLVT---LKVE-NKNRAL
71 TLNFTKNTTRYSVQHMYFTYNLSDTEHFPNAISKEI-YTMDSTTDIKADINKAYRCVSDIRVYM-KNVTV
252 ELQFGMNASS----SLFFLQGVRLNMTLPDALVPTFSISNHSLKALQATVGNSYKCNTEEHIFVSKMLSL
139 VLRDATIQAY-LSSGNFSKEETHCTQDG 165
318 NVFSVQVQAFKVDSDRFGSVE-ECVQDG 344

FIGURE 3

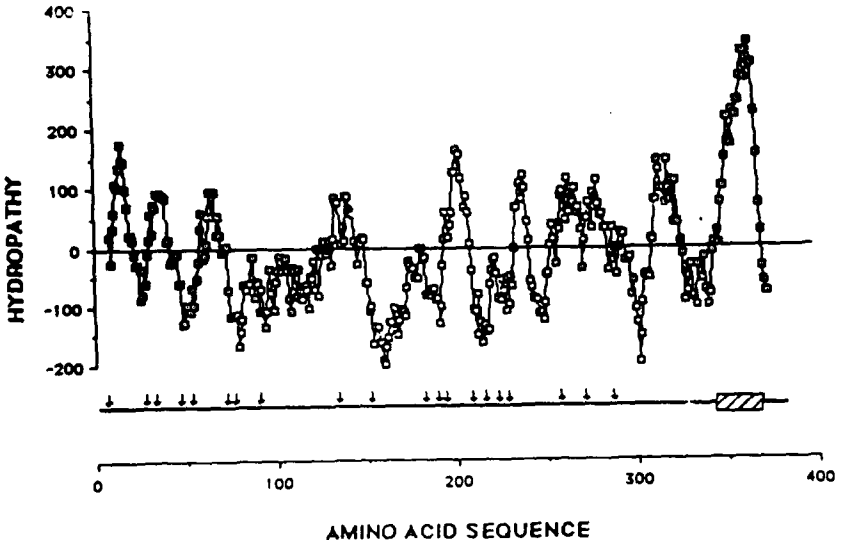
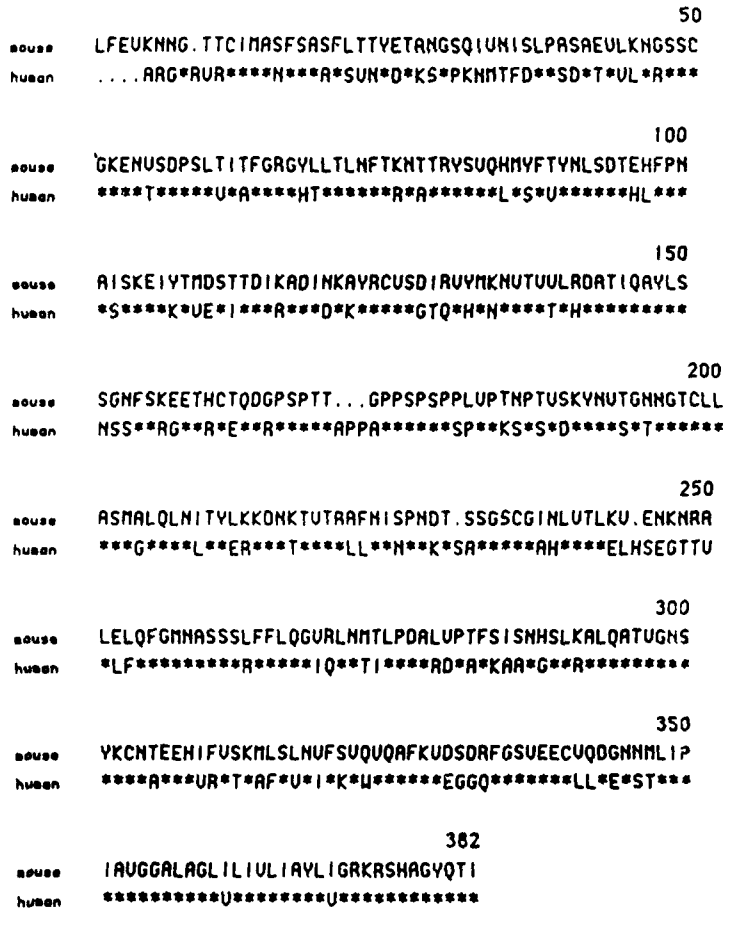

400
300
200
100
0
-100
-200
HYDROPATHY
0
100
200
300
400
AMINO ACID SEQUENCE

FIGURE 4


50
mouse LFEVKNNG.TTCIMASFSASFLTTVETANGSQIVNISLPASAEVLKNGSSC
human ....ARG*RVR****N***A*SVN*D*KS*PKMMTFD**SD*T*VL*R***
100
mouse GKENVSDPSLTITFGRGYLLTLNFTKNTTRYSVQHMYFTYNLSDTEHFPN
human ****T*****V*A****HT******R*R******L*S*V******HL***
150
mouse AISKEIYTMDSTTDIKADINKAYRCVSDIRVYMKNVTVVLRDATIQAVLS
human *S****K*VE*I***R***D*K*****GTQ*H*N****T*H*********
200
mouse SGNFSKEETHCTQDGPSPTT...GPPSPSPPLVPTNPTVSKYNVTGNNGTCLL
human NSS**RG**R*E**R*****APPA******SP**KS*S*D****S*T******
250
mouse ASMALQLNITYLKKDNKTVTRAFNISPNDT.SSGSCGINLVTLKV.ENKNRA
human ***G****L**ER***T****LL**N**K*SA*****AH****ELHSEGTTV
300
mouse LELQFGMNASSSLFFLQGVRLMMTLPDALVPTFSISNHSLKALQATVGNS
human *LF*********R******IQ**TI****RD*A*KAA*G**R*********
350
mouse YKCNTEEHIFVSKMLSLNVFSVQVQAFKVDSDRFGSVEECVQDGNNMLIP
human ****A***VR*T*AF*V*I*K*W******EGGQ*******LL*E*ST***
382
mouse IAVGGALAGLILIVLIAYLIGRKRSHAGYQTI
human **********V********V**********