

# Nucleotide sequence of a cDNA encoding mouse cathepsin D

J.F.Diedrich, K.A.Staskus, E.F.Retzels and A.T.Haase\*

Department of Microbiology, University of Minnesota, Minneapolis, MN 55455, USA

Submitted June 13, 1990

EMBL accession no. X53337

We report the nucleotide sequence of mouse cathepsin D mRNA derived from clones in a  $\lambda$ gt10 cDNA library constructed from C57BL/6J mouse brain. Cathepsin D is a lysosomal aspartyl protease, and this cDNA was isolated while screening for upregulated genes in scrapie (manuscript in preparation). The sequence is 1893 bp and contains an open reading frame of 1230 bp with an initiation codon at nucleotide position 7 and a termination codon at 1237. The predicted amino acid sequence (410 aa) shows 81% similarity with the human kidney cathepsin D (1) and 82% with the porcine spleen cathepsin D (2, 3). Active-

site aspartyl residues and the presumed polyadenylation signal within the 3' untranslated region are underlined.

## REFERENCES

1. Faust, P.L., Kornfeld, S. and Chirgwin, J.M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4910–4914.
2. Shewale, J.G. and Tang, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3703–3707.
3. Erickson, A.H., Conner, G.E. and Blobel, G. (1981) *J. Biol. Chem.* **256**, 11224–11231.

```

M K T P G V L L L I L G L L A S S S F A I I R I P L R K F T S I R R T M T E
CGGACCATGAGACTCCCGGGGCTTCTGCTCCTCATTCTCGGCTCCTGGCTTCTGCTCCTTCGGGATTATCAGAAATCCCTCTGGGCAAGTTACACATCTATCCGTCGGACTATGACGGAG 120
V G G S V E D L I L K G P I T K Y S M Q S S P K T T E P V S E L L K N Y L D A Q
GTGGCGGGCTCTGTGAGAGCTGATCCTCAAGGGCCCATCAACAGTACTCCATGCACTCATCGCTAAGACCAGCGGAGCCAGTGTGAGAGTACTCAAAAATACCTGGATGCCAG 240
Y Y G D I G I G T P P Q C F T V V F D T G S S N L W V P S I H C K I L D I A C W
TACTATGGGATATGGGATCGGAGCCCGCCGAGTGTTCACAGTGGCTTTGATACTGGCTCCTCAACCTGTGGGTCCCTCCATTCAATGCAAGATACTTGACATAGCCTGCTGG 360
V H H K Y N S D K S S T Y V K N G T S F D I H Y G S G S L S G Y L S Q D T V S V
GTCCACCATAAGTACAACAGTGCAGATCCAGCACCTATGTGAGAACGGCAGCTCCTTTGACATCCACTACGGCTCAGGAGCCCTCTCTGGGTACTCGAGCCAGGACACTGTATCGGTT 480
P C K S D Q S K A R G I K V E K Q I F G E A T K Q P G I V F V A A K F D G I L G
CCATGTAGTCTGACCACTCAAGGCCAAGAGGTATCAAGGTCCGAGAAACAGATCTTTGGAGAGCCCAACCAAGCCAGCCCTGGAAATCGTATTGTTGACAGCCAAAGTTGATGGCATCTGGGGC 600
M G Y P H I S V N N V L P V F D N L M Q Q K L V D K N I F S F Y L H R D P E G Q
ATGGGCTAGCCTCATATCTCTGTAAACAGGTGCTTCCGGTCTTTGACAACTGATGCAACAGAGCTGGTGACAGAACATCTTCTCCTTCTACCTGAACAGGAGCCAGCCAGGGCCAA 720
P G G G E L M L G G T D S K Y Y H G E L S Y L N V T R K A Y W Q V H M D Q L E V G
CCCGGAGGAACTAAATGCTTGGTGGCACTGACTCCAACTACTACCGGGGAGCTGTCCTACTGAAACCTCACTCGAAGGCTACTGGCAGGTGCACATGGACAGTGGAGGTGGGC 840
N E L T L C K G G C E A I V D T G T S L L V G P V E E V K E L Q K A I G A V P L
AATGAGCTGACCTCTGCAAGGGAGGCTGTGAGGCTATTGTGACACAGGACATCTCTCTGCTGGGCTCTGTGAGAGCTGAAAGGCTGCAGAGGCCATGGGGAGTGCCTCTT 960
I Q G E Y M I P C E K V S S L P T V Y L K L G G K N Y E L H P D K Y I L K V S Q
ATCCAGGGTGAATATGATTCCTTGTGAGAGGTGTCCAGCCTGCCACGGTCTACCTGAAGTAGGAGCCAAAATATGAACTACACCCAGACAGTATGACTCAAGGTATGCCAG 1080
G G K T I C L S G F M G M D I P P P S G P L W I L G D V F I G S Y Y T V F D R D
GGTGGAAAGCAATCTGCTGATGGCTTCATGGGAATGGACATACCCCTCCAGTGGGCGCTCTGGATCCCTGGGGATGCTTCAITGGTTCCTACTACTACTGTGTTGACAGAGAC 1200
H N R V G F A N A V V L #
AACAATAGGGTGGCTTTGGCAATGCTGTGCTACTCTCACTTCTCTCTCTCACTGTCAAGGAACTGGATCAGAGTCCAGTAGAGGAAAGCCAGCCAGCCCAATCCCTCCACTGCCCCA 1320
CTCACACATAGTACACTGCTTACTGTGCTGCAACCTTGGGGGAGACCTGGCTGGAGCTTGTCCAGCTGTCTGTGTTCTGTGTTCTCCTCAGCCCTGGGTTGAGATTGCTGCCCTCTGCT 1440
GTCTGAAGGAGCCAAAGGCCACCCAGTACACGAGGCTGCTTCAAGGGCCCTACTGGTTAAATAGCTGCTGAGATGGATTGCTTGTCTGCTGCCGCCCTTTGCTGTGTGGCAGTACTC 1560
TGAAGCAGGCAAAATGGCTTATGATCCCTCCAGAAACCTGCTCTGACCAAGCCATCACCCAGCTTGGGATGGCCACCAAGTTCCTACTGCCCCCTCCAACTCTGGCCCTGGCAAAGGCC 1680
AGAAGGTGACGAGGAGGAGCAGAGCAGAACTATGAACTGGGGGCTTACTAGGGCTTACCCCGCCCTCTGGGAGGGCAGCCCTCAGCCTGGGGTAGAGGTAGGATGA 1800
CTGACTGCTTTTGGCTGGCTCTGCTGCCCTCATCTGGGCTAGGCAGCTGGGAGCCAAAGTTCGAATACAAATAAAGTCTTTTGGGCTTG 1893

```

*Note added in proof:* Since the acceptance of this manuscript we have become aware of results by Grusby *et al.* (*Nucl. Acids Res.*, vol. **18**, p. 4008, 1990) concerning the cloning of a mouse cathepsin D cDNA. Their sequence shows one nucleotide difference (position 1681, A to T) in the 3' untranslated region when compared to the sequence given here. This sequence was accepted July 16, 1990.

\* To whom correspondence should be addressed