Volume 237, number 1,2, 191-195 FEB 06304 September 1988

Cloning and sequence analysis of a cDNA encoding the β-subunit of mouse β-hexosaminidase

Bharati Bapat*, Marguerite Ethier*+, Kuldeep Neote*+, Don Mahuran and Roy A. Gravel*+

*Research Institute, Hospital for Sick Children, 555 University Ave, Toronto, +Department of Medical Genetics and Department of Clinical Biochemistry, University of Toronto, Ontario, Canada

Received 1 August 1988

A cDNA encoding the prepro-β-polypeptide of mouse β-hexosaminidase (Hex) was isolated from a mouse lymphoblast cDNA library. The cDNA contains an open reading frame corresponding to a polypeptide of 536 amino acids which shows 74% homology with the human prepro-β-polypeptide. An examination of the amino acid sequence identifies a putative signal peptide and five possible glycosylation sites, two of which are identical to the confirmed glycosylation sites of the human β-chain. The amino acid sequence also shows a structurally similar though not identical site for internal cleavage responsible for the generation of mature β_a- and β_b-polypeptides.

cDNA cloning; Lysosome; Glycosylation; Polypeptide processing; Sequence homology; (Mouse lymphoid cell line)

1. INTRODUCTION

β-Hexosaminidase (Hex) is a lysosomal enzyme that catalyses the hydrolysis of GM_2 ganglioside and a variety of other molecules containing terminal *N*-acetyl hexosamines. The human enzyme occurs as two major isozymes, Hex A ($\alpha(\beta_a\beta_b)$) and HexB ($2(\beta_a\beta_b)$). The mature α-subunit consists of a single polypeptide chain. The mature β-subunit consists of two nonidentical polypeptides, β_a and β_b, formed through proteolytic cleavage of the pro-β-precursor in the lysosome [1]. We and others have previously reported the isolation of genomic and cDNA clones encoding the α- and β-subunits of human Hex [2-6] and defined the sites of proteolytic processing and glycosylation generating the mature α-, β_a- and β_b-polypeptide chains [7-9].

In mammalian species, Hex is expressed in all tissues. However, differences in the tissue-specific levels of Hex have been reported and regulation by steroid hormones has been suggested [10-12]. In order to study the expression and regulation of Hex at the molecular level, we have isolated a cDNA encoding the mouse β-subunit.

Here we report its nucleotide sequence and a comparison of the deduced amino acid sequence with that of the human α- and β-chains.

Correspondence address: R.A. Gravel, Research Institute, Hospital for Sick Children, 555 University Ave, Toronto, Ontario M5G 1X8, Canada

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no. Y00964

2. MATERIALS AND METHODS

A mouse lymphoid cell line (70Z/3) cDNA library, constructed in λgt11, was screened for Hex sequences with a 1 kb cDNA fragment (pHexB43) encoding the human β-subunit [4,13]. Filters were washed at a final stringency of 0.1×SSC, 0.1% SDS at 55°C. The library screening, RNA isolation and Northern blot analysis were performed as described [4]. The isolated cDNAs were subcloned into pBS (Stratagene) and restriction mapped. Sequencing was performed using the Sequenase kit (US Biochemical) according to the manufacturer's instructions.

3. RESULTS AND DISCUSSION

Northern blot analysis of mouse liver RNA using pHexB43 as a probe revealed the presence of a 2.0 kb mRNA (not shown) indicating that the clone could be used to isolate mouse β Hex cDNA. Ap-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

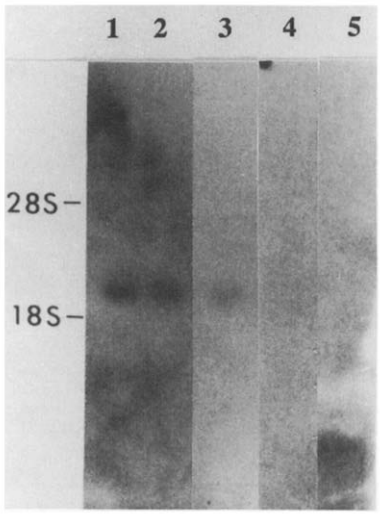

Fig.1. Northern blot analysis of mhex54 with human fibroblast RNA. Total cellular RNA (20 μg) from Tay-Sachs disease variant cell lines GM2968 (lane 1) and GM595A (lane 2), a normal cell line (lane 3) and Sandhoff disease variant cell lines 1954 and 1039 (lanes 4,5, respectively) was hybridized to the mouse cDNA clone.

prox. 10^6 plaques from the mouse fibroblast cDNA library were screened and seven positive clones containing inserts > 1.7 kb were isolated. One of them, mhex54 was selected for further analysis. Because of the considerable homology (55%) between the human α and β clones [4], Northern blots of total RNA from Tay-Sachs and Sandhoff disease cell lines (previously shown to have no detectable mRNA for the α- and β-subunits of Hex, respectively) were used to confirm the identity of mhex54. The mouse cDNA hydridizes to a 2.0 kb mRNA in the Tay-Sachs cell lines (containing β-subunit mRNA) and not to mRNA from Sandhoff cell lines (containing α-subunit mRNA). This confirms the identity of mHex54 as coding for the β-subunit of the mouse enzyme (fig.1).

The complete nucleotide and deduced amino acid sequence of mhex54 cDNA are shown in fig.2. The nucleotide sequence contains an open reading frame (ORF) of 1608 bp encoding a polypeptide of 536 amino acids starting from the first Met residue (see below) through to a TAG termination codon. The 3′-untranslated region of 130 bp contains a polyadenylation signal (AATAAA) preceding the poly(A) tail. Since the ORF continues to the 5′-end, the possibility of additional coding sequences cannot be completely ruled out. However, we suggest that the first Met residue likely corresponds to the site of translation initiation because: (i) the site aligns closely with the proposed initiation codon (fig.3) of the human prepro-β polypeptide (fig.3) [8]; (ii) the sequence preceding the ATG (CAGTC*ATG*) is a close match to the consensus sequence (CC{A or G}C*CATG*) found near initiation codons of eukaryotic mRNAs [14]; and (iii) as for all glycoproteins, the amino acids following the ATG in mhex54 form a characteristic signal peptide sequence with potential cleavage sites at residues 23:24, 24:25, and 28:29 (S=7.6, 8.2 and 7.7, respectively) [15].

A comparison of the deduced amino acid sequence of the mouse cDNA with those of the human α- and β-subunit cDNAs is presented in fig.3. The overall nucleotide and amino acid sequence homology is 78 and 72%, respectively, for the mouse and human β-subunit cDNAs. The mouse β- and human α-subunit cDNAs show 59 and 51% identity, respectively. We have previously identified the glycosylation sites of α- and β-subunits of the human enzyme [9]. Glycosylated asparagines appear as a part of the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except Pro or Asp [16]. In the mouse β-subunit sequence, there are five possible glycosylation sites (fig.3). Two of these, at positions 186–188 (NES) and 323–325 (NTT) are identical to the confirmed glycosylation sites of the human β-subunit, suggesting that the mouse protein is likely glycosylated at least at these sites.

In the lysosome, the human pro-β precursor is processed to β_a and β_b polypeptides by an internal cleavage which removes 3 or 4 amino acids [9]. We considered whether a similar cleavage might occur in the mouse sequence. Examination of the amino acid sequence of mouse Hex reveals lack of homology at this site (fig.3). However, there is an overall maintenance of positive charge, hydrophilicity and secondary structure across the region suggesting that the mouse polypeptide might also be subject to such a cleavage. To determine this, a small amount of Hex from mouse kidney was affinity purified as previously described for the human enzyme [1]. An examination of the reduced protein by SDS-PAGE revealed the presence of two bands corresponding to the positions of human β_a (~26 kDA) and β_b (30 kDa) (not

```
GlyAlaValMETProGlnSerProArgSerAlaProGlyLeuLeuLeuLeuGlnAlaLeuValSerLeuValSerLeuAlaLeuValAlaProAlaArgLeuGlnProAlaLeu   38
GGAGCAGTCATGCCGCAGTCCCCGCGTAGCGCCCCCGGGCTGCTGCTGCTGCAGGCGCTGGTGTCGCTAGTGTCGCTGGCCCTAGTGGCCCCGGCCCGACTGCAACCTGCGCTA  114

TrpProPheProArgSerValGlnMetPheProArgLeuLeuTyrIleSerAlaGluAspPheSerIleAspHisSerProAsnSerThrAlaGlyProSerCysSerLeuLeu   76
TGGCCCTTCCCGCGCTCGGTGCAGATGTTCCCGCGGCTGTTGTACATCTCCGCGGAGGACTTCAGCATCGACCACAGTCCCAATTCCACAGCGGGCCCTTCCTGCTCGCTGCTA  228

GlnGluAlaPheArgArgTyrTyrAsnTyrValPheGlyPheTyrLysArgHisHisGlyProAlaArgPheArgAlaGluAlaGlnLeuGlnLysLeuLeuValSerIleThr  114
CAGGAGGCGTTTCGGCGATATTACAACTATGTTTTTGGTTTCTACAAGAGACATCATGGCCCTGCTAGATTTCGAGCTGAGGCACAGTTGCAGAAGCTCCTGGTCTCCATTACC  342

LeuGluSerGluCysGluSerPheProSerLeuSerSerAspGluThrTyrSerLeuLeuValGlnGluProValAlaValLeuLysAlaAsnSerValTrpGlyAlaLeuArg  152
CTCGAGTCAGAGTGCGAGTCCTTCCCTAGTCTGTCTTCAGATGAAACCTATTCTCTGCTTGTACAAGAACCAGTAGCCGTCCTCAAGGCCAACAGCGTTTGGGGAGCGTTACGA  456

GlyLeuGluThrPheSerGlnLeuValTyrGlnAspSerPheGlyThrPheThrIleAsnGluSerSerIleAlaAspSerProArgPheProHisArgGlyIleLeuIleAsp  190
GGTTTAGAGACGTTTAGCCAGTTAGTTTACCAAGACTCTTTCGGGACTTTCACCATCAATGAATCCAGTATAGCTGATTCTCCACGATTCCCTCATAGAGGAATTTTAATTGAT  570

ThrSerArgHisLeuLeuProValLysThrIlePheLysThrLeuAspAlaMetAlaPheAsnLysPheAsnValLeuHisTrpHisIleValAspAspGlnSerPheProTyr  228
ACATCTAGACATCTCCTGCCTGTGAAGACAATTTTTAAAACTCTGGATGCCATGGCTTTTAATAAGTTTAATGTTCTTCACTGGCACATAGTGGACGACCAGTCTTTCCCTTAT  684

GlnSerThrThrPheProGluLeuSerAsnLysGlySerTyrSerLeuSerHisValTyrThrProAsnAspValArgMetValLeuGluTyrAlaArgLeuArgGlyIleArg  266
CAGAGTACCACTTTTCCTGAGCTAAGCAATAAGGGAAGCTACTCTTTGTCTCATGTCTATACACCAAACGATGTCCGGATGGTGCTGGAGTACGCCCGGCTCCGAGGGATTCGA  798

ValIleProGlyPheAspThrProGlyHisThrGlnSerTrpGlyLysGlyGlnLysAsnLeuLeuThrProCysTyrAsnGlnLysThrLysThrGlnValPheGlyProVal  304
GTCATACCAGGATTTGATACCCCTGGCCATACACAGTCTTGGGGCAAAGGACAGAAAAACCTTCTAACTCCATGTTACAATCAAAAAACTAAAACTCAAGTGTTTGGGCCTGTA  912

AspProThrValAsnThrThrTyrAlaPhePheAsnThrPhePheLysGluIleSerSerValPheProAspGlnPheIleHisLeuGlyGlyAspGluValGluPheGlnCys  342
GACCCAACTGTAAACACAACGTATGCATTCTTTAACACATTTTTCAAAGAAATCAGCAGTGTGTTTCCAGATCAGTTCATCCACTTGGGAGGAGATGAAGTAGAATTTCAATGT 1026

TrpAlaSerAsnProAsnIleGlnGlyPheMetLysArgLysGlyPheGlySerAspPheArgArgLeuGluSerPheTyrIleLysLysIleLeuGluIleIleSerSerLeu  380
TGGGCATCAAATCCAAACATCCAAGGTTTCATGAAGAGAAAGGGCTTTGGCAGCGATTTTAGAAGACTAGAATCCTTTTATATTAAAAAGATTTTGGAAATTATTTCATCCTTA 1140

LysLysAsnSerIleValTrpGlnGluValPheAspAspLysValGluLeuGlnProGlyThrValValGluValTrpLysSerGluHisTyrSerTyrGluLeuLysGlnVal  418
AAGAAGAACTCCATTGTTTGGCAAGAAGTTTTTGATGATAAGGTGGAGCTTCAGCCGGGCACAGTAGTCGAAGTGTGGAAGAGTGAGCATTATTCATATGAGCTAAAGCAAGTC 1254

ThrGlySerGlyPheProAlaIleLeuSerAlaProTrpTyrLeuAspLeuIleSerTyrGlyGlnAspTrpLysAsnTyrTyrLysValGluProLeuAsnPheGluGlySer  456
ACAGGCTCTGGCTTCCCTGCCATCCTTTCTGCTCCTTGGTACTTAGACCTGATCAGCTATGGGCAAGACTGGAAAAACTACTACAAAGTTGAGCCCCTTAATTTTGAAGGCTCT 1368

GluLysGlnLysGlnLeuValIleGlyGlyGluAlaCysLeuTrpGlyGluPheValAspAlaThrAsnLeuAspSerLysIleMetProArgAlaSerAlaValGlyGluArg  494
GAGAAGCAGAAACAACTGGTTATTGGTGGAGAAGCTTGCCTGTGGGGAGAATTTGTGGATGCAACTAACCTTGACTCCAAGATTATGCCTCGGGCAAGCGCTGTTGGTGAGAGA 1482

LeuTrpSerProLysThrValThrAspLeuGluAsnAlaTyrLysArgLeuAlaValHisArgCysArgMetValSerArgGlyIleAlaAlaGlnProLeuTyrThrGlyTyr  532
CTCTGGAGCCCTAAAACTGTCACTGACCTAGAAAATGCCTACAAACGACTGGCCGTGCACCGCTGCAGAATGGTCAGCCGTGGAATAGCTGCACAACCTCTCTATACTGGATAC 1596

CysAsnTyrGluAsnLysIle***                                                                                              539
TGTAACTATGAGAATAAAATATAGAAGTGACAGAACGTCTACAGCATTCCAGCTATGATCAATGTTGATTCTGAAATCATGTAAATTAAGATTTGTTAGGCTGTTTTTTTTT  1713

AAATAAACCATCTTTTTATTGATTGAATCTTTCTAAAAAAAAA                                                                          1756
```

Fig.2. The nucleotide and deduced amino acid sequence of cDNA clone for mouse β-hexosaminidase. Numbers on the right indicate nucleotide and amino acid residues in 5′ to 3′ direction. The putative initiation codon is highlighted (position 10-12). TAG (position 1618-1620) is the termination codon. The polyadenylation site is underlined.

Fig.3. Comparison of amino acid sequences of the prepro-β chain of mouse β-hexosaminidase (mβ) with those of the α- and β-prepro polypeptides of the human isozymes. Homology is denoted by (:). The putative (-) and the confirmed (=) glycosylation sites are shown. The internal cleavage site for the human β polypeptide subunit is marked, ( ). The signal peptide sequences for the human proteins [8,17] and the putative sequence for the mouse preproβ-chain

shown), indicating that the proteolytic clip does occur. The exact site of processing has yet to be determined.

We conclude that the β-subunit of mouse Hex is comparable to that of the human enzyme and that it undergoes similar proteolytic and oligosaccharide processing during its biosynthesis and delivery to the lysosome.

REFERENCES

[1] Mahuran, D.J., Tsui, F., Gravel, R.A. and Lowden, J.A. (1982) Proc. Natl. Acd. Sci. USA 79, 1602–1605.

[2] O'Dowd, B.F., Quan, F., Willard, H.F., Lamhonwah, A.M., Korneluk, R.G., Lowden, J.A., Gravel, R.A. and Mahuran, D.J. (1985) Proc. Natl. Acad. USA 82, 1184–1188.

[3] Myerowitz, R., Piekarz, R., Neufeld, E.F., Shows, T.B. and Suzuki, K. (1985) Proc. Natl. Acad. Sci. USA 82, 7830–7834.

[4] Korneluk, R.G., Mahuran, D.J., Neote, K., Klavins, M.H., O'Dowd, B.F., Tropak, M., Willard, H.F., Anderson, M.-J., Lowden, J.A. and Gravel, R.A. (1986) J. Biol. Chem. 261, 8407–8413.

[5] Proia, R.L. and Soravia, E. (1987) J. Biol. Chem. 262, 5766–5681.

[6] Proia, R.L. (1988) Proc. Natl. Acad. Sci. USA 85, 1183–1187.

[7] Mahuran, D.J., Neote, K., Klavins, M.H., Leung, A. and Gravel, R.A. (1988) J. Biol. Chem. 263, 4612–4618.

[8] Stirling, J., Leung, A., Gravel, R.A. and Mahuran, D.J. (1988) FEBS Lett. 231, 47–50.

[9] O'Dowd, B.F., Cummings, D.A., Gravel, R.A. and Mahuran, D.J. (1988) Biochemistry, in press.

[10] Conchi, J., Findlay, J. and Levvy, G.A. (1959) Biochem. J. 71, 318–325.

[11] Rahi, H. and Srivastawa, P.N. (1983) J. Reprod. Fertil. 67, 447–455.

[12] Oberkotter, L.V., Temnore, A. and Koldovsky, D. (1980) Early Hum. Dev. 4, 325–332.

[13] Ben-Neriah, Y., Bernards, A., Paskind, M., Daley, G.Q. and Baltimore, D. (1986) Cell 44, 577–586.

[14] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.

[15] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.

[16] Kornfeld, R. and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631–664.

[17] Little, L.E., Lau, M.M.L., Quon, D.V.K., Fowler, A.V., Neufeld, E.F. (1988) J. Biol. Chem. 263, 4288–4292.