Cloning and sequence analysis of a cDNA encoding the α -subunit of mouse β -N-acetylhexosaminidase and comparison with the human enzyme

Tommaso BECCARI,* John HOADE,† Aldo ORLACCHIO* and John L. STIRLING††

*Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università di Perugia, via del Giochetto, cp 37 succ. 3, 06100 Perugia, Italy, and †Division of Biomolecular Sciences, King's College London, Kensington Campus, Campden Hill, London W8 7AH, U.K.

cDNAs encoding the mouse β -N-acetylhexosaminidase α -subunit were isolated from a mouse testis library. The longest of these (1.7 kb) was sequenced and showed 83 % similarity with the human α -subunit cDNA sequence. The 5' end of the coding sequence was obtained from a genomic DNA clone. Alignment of the human and mouse sequences showed that all three putative N-glycosylation sites are conserved, but that the mouse α -subunit has an additional site towards the C-terminus. All eight cysteines in the human sequence are conserved in the mouse. There are an additional two cysteines in the mouse α -subunit signal peptide. All amino acids affected in Tay–Sachs-disease mutations are conserved in the mouse.

INTRODUCTION

β -N-Acetylhexosaminidase is a lysosomal enzyme that occurs in two main isoenzyme forms in mammalian tissues. The human isoenzymes A ($\alpha\beta$) and B ($\beta\beta$) and the genes that encode the α - and β -subunits of which they are composed have been analysed at the level of cDNA (Myerowitz *et al.*, 1985; O'Dowd *et al.*, 1985; Korneluk *et al.*, 1986), and the exon and intron organization of *HEX A* and *HEX B* genes have been determined (Proia & Soravia, 1987; Proia, 1988; Neote *et al.*, 1988). Mutations of the *HEX A* gene cause Tay–Sachs disease, a lysosomal storage disorder that results from an inability to hydrolyse the ganglioside G_{M2} . A variety of Tay–Sachs mutations have now been described and these have been reviewed by Mahuran (1991). Some of these are point mutations or small deletions causing single amino acid substitutions or deletions that may be informative about the location of the active site and processing of the polypeptide.

β -N-Acetylhexosaminidases of the mouse also exist as two major isoenzymes but there is greater variability in the proportions of these in different tissues than is the case in man (Beccari *et al.*, 1988). cDNA for the β -subunit of mouse β -N-acetylhexosaminidase has been cloned and the sequence compared with those of the α - and β -subunit cDNAs for the human enzymes (Bapat *et al.*, 1988).

We have isolated cDNA clones encoding the α -subunit of mouse β -N-acetylhexosaminidase with a view to sequence comparison with the human α -subunit and identification of conserved regions that might contribute to the active site or be involved in interactions between the subunits. In this paper we describe the cloning and sequence analysis of cDNAs from a mouse testis library and compare the deduced amino acid sequence with that of the human enzyme. All Tay–Sachs mutations described so far that cause single amino acid changes in the sequence of the human α -subunit are in residues conserved between mouse and man.

EXPERIMENTAL

Materials

A mouse testis cDNA library constructed in λ ZAP was kindly

provided by Dr. R. K. Dudley. We are grateful to Dr. R. A. Gravel for providing the α -14 cDNA clone for the α -subunit of human β -N-acetylhexosaminidase. [α - ^{32} P]dCTP (3000 Ci/mmol), [α - ^{35} S]thio]dATP (1000 Ci/mmol) and the Multi-prime DNA-labelling system were from Amersham International p.l.c., Amersham, Bucks., U.K. Nitrocellulose membranes (Schleicher & Schuell) were from Anderman & Co. Ltd., Kingston-upon-Thames, Surrey, U.K. Pharmacia Biosystems Ltd. supplied the nested deletion kit, and a variety of restriction endonucleases. Other restriction endonucleases, DNA ligase and alkaline phosphatase were from well-known suppliers. Sequenase sequencing kit was from Cambridge Bio Science, Cambridge, U.K.

Methods

Screening of the cDNA library, subcloning of DNA fragments, gel electrophoresis, Northern blotting and other general methods were performed according to methods in Maniatis *et al.* (1982). Automated excision of pBluescript from λ ZAP was carried out according to the method supplied by Stratagene Ltd. (Cambridge, U.K.). DNA sequencing was performed according to the method of Sanger *et al.* (1977) using oligonucleotide primers designed to hybridize with the T₇ or T₃ promoters of the pBluescript vector, or oligonucleotides specific for the mouse α -subunit sequence. Sequencing reactions were performed using denatured supercoiled plasmid DNA according to the method of Hatturi & Sakaki (1986). Nested deletions of the cDNA insert were made using the protocol supplied by Pharmacia. Completion of the coding sequence that was missing from the longest cDNA clone necessitated the isolation, from an EMBL-3 library, of a genomic DNA clone containing exon 1 of the gene. This was contained in a 4.8 kb *EcoRI*–*Bam*HI fragment which was subcloned into pBluescript and sequenced using an oligonucleotide primer that hybridized towards the 5' end of the cDNA.

RESULTS

Screening the cDNA library

A total of 8×10^5 plaques from the mouse testis cDNA library

† To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL and GenBank Nucleotide Sequence Databases under the accession number X64331 (MMHEXA).

were screened using the human α -14 cDNA as a probe. Five positive plaques were identified and the longest cDNA insert in these (1.7 kb) was selected for characterization and sequencing.

Northern blotting

Northern blotting of total RNA from mouse brain and testis revealed two mRNA species with estimated sizes of 2.1 and 2.7 kb. The smaller of these was very much more abundant than the larger which was only visible after exposure of the autoradiograph for several days (Fig. 1).

Fig. 1. Northern blotting of RNA from mouse tissues

Total RNA (10 μ g) was denatured and run on an agarose gel containing formaldehyde. RNA was blotted on to nitrocellulose by capillary transfer and probed with the 1.7 kb cDNA for the α -subunit of β -N-acetylhexosaminidase. Lane 1, brain; lane 2, testis.

Fig. 2. Restriction map and sequencing strategy for the mouse β -N-acetylhexosaminidase cDNA

Positions of the restriction sites *SacI* and *ApaI* used for subcloning of fragments are shown together with other selected restriction sites. The cDNA was sequenced in both directions with subcloned fragments or from subclones generated by nested deletions. $\square$, indicates where sequencing primers specific for the cDNA were used.

Mouse	1	MAGCRLWVSLLLAALACLAATALWPFQYQTYHRRYLYPNFQFRYHV	50
Human	1	MTSSRLWFSLLAAAFAGRATALWPFQNFQSDQRVLYPNFQFQYDV	50
Mouse	51	SSAAQGGCVLDEAFRRYRNLLFGSGSWPRPFSFNKQQTGLKNILVSVV	100
Human	51	SSAAQPGCSVLDEAFQRYRDLLEFGSGSWPRPYLTGKRHTLEKNILVSVV	100
Mouse	101	TAECNEFPNLESVENYTLTINDDQCLLASETVWGALRGLETFSQLVWKS	150
Human	101	TPGCNQLPTLESVENYTLTINDDQCLLSETVWGALRGLETFSQLVWKS	150
Mouse	151	EGTFFINKTKIKDFPRFPHRGVLLDTSRHYLPLSSILDTLDVMAYNKFN	200
Human	151	EGTFFINKTEIEDFPRFPHRGVLLDTSRHYLPLSSILDTLDVMAYNKLN	200
Mouse	201	FHWHLVDDSSFPYESFTFPELTKRGSFNPTVTHIYAQDVKEVIEYARLG	250
Human	201	FHWHLVDDSPFPYESFTFPELMRKGSYNPTVTHIYAQDVKEVIEYARLG	250
Mouse	251	IRVLAEDFTPGHTLSWGPAGPGLLTPCYSGSHLSGTGFPVNPNSLNTYDF	300
Human	251	IRVLAEDFTPGHTLSWGPAPGLLTPCYSGSESGTGFVNPNSLNTYEF	300
Mouse	301	MSTLFLEISSVFPDFYLHLGGDEVDFTCWKSNEPNIQAFMKKKGFT.DFKQ	349
Human	301	MSTFLELVSVFPDFYLHLGGDEVDFTCWKSNEPNIQDFMRKKGFGEDEFKQ	350
Mouse	350	LESFYIQTLLDIVSDYDKGVVWQEVFDNKVKVRPDTIIQVWREEMPVEY	399
Human	351	LESFYIQTLLDIVSSYGKGVVWQEVFDNKVKIQPDTIIQVWREDIPVNY	400
Mouse	400	MLEMODITRAGFRALLSAPWYLNVRKYGPDKMDYKVEPLAFHGTPEQKA	449
Human	401	MKELELVTKAGFRALLSAPWYLNRIISYGPDKDFYVVEPLAFEGTPEQKA	450
Mouse	450	LVIIGGEACMWGEYVDSTNLVPLRWPRAGAVAERLWSNLLTNIDFAFKRL	499
Human	451	LVIIGGEACMWGEYVDNTNLVPLRWPRAGAVAERLWSNKLTSIDLTFAYERL	500
Mouse	500	SHFRCELVRGGIAQAPISVGCYCEQFEQT	528
Human	501	SHFRCELLRRGVQAQPLNVGFCEQFEQT	529

Fig. 3. Alignment of the deduced amino acid sequences of mouse and human β -N-acetylhexosaminidase α -subunits

Similarity of residues based on evolutionary distance uses the values of Gribskov & Burgess (1986) in which identity (I) has a value of 1.5, (:) > 1.0 and (.) > 0.5. Conserved cysteines are shown in bold, putative N-glycosylation sites are underlined and amino acids affected in Tay-Sachs disease are double underlined.

Characterization and sequencing of the 1.7 kb cDNA clone

A simplified restriction map and the sequencing strategy is shown in Fig. 2. The sequence, which has been deposited with EMBL, had an open reading frame of 1455 bp followed by a 3' non-translated sequence of 226 bp, in which there was a single polyadenylation signal (AATAAA) 17 bp upstream from the poly(A) tail.

Alignment of the mouse and human α -subunit cDNA sequences using University of Wisconsin Genetic Computer Group (UWGCG) GAP alignment program showed 83% similarity with a single gap of 3 bp in the mouse sequence. The longest mouse cDNA clone was 296 bp shorter at the 5' end than the human sequence numbered from the initiation codon, and does not represent the full-length of the α -subunit mRNA. This part of the sequence was obtained from an EMBL-3 mouse genomic DNA clone isolated by using the cDNA as a probe.

Comparison of the deduced amino acid sequences of the mouse and human α -subunit using the UWGCG GAP program (Devereux *et al.*, 1984) showed that they had 85% identity and 92% similarity using the weightings based on evolutionary distance described by Gribskov & Burgess (1986) (Fig. 3). All eight cysteines of the human sequence are conserved in the mouse but there are two additional cysteines in the mouse signal peptide that are absent from man. The human prepro β -subunit also has cysteine at position 4 (Bapat *et al.*, 1988). There are three putative N-glycosylation sites in the human α -subunit sequence. These are conserved in the mouse but there is an additional putative N-glycosylation site in the mouse sequence (Asn-444) that is absent from the human α -subunit. The gap in the alignment

Table 1. Tay–Sachs mutations causing single amino acid changes or deletions in the human β -N-acetylhexosaminidase α -subunit

DNA and amino acid sequences are numbered from the initiation codon (Myerowitz *et al.*, 1985).

Exon	Base change	Amino acid change	Position	Onset of Tay–Sachs disease*	Reference
5	G-509 to A	Arg to Gln	170	I	Nakano <i>et al.</i> (1990)
	C-532 to T	Arg to His	178	I	Tanaka <i>et al.</i> (1990)
	G-533 to A	Arg to His	178	I	Ohno & Suzuki (1988)
6	C-629 to T	Ser to Phe	210	I	Akli <i>et al.</i> (1991)
7	G-749 to A	Gly to Asp	250	J	Trop <i>et al.</i> (1990)
	G-805 to A	Gly to Ser	269	A	Navon & Proia (1989)
8	del TTC-910–915†	Phe deletion	304 or 305	I	Navon & Proia (1991)
13	G-1444 to A	Glu to Lys	482	I	Nakano <i>et al.</i> (1988)
	G-1496 to A	Arg to His	449	J	Paw <i>et al.</i> (1990)
	G-1511 to A	Arg to His	504	J	Paw <i>et al.</i> (1990)

* I, infantile; J, juvenile; A, adult.

† Deletion of TTC between bases 910 and 915 of the sequence.

of mouse and human sequences results in the replacement of glycine and glutamic acid with a single threonine in the mouse.

Amino acids affected in Tay–Sachs mutations

Ten Tay–Sachs mutations have been described in which there are single amino acid changes or deletions in the protein sequence (Table 1). In each case amino acids affected in Tay–Sachs disease are conserved in the mouse.

DISCUSSION

As expected, the mouse cDNA for the α -subunit of β -N-acetylhexosaminidase has a high degree of similarity with the human sequence. In man there are two mRNA molecules (2.1 and 2.6 kb) encoding the α -subunit that differ in the lengths of their 3' untranslated regions (Myerowitz *et al.*, 1985; Korneluk *et al.*, 1986). In the mouse 1.7 kb cDNA clone, and in a shorter one (1.5 kb), the untranslated region was intermediate in size between those found in man and contained a single polyadenylation signal. Northern blotting, however, showed that although there was a relatively abundant species of α -subunit mRNA in brain and testis (2.1 kb), there was also another transcript of low abundance (2.7 kb).

All three of the putative N-glycosylation sites in the human sequence are conserved in the mouse α -subunit sequence. Only one of these (Asn-114) appears to be utilized in man (O'Dowd *et al.*, 1988) and it remains to be established whether this is also used in the mouse. There is much greater variability in the β -N-acetylhexosaminidase isoenzyme profiles of mouse tissues than those of man (Beccari *et al.*, 1988) and it is possible that either the utilization of glycosylation sites that are cryptic in humans or the putative site towards the C-terminus of the mouse sequence (Asn-444) could contribute to the greater complexity of the isoenzyme profiles in mouse tissues.

Comparison of amino acids affected in Tay–Sachs mutations shows that in each case they are conserved in the mouse. Navon & Proia (1991) have noted that the two adjacent phenylalanine residues, one of which is deleted in Tay–Sachs disease in Moroccan Jews, are conserved in the mouse, but in the clone we sequenced the first of these was replaced by leucine. Conservation of the mouse residue corresponding to Asp-178 of the human sequence is of interest because of its involvement in the B1 variant of Tay–Sachs disease. In this form of the disease the α -subunit retains its ability to associate with β -subunits to make β -N-acetylhexosaminidase A, but this lacks activity towards the

ganglioside G_{M2} and the synthetic substrate 4-methylumbelliferyl- β -acetylglucosamine 6-sulphate. Introduction of this mutation into the corresponding amino acid of the β -subunit results in loss of activity by this subunit too (Brown *et al.*, 1989). An adult-onset variant of Tay–Sachs disease found in Ashkenazi and non-Jewish patients is thought to result from a single amino acid change Gly-269 to Ser (Navon & Proia, 1989; Paw *et al.*, 1989; Navon *et al.*, 1990). Kappler *et al.* (1990) have raised the issue of whether this mutation might represent a β -N-acetylhexosaminidase pseudodeficiency. They make a comparison with arylsulphatase A pseudodeficiency which is allelic to the locus for metachromatic leucodystrophy. Proia *et al.* (1990) have rejected this possibility. Comparison with the mouse sequence does not resolve the question of whether individuals homozygous for adult-onset mutation have a pseudodeficiency or not but it is interesting to note that the glycine residue in question is conserved in the mouse.

We are grateful to C.N.R. (P.F. Ingeneria Genetica 92.00449, P.F. 99/115, 02363) for financial support and to the C.N.R. British Council agreement for a travel grant T.B.

REFERENCES

- Akli, S., Chelly, J., Lacorte, J.-M., Poenaru, L. & Khan, A. (1991) *Genomics* **11**, 124–134
- Bapat, B., Ethier, M., Neote, K., Mahuran, D. & Gravel, R. A. (1988) *FEBS Lett.* **237**, 191–195
- Beccari, T., Orlacchio, A. & Stirling, J. L. (1988) *Biochem. J.* **252**, 617–620
- Brown, C. A., Neote, K., Leung, A., Gravel, R. A. & Mahuran, D. J. (1989) *J. Biol. Chem.* **264**, 21705–21710
- Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Gribskov, A. & Burgess, R. R. (1986) *Nucleic Acids Res.* **14**, 6745–6763
- Hatturi, M. & Sakaki, Y. (1986) *Anal Biochem.* **152**, 232–238
- Kappler, J., Gieselmann, V. & Propping, P. (1990) *Am. J. Hum. Genet.* **47**, 880–881
- 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. & Gravel, R. A. (1986) *J. Biol. Chem.* **261**, 8407–8413
- Mahuran, D. J. (1991) *Biochim. Biophys. Acta* **1096**, 87–94
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Myerowitz, R., Piekarz, R., Neufeld, E., Shows, T. B. & Suzuki, K. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7830–7834
- Nakano, T., Muscillo, M., Ohno, K., Hoffman, A. J. & Suzuki, K. (1988) *J. Neurochem.* **51**, 984–987

- Nakano, T., Namba, E., Tanaka, A., Ohno, K., Suzuki, Y. & Suzuki, K. (1990) *Ann. Neurol.* **27**, 465–473
- Navon, R. & Proia, R. L. (1989) *Science* **243**, 1471–1474
- Navon, R. & Proia, R. L. (1991) *Am. J. Hum. Genet.* **48**, 412–419
- Navon, R., Kolodny, E. H., Mitumoto, H., Thomas, G. H. & Proia, R. L. (1990) *Am. J. Hum. Genet.* **46**, 817–821
- Neote, K., Bapat, B., Dumbrille-Ross, A., Troxel, C., Schuster, S. M., Mahuran, D. J. & Gravel, R. A. (1988) *Genomics* **3**, 279–286
- O'Dowd, B., Quan, F., Willard, H., Lamhonwah, A. M., Korneluk, R., Lowden, J. A., Gravel, R. A. & Mahuran, D. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1184–1188
- O'Dowd, B. F., Cumming, D. A., Gravel, R. A. & Mahuran, D. (1988) *Biochemistry* **27**, 5216–5226
- Ohno, K., Suzuki, K. (1988) *J. Neurochem.* **50**, 316–318
- Paw, B., Kaback, M. M. & Neufeld, E. F. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2413–2417
- Paw, B., Moskowitz, S. M., Uhrhammer, N., Wright, N., Kaback, M. M. & Neufeld, E. F. (1990) *J. Biol. Chem.* **265**, 9452–9457
- Proia, R. L. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1883–1887
- Proia, R. L. & Soravia, E. (1987) *J. Biol. Chem.* **262**, 5677–5681
- Proia, R. L., Kolodny, E. H. & Navon, R. (1990) *Am. J. Hum. Genet.* **47**, 811–882
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Tanaka, A., Ohno, K., Sandhoff, K., Maire, I., Kolodny, E. H., Brown, A. & Suzuki, K. (1990) *Am. J. Hum. Genet.* **46**, 329–339
- Trop, L., Kaplan, F. & Hechtman, P. (1990) *Am. J. Hum. Genet.* **47**, A168

Received 13 December 1991; accepted 8 January 1992