Cloning and sequence analysis of a cDNA clone coding for the mouse G_{M2} activator protein

Gianluca BELLACHIOMA,* John L. STIRLING,† Aldo ORLACCHIO* and Tommaso BECCARI*‡

*Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Universita di Perugia, via del Giochetto, cp 37 succ. 3, 06126 Perugia, Italy, and †Division of Life Sciences, King's College London, Kensington Campus, Campden Hill, London W8 7AH, U.K.

A cDNA (1.1 kb) containing the complete coding sequence for the mouse G_{M2} activator protein was isolated from a mouse macrophage library using a cDNA for the human protein as a probe. There was a single ATG located 12 bp from the 5′ end of the cDNA clone followed by an open reading frame of 579 bp. Northern blot analysis of mouse macrophage RNA showed that there was a single band with a mobility corresponding to a size of 2.3 kb. We deduce from this that the mouse mRNA, in common with the mRNA for the human G_{M2} activator protein, has a long 3′ untranslated sequence of approx. 1.7 kb. Alignment of the mouse and human deduced amino acid sequences showed 68% identity overall and 75% identity for the sequence on the C-terminal side of the first 31 residues, which in the human G_{M2} activator protein contains the signal peptide. Hydropathicity plots showed great similarity between the mouse and human sequences even in regions of low sequence similarity. There is a single N-glycosylation site in the mouse G_{M2} activator protein sequence (Asn^{151}-Phe-Thr) which differs in its location from the single site reported in the human G_{M2} activator protein sequence (Asn^{63}-Val-Thr).

INTRODUCTION

Hydrolysis of the ganglioside G_{M2} in human tissues requires two proteins, the lysosomal enzyme β-*N*-acetylhexosaminidase A (EC 3.2.1.52) and the G_{M2} activator protein (Conzelmann and Sandhoff, 1978). G_{M2} activator protein is a low-molecular-mass (22 kDa) soluble monomeric glycoprotein that binds to G_{M2} and stabilizes it as an activator/lipid complex in a form accessible to the active site of the α-subunit of β-*N*-acetylhexosaminidase A (for a review see Furst and Sandhoff, 1992).

Mutations affecting the gene encoding the G_{M2} activator protein, which has been mapped by analysis of somatic cell hybrids to human chromosome 5 (Burg et al., 1985; Xie et al., 1992a), cause a form of G_{M2} gangliosidosis that is clinically indistinguishable from Tay–Sachs disease. cDNAs encoding the human G_{M2} activator protein have been isolated and sequenced (Schroder et al., 1989; Xie et al., 1991; Nagarajian et al., 1992), the protein sequence has been determined independently (Furst et al., 1990) and the exon–intron organization of the gene determined (Klima et al., 1991). A processed G_{M2} activator pseudogene is located on chromosome 3 (Xie et al., 1992a).

Fuller analysis of biological functions and developmental changes in the α- and β-subunits of β-*N*-acetylhexosaminidase and the G_{M2} activator protein require that the genes are cloned from a suitable experimental animal. G_{M2} activator proteins have been detected in rat, mouse, pig and bovine tissues (Burg et al., 1983), and in the guinea pig (Kasakura et al., 1985), and have similar properties to those of the human G_{M2} activator. Mice are the preferred experimental animal for the analysis of mammalian development and have the additional advantage that models of human diseases can be generated by homologous recombination. A cDNA encoding the β-subunit of mouse β-*N*-acetylhexosaminidase has been cloned and sequenced (Bapat et al., 1988) and we have recently isolated and sequenced a cDNA encoding the α-subunit (Beccari et al., 1992). In this paper we report the cloning and sequencing of a cDNA encoding the mouse G_{M2} activator protein and compare the deduced protein sequence with that of the human G_{M2} activator protein.

Two mutations of the G_{M2} activator protein gene causing the AB variant of G_{M2} gangliosidosis (Schroder et al., 1991; Xie et al., 1992b; Furst and Sandhoff, 1992) result in changes of amino acids that are conserved in the mouse.

EXPERIMENTAL

Materials

A mouse macrophage cDNA library constructed in λgt11 was purchased from Clontech Laboratories, Palo Alto, CA, U.S.A. We are grateful to Professor Konrad Sandhoff for providing the cDNA clone for the human G_{M2} activator protein. [α-^{32}P]dCTP (3000 Ci/mmol), [α-^{35}S]dATP (1000 Ci/mmol), Hybond-C membranes and the Multiprime DNA labelling system were from Amersham International p.l.c., Amersham, Bucks., U.K. Restriction endonucleases, DNA ligase and alkaline phosphatase were from Boehringer Mannheim U.K., Lewes, E. Sussex, U.K. Sequenase sequencing kit was from Cambridge BioScience, Cambridge, U.K. All other reagents were from BDH, Poole, Dorset, 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 Sambrook et al. (1989). cDNA inserts from λgt11 were excised by digestion with *Eco*RI and separated from the vector arms by agarose gel electrophoresis. They were then subcloned into the *Eco*RI site of pBluescript. A Bluescript subclone containing the longest of

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported have been deposited in the EMBL, DDBJ and GenBank Nucleotide Sequence Databases under accession no. L19526 (MUSGM 2ACT).

these inserts (1.1 kb) was digested with *Kpn*I and the products (0.445 and 0.65 kb) were separated on an agarose gel. The larger fragment was recovered from the gel and ligated with T4 DNA ligase before transformation of *Escherichia coli* strain BB4. The smaller fragment was cloned into the *Kpn*I site of Bluescript. Sequencing of the inserts in these plasmids was by the double-stranded sequencing protocol supplied with Sequenase, using either vector-specific oligonucleotides or oligonucleotides specific for the mouse G_{M2} activator protein cDNA. The sequencing strategy is shown in Figure 2.

For Northern blot analysis of mRNA, total cellular RNA was isolated from mouse peritoneal macrophages using the method of Chomczynski and Sacchi (1987).

RESULTS

Screening the mouse macrophage cDNA library

A total of 1×10^6 plaques from the mouse macrophage cDNA library were screened using a 0.8 kb human G_{M2} activator protein cDNA, pGAP1 (Schroder et al., 1989), as a probe. Twenty positive plaques were identified and the longest cDNA insert in these (1.1 kb) was selected for characterization and sequencing.

Northern blotting

Northern blotting of total RNA from mouse macrophages showed a single mRNA species with an estimated size of 2.3 kb (Figure 1).

Characterization and sequencing of the 1.1 kb cDNA clone

The nucleotide sequence (Figure 2), which has been deposited with EMBL, had an open reading frame of 579 bp followed by a 3′ non-translated sequence of approx. 500 bp. Neither a poly(A) tail nor a polyadenylation signal was found at the 3′ end of the clone. This, together with the mRNA length estimated from the Northern blot, shows that although it contains the full coding sequence for the G_{M2} activator protein, the clone is incomplete at the 3′ end.

Comparison of the deduced amino acid sequences of the mouse and human G_{M2} activator proteins, using the UWGCG Gap program (Figure 3), showed that they had 68 % identity and 81 % similarity using the weightings based on evolutionary distance described by Gribskov and Burgess (1986). All eight cysteines of the human sequence are conserved in the mouse. There is a single putative N-glycosylation site in the mouse sequence (Asn^{151}), which differs in its location from the single site found in the human G_{M2} activator protein (Asn^{63}).

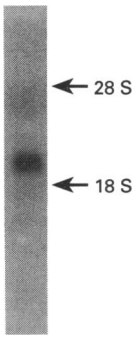

Figure 1 Northern blotting of mouse macrophage RNA

Total RNA (10 μg) was denatured and run on an agarose electrophoresis gel containing formaldehyde. RNA was blotted on to nitrocellulose by capillary transfer and probed with the cDNA encoding the mouse G_{M2} activator protein.

Hydropathicity analysis of mouse and human G_{M2} activator protein sequences

Deduced amino acid sequences of the mouse and human G_{M2} activator proteins were analysed by Kyte and Doolittle hydropathicity plots using the GeneJockey Sequence Processor Software supplied by BIOSOFT, Cambridge, U.K. (Figure 4).

DISCUSSION

Isolation of cDNA clones containing the complete coding sequence of the human G_{M2} activator protein has previously presented difficulties that may stem from the low abundance of mRNA (Schroder et al., 1989; Xie et al., 1991; Nagarajian et al., 1992). We were somewhat surprised to find a relatively high number of positive plaques when screening the mouse macrophage library, and to isolate one containing the whole coding sequence, although it was incomplete at the 3′ end. The commercial cDNA library was made using mRNA from macrophages activated with linoleic acid, and it is possible that this treatment affects the expression of the G_{M2} activator protein gene. Human tissues have two mRNAs coding for the human G_{M2} activator protein that differ in their long 3′ untranslated sequences of approx. 1.7 kb. The mouse G_{M2} activator protein mRNA is likely to be similar in its overall structure to that of the human mRNAs in that it is of a similar length and the coding sequence is also at the 5′ end. The function of the long 3′ untranslated sequences in the human G_{M2} activator protein mRNAs is unknown, but it may be of significance that this feature is conserved in the mouse. Comparison of mouse and human sequences downstream from the termination codon using the UWGCG BESTFIT program showed that 67 % of the nucleotides were identical, although 19 gaps were inserted to make the best alignment. A high degree of similarity was found when the deduced amino acid sequences of the mouse and human G_{M2} activator proteins were compared. Overall, 68 % of the residues were identical. Greatest divergence between the sequences was found in the first 31 amino acids, with a higher degree of similarity between them from residue 32 to the C-terminus, a region in which 75 % of the amino acids were identical. This N-terminal divergence between the sequences may not have functional significance because in the human G_{M2} activator the first 31 residues contain the signal peptide sequence and are not retained in the mature protein. We have no direct evidence that these residues in the mouse sequence function as a signal peptide, but they have similar hydrophobicity to those of the human sequence, as revealed by the hydrophobicity plot (Figure 4).

The structures of the human and mouse G_{M2} activator proteins are likely to be very similar. All eight cysteines of the human sequence are conserved in the mouse and the importance of at least one of these, Cys^{138}, is demonstrated in the mutation described by Schroder et al. (1991) and Xie et al. (1992b) in which the transition T^{412} to C causes the incorporation of Arg at this point, and the loss of function. A second mutation of the G_{M2} activator protein, i.e. G^{505} to C (Arg^{169} to Pro) reported by Furst and Sandhoff (1992) also affects an amino acid that is conserved in the mouse sequence.

```
-12 AGGACCATCGCCATGCACCGTCTACCGCTGCTGCTCCTGCTGGGCTTGCTGCTCGCAGGC   48
  1             MetHisArgLeuProLeuLeuLeuLeuGlyLeuLeuLeuAlaGly    16

 49 TCCGTCGCCCCTGCGCGCCTCGTCCCGAAGCGCCTTTCCCAACTTGGTGGCTTCTCCTGG  108
 17 SerValAlaProAlaArgLeuValProLysArgLeuSerGlnLeuGlyGlyPheSerTrp   36

109 GATAACTGTGATGAAGGAAAGGACCCTGCAGTGATCAAAAGCCTCACGATCCAACCTGAC  168
 37 AspAsnCysAspGluGlyLysAspProAlaValIleLysSerLeuThrIleGlnProAsp   56

169 CCCATTGTGGTTCCTGGAGATGTAGTCGTCAGCCTTGAGGGCAAGACCAGCGTTCCCCTC  228
 57 ProIleValValProGlyAspValValValSerLeuGluGlyLysThrSerValProLeu   76

229 ACTGCTCCTCAGAAGGTGGAGCTCACCGTGGAGAAGGAAGTGGCTGGCTTCTGGATCAAG  288
 77 ThrAlaProGlnLysValGluLeuThrValGluLysGluValAlaGlyPheTrpIleLys   96

289 ATTCCTTGTGTAGAACAGCTAGGCAGCTGTAGCTACGAGAACATCTGTGACCTGATAGAC  348
 97 IleProCysValGluGlnLeuGlySerCysSerTyrGluAsnIleCysAspLeuIleAsp  116

349 GAATACATTCCTCCTGGAGAGAGCTGCCCAGAGCCCCTGCACACCTACGGGCTGCCCTGC  408
117 GluTyrIleProProGlyGluSerCysProGluProLeuHisThrTyrGlyLeuProCys  136

409 CACTGTCCCTTCAAGGAAGGTACCTACTCACTACCCACGAGCAACTTCACAGTGCCTGAC  468
137 HisCysProPheLysGluGlyThrTyrSerLeuProThrSerAsnPheThrValProAsp  156

469 CTAGAGCTACCGAGCTGGCTGAGCACAGGCAACTACCGCATCCAGAGTATCTTGAGCAGT  528
157 LeuGluLeuProSerTrpLeuSerThrGlyAsnTyrArgIleGlnSerIleLeuSerSer  176

529 GGTGGGAAGCGCCTGGGCTGCATCAAGATTGCTGCCTCTCTCAAGGGCAGATAACCTGGC  588
177 GlyGlyLysArgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyArgEnd         193

589 AGCAGCACCACAGAATGAAGGGGCACCAGCACGGCAGGCCTCTTCTTCTGTGTTCCGGTG  648

649 CCCAGGCCTGCCTGCCTCTAACTCTATCCTCTACCGGTCTCTACCCTCTCTGGAAATCAT  708

709 TGCAACCTGACCTGAAGGGAGGATGGACCAAGGCCCAACATCTGATGGCTGTAAGCCATG  768

769 CTCCTTTGCACCCCTACCAACTTTCTAAGAATTTCTTTCGTATCCAAAACAGTCAAGGAA  828

829 TGGGAACCAACATGTTTGGGAATCTAAGGTCAAACTGGCCAACCTGGACCTCCTGGAGTT  888

889 ATGCCCACCCCACTTCTTTGTTTTAGCCACCAACTTTAAAACCAGGATAACCGTATTAAT  948

949 TTCTCATTCCATCCTGAGCCTCTCTTCCTGCAAAGAGGTCTAGGTAGTCAGTCTTGGGTC 1008

1009 CCTCTATTAACTTCTGTGATTAACTGCATTCTCTTTGCAACTTAACTTGTCAACACTGGG 1068

1069 GCTGACAGATCCTGAAGAGGCAGGGTGGTGTG 1100
```

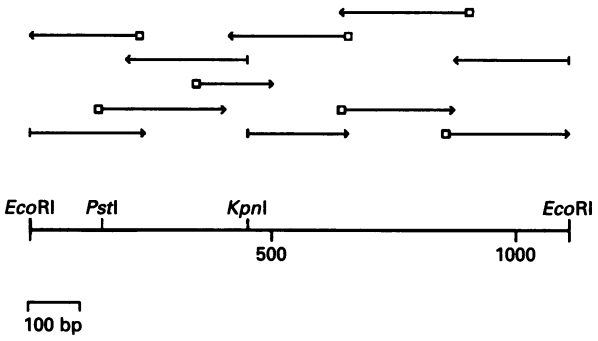

Figure 2 Nucleotide sequence of a cDNA encoding the mouse G_{M2} activator protein and the deduced amino acid sequence

The first 600 bp at the 5′ end of the cDNA contain the putative initiation codon (underlined) and an open reading frame of 579 nucleotides ending with the termination codon TAA. The simplified restriction map in the lower part of the Figure shows the strategy used to sequence the clone.

It is expected that the binding site of the G_{M2} activator protein will have hydrophobic character consistent with its function in solubilizing the amphipathic G_{M2} molecule. Three hydrophobic α-helices were tentatively assigned this function in the human G_{M2} activator protein (Furst et al., 1990). Secondary structure predictions according to Chou and Fasman (1978) and Garnier et al. (1978) showed that only one of these α-helices was predicted by both methods in both mouse and human sequences. This lies in the sequence bounded by Pro^{79} and Pro^{98} of the human and mouse proteins. There is a helix-breaking Gly^{92} in both sequences. The hydrophobicity plots of the human and mouse deduced amino acid sequences resemble each other closely even in those parts of the sequence where there is low similarity, but the most highly conserved sequences showing hydrophobic character are towards the C-termini, and it is possible that these participate in binding the lipid tail of G_{M2}.

In the sequence of the human G_{M2} activator protein there are two possible initiating ATGs, corresponding to Met^1 and Met^5 in

```
Mouse    1 MHRLPLLLLLGLLLAGSVAPARLVPKRLSQLGGFSWDNCDEGKDPAVIKS  50
           |:.|    ||  |    ..||.   |: |||::||||||||||||||:|
Human    1 MQSLMQAPLLIALGLLLATPAQAHLKKPSQLSSFSWDNCDEGKDPAVIRS  50

Mouse   51 LTIQPDPIVVPGDVVVSLEGKTSVPLTAPQKVELTVEKEVAGFWIKIPCV 100
           ||::|||||||:|.:|: |.||||..| ||:|.:||||||:||||||.
Human   51 LTLEPDPIVVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100

Mouse  101 EQLGSCSYENICDLIDEYIPPGESCPEPLHTYGLPCHCPFKEGTYSLPTS 150
           : :|||.:|::||::| .||.||.|||||:|||||||||||||||.|
Human  101 DYIGSCTFEHFCDVLDMLIPTGEPCPEPLRTYGLPCHCPFKEGTYSLPKS 150

Mouse  151 NFTVPDLELPSWLSTGNYRIQSILSSGGKRLGCIKIAASLKGR 193
           :|.||||||||||.||||||:|:|||:|||||||||||||
Human  151 EFVVPDLELPSWLTTGNYRIESVLSSSGKRLGCIKIAASLKGI 193
```

Figure 3 Alignment of the deduced amino acid sequence of the mouse G_{M2} activator protein with the sequence of the human G_{M2} activator protein

The similarity of residues in the mouse and human sequences is based on evolutionary distance using the values of Gribskov and Burgess (1986). Identical amino acids (|) have a value of 1.5; (:) > 1.0 and (.) > 0.5. Cysteines are shown in bold type. The only *N*-glycosylation site in the human sequence is Asn^{63}-Val-Thr and in the mouse it is Asn^{151}-Phe-Thr.

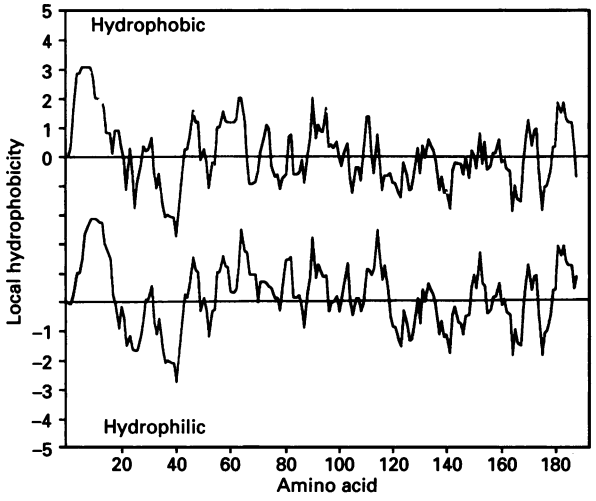

Figure 4 Kyte and Doolittle hydrophobicity plots for the deduced amino acid sequence of the mouse G_{M2} activator protein (upper block) and the protein sequence of the human G_{M2} activator protein (lower block)

Profiles were calculated according to Kyte and Doolittle using GeneJockey Sequence Processor Software. Positive values are hydrophobic and negative values hydrophilic. Values below the horizontal line in the upper block are negative and values above the horizontal line in the lower block are positive.

Figure 3. Inspection of the sequence immediately preceding these ATGs showed that they both matched poorly with the translation initiation consensus sequence (Xie et al., 1991). The mouse sequence has only one ATG, towards the 5′ end of the cDNA, and that corresponds with Met^1 of the human sequence. There is close similarity between the sequence immediately upstream from the ATG, which is TCGCCATG, and the translation initiation consensus sequence CC(A or G)CCATG described by Kozak (1984), so that the context for initiation at this point in the mouse sequence is more favourable than the corresponding sequence of the human mRNA.

G_{M2} activator protein requires N-linked glycosylation for transport to lysosomes (Kornfeld, 1989), and the human protein must therefore be glycosylated at the only *N*-glycosylation site (Asn^{63}-Val-Thr) within the sequence (Xie et al., 1991). This putative *N*-glycosylation site is missing from the deduced amino acid sequence of the mouse G_{M2} activator protein and it is likely that the only *N*-glycosylation motif (Asn^{151}-Phe-Thr) is used instead. G_{M2} activator protein from mouse kidney is able to stimulate hydrolysis of G_{M2} by human hexosaminidase A (Burg et al., 1983) and it is unlikely, therefore, that amino acid residues close to Asn^{151} are involved in specific interactions with the α-subunit.

We are grateful to C.N.R. (P.F. Ingeneria Genetica 93.000.32, P.F. 99/115, 02363) for financial support, to Dr. M. M. Anceschi, II Istituto di Clinica Ginecologia e Ostetrica, University of Rome, for providing the oligonucleotides used for sequencing, and to the C.N.R. British Council agreement for a travel grant to T.B.

REFERENCES

Bapat, B., Ethier, M., Neote, K., Mahuran, D. and Gravel, R. A. (1988) FEBS Lett. **237**, 191–195

Beccari, T., Hoade, J., Orlacchio, A. and Stirling, J. L. (1992) Biochem. J. **285**, 593–596

Burg, J., Banerjee, B., Conzelmann, E. and Sandhoff, K. (1983) Hoppe-Seyler's Z. Physiol. Chem. **364**, 821–829

Burg, J., Conzelmann, E., Sandhoff, K., Solomon, E. and Swallow, D. M. (1985) Ann. Hum. Genet. **49**, 41–45

Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159

Chou, P. Y. and Fasman, G. D. (1978) Adv. Enzymol. **47**, 45–148

Conzelmann, E. and Sandhoff, K. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**, 3979–3983

Furst, W. and Sandhoff, K. (1992) Biochim. Biophys. Acta **1126**, 1–16

Furst, W., Schubert, J., Machleidt, W., Meyer, H. E. and Sandhoff, K. (1990) Eur. J. Biochem. **192**, 709–714

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

Gribskov, A. and Burgess, R. R. (1986) Nucleic Acids Res. **14**, 6745–6763

Kasakura, H., Hirabayashi, Y. and Matsumoto, M. (1985) in Glycoconjugates: Proceedings of the VIIIth International Symposium on Glycoconjugates, Houston, Texas (Davidson, E. A., Williams, J. C. and DiFerrante, N., eds.), vol. 2, p. 600, Praeger, New York

Klima, H., Tanaka, A., Schnabel, D., Nakano, T., Schroder, M., Suzuki, K. and Sandhoff, K. (1991) FEBS Lett. **289**, 260–264

Kornfeld, S. (1989) Annu. Rev. Cell. Biol. **5**, 483–525

Kozak, M. (1984) Nucleic Acids Res. **12**, 857–872

Nagarajian, S., Chen, H. C., Li, S. C., Li, Y. T. and Lockyer, J. M. (1992) Biochem. J. **282**, 807–813

Sambrook, J. Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Schroder, M., Klima, H., Nakano, T., Kwan, H., Quintern, L. E., Gartner, S., Suzuki, K. and Sandhoff, K. (1989) FEBS Lett. **251**, 197–200

Schroder, M., Schnabel, D., Suzuki, K. and Sandhoff, K. (1991) FEBS Lett. **290**, 1–3

Xie, B., McInnes, B., Neote, K., Lamhonwah, A. M. and Mahuran, D. (1991) Biochim. Biophys. Acta **177**, 1217–1223

Xie, B., Kennedy, J. L., McInnes, B., Auger, D. and Mahuran, D. (1992a) Genomics **14**, 796–798

Xie, B., Wang, W. and Mahuran, D. (1992b) Am. J. Hum. Genet. **50**, 1046–1052

Received 3 December 1992/29 March 1993; accepted 14 April 1993