

Complete cDNA sequence encoding human β -galactoside α -2,6-sialyltransferase

Ulrich Grundmann, Claudia Nerlich, Thomas Rein and Gerd Zettlmeissl
Department of Molecular Biology, Research Institutes, Berhringwerke AG, Postfach 1140, 3550 Marburg, FRG

Submitted January 5, 1990

EMBL accession no. X17247

Here we present the cDNA sequence of human beta-galactoside alpha-2,6-sialyltransferase. The protein mediates the attachment of sialic acid to glycoprotein sugar chains. The cDNA was isolated from a human placental lambda gt10 library (1) by screening with two oligonucleotides generated from the rat liver cDNA encoding sialyltransferase (2). The human cDNA with 2188 bp (including EcoRI linkers, lower case) encodes the complete protein with 406 aa (MW 46604) including a typical hydrophobic signal anchor sequence (aa 11 to 25) and two potential N-linked glycosylation sites (aa 149 and 161; marked by asterisks). The protein sequence shows 87.6% similarity to

the rat protein sequence with 403 residues. The human cDNA encodes three additional residues E-K-K near the amino acid position 30.

REFERENCES

1. Grundmann,U., Amann,E., Zettlmeissl,G. and Kuepper,H. (1986) Characterization of cDNA coding for human factor XIIIa. *Proc. Natl. Acad. Sci. USA* **83**, 8024-8028.
2. Weinstein,J., Lee,E.U., McEntee,K., Lai,P.-H. and Paulson,J.C. (1987) *J. Biol. Chem.* **262** No. 36, 17735-17743.

```

      10              30              50              70              90              110
aatctcGCCCGCGTTAACAAAGGGAGCCGATACCGACCGCGTGGGCGCGGAGCGGGCGGCCCCAGCGCTGCTGAGCAACCGCAGCTCCCGCGCCGAGAGTGCGAGCGAAGG
      130              150              170              190              210              230
GGAGAGCCAGTTGCGCAGAGCCCTGCAACAGCAGTCCAGGGAGAAGTGGTGAATGTCTATGGAGCCAGCTGAAATGGACTGGCCCCCTTGAGCCCTGTCACAGCCCTGGTGCCAGGTGT
      250              270              290              310              330              350
CCATCCCCGTGCTGAGATGAGTTTGTGATCATCTGAGAAAAATGGGCCCTGGCCCTGCAGACCCAATAAACCTTCCCTCCCATGGATAATAGTGCTAATCTCTGAGGACCTGAAGGCCCTGC
      370              390              410              430              450              470
CGCCCCTGGGGGATTAGCCAGAAGCAGGCTTGTCTTCTGCTCAGAACAAAGTGACTTCCCTGAAACATCTTCATTATGATTCACACCAACCTGAAGAAAAAGTTCAGCTGCTGCGTCC
      490              510              530              550              570              590
TGGTCTTTCTTGTGTTGAGTCTGTTGGTGAAGAAAAAGAAAGGGAGTTACTATGATTCCTTTAAATGCAACCAAGGAATTCAGGTGTTAAAGAGTCTGGGGAAATGG
      610              630              650              670              690              710
V F L L F A V I C V W K E K R K G S Y Y D S F K L Q T K E F Q V L K S L G K L A
CCATGGGGTCTGATCCAGCTGTATCTCAAGCAGCACCAGGACCCACAGGGCCCGCAGACCTCGGCAGTCTCAGAGCCCTAGCCAAGGCCAAACAGAGGCCCTCTCCAGG
      730              750              770              790              810              830
M G S D S Q S V S S S T Q D F H R G R Q T L G S L R G L A K A K P E A S F Q V
TGTGGAACAAGGACAGCTTCCAAAAACCTTATCCCTAGGCTGCAAAAGATCTGGAAGAATTACCTAAGCATGAACAAGTCAAAAGTCTCTACAGGGGCCAGGACAGGCATCAAGT
      850              870              890              910              930              950
W N K D S S S K N L I P R L Q K I W K N Y L S M N K Y K V S Y K G F G P F G I K F
TCAGTGCAGAGGCCCTGCGCTGCCACCTCCGGACCATGTGAATGTATCCATGGTAGAGGTACAGATTTCCCTTCAATACCTCTGAAATGGGAGGGTATCTGCCAAGGAGGACATTA
      970              990              1010              1030              1050              1070
S A E A L R C H L R D H V N * V S H V E V T D F P F N * T S E W E G Y L P K E S I R
GGACCAAGGCTGGGCCTGGGGCAGGTGTGCTGTTGTGTCGTCAGCGGGATCTCTGAAGTCTCCCAACTAGGCAGAGAAATCGATGATCATGACGCGCTCTGAGGTTAATGGGGCAC
      1090              1110              1130              1150              1170              1190
T K A G P W G R C A V V S S A G S L K S S Q L G R E I D D H D A V L R F N G A P
CCACAGCCAACTTCCAACAAGATGTGGGCACAAAACTACCATTCCGCTGATGAATCTCAGTTGGTTACCACAGAGAAGCGCTTCCCTCAAAGACAGTTTGTACAATGAAGGAATCCTAA
      1210              1230              1250              1270              1290              1310
T A N F Q Q D V G T K T T I R L M N S Q L V T T E K R F L K D S L Y N E G I L I
TTGTATGGACCCATGTTATACCACTCAGATATCCAAAGTGGTACCAGAAATCCGGATTATAATTTCTTTAACTACAAGACTTATCGTAAGCTGCACCCCAATCAGCCCTTTTACA
      1330              1350              1370              1390              1410              1430
V W D P S V Y H S D I P K W Y Q N P D Y N F P N N Y K T Y R K L H P N Q P F Y I
TCCTCAAGCCCAAGTGCCTTGGGAGCTATGGGACATTCTTCAAGAAATCTCCCAAGAGATTACGCCAAACCCCATCTCTGGGATGCTTGGTATCATCATGATGACGCTGT
      1450              1470              1490              1510              1530              1550
L K P Q H P W E L W D I L Q E I S P E E I Q P N P P S S G M L G I I I M M T L C
GTGACCAGGTGGATATTTATGAGTCTCTCCATCCAAGCGCAAGACTGAGTGTGCTACTACTACAGAAATTTCTCGATAGTGCCTCAGCATGGGTGCTTACCACCGCTGCTATG
      1570              1590              1610              1630              1650              1670
D Q V D I Y E F L F S K R K T D V C Y Y Y Q K F P D S A C T M G A Y H P L L Y E
AGAAGAAATTTGGTGAAGCATCTCAACAGGGCACAGATGAGGACATCTACTCTGGAAAAAGCCACACTGCTGGCTCCGGACCATCTCACTGCTAAGCACAGGCTCTCACTCTTCTC
      1690              1710              1730              1750              1770              1790
K N L V K H L N Q G T D E D I Y L L G K A T L P G F R T I H C
CATCAGGCATTAATGAATGGTCTCTTGGCCACCCAGCCTGGGAAGAACTTTCTGAAACAATTCAGCCTGCTCCTTTTACTCTAGGGCCCTGTGTCAGCAAGACCATGGGACTTCA
      1810              1830              1850              1870              1890              1910
AGAGCGTGGTTCAGGAAATCAGGTCAGCCCTTCCCTGTAGCCAGACAGTTTATGAGCCAGAGCCTCTGCCACACACATGCACACATATCTAGCATTCTTTCAAGACAGCATCTCC
      1930              1950              1970              1990              2010              2030
CCGCCCTTCCACCTTGTAGATGCAAGGTCTATCTCTCCATCAGGGCTGCCAAAGCTGGGCCCTTGTGTTTTTCCAGCAGAATGATGCCATTCTCAAAACCAATGCTTATATTTGCTGAAG
      2050              2070              2090              2110              2130              2150
TCTGCATCTAAATATTGATTTACAGTTTTAAAGAAATCTCTTAAATTACAATGTGCCCAATGCAGGGTGGCTCTGGGGGCAAGTAGGTGTACAGGGGATTGGAACAATCGTCCGC
      2170              2188
GCCTCCAGAAAAAGTTGCTCCCGAGag

```