ELSEVIER

Biochimica et Biophysica Acta 1271 (1995) 369–373

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

Cloning and sequence analysis of caprine *N*-acetylglucosamine 6-sulfatase cDNA ☆

Karen Friderici [a], Kevin T. Cavanagh [a], Jeffrey R. Leipprandt [a], Christine E. Traviss [a], Donald S. Anson [b], John J. Hopwood [b], Margaret Z. Jones [a,*]

[a] *Department of Pathology, Michigan State University, 618 East Fee Hall, East Lansing, MI 48824-1314, USA*
[b] *Department of Chemical Pathology, Adelaide Children's Hospital, North Adelaide, South Australia*

Received 7 December 1994; revised 14 February 1995; accepted 22 February 1995

Abstract

Mucopolysaccharidosis IIID results from the deficiency of *N*-acetylglucosamine 6-sulfatase activity. A Nubian goat with this lysosomal storage disease has been identified. As a first step in developing this animal model for testing treatment methods, we cloned and sequenced the caprine *N*-acetylglucosamine 6-sulfatase cDNA coding region. Overall there is 88% nucleotide homology between the goat and human sequence and 94% homology of the deduced amino acid sequence. The human and two ruminant species differ by the presence of an imperfect trinucleotide (CCG) repeat in the ruminant signal sequence.

Keywords: Caprine; *N*-Acetylglucosamine 6-sulfatase; cDNA; Mucopolysaccharidosis IIID; Comparative genomics; Trinucleotide repeat; (Human)

The mucopolysaccharidoses (MPS) are lysosomal storage diseases characterized by inherited deficiencies of lysosomal enzymes which degrade glycosaminoglycans. Ten different enzyme deficiencies that result in MPS disorders have been identified; four of these enzymes are glycosidases, five are sulfatases and one is a non-hydrolytic transferase. Animal models of four of the human MPS have been described. One of these models is caprine *N*-acetylglucosamine 6-sulfatase (G6S, EC 3.1.6.14) deficiency [1] which represents the only animal analog of Sanfilippo D syndrome (MPS IIID).

The goat with MPS IIID demonstrated delayed motor development and growth retardation but reached sexual maturity before dying suddenly at 19 months of congestive heart failure [1]. Biochemical and histochemical analyses showed increased urinary accumulation of heparan sulfate and free *N*-acetylglucosamine sulfate with glycosaminoglycan storage in the liver. GM_3 ganglioside accumulation was demonstrated in the brain. Cloning and sequencing of the G6S cDNA was undertaken for development of this animal model for testing of treatment strategies.

Human G6S cDNA has been isolated and sequenced for the entire coding region of this gene product [2,3]. The human cDNA was used to isolate and sequence the goat cDNA.

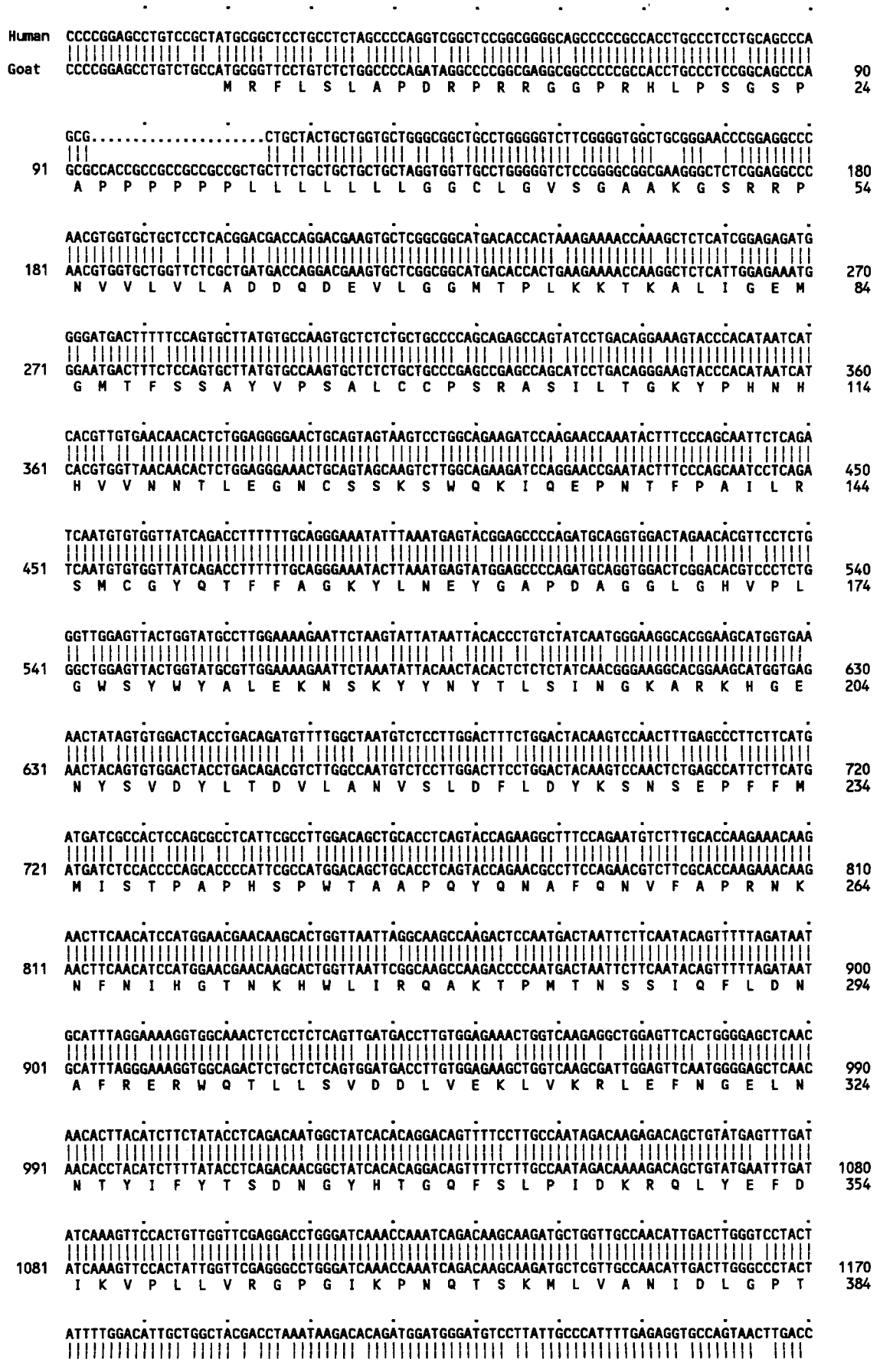
The goat cDNA was obtained by combining two approaches. The major portion of the cDNA was obtained by screening a custom made λZAP caprine kidney cDNA library (Clontech) using two human G6S probes. Four clones were isolated using a 1700 bp fragment of a human G6S clone [2], but none of these clones contained 5′ end coding information. Therefore, to obtain clones representing additional 5′ sequence, a 400 bp human amplicon of the 5′ coding region [3] was used as a probe to isolate three more clones. These clones provided partial additional 5′ end information. All clones were sequenced from both strands and assembled into a consensus sequence. In total, library screening yielded 80% of the coding sequence and 708 bp of the 3′ untranslated region.

Since the clones did not extend to the extreme 5′ end of the coding region or the 5′ untranslated region, the sequence of this region was obtained using RT-PCR. Total RNA was obtained from goat brain following necropsy and prepared using standard techniques [4]. First strand cDNA synthesis was performed using Moloney Murine Leukemia Virus Reverse Transcriptase (Gibco BRL) using antisense primers determined from the cloned goat se-

☆ The sequence data reported in this paper have been submitted to the EMBL/GenBank Data Libraries under the accession number U17694.

* Corresponding author. E-mail: jones@pathology.cvm.msu.edu. Fax: +1 (517) 4321053.

SSDI 0925-4439(95)00054-2


Human CCCCGGAGCCTGTCCGCTATGCGGCTCCTGCCTCTAGCCCCAGGTCGGCTCCGGCGGGGCAGCCCCCGCCACCTGCCCTCCTGCAGCCCA
Goat CCCCGGAGCCTGTCTGCCATGCGGTTCCTGTCTCTGGCCCCAGATAGGCCCCGGCGAGGCGGCCCCCGCCACCTGCCCTCCGGCAGCCCA 90
M R F L S L A P D R P R R G G P R H L P S G S P 24
GCG.....CTGCTACTGCTGGTGCTGGGCGGCTGCCTGGGGGTCTTCGGGGTGGCTGCGGGAACCCGGAGGCCC
91 GCGCCACCGCCGCCGCCGCCGCTGCTTCTGCTGCTGCTGCTAGGTGGTTGCCTGGGGGTCTCCGGGGCGGCGAAGGGCTCTCGGAGGCCC 180
A P P P P P P L L L L L L L G G C L G V S G A A K G S R R P 54
AACGTGGTGCTGCTCCTCACGGACGACCAGGACGAAGTGCTCGGCGGCATGACACCACTAAAGAAAACCAAAGCTCTCATCGGAGAGATG
181 AACGTGGTGCTGGTTCTCGCTGATGACCAGGACGAAGTGCTCGGCGGCATGACACCACTGAAGAAAACCAAGGCTCTCATTGGAGAAATG 270
N V V L V L A D D Q D E V L G G M T P L K K T K A L I G E M 84
GGGATGACTTTTTCCAGTGCTTATGTGCCAAGTGCTCTCTGCTGCCCCAGCAGAGCCAGTATCCTGACAGGAAAGTACCCACATAATCAT
271 GGAATGACTTTCTCCAGTGCTTATGTGCCAAGTGCTCTCTGCTGCCCGAGCCGAGCCAGCATCCTGACAGGGAAGTACCCACATAATCAT 360
G M T F S S A Y V P S A L C C P S R A S I L T G K Y P H N H 114
CACGTTGTGAACAACACTCTGGAGGGGAACTGCAGTAGTAAGTCCTGGCAGAAGATCCAAGAACCAAATACTTTCCCAGCAATTCTCAGA
361 CACGTGGTTAACAACACTCTGGAGGGAAACTGCAGTAGCAAGTCTTGGCAGAAGATCCAGGAACCGAATACTTTCCCAGCAATCCTCAGA 450
H V V N N T L E G N C S S K S W Q K I Q E P N T F P A I L R 144
TCAATGTGTGGTTATCAGACCTTTTTTGCAGGGAAATATTTAAATGAGTACGGAGCCCCAGATGCAGGTGGACTAGAACACGTTCCTCTG
451 TCAATGTGTGGTTATCAGACCTTTTTTGCAGGGAAATACTTAAATGAGTATGGAGCCCCAGATGCAGGTGGACTCGGACACGTCCCTCTG 540
S M C G Y Q T F F A G K Y L N E Y G A P D A G G L G H V P L 174
GGTTGGAGTTACTGGTATGCCTTGGAAAAGAATTCTAAGTATTATAATTACACCCTGTCTATCAATGGGAAGGCACGGAAGCATGGTGAA
541 GGCTGGAGTTACTGGTATGCGTTGGAAAAGAATTCTAAATATTACAACTACACTCTCTCTATCAACGGGAAGGCACGGAAGCATGGTGAG 630
G W S Y W Y A L E K N S K Y Y N Y T L S I N G K A R K H G E 204
AACTATAGTGTGGACTACCTGACAGATGTTTTGGCTAATGTCTCCTTGGACTTTCTGGACTACAAGTCCAACTTTGAGCCCTTCTTCATG
631 AACTACAGTGTGGACTACCTGACAGACGTCTTGGCCAATGTCTCCTTGGACTTCCTGGACTACAAGTCCAACTCTGAGCCATTCTTCATG 720
N Y S V D Y L T D V L A N V S L D F L D Y K S N S E P F F M 234
ATGATCGCCACTCCAGCGCCTCATTCGCCTTGGACAGCTGCACCTCAGTACCAGAAGGCTTTCCAGAATGTCTTTGCACCAAGAAACAAG
721 ATGATCTCCACCCCAGCACCCCATTCGCCATGGACAGCTGCACCTCAGTACCAGAACGCCTTCCAGAACGTCTTCGCACCAAGAAACAAG 810
M I S T P A P H S P W T A A P Q Y Q N A F Q N V F A P R N K 264
AACTTCAACATCCATGGAACGAACAAGCACTGGTTAATTAGGCAAGCCAAGACTCCAATGACTAATTCTTCAATACAGTTTTAGATAAT
811 AACTTCAACATCCATGGAACGAACAAGCACTGGTTAATTCGGCAAGCCAAGACCCCAATGACTAATTCTTCAATACAGTTTTAGATAAT 900
N F N I H G T N K H W L I R Q A K T P M T N S S I Q F L D N 294
GCATTTAGGAAAAGGTGGCAAACTCTCCTCTCAGTTGATGACCTTGTGGAGAAACTGGTCAAGAGGCTGGAGTTCACTGGGGAGCTCAAC
901 GCATTTAGGGAAAGGTGGCAGACTCTGCTCTCAGTGGATGACCTTGTGGAGAAGCTGGTCAAGCGATTGGAGTTCAATGGGGAGCTCAAC 990
A F R E R W Q T L L S V D D L V E K L V K R L E F N G E L N 324
AACACTTACATCTTCTATACCTCAGACAATGGCTATCACACAGGACAGTTTTCCTTGCCAATAGACAAGAGACAGCTGTATGAGTTTGAT
991 AACACCTACATCTTTTATACCTCAGACAACGGCTATCACACAGGACAGTTTTCTTTGCCAATAGACAAAAGACAGCTGTATGAATTTGAT 1080
N T Y I F Y T S D N G Y H T G Q F S L P I D K R Q L Y E F D 354
ATCAAAGTTCCACTGTTGGTTCGAGGACCTGGGATCAAACCAAATCAGACAAGCAAGATGCTGGTTGCCAACATTGACTTGGGTCCTACT
1081 ATCAAAGTTCCACTATTGGTTCGAGGGCCTGGGATCAAACCAAATCAGACAAGCAAGATGCTCGTTGCCAACATTGACTTGGGCCCTACT 1170
I K V P L L V R G P G I K P N Q T S K M L V A N I D L G P T 384
ATTTTGGACATTGCTGGCTACGACCTAAATAAGACACAGATGGATGGGATGTCCTTATTGCCCATTTTGAGAGGTGCCAGTAACTTGACC

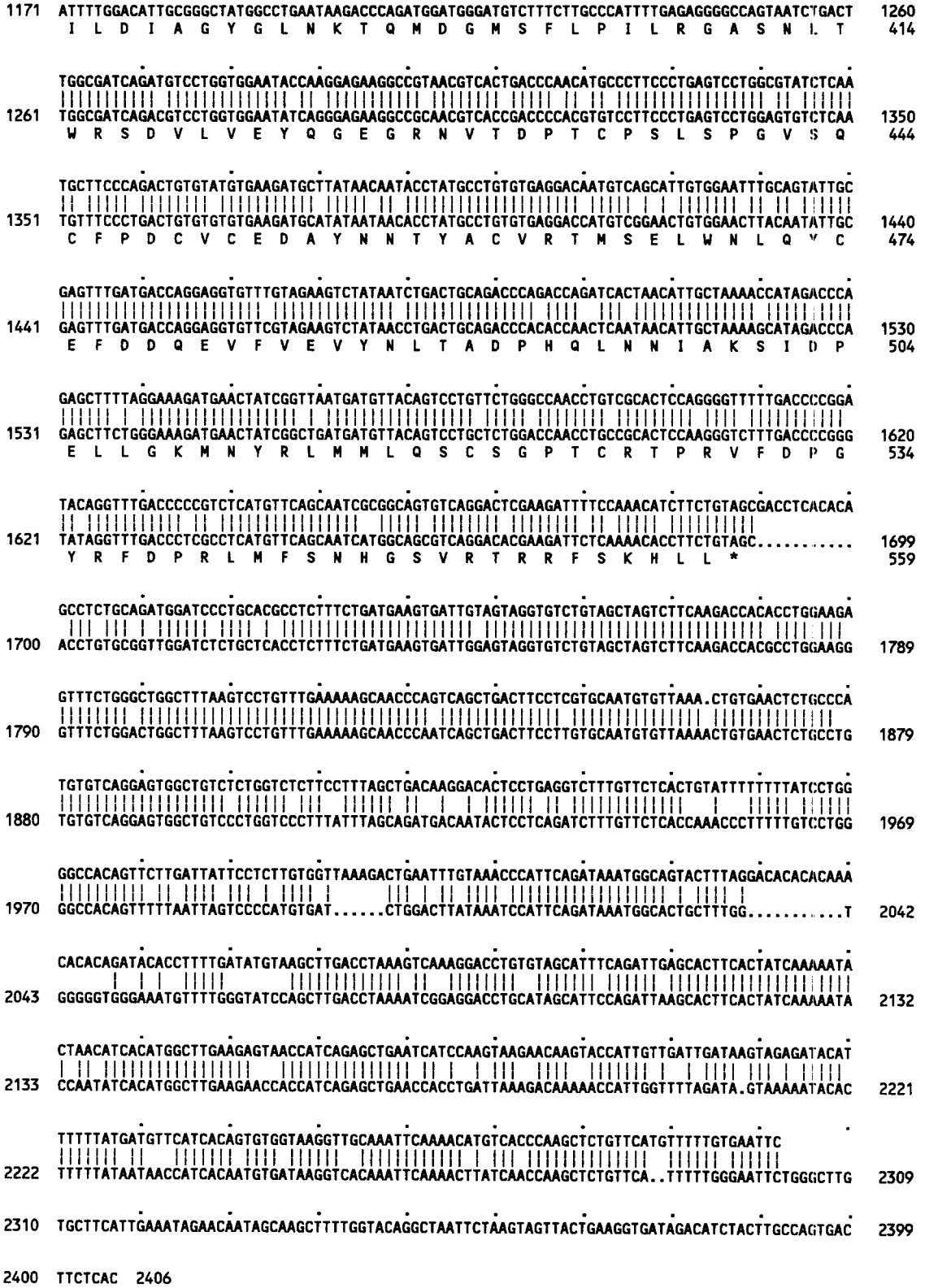


Fig. 1. Caprine *N*-acetylglucosamine 6-sulfatase cDNA with its deduced amino acid sequence and comparison to the human sequence [3]. Dots (...) indicate where gaps have been placed for optimal alignment.

quence. This area was PCR amplified using an antisense primer based on the human 5′ untranslated region [3] and sense primers based on the 5′ end of the caprine clones. This region was very GC rich, thus requiring use of 7.5% DMSO to amplify this segment. PCR reaction products were separated by 2% agarose (Nusieve 3:1, FMC Bio-

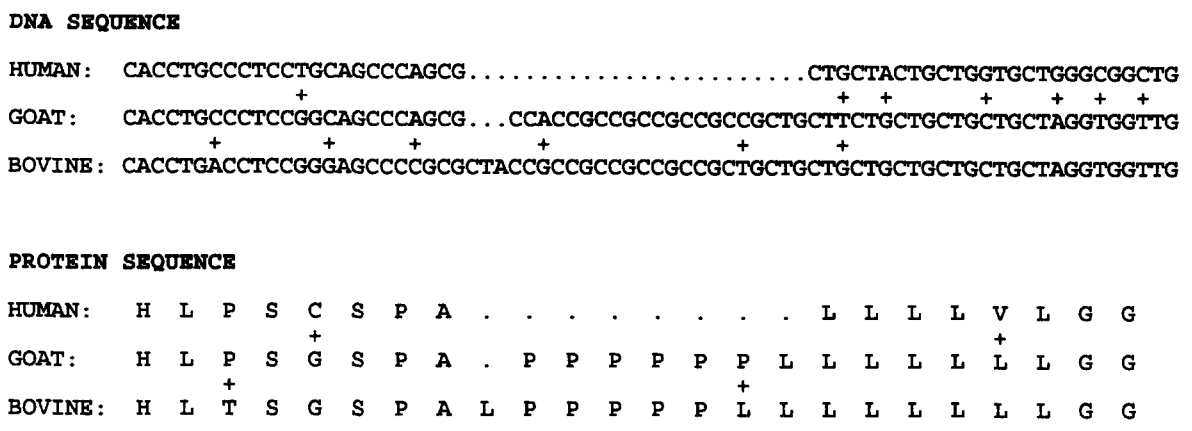


Fig. 2. Comparison of the caprine proline-rich segment of the signal peptide with the bovine and human counterparts and its associated nucleotide sequence. Plus signs (+) indicate a difference in the goat vs. human or bovine sequence; dots (...) indicate where gaps have been placed for optimal alignment.

products) electrophoresis and sequenced. Amplicons from several separate PCR reactions were sequenced for three goats. The 5′ sequence was also obtained for bovine cDNA using a similar protocol described above.

The entire coding region of G6S was verified by sequencing RT-PCR products from goat RNA. Either oligo(dT) or antisense primers were used for cDNA synthesis. Typically, 300–400 bp amplicons were produced using standard reaction conditions with AmpliTaq DNA polymerase (Perkin Elmer Cetus) and primers based on caprine G6S clone sequences. Some amplicons obtained in low yield required reamplification or nested PCR before sequencing. These amplicons were either diluted and applied directly to the subsequent PCR reaction or they were first separated by electrophoresis (2% Nusieve GTG,FMC Bioproducts) and capillary gel plugs of the bands interest used as a template for the PCR.

Goat G6S cDNA and its deduced amino acid sequence are shown in Fig. 1. This goat cDNA codes for 559 amino acids as compared to 552 amino acids for the human G6S. Homology between the goat and human cDNA sequences was 88% and 94% from the nucleotide sequence and amino acid sequences, respectively.

Analysis of the caprine and human [3] amino acid sequences using the weight-matrix approach [5] predicted that the most likely signal peptide cleavage site for the goat peptide was on the N-terminal side of Ser-51. Comparison of the caprine nucleotide and deduced amino acid sequences of the signal peptide with those of the human counterpart yielded homologies of 84% and 76%, respectively. This lower homology was accompanied by the presence of a proline repeat with an extra leucine in the goat sequence, Pro-6,Leu, that was not present in the human sequence (Fig. 2). This proline-rich stretch was confirmed by amplifying and sequencing this same cDNA segment in two other goats.

The signal sequence of another ruminant was studied for comparison with the goat. The bovine sequence in this same cDNA segment was Leu,Pro-5,Leu-2 (Fig. 2). Comparison of the caprine G6S signal sequence to the bovine sequence yielded a 95% nucleotide homology and a 92% deduced amino acid homology.

The significance of this proline stretch is unclear. Many transcription factors contain proline and glutamine-rich domains [6]. However, it seems unlikely that *N*-acetylglucosamine 6-sulfatase plays a role in modulating transcription. The non-conservation of a trinucleotide repeat loci between species is in agreement with the data of Stallings [7] obtained by GenBank database searches. Even in coding sequences the length of the trinucleotide repeat was not maintained, suggesting that long tracts of homopolymers ($n > 7$) are frequently not critical for protein function [7].

In the coding region for caprine G6S, the nucleotide homology was 91% compared to the human gene with an 96% homology for the deduced amino acid sequences. In the noncoding region, the nucleotide homology was 83%.

Northern analysis of the caprine G6S mRNA shows two species of about 4.0 and 4.9 kb (data not shown) similar to the two large human G6S transcripts [8].

Knowledge of the entire cDNA sequence for G6S enabled determination of the DNA defect for caprine Sanfilippo D syndrome. A stop codon in the 5′ coding regions was identified [9].

This work was funded by NS 16886 (M.Z.J.) and National Health and Medical Research Council of Australia (J.J.H.).

References

[1] Thompson, J.N., Jones, M.Z., Dawson, G. and Huffman, P.S. (1992) J. Inher. Metab. Dis. 115, 760–768.

[2] Robertson, D.A., Freeman, C., Nelson, P.V., Morris, C.P. and Hopwood, J.J. (1988) Biochem. Biophys. Res. Commun. 157, 218–224.

[3] Robertson, D.A., Freeman, C., Morris, C.P. and Hopwood, J.J. (1992) Biochem. J. 288, 539–544.
[4] Ausbel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (1987) Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience, New York.
[5] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4691.
[6] Gerber, H.P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S. and Schaffner, W. (1994) Science 263, 808–811.
[7] Stallings, R.L. (1994) Genomics 21, 116–121.
[8] Siciliano, L., Giumara, A. Pavone, L., Freeman, C., Robertson, D., Morris, C.P., Hopwood, J.J., Di Natale, P., Musumeci, S. and Horwitz, A.L. (1991) J. Med. Genet. 28, 402–405.
[9] Cavanagh, K.T., Leipprandt, J.R., Jones, M.Z. and Friderici, K.F. (1995) J. Inher. Metab. Dis. 18, in press.