Vol. 157, No. 1, 1988
November 30, 1988
BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 218-224

HUMAN GLUCOSAMINE-6-SULFATASE cDNA REVEALS HOMOLOGY WITH STEROID SULFATASE

Daniel A. Robertson, Craig Freeman,
Paul V. Nelson, C. Phillip Morris, and John J. Hopwood*

Lysosomal Diseases Research Unit, Department of Chemical Pathology,
Adelaide Children's Hospital, 72 King William Road, North Adelaide,
South Australia 5006, Australia

Received September 19, 1988

Glucosamine-6-sulfatase is a lysosomal enzyme which degrades glycosaminoglycans and is deficient in mucopolysaccharidosis type IIID. Human liver contains two major active forms of glucosamine-6-sulfatase, form A which has a single 78 kDa polypeptide and form B which has two polypeptides of 48 kDa and 32 kDa. A 1761 base pair cDNA clone encoding the complete 48 kDa polypeptide of form B was isolated. Form A is shown to be processed to form B with the 48 kDa polypeptide C-terminal to the 32 kDa polypeptide, and it is shown that C-terminal processing is limited to a region of thirty amino acids. The glucosamine-6-sulfatase sequence reveals homology with steroid sulfatase, a microsomal enzyme.

The human genome encodes at least nine separate enzymes which hydrolyse sulfate esters. Steroid sulfatase (STS) is a microsomal enzyme which acts upon a number of 3β-hydroxysteroids (1), while the others are lysosomal *exo*hydrolases (2,3). Five of these lysosomal sulfatases are involved in the stepwise degradation of heparan sulfate and heparin (3). Though they act upon a common substrate and catalyse very similar reactions, they exhibit an exquisite specificity and thus make an attractive model for investigations into the molecular details of enzyme/substrate interactions and sulfate ester hydrolysis. Glucosamine-6-sulfatase (G6S) cleaves the O-linked sulfate group at the 6-position of the non-reducing-end glucosaminide residues of heparan sulfate, heparin and keratan sulfate (4-6). A deficiency in the activity of G6S results in the accumulation of partially degraded heparan sulfate in lysosomes causing organelle, cell and tissue distortion, ultimately leading to the lysosomal storage disorder, mucopolysaccharidosis type IIID, or Sanfilippo D syndrome (4). G6S has been purified from normal human liver in two major forms, one with a single subunit of M_r of 78 kDa and the other with two

* To whom reprint requests should be addressed.

Abbreviations: G6S, glucosamine-6-sulfatase; STS, steroid sulfatase; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis.

0006-291X/88 $1.50

subunits of M_r 48 kDa and 32 kDa (7). The gene for G6S has been localized to human chromosome 12q14 (8).

In this report we describe the isolation and identification of a cDNA clone coding for human G6S using oligonucleotides directed against peptide sequence data generated from pure G6S. Comparison of the amino acid sequence of the encoded protein with that of STS (1), the only other sulfatase with gene sequence available, revealed regions of significant homology.

MATERIALS AND METHODS

Materials. G6S was purified from human liver as previously described (7). The human liver cDNA library in λgt10 was kindly donated by G. Howlett (University of Melbourne).

Peptide isolation and sequencing. Peptides for amino acid sequencing were produced in three ways: Approximately 50 μg of form A and form B human liver G6S (Fig. 1a) were subjected to SDS-PAGE, the major bands were isolated and subjected to N-terminal amino acid sequencing (9); approximately 150 μg of form A was digested with trypsin (10); approximately 150 μg of form B was cleaved with cyanogen bromide (11); total tryptic and cyanogen bromide digests were applied to an Aquapore RP-300 (Brownlee Labs Inc) cartridge (30x4.6 mm), and the peptides eluted with a linear 180 min gradient from 0-30% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid at a flowrate of 1.0 ml/min. Absorbance at 216 nm was followed and fractions were collected manually and stored at -75°C. Amino acid sequencing and oligonucleotide synthesis were services of BRESATEC (Adelaide, S.A., Australia).

Oligonucleotide Screening. Two oligonucleotide mixtures were used to screen a human liver cDNA library in λgt10 (12,13). The bacterial host used was C600 (14) and approximately 250,000 recombinants were plated and screened with the oligonucleotide mixtures. Each species in the probe mixture was present at 180 pM. Hybridization was conducted overnight at 37°C, and washing conditions were established empirically for each probe mix.

General Methods. Standard DNA techniques were used throughout (14). Probe preparation and DNA blot analysis are described in (8).

DNA Sequencing. Insert cDNA was subcloned into the Eco RI site of pUC19 for restriction endonuclease analysis. Selected fragments were subcloned into M13mp19 for sequencing *via* the dideoxy chain termination method (15) using the Klenow fragment of *E. coli* DNA polymerase I at 47°C.

Sequence Analysis. The nucleotide sequence was screened against the GenBank database and the encoded protein sequence against the NBRF protein database. General sequence analysis and comparisons were performed using a package of programs supplied by A. Reisner (16).

RESULTS AND DISCUSSION

G6S from human liver can be purified as two major forms which together account for some 90% of G6S activity. These two forms, designated A and B, both have a native M_r of approximately 75 kDa and a pI greater than 9.5, they have similar activity towards heparan sulfate and keratan sulfate substrates (17), but they have different subunit compositions. Form A contains a single polypeptide with an apparent M_r of 78 kDa, while form B consists of two

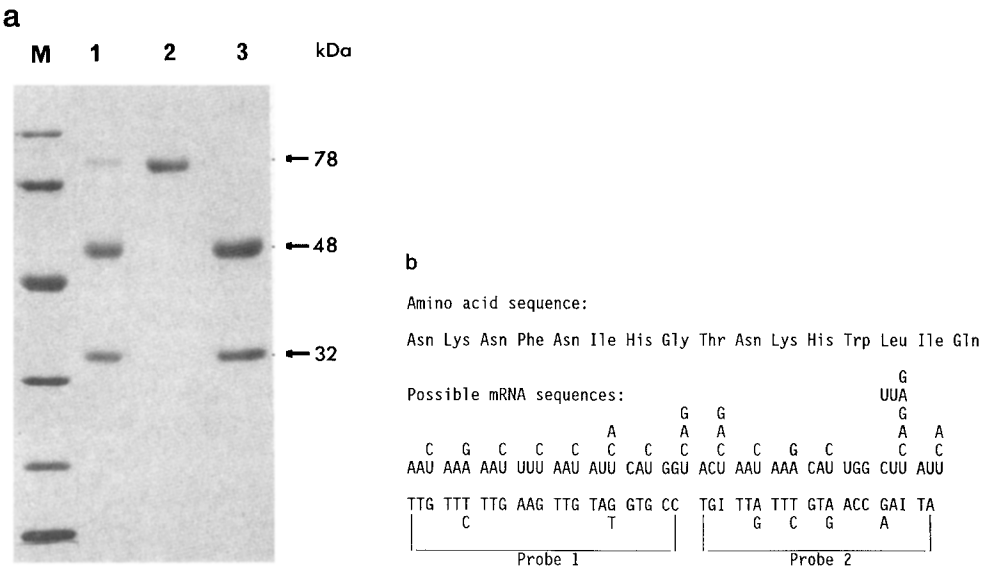

FIG. 1. a; SDS-PAGE of human liver G6S. All samples were reduced with dithioerythritol. Track 1, a mixture of form A and form B G6S; track 2, form A G6S; track 3, form B G6S; track M, molecular weight markers. The molecular weights of the major species present are indicated at the side of the diagram in kDa. b; Oligonucleotides based on amino acid sequence of tryptic peptide 45 derived from form A G6S (A/tryp45 in Table 1). I = inosine.

polypeptides with apparent M_r of 48 kDa and 32 kDa (Fig. 1a). It has been proposed that form B is produced by proteolytic processing of form A (7).

The three polypeptides were subjected to direct N-terminal amino acid sequencing. Both forms were also digested to produce peptides that were separated by HPLC. Several of these peptides which fractionated as large well isolated peaks were sequenced (Table 1). One of these peptides, A/tryp45, was

TABLE 1. Amino acid sequences of peptides derived from G6S. Peptide designations indicate which form of G6S the peptide was derived from [A/B], how it was derived [trypsin/cyanogen bromide/N-terminal sequencing] and which HPLC fraction/polypeptide species was used. The location of peptides in the sequence (Fig. 2) is indicated in brackets. "Xaa" indicates that the amino acid at that position was not identified, "?" indicates a tentative identification.

A/tryp45 (79-94)
Asn Lys Asn Phe Asn Ile His Gly Thr Asn Lys His Trp Leu Ile Gln
A/tryp52 (245-261)
Ser Val Thr Asp Pro Thr Cys Pro Ser Leu Ser/Leu Pro Gly Val/Ser Val? Gln Cys
B/CNBr59 (327-345)
Asn Tyr Arg Leu Met Met Leu Gln Ser Xaa Ser Gly Arg/Pro Thr Xaa Arg Thr Pro Gly
B/Nter48 (45-74)
Phe Glu Pro Phe Phe Met Met Ile Ala Thr Pro Ala Pro His Ser Pro Trp Thr Ala Ala Pro Gln Tyr Gln Lys Ala Phe Gln Asn Val

used to design two oligonucleotide mixtures (Fig. 1b) that were used to screen a human liver cDNA library in λgt10. Six positive clones were isolated after screening with probe 1. Of these clones only one was positive when screened with probe 2. The locations of cleavage sites for several restriction endonucleases were determined, and fragments of the cDNA insert were subcloned into M13mp19 for nucleotide sequencing.

The nucleotide sequence of the cDNA (Fig. 2) revealed a continuous open reading frame of 1125 bases which encodes a protein of 375 amino acids. Within the sequence four regions display colinearity with the sequences of peptides derived from pure G6S (Table 1). There are three apparent discrepancies between the direct and predicted amino acid sequence data. These are at amino acids 94, 245, and 259. We believe these to be amino acid sequencing errors. The nucleotide sequence data unambiguously predicts amino acid 245 to be an asparagine located in a consensus site for N-linked glycosylation. Such a modification, particularly at the first residue, would conceivably cause misidentification of the residue during direct amino acid sequencing. Indeed,

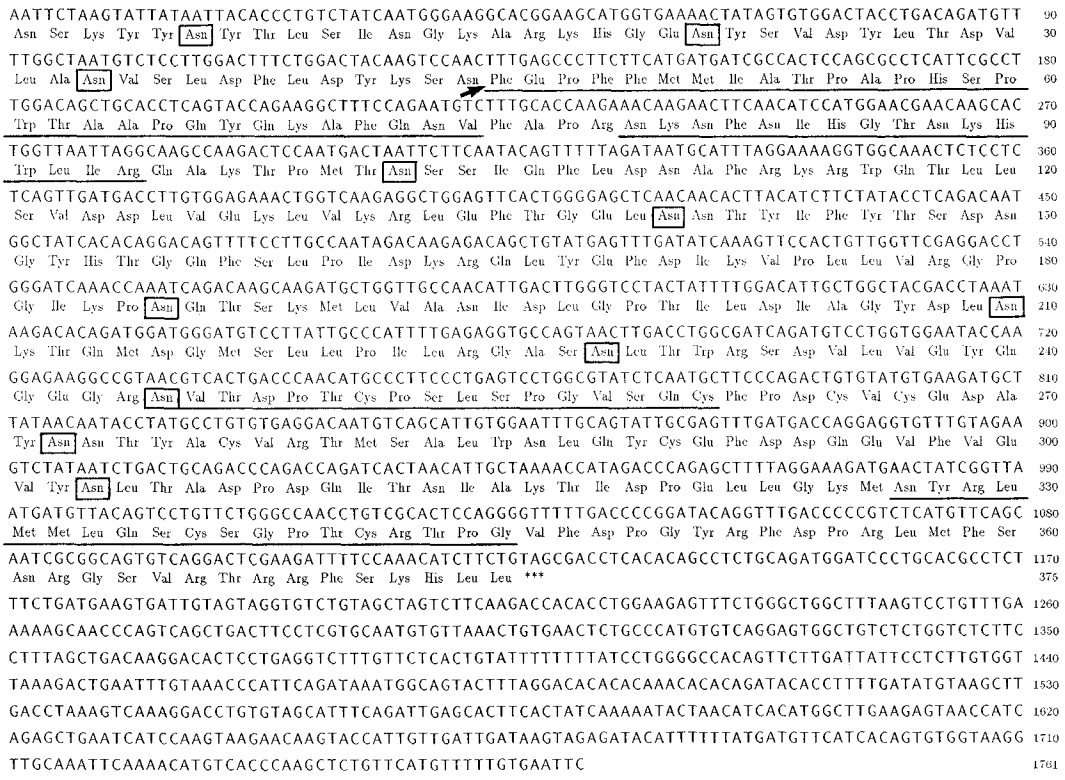

FIG. 2. Nucleotide and amino acid sequences of the G6S cDNA clone. Regions corresponding to peptides in Table 1 are underlined, possible sites for N-linked asparagine glycosylation are boxed, the N-terminus of the 48 kDa form B polypeptide is marked with an arrow, and the translation-stop codon is marked "***"

this data may be interpreted as evidence for glycosylation at this site. The otherwise consistent colinearity of the direct amino acid sequence data with the predicted amino acid sequence over more than eighty amino acids in four different peptides confirms that this clone encodes human G6S.

A translation-stop codon is found at nucleotide position 1126, leaving a long 3'-untranslated region of at least 636 bases. The N-terminus of the 48 kDa species is found at amino acid 45 (Fig. 2). These two points define the entire 48 kDa polypeptide of form B G6S as a polypeptide of 331 amino acids with a predicted M_r of 37.8 kDa on the basis of amino acid composition alone. Comparison of this predicted value with the M_r estimation by SDS-PAGE indicates that about 20% of the apparent M_r is contributed by modifications to the primary amino acid sequence. There are eight sites for possible N-linked asparagine glycosylation within the defined protein (Fig. 2). Glycosylation with oligosaccharides containing eight mannose residues at four or five of these sites would account for the apparent discrepancy in M_r.

The clone encodes peptides derived from the 78 kDa form A G6S and the 48 kDa species of form B. In addition, the N-terminal amino acid sequence of the 78 kDa form A and of the 32 kDa form B species are identical (C.F. unpublished data). These results indicate that, allowing for minimal inaccuracies in SDS-PAGE M_r estimation, a 78 kDa polypeptide (form A) is processed by internal peptidase cleavage to a 32 kDa N-terminal species and a 48 kDa C-terminal species (form B). If there is consistent modification throughout the polypeptide, then the G6S cDNA clone encodes approximately 70% of the mature 78 kDa G6S protein.

Proteolytic processing is observed in other lysosomal enzymes. It is suggested (18,19) that the sites of proteolytic processing of the β-hexosaminidase precursors are hydrophilic regions that expose sites for thiol and/or trypsin-like endopeptidase action, which is followed by exopeptidase trimming. There is a hydrophilic region of the G6S sequence from amino acids 14-23 which is rich in basic amino acids and may be an analogous site involved in the proteolytic processing that exposes the N-terminus of the 48 kDa form B polypeptide. There is also a lysine two residues before the 48 kDa N-terminus which may be a site for trypsin-like activity. Lysosomal enzymes generally are synthesized with transient carboxyl sequences that are removed during maturation. The extent of C-terminal processing in the maturation of form B, and presumably form A, G6S is confined to a region of 30 amino acid residues, since peptide B/CNBr59 has been directly sequenced to within 30 amino acids of the C-terminus of the precursor.

The nucleotide and encoded protein sequences of the G6S clone were screened against the GenBank and NBRF databases, and those lysosomal protein and gene sequences which are available. They were also compared to those of the only other sulfatase for which such information is available, STS (1,20).

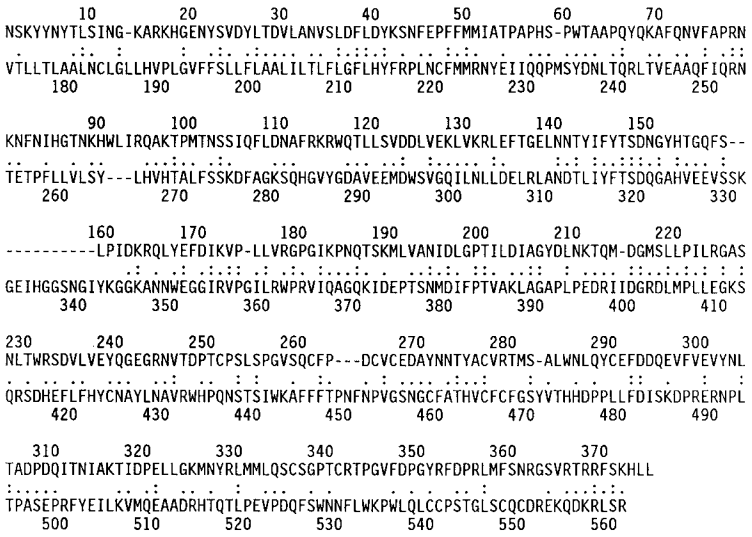

FIG. 3. Alignment of G6S, upper sequence, from amino acid 1 to the C-terminus with STS, lower sequence, from amino acid 173 to the C-terminus, both shown in single letter amino acid code. Positions with identical residues are marked (:), positions of conservative substitution are marked (.), and insertions are indicated (-). STS numbering is from reference 1.

No significant homology was detected with any species except the STS amino acid sequence. A region of local homology was found which aligned G6S residues 171-230 with STS residues 354-415. Using the algorithm of Needleman and Wunsch (21), comparison of the score of that alignment with the scores of 100 randomized sequences of the same lengths and compositions showed that the similarity between STS and G6S is statistically significant. The alignment of the authentic sequences generated a score that was 9.7 standard deviations greater than the score generated by the random sequences. This local alignment was used to produce a global alignment that illustrates a possible overall similarity between G6S and STS (Fig. 3). About 20% of amino acids match, and if conservative substitutions according to the mutation matrix scores of Dayhoff (22) are included, the homology is over 50%. There are regions of still higher local homology, notably the region that gave the original alignment, from G6S amino acids 170 to 230, where some 75% of the amino acids match or are conservative substitutions. It is tempting to speculate that this homology indicates that the two enzymes are related to an ancestral sulfohydrolase, and that their divergence reflects their different locations and substrates. The biological significance of this homology may become clearer with the cloning of genes for other sulfatases and lysosomal enzymes.

ACKNOWLEDGMENTS

We would like to thank L.J. Shapiro, Department of Pediatrics, UCLA School of Medicine for helpful discussions on STS homology, P.R. Clements, Department of Chemical Pathology, Adelaide Children's Hospital for advice on N-terminal sequencing of G6S and G. Howlett, Department of Biochemistry, University of Melbourne for providing the human liver cDNA library. This work was supported by The Australian National Health and Medical Research Council and the Adelaide Chidren's Hospital Research Foundation.

REFERENCES

1. Yen, P.H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R.T. and Shapiro, L.J. (1987) Cell **49**, 443-454.
2. McKusick, V.A. and Neufeld, E.F. (1983) in The Metabolic Basis of Inherited Disease, eds. Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. and Brown, M.S. (McGraw-Hill, New York), pp. 751-777.
3. Hopwood, J.J., (1988) in Heparin, eds. Lane, D. and Lindahl, U. (Arnold, London), in press.
4. Kresse, H., Paschke, E., von Figura, K., Gilbert, W. and Fuchs, W. (1980) Proc. Natl. Acad. Sci. USA **77**, 6822-6826.
5. Hopwood, J.J. and Elliott, H. (1983) Biochem. Int. **6**, 141-148.
6. Fuchs, W., Beck, M. and Kresse, H. (1985) Eur. J. Biochem. **151**, 551-556.
7. Freeman, C., Clements, P.R. and Hopwood, J.J. (1987) Biochem. J. **246**, 347-354.
8. Robertson, D.A., Callen, D.F., Baker, E.G., Morris, C.P. and Hopwood, J.J. (1988) Hum. Genet. **79**, 175-178.
9. Matsudaira, P. (1987) J. Biol. Chem. **262**, 10035-10038.
10. van Driel, I.R., Stearne, P.A., Grego, B., Simpson, R.J. and Goding, J.W. (1984) J. Immunol. **133**, 3220-3224.
11. O'Dowd, B.F., Quan, F., Huntington, F.W., Lamhonwah, A., Korneluk, R.G., Lowden, J.A., Gravel, R.A. and Mahuran, D.J. (1985) Proc. Natl. Acad. Sci. USA **82**, 1184-1188.
12. Whitehead, A.S., Goldberg, G., Woods, D.E., Markham, A.F. and Colten, H.R. (1983) Proc. Natl. Acad. Sci. USA **80**, 5387-5391.
13. Huynh, T.V., Young, R.A., and Davis, R.W. (1985) in DNA cloning, ed. D.M. Glover (IRL Press Limited, Oxford) Vol. 1, pp. 49-78.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbour Laboratory, Cold Spring Harbour, New York).
15. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
16. Reisner, A.H. and Bucholtz, C. (1985) Nature **314**, 310.
17. Freeman, C. and Hopwood, J.J. (1987) Biochem. J. **246**, 355-365.
18. Little, L.E., Lau, M.M.H., Quon, D.V.K., Fowler, A.V. and Neufeld, E.F. (1988) J. Biol. Chem. **263**, 4288-4292.
19. Mahuran, D.J., Neote, K., Klavins, M.H., Leung, A. and Gravel, R.A. (1988) J. Biol. Chem. **263**, 4612-4618.
20. L.J. Shapiro, personal communication.
21. Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. **48**, 443-453.
22. Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in Atlas of Protein Sequence and Structure, ed. Dayhoff, M.O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, Supplement 3, pp. 345-352.