A cDNA clone for human glucosamine-6-sulphatase reveals differences between arylsulphatases and non-arylsulphatases

Daniel A. ROBERTSON, Craig FREEMAN, 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

Glucosamine-6-sulphatase is an *exo*-hydrolase required for the lysosomal degradation of heparan sulphate and keratan sulphate. Deficiency of glucosamine-6-sulphatase activity leads to the lysosomal storage of the glycosaminoglycan, heparan sulphate and the monosaccharide sulphate *N*-acetylglucosamine 6-sulphate and the autosomal recessive genetic disorder mucopolysaccharidosis type IIID. Glucosamine-6-sulphatase can be classified as a non-arylsulphatase since, relative to arylsulphatase B, it shows negligible activity toward 4-methylumbelliferyl sulphate. We have isolated human cDNA clones and derived amino acid sequence coding for the entire glucosamine-6-sulphatase protein. The predicted sequence has 552 amino acids with a leader peptide of 36 amino acids and contains 13 potential *N*-glycosylation sites, of which it is likely that 10 are used. Glucosamine-6-sulphatase shows strong sequence similarity to other sulphatases such as the family of arylsulphatases, although the degree of similarity is not as high as that between members of the arylsulphatase family. This pattern of inter- and intra-family similarity delineates regions and amino acid residues that may be critical for sulphatase function and substrate specificity.

INTRODUCTION

In humans nine sulphatase activities have been identified which are now classified according to the structure of their natural substrates. Sulphatase activities were discovered by their ability to desulphate the artificial aryl sulphates. These were called arylsulphatases A, B and C before their natural substrates were identified [1]. However not all sulphatases are able to effectively desulphate aryl sulphates. Thus sulphatases can be classified as arylsulphatases or non-arylsulphatases according to their activity toward the aryl sulphate 4-methylumbelliferyl sulphate. Glucosamine-6-sulphatase (G6S, EC 3.1.6.14) cleaves the *O*-linked sulphate group at the 6-position of the non-reducing-end glucosamine residues of heparan sulphate and keratan sulphate [2–4], but its activity toward aryl sulphate substrates has not previously been characterized. A deficiency in the activity of G6S results in the accumulation of partially degraded heparan sulphate in lysosomes and its excretion in urine along with the monosaccharide *N*-acetylglucosamine 6-sulphate, the product of the block in the keratan sulphate degradative pathway [5,6]. The lysosomal accumulation of heparan sulphate causes organelle, cell and tissue distortion, ultimately leading to the lysosomal storage disorder mucopolysaccharidosis type IIID, or Sanfilippo D syndrome [2]. This disorder is characterized by relatively mild somatic mucopolysaccharidosis features, with severe progressive mental retardation becoming evident after 2–3 years of seemingly normal development [5]. Some phenotypic variation is apparent, in that less severe cases have been reported where later onset or only mild intellectual impairment was evident [7,8].

Human liver G6S exists as two major forms, form A with a single subunit of molecular mass 78 kDa, and form B, with two subunits of 48 and 32 kDa [9]. We have previously used purified G6S from normal human liver to isolate a cDNA clone covering

the C-terminal half of G6S [10]. This clone was used to localize the G6S gene to human chromosome 12q14 [11], and sequence from this clone provided the first demonstration of amino acid sequence similarity between sulphatases [10].

Several other sulphatase genes have been isolated and sequenced [12–18]. Here we report the isolation of cDNA clones encoding the complete amino acid sequence of the entire G6S enzyme. We show that G6S is a member of the non-arylsulphatase family and use a comparison of G6S with the arylsulphatases to identify regions and residues that may be critical for the catalysis and substrate specificity of these enzymes.

MATERIALS AND METHODS

Materials

G6S was purified from human liver as previously described [9]. Arylsulphatase B (*N*-acetylgalactosamine-4-sulphatase) was purified from human liver by immunoaffinity chromatography [19]. Restriction endonucleases, polynucleotide kinase, T4 DNA ligase, Klenow enzyme and M13 sequencing kits were purchased from Boehringer Mannheim (Mannheim, Germany); Gene-Screen-Plus nylon filters were from du Pont/NEN (Boston, MA, U.S.A.); [γ -³²P]ATP (5000 Ci/mmol), [α -³²P]dCTP (3000 Ci/mmol) and Multiprime DNA Labelling kits were purchased from Amersham International, and the λ gt10 random primed human colon cDNA library (1.5×10^6 independent clones) was supplied by Clontech (Palo Alto, CA, U.S.A.). Amino acid sequencing and oligonucleotide synthesis were services of Bresatec (Adelaide, Australia).

Polypeptide isolation and sequencing

Approx. 50 μ g of form A and form B human liver G6S were subjected to SDS/PAGE; the major bands were isolated

Abbreviations used: G6S, glucosamine-6-sulphatase; GlcNAc6S-IdoA2S-[³H]anM6S, α -*N*-acetylglucosamine-6-sulphate-(1–4)-*L*-*O*-(α -iduronic acid 2-sulphate)-(1–4)-*D*-*O*-2,5-anhydro[1-³H]mannitol 6-sulphate; GalNAc4S-GlcA-GalNAcitol4S, β -*N*-acetylgalactosamine-4-sulphate(1–4)-*D*-*O*-(β -glucuronic acid)-(1–3)-*D*-*O*-*N*-acetyl[1-³H]galactosaminitol-4-sulphate.

* To whom correspondence and reprint requests should be sent.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number M23657.

Table 1. Peptide sequence of the N-terminus of 78, 45 and 32 kDa polypeptides of human liver G6S in single-letter code and the sequence of the oligonucleotide probe mixture used to detect N-terminal coding sequence

The underlining represents the position of the oligonucleotide in the 78 and 32 kDa peptide sequences; 'X' indicates an unknown amino acid residue, and 'I' indicates the presence of an inosine residue in the oligonucleotide. Abbreviation: *M*, molecular mass.

<i>M</i> (kDa)		Sequence													
Human liver G6S polypeptides	78	T	R	R	P	N	V	V	L	L	L	T	D	D	Q
	45	F	E	P	F	F	M	M	I	A	T	I	A	T	P
	32	X	R	R	P	N	V	V	X	L	L	T	D	D	Q
Oligonucleotide probe	—	TGI	CTG	CTG	GTC	CTG	CTC	CT							
			A	A	T	A	T								

and subjected to N-terminal amino acid sequencing as previously described [17].

Isolation of cDNA clones

A 20-mer oligonucleotide sequence was designed (Table 1) from the N-terminal amino acid sequence of the 78 kDa polypeptide of human liver G6S. This oligonucleotide mixture was end-labelled and used to screen clones isolated from the colon library using a 1762 bp G6S cDNA clone (p6S18) which was previously isolated from a human liver cDNA library [10]. The bacterial host used was *E. coli* C600, and approx. 5×10^6 recombinants were plated and screened with p6S18. Clones extending further in the 5' direction were isolated by screening a λ gt11 human endothelial cDNA library (Clontech) with a pair of oligonucleotides from the 5' end of the longest colon cDNA clone (6SN1 and 6SN2).

Nucleotide sequencing

*Eco*RI insert fragments were subcloned into M13mp19 and sequenced using the dideoxy-chain-termination method [20]. Forward and reverse λ gt10 sequencing primers (New England Biolabs; Beverly, U.S.A.) were used for PCR amplification of inserted DNA [21]. The amplified DNA was digested with *Rsa*I, subcloned into M13mp19 and sequenced. Sequencing of the longest endothelial cDNA clone was performed by sequencing the ends after cloning into M13mp18 then sequencing the extra region of DNA by amplification from the lambda DNA clone using the reverse sequencing primer of λ gt11 (gt11R) in combination with the G6S primer 6SN2. This PCR product was then sequenced in both directions by direct cycled sequencing using gt11R and 6SN2.

Sequence analysis

The nucleotide sequence was screened against the GenBank database (Release 71, 15 March 1992), and the encoded protein sequence against the Swissprot (Release 21, March 1992) and GenBank Peptide (Release 71, 15 March 1992) protein databases.

Multiple amino-acid-sequence alignment of the arylsulphatases was performed using the method of Lipman *et al.* [22]. This alignment maximized sequence similarities by taking into account conservative amino acid substitutions, using the mutational data of Dayhoff *et al.* [23] while minimizing gaps and preventing bias toward the more closely related sequences. Optimal alignment for the arylsulphatases (A, B and C) and sea-urchin arylsulphatase were obtained by dividing the sequences into overlapping segments. The G6S sequence was added to the alignment using programs from a package described by Devereux *et al.* [24].

Glucosamine-6-sulphatase and arylsulphatase activity assays

A radiolabelled trisaccharide, α -N-acetylglucosamine-6-

sulphate)-(1-4)-L-O-(α -iduronic acid-2-sulphate)-(1-4)-D-O-2,5-anhydro[1-³H]mannitol-6-sulphate (GlcNAc6S-IdoA2S-[³H]-anM6S), was used as a substrate for the specific assay of G6S activity [25]. Arylsulphatase B and G6S activity toward 4-methylumbelliferyl sulphate was determined using previously described methods [19]. Incubation of G6S with 4-methylumbelliferyl sulphate was achieved by using a standard assay mixture containing 1.92 μ g of purified G6S, 5 mM-4-methylumbelliferyl sulphate and BSA (50 μ g/ml) in 60 mM-sodium acetate buffer, pH 4.8, in a total volume of 120 μ l. The mixture was incubated for 1 h at 37 °C, and the level of released 4-methylumbelliferone was determined [19].

TSK G3000 SW Ultrapac chromatography

Purified G6S (120 μ g in 250 μ l of buffer A containing 15 mM-sodium dimethylglutarate, pH 6.0, and 0.5 M-NaCl) was applied at room temperature to a LKB Ultrachrom GTi f.p.l.c. system with a TSK G3000 SW Ultrapac column (30 cm $\times$ 0.8 cm), equilibrated and eluted in buffer A at a flow rate of 0.5 ml/min.

Arylsulphatase enzyme kinetics

K_m and catalytic efficiency were obtained from Lineweaver-Burk plots with 4-methylumbelliferyl sulphate assayed within a concentration range of 0.4–6.3 mM in 60 mM-sodium acetate buffer (pH 5.5)/BSA (50 μ g/ml) in a total volume of 120 μ l. Incubations were for 20 min at 37 °C.

RESULTS AND DISCUSSION

The classification of G6S as an arylsulphatase or nonaryl-sulphatase can be achieved by comparison of its substrate specificity with arylsulphatase B. The natural substrates for arylsulphatase B are non-reducing-end N-acetylgalactosamine-4-sulphate residues in dermatan sulphate oligosaccharides [19]. Arylsulphatase B has no detectable activity toward GlcNAc6S-IdoA2S-anM6S and a catalytic efficiency of 285 min⁻¹·M⁻¹ toward 4-methylumbelliferyl sulphate and 1900 min⁻¹·M⁻¹ toward its natural substrate (Table 2). Under similar conditions G6S has a catalytic efficiency of 9315 min⁻¹·M⁻¹ toward GlcNAc6S-IdoA2S-anM6S compared with a catalytic efficiency of 0.36 min⁻¹·M⁻¹ toward 4-methylumbelliferyl sulphate. Waheed & van Etten [26] reported that purified rabbit arylsulphatase A has a K_m of 60 μ M and a catalytic efficiency of 467 min⁻¹·M⁻¹ toward cerebroside sulphate, its natural substrate, compared with a K_m of approx. 4 mM and catalytic efficiency of 134 min⁻¹·M⁻¹ toward 4-methylumbelliferyl sulphate. Incubation of the purified G6S with 5 mM-4-methylumbelliferyl sulphate in the presence of 5 mM-GlcNAc-6-sulphate, a poor substrate [25] for G6S but with a high substrate affinity (K_m 7.1 μ M), competitively inhibited G6S arylsulphatase activity by 90%. Under similar conditions purified arylsulphatase B was not inhibited.

Table 2. Comparison of arylsulphatase B and G6S activities toward 4-methylumbelliferyl sulphate and their natural substrates

Substrate and parameter	Parameter value	
	Arylsulphate B	G6S
4-Methylumbelliferyl sulphate		
K_m (mM)	4.0	4.2
k_{cat}/K_m^*	285.0	0.36
pH optimum	5.5	4.8
GlcNAc6S-IdoA2S-anM6S		
K_m (μ M)	nd†	0.76
k_{cat}/K_m	nd	9315.0
pH optimum	nd	4.3
GalNAc4S-GlcA-GalNAcitol4S		
K_m (μ M)	60.0	nd
k_{cat}/K_m	1900.0	nd
pH optimum	4.0	nd

* k_{cat}/K_m , catalytic efficiency ($\text{min}^{-1} \cdot \text{M}^{-1}$).

† k_{cat} , catalytic-centre activity (mol/min per mol of enzyme).

‡ nd, no activity detected.

Fig. 1. G6S cDNA structure and sequencing strategy

(a) Coding sequences are represented by the open box and 5'- and 3'-untranslated sequences by a thin line. (b) Three overlapping G6S cDNA clones showing restriction sites used for subcloning (E, *EcoRI*; R, *RsaI*). p6S18 has been previously sequenced [10], and p6S41 and p6S50 are described in the present paper. (c) Sequencing strategy for p6S41 and p6S50. The scale is in kbp.

Size filtration of the purified G6S on TSK G3000 failed to separate the arylsulphatase activity and the specific sulphatase activity toward GlcNAc6S-IdoA2S-anM6S. These results demonstrate that G6S binds 4-methylumbelliferyl sulphate with binding efficiencies similar to those of arylsulphatase A and B but, relative to these two arylsulphatases and particularly to its natural substrate, catalytic desulphation of 4-methylumbelliferyl sulphate is extremely low. We therefore conclude that G6S is not an arylsulphatase.

The *N*-terminal amino acid sequences of the 32, 48 and 78 kDa polypeptides of G6S were determined after separation on SDS/PAGE (Table 1). The 78 kDa and 32 kDa polypeptides have identical *N*-terminal amino acid sequences, consistent with the proposal that form B is derived from form A by internal proteolysis to produce a 32 kDa *N*-terminal species and a 48 kDa C-terminal species [9,10].

Approx. 5×10^5 plaques of the colon cDNA library were screened with the previously isolated 1.8 kb G6S cDNA (p6S18). In all, 30 positive clones were isolated and rescreened with the

oligonucleotide probe mixture designed from the 78 kDa *N*-terminal amino acid sequence. A G6S cDNA clone (p6S41) that was positive to both p6S18 and the oligonucleotide to the *N*-terminal amino acid sequence was isolated and sequenced according to the strategy shown in Fig. 1. The sequence obtained extended to encode all but the first six amino acid residues of the *N*-terminal peptide sequence from the 78 kDa and 32 kDa polypeptides, but did not contain any sequence from the leader peptide or the 5' untranslated region of G6S mRNA. In order to obtain a full-length clone, two oligonucleotides (6SN1 and 6SN2) were synthesized and used to screen a human endothelial cDNA library. One clone that was positive to both oligonucleotides (p6S50) was sequenced as shown in Fig. 1.

The 2379 bp of nucleotide sequence derived from three overlapping cDNA clones (p6S18, p6S41 and p6S50; Fig. 1) that is presented in Fig. 2 contains the full coding sequence (552 amino acids) of G6S as well as 87 bp of 5' untranslated sequence and 634 bp of 3' untranslated region, but did not contain a polyadenylation signal or poly(A) sequence. The authenticity of the encoded protein was confirmed by co-linearity of the sequence with four tryptic peptides [10] and the 78, 48 and 32 kDa *N*-terminal amino acid sequences (Fig. 2), and by similarity to other sulphatases (Fig. 3). The position of the 48 kDa polypeptide was identified beginning at amino acid position 222 and the positions of the 32 and 78 kDa *N*-termini were found at amino acid position 44. The amino acid sequence preceding the mature *N*-terminus of G6S contains features consistent with a signal peptide, i.e. a basic *N*-terminus, a hydrophobic core and a more polar *C*-terminal region. Analysis of the amino acid sequence using the weighted matrix method of von Heijne [27] predicted that the most likely signal peptidase cleavage site was *N*-terminal to Val³⁷ (von Heijne score 9.8), although other high-scoring amino acids were found in this region (Val⁴⁰; score 9.6; Gly⁴³; score 4.2; and Thr⁴⁴; score 3.5). G6S appears to be atypical in that the position of signal peptidase cleavage is normally well defined, with no other high-scoring sites in the region. When the *N*-termini of the 78 and 32 kDa polypeptides were sequenced, there was no preview sequences, i.e. ragged peptide ends. There are, therefore, two possibilities for signal peptidase cleavage in G6S, i.e., either Val³⁷ is the site of cleavage and the remaining seven amino acids are specifically removed during protein maturation to reveal the Thr⁴⁴ at the mature *N*-terminus, or the Thr⁴⁴ is always cleaved by the signal peptidase to produce the mature *N*-terminus.

The molecular mass of the deduced polypeptide for the 78 kDa G6S component, i.e. mature G6S, was calculated as 57.7 kDa, suggesting that ten of the possible thirteen *N*-glycosylation sites may be occupied with a complex oligosaccharide of approx. 2 kDa, to give a total molecular size close to the observed size (78 kDa). The predicted molecular mass of the 48 kDa and 32 kDa components of form B are 37.8 kDa and 19.9 kDa respectively, suggesting that five of the eight *N*-glycosylation sites in the 48 kDa species and all five *N*-glycosylation sites in the 32 kDa species may be used. Thus it would appear that G6S is more heavily substituted with complex carbohydrate residues than the other sulphatases, which generally have two to four sites of *N*-glycosylation [12–18].

Amino-acid-sequence alignment for five sulphatases is shown in Fig. 3. Incorporation of the G6S sequence into the arylsulphatase group alignment identified regions of reduced sequence similarity and regions which were highly conserved. In evolutionary terms this suggests that G6S has diverged from the relatively more highly conserved arylsulphatase group. For instance, according to this alignment there are 90 identical residues between the mature protein sequences of the human arylsulphatases (this becomes 73 residues when sea-urchin

```

                                M R L L P L A P G R L 11
GGAATTCCGGTCGGCCTCTCGCCCTTCAGCTACCTGTGCGTCCCTCCGTCGCCGTCGCCGGGTACCCCGAGCCTGTCCGCTATGCGGCTCCTGCCTTAGCCCCAGGTCGGCTC 120

R R G S P R H L P S C S P A L L L L L V L G G C L G V F G V A A G T R R P N V V L 51
CGGGGGGGCAGCCCCGCCACCTGCCCTCTGCAGCCAGCGTGTACTGTGTGTGGCGGCTGCCTGGGGGTCTTCGGGGTGGCTGCGGGAAACCGGAGGCCAACGTGGTGTCT 240

L L T D D Q D E V L G G M T P L K K T K A L I G E M G M T F S S A Y V P S A L C 91
CTCTCAGCGACGACGAGGACGAAGTGTGCGGGCATGACACCATAAAGAAAACCAAGCTCTCATCGGAGAGATGGGGATGACTTTTCCAGTGCTTATGTGCCAAGTGTCTCTGC 360

C P S R A S I L T G K Y P H N H V V N N T L E G N C S S K S W Q K I Q E P N T 131
TGCCCCAGCAGAGCCAGTATCTGACAGGAAAGTACCCACATAATCATCAGTTGTGAACAACACTCTGGAGGGGAACGACAGTAGTAAGTCTGGCAGAAGATCCAAGAACCAATACT 480

F P A I L R 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 E H V P L G W S Y 171
TTCCAGCAATCTCAGATCAATGTGTGGTATCAGACCTTTTTCAGGGAAATATTAAATGAGTACGGAGCCAGATGCAGGTGGAGTAAACACGTTCTCTGGGTGGAGTTAC 600

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 N Y S V D Y L T D V L A N V 211
TGGTATGCCTTGGAAAAGAAATCTAAGTATTATAATTACACCTGTCTATCAATGGGAAGGCACGGAAGCATGGTGAAGTATAGTGTGGACTACCTGACAGATGTTTGGCTAATGTC 720

S L D F L D Y K S N F E P F F M M I A T P A P H S P W T A A P Q Y Q K A F Q N V 251
TCCTTGACTTTCTGACTACAAGTCAACTTTGAGCCCTTCTTCATGATGATCGCCACTCCAGCGCCTCATTGCGCTTGACAGCTGCACCTCAGTACCAGAAGGCTTCCAGAATGTC 840

F A P R N K 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 A F R K 291
TTTGACCAAGAAACAAGAACTTCAACATCCATGGAACGAACAAGCACTGGTTAATTAGGCAAGCAAGACTCCAATGACTAATCTTCAATACAGTTTTATAGTAATGCATTAGGAAA 960

R W Q T L L S V D D L V E K L V K R L E F T G E L N N T Y I F Y T S D N G Y H T 331
AGGTGCAAACTCTCTCTCAGTTGATGACCTTGTGGAGAACTGGTCAAGAGGGCTGGAGTCTACTGGGAGCTCAACAACACTTACATCTTCTATACCTCAGACAAATGGCTATCACACA 1080

G Q F S L P I D K R Q L Y E F D I K V P L L V R G P G I K P N Q T S K M L V A N 371
GGACAGTTCCTTGCATAGACAAGAGACAGCTGTATGAGTTGATATCAAAGTCCACTGTTGGTTCGAGGACCTGGGATCAACCAAAATCAGACAAGCAAGATGCTGGTGGCAAC 1200

I D L G P T I L D I A G Y D L N K T Q M D G M S L L P I L R G A S N L T W R S D 411
ATTGACTTGGGTCTACTATTTTGGACATTGCTGGCTACGACCTAAATAAGACACAGATGGATGGGATGTCTTATTGCCCATTTTGAAGGTGCCAGTAACCTGACCTGGCGATCAGAT 1320

V L V E Y Q G E G R N V T D P T C P S L S P G V S Q C F P D C V C E D A Y N N T 451
GTCTGGTGAATACCAAGGAGAAGGCCGTAACGTCACTGACCCAACATGCCCTTCCCTGAGTCTGGCGTATCTCAATGCTTCCAGACTGTGTATGTGAAGATGCTTATAACAATACC 1440

Y A C V R T M S A L W N L Q Y C E F D D Q E V F V E V Y N L T A D P D Q I T N I 491
TATGCTGTGTGAGGACAAATGTGACGATTGTGGAATTTGAGTATTGCGAGTTGATGACCAAGAGGTGTTGTAGAACTCTAATCTGACTGCAGACCCAGACAGATCACTAACATT 1560

A K T I D P E L L G K M N Y R L M M L Q S C S G P T C R T P G V F D P G Y R F D 531
GCTAAAACATAGACCCAGAGCTTTTAGGAAAGATGAAGTATCGGTTAATGATGTTACAGTCTGTTCTGGGCCAACCTGTGCACTCCAGGGGTTTGTAGCCCGGATACAGGTTTGAC 1680

P R L M F S N R G S V R T R R F S K H L L 552
CCCCGTCTCATGTTTCAGCAATCGCGGAGTGTGAGGACTCGAAGATTTTCAAAACATCTTCTGTAGCGACCTCACACAGCCTCTGCAGATGGATCCCTGCACGCTCTTTCTGATGAAGT 1800

GATTGTAGTAGGTGTCTGTAGTAGTCTTCAAGACCACACCTGGAAGATTTCTGGGCTGGCTTTAAGTCTGTTTGAAGCAACCCAGTCAAGTCACTTCTCGTGCAATGTGTAA 1920

ACTGTGAACCTCTGCCATGTGTGAGGAGTGGCTGTCTGGTCTCTTCTTTAGCTGACAAGGACACTCTGAGGCTTTTGTCTCACTGTATTTTTTATCCTGGGGCCACAGTCTCT 2040

GATTATTCTCTTGTGGTTAAAGACTGAATTTGTAACCCATTGAGATAAATGGCAGTACTTTAGGACACACACAACACACAGATACACCTTTTATGATGTAAGCTTGACCTAAAGTCA 2160

AAGGACCTGTGTAGCATTTGAGATTGAGCACTTCACTATCAAAATACTAACATCACATGGCTTGAAGAGTAACCATCAGAGCTGAATCATCAAGTAAGAACAAGTACCATTGTTGATT 2280

GATAAGTAGAGATACATTTTTATGATGTTTCATCACAGTGTGGTAAGTTGCAAAATCAAAACATGTACCCAAGCTGTTTCATGTTTTGTGAATTC 2379

```

Fig. 2. Compiled nucleotide sequence of G6S cDNA clones and the deduced amino acid sequence of the encoded protein

The amino acid sequence is shown in the single-letter code above the nucleotide sequence. Regions corresponding to the amino acid sequence of the N-terminus of the 78, 48 and 32 kDa G6S polypeptides shown in Table 1 are underlined. Potential N-glycosylation sites are marked by asterisks. Nucleotide and amino acid residue numbers are shown in the right margin.

sulphatase is included), with 158 residues having three out of four identical residues between the four arylsulphatases. In contrast there are only 35 identical residues between G6S and the arylsulphatases. The conservation of identity at some of these positions will no doubt have some functional significance in the way that sulphatases catalyse the hydrolysis of sulphate ester bonds. Of these 35 rigidly conserved residues, 13 are either glycine or proline residues, which commonly are important structural determinants in protein folding.

A histidine residue and two arginine residues have been reported to be essential for the catalytic activity of arylsulphatase A [28,29]. Peters *et al.* [15] have speculated that the histidine residue in the highly conserved GKWHLG region (Fig. 3; with asterisks and underlined) in the arylsulphatases, together with the conserved arginine residues in the CTPSR region (Fig. 3, with asterisks and underlined), and the nearby TGR region may contain the catalytic residues essential for sulphate ester hydrolysis. Alignment of the G6S sequence with the aryl-

sulphatases does not support this speculation. For instance, the GKWHLG, CTPSR and TGR arylsulphatase regions are changed to GKYLNE, CCPSR and TGK in G6S (Fig. 3). It is possible that the GKWHLG and CTPSR sequence blocks may enable the arylsulphatases to efficiently hydrolyse the non-specific arylsulphate substrates.

Substrate structure has an important influence on the specificity of action and the catalytic efficiency of G6S [25]. G6S is highly specific toward glucosamine 6-sulphate and glucose 6-sulphate residues such that other monosaccharide sulphates are not hydrolysed to any measurable extent [25]. The C-4 hydroxy group of the glucosamine 6-sulphate residue under attack is involved in the binding and/or catalytic mechanism [25]. Similarly, aglycone structure such as C-5 carboxy and C-2 sulphate ester groups on penultimate iduronic acid residues increase the catalytic efficiency of G6S by several-thousand-fold [25]. Substrate and aglycone substrate structure also affect the catalytic efficiency and specificity of other sulphatases (e.g. [18,28-31]).

```

6 .....mrllplapgrlrrgsprhlpscaPaillLVlggclgvfgvaagtRrPnVLLltDDqd.evLGmtplkktkal
A .....mgafPrallLal.....aaglavaRPFNlVlifaDDLGyDLGcYGHpsstTPn
B .....mgprgaasLprgpgprlllpvvlPllLlLl.....appgsgagasRPFhVfLlADDLgwnDvc.fhgserirTPh
C .....mPlrkmkLlPllLlLl.....weashaasRPFNlLvmADDLgIDpGcYGNktirTPh
U mmsapflfllglglttaqtqdpallDLrenPllLlLl.....qsnehraplrvkPnVLLvADDLgCsDLtsYGHptqeaagf

6 Igem...GatFssaYVpsalCPSRAAIlTCkYPhnhhv.....nntlegncsskswqkiqepnTfpailRsmcGyHfffaGKy
A lDqLAAGClrTfdYVpvsLCPSPRAAIlTCRlPVRmGm.....ypgVlv...PssrgGLPleEvTvAevLaar.GYHngMaGKW
B lDaLAAGClrldnyYtq.pLCPSPRAAIlTCRYqIRtG1.....qhqiiv...PcqpscvPldeklipqllKea.GYHnmMaGKW
C lDrLASGGvklTqlhaasplCPSRAAIlTCRYPVRsGmaswstrtgVfl.ftassgGLPtdEITfAkllKdq.GYHnllCKW
U lDkmaAAGClrTfngYVgdavCPSRAAIlTCRlPVRIGt..fgetrVfl...PwtktGLPksElTlAeamKea.GYHngMaGKW

6 lneygapdaggLeHvPl...lwsywyalekns..kyynytlslngkArk.....hgenysvdyldtdvLa...
A HLGvg...pegafLpPhqCFhrFlGipysh.dagpcqnlctfppAtp.....cdggCdglvpiPl....
B HLGmy...rkec.LPrrrPtdtyfGylLgs.eddysherctlidAl.....nvtrC.....ad...
C HLGmschsktdfchHhGfnyFyGisLtnlrdckpgegsvfttgfkrlvflplqivgtliltlaalnC.lgllhvpLgvyffs
U HLGInensstdgahLPEnhCFd.FvGhnLpf.tnswscddtghlkdfp.....dsqrC.....yl.....

6 .....nvsldfldyksenfepffmmiatpaphspwtAaPyqkafqnvFaprnknFnhgtN..khwli....rqakTPmt
A .....laNlsveAQpwlpgLearymafAhdmladaqrqdrPFFLYyAshHthYpQ.
B .....frdgeevAtgyknmysTnIFtkraialltnhnp.ekPILFLYlAlqsvHePlQ
C lflaalliltlflglfhyfrplncf.....mmrNyeiiQPMsydnLTqrlTveAaqFIqrntc.etPFFlLvsylHvHTal.
U .....yvNatlvsQPyqhgkLTqlFTddAlGFiednh..adPFFLYvafahMHTsL..

6 n....SSicPldnafRkrw.qtLlsvDdlPeklvkrLeftCeLNMyIfyTSNGYHTgqfs.....LpIdKtqly
A .....FSggsPaerSgrGpGDSLmElDAAGVtmtaigdlGLleellviFTaDNGoeImraS.....rGGcsgilRcGKetty
B vpeeylkyPldfkdRhhYagmvsLMDAAGVntaaLkssGLwNNvfiFstDNGeqTl.....aGcnnwPLRGKwslW
C .....FSSkDfagkSghGvYGDaveEMDwaVcQilnlLdelrLaNdllyFTSDGahveevSskgeihGGnglykGGKannW
U .....FSSDfPactSrRGryGdnLlEMhdaVqkivdkLeenniseNlIfiFSDGphreyce....eGGdasifRGGKshsW

6 EfdikvPl1VrgGikpnqts.kmLvanidLgPllidAGydLnkt.qmDGasLlPlrGaSnltwrsdvlveyqgegrnVtdp
A ECGvRePalafWpChIapGvt.hELassLDlPllAaLAGapLPnv.tlDGfDLrPpaAChrqePsavsLl.....Lpvlp..
B ECGvRvgvFvaslilkqkGvknrELihiswLFlvlkLarghtngtkplDGfDvwtisegSpsPriellhndpnfvdspcp
C ECGIRvPgllrWpViqGqkidEptsnmDfPllvAKLAGapLPedriDGfDLmPlLeGkSqrshdelflhyenayLnaVr..
U ECGhRiPyiVwPgtIspGis.nEivtsmDhlaAdLgGtCLPtdriyDGksikdvLlegSasPhasffy.yckdnLmaVr..

6 tcpslspgvsqcfpdcvedaynnntyacvrtsalwnlqyCef.....DdqvfevynLtaDpDqitniakt
A r....rgpgfcadwkvgslhpgsahsdttad..pach.....asSslTaHePP...L.LyDLsLdPGEEnylLlg.
B rnsmapakddsslpaysafntsvhaairhGnwklitgypgCgywfpppsnyvSeipssDPPTktLwLFDldidPPErhdLsr.
C w....hpqntsiwkaffftpnfnpgsnGcfath...vce.....cfgySvThHDDP...L.LFDiskDPErnpLtp.
U .....vgkyahtfrtqrsvsdeylecaGefple.dyfch.....dcegdvTeHDDP...L.LFDLhDPEEaypleac

6 idpellgkmnyrLmLQscsgptcrtg.gvfdpg...yrfDPrLmfs.....nrgsvrtrrfskhl1
A gvagatpeVlqaLkLQllkaqldaav..tPpsqvargdPaLqicchpgCtPrpacChcpdpha
B eyphi...VtkLsrLQfyhkhsvPvy...Ppaq.....DPr.....CdPkatgvwgpwm
C aseprfyelkvmqeaadrhtqtlPevpdqFswnnflw..kPwLqlc...CpstglscQcdrekqdkrlsr
U ghedvltVkstveehkaalvkgtPll.dsFhshivpcc.nPangicnyvhePgmpCyqdvataarhyrp

```

Fig. 3. Alignment of sulphatase amino acid sequences

Abbreviations: 6, G6S; A, arylsulphatase A [13]; B, arylsulphatase B [15]; C, steroid sulphatase [12]; U, sea-urchin arylsulphatase [14]. Upper-case letters indicate positions of identity in three of the five sulphatases. Positions where there are identical residues in the four arylsulphatase sequences (A, B, C and U) have asterisks underneath the sequences. Positions where residues are identical in all sulphatases are boxed.

The addition of sequences of other sulphatases to the already established family of sulphatases is likely to aid in the identification of the specific amino acid residues involved in their catalytic process. Studies involving either site-directed mutagenesis or the characterization of mutations causing sulphatase deficiencies, will lead to an improved understanding of the functional importance of specific sulphatase sequences.

This work was supported by Program Grants from the National Health and Medical Research Council of Australia and the Adelaide Children's Hospital Research Foundation.

REFERENCES

- Roy, A. B. (1971) *Enzymes* 3rd Ed. 5, 1-19
- Kresse, H., Paschke, E., von Figura, K., Gilberg, W. & Fuchs, W. (1980) *Proc. Natl. Acad. Sci. U.S.A.* 77, 6822-6826
- Hopwood, J. J. & Elliott, H. (1983) *Biochem. Int.* 6, 141-148
- Fuchs, W., Beck, M. & Kresse, H. (1985) *Eur. J. Biochem.* 151, 551-556
- Neufeld, E. F. & Muenzer, J. (1989) in *The Metabolic Basis of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D., eds.), pp. 1565-1587, McGraw-Hill, New York
- Hopwood, J. J. (1989) in *Heparin: Chemical and Biological Properties Clinical Applications* (Lane, D. & Lindahl, U., eds.), pp. 191-227, Edward Arnold, London
- Coppa, G. V., Giorgi, P. L., Felici, L., Gabrielli, O., Donti, E., Bernasconi, S., Kresse, H., Paschke, E. & Mastrapaolo, C. (1983) *Eur. J. Pediatr.* 140, 130-133
- Siciliano, L., Fiumara, A., Pavone, L., Freeman, C., Robertson, D., Morris, C. P., Hopwood, J. J., Di Natale, P., Musumeci, S. & Horwitz, A. L. (1991) *J. Med. Genet.* 28, 402-405
- Freeman, C., Clements, P. R. & Hopwood, J. J. (1987) *Biochem. J.* 246, 347-354
- Robertson, D. A., Freeman, C., Nelson, P. V., Morris, C. P. & Hopwood, J. J. (1988) *Biochem. Biophys. Res. Commun.* 157, 218-224
- Robertson, D. A., Callen, D. F., Baker, E. G., Morris, C. P. & Hopwood, J. J. (1988) *Hum. Genet.* 79, 175-178
- Yen, P. H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R. T. & Shapiro, L. J. (1987) *Cell* 49, 443-454
- Stein, C., Gieselmann, V., Kroysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H. E., O'Brien, J. S. & von Figura, K. (1989) *J. Biol. Chem.* 264, 1252-1259
- Sasaki, H., Yamada, K., Akasaka, K., Kawasaki, H., Suzuki, K., Saito, A., Sato, M. & Shimada, H. (1988) *Eur. J. Biochem.* 177, 9-13
- Peters, C., Schmidt, B., Rommerskirch, W., Rupp, K., Zuhlsdorff, M., Vingron, M., Meyer, H. F., Pohlmann, R. & von Figura, K. (1990) *J. Biol. Chem.* 265, 3374-3381

16. Schuchman, E. H., Jackson, C. E. & Desnick, R. J. (1990) *Genomics* **6**, 149–158
17. Wilson, P. J., Morris, C. P., Anson, D. S., Occhiodoro, T., Bielicki, J., Clements, P. R. & Hopwood, J. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8531–8535
18. Tomatsu, S., Fukuda, S., Masue, M., Sukegawa, K., Fukao, T., Yamagishi, A., Hori, T., Iwata, H., Ogawa, T., Nakashima, Y., Hanyu, Y., Hashimoto, T., Titani, K., Oyama, R., Suzuki, M., Yagi, K., Hayashi, Y. & Orii, T. (1991) *Biochem. Biophys. Res. Commun.* **181**, 677–683
19. Gibson, G. J., Saccone, G. T. P., Brooks, D. A., Clements, P. R. & Hopwood, J. J. (1987) *Biochem. J.* **248**, 755–764
20. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
21. Friedman, K. D., Rosen, N. L., Newman, P. J. & Montgomery, R. R. (1988) *Nucleic Acids Res.* **16**, 8718
22. Lipman, D. J., Altschul, S. F. & Kececiloglu, J. D. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4412–4415
23. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), vol. 5, Supp 3, pp. 345–352, National Biomedical Research Foundation, Washington, DC
24. Devereux, D. F., Haeberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387–395
25. Freeman, C. & Hopwood, J. J. (1987) *Biochem. J.* **246**, 355–365
26. Waheed, A. & van Etten, R. L. (1980) *Biochim. Biophys. Acta* **614**, 92–101
27. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
28. Lee, G. D. & van Etten, R. L. (1975) *Arch. Biochem. Biophys.* **171**, 424–434
29. James, G. T. (1979) *Arch. Biochem. Biophys.* **197**, 57–62
30. Freeman, C. & Hopwood, J. J. (1986) *Biochem. J.* **234**, 83–92
31. Hopwood, J. J., Elliott, H., Muller, V. & Saccone, G. T. P. (1986) *Biochem. J.* **234**, 507–514
32. Bielicki, J., Freeman, C., Clements, P. R. & Hopwood, J. J. (1990) *Biochem. J.* **271**, 75–86

Received 28 April 1992/9 June 1992; accepted 16 June 1992