View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Cell, Vol. 113, 435–444, May 16, 2003, Copyright ©2003 by Cell Press

Multiple Sulfatase Deficiency Is Caused by Mutations in the Gene Encoding the Human C_α-Formylglycine Generating Enzyme

Thomas Dierks,[1] Bernhard Schmidt,[1] Ljudmila V. Borissenko,[1] Jianhe Peng,[1] Andrea Preusser, Malaiyalam Mariappan, and Kurt von Figura*
Biochemie II
Universität Göttingen
Heinrich-Düker-Weg 12
37073 Göttingen
Germany

Summary

C_α-formylglycine (FGly) is the catalytic residue in the active site of eukaryotic sulfatases. It is posttranslationally generated from a cysteine in the endoplasmic reticulum. The genetic defect of FGly formation causes multiple sulfatase deficiency (MSD), a lysosomal storage disorder. We purified the FGly generating enzyme (FGE) and identified its gene and nine mutations in seven MSD patients. In patient fibroblasts, the activity of sulfatases is partially restored by transduction of FGE encoding cDNA, but not by cDNA carrying an MSD mutation. The gene encoding FGE is highly conserved among pro- and eukaryotes and has a paralog of unknown function in vertebrates. FGE is localized in the endoplasmic reticulum and is predicted to have a tripartite domain structure.

Introduction

Sulfatases in eukaryotes and prokaryotes contain a unique amino acid derivative, $C\alpha$-formylglycine (FGly), which is posttranslationally generated from cysteine (eukaryotes and prokaryotes) or serine (prokaryotes) (Schmidt et al., 1995; Selmer et al., 1996; Miech et al., 1998; Dierks et al., 1998a, 1997, 1998b; Szameit et al., 1999). FGly is essential for catalytic activity of sulfatases. It is present in the active site as an aldehyde hydrate (Boltes et al., 2001). During catalysis, one of its geminal hydroxyl groups performs a nucleophilic attack on the sulfur of the sulfate group of the substrate and becomes covalently sulfated in the enzyme intermediate. The other hydroxyl group is required for an elimination reaction by which the sulfated enzyme intermediate is desulfated and the FGly regenerated (Lukatela et al., 1998; von Bülow et al., 2001).

The key role of the FGly residue for the biological function of sulfatases is illustrated by multiple sulfatase deficiency (MSD), a rare autosomal recessive disorder in man. In MSD, the activity of all sulfatases is severely decreased due to the inability to generate the FGly residues (Schmidt et al., 1995). The phenotype of MSD combines the clinical features of disorders caused by the deficiency of single sulfatases, such as metachromatic leukodystrophy, mucopolysaccharidoses, X-linked ichthyosis, and chondrodysplasia punctata (Hopwood and Ballabio, 2001).

Mammalian cells synthesize sulfatases at ribosomes bound to the endoplasmic reticulum. During or shortly after protein translocation and while the sulfatase polypeptides are still largely unfolded, FGly residues are generated (Dierks et al., 1997, 1998a) by lumenal components of the endoplasmic reticulum (Fey et al., 2001). A linear sequence motif following the cysteine residue to be modified (CXPSRXXX[L/M]TG[R/K/L]) posttranslationally directs the FGly generation in sulfatases (Dierks et al., 1999). This FGly modification motif is located 50–80 residues downstream of the signal peptidase cleavage site. A 16-mer sequence of arylsulfatase A (ASA) encompassing this motif was shown to be sufficient to direct FGly formation when inserted into a heterologous polypeptide background (Dierks et al., 1999). Using ribosome-associated nascent ASA fragments as substrate and the soluble content of the endoplasmic reticulum as enzyme source FGly formation could be demonstrated in an in vitro system (Fey et al., 2001).

In the present study we have purified the FGly generating enzyme (FGE) from bovine testis to homogeneity, identified its molecular nature, and characterized mutations in patients with multiple sulfatase deficiency.

Results

A Rapid Peptide-Based Assay for FGE Activity

We had developed an assay for determining FGE activity in microsome extracts using in vitro synthesized [^{35}S]ASA fragments as substrate. The fragments were added to the assay mixture as ribosome-associated nascent chain complexes. The quantitation of the product included tryptic digestion, separation of the peptides by RP-HPLC, and identification and quantitation of the ^{35}S-labeled FGly containing tryptic peptide by a combination of chemical derivatization to hydrazones, RP-HPLC separation, and liquid scintillation counting (Fey et al., 2001). For monitoring the enzyme activity during purification, this cumbersome procedure needed to be modified. A new assay based on a peptidic substrate and matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry was developed.

A synthetic 16-mer peptide corresponding to ASA residues 65–80 and containing the sequence motif required for FGly formation inhibited the FGE activity in the in vitro assay (Fey et al., 2001). This suggested that peptides such as ASA65–80 may serve as substrates for FGE. We synthesized the 23-mer peptide P23, which corresponds to ASA residues 60–80, with an additional N-acetylated methionine and a C-amidated serine residue to protect its N- and C terminus, respectively. The cysteine and the FGly containing form of P23 could be identified and quantified by MALDI-TOF mass spectrometry. The presence of the FGly residue in position 11 of P23 was verified by MALDI-TOF post source decay mass spectrometry (see Peng et al., 2003). Incubation of P23 with extracts from microsomes of bovine pan-

*Correspondence: kfigura@gwdg.de
[1]These authors contributed equally to this work.

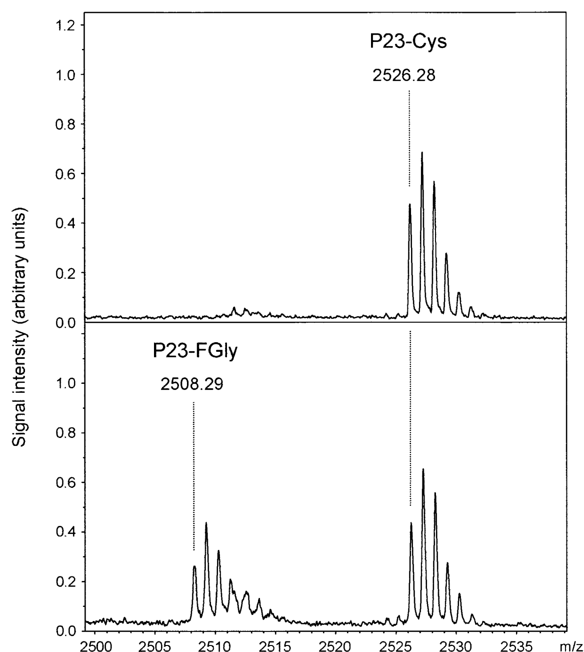

Figure 1. FGly Modification of P23

6 pmol of P23 were incubated under standard conditions for 10 min at 37°C in the absence (top) or presence (bottom) of 1 μl microsomal extract. The samples were prepared for MALDI-TOF mass spectrometry as described in Experimental Procedures. The monoisotopic masses MH^+ of P23 (2526.28) and its FGly derivative (2508.29) are indicated.

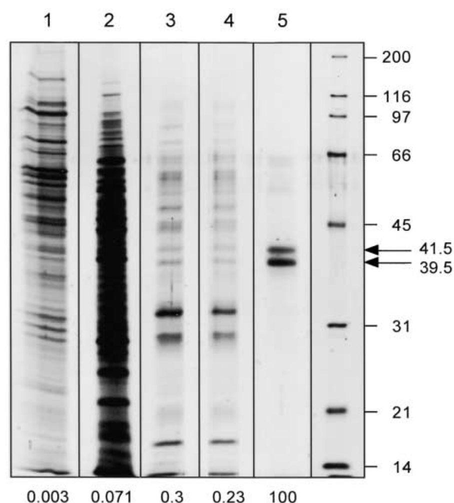

Figure 2. Purification of FGE from Bovine Testis

Aliquots of the soluble extract from microsomes (lane 1), of the pooled fractions after chromatography on MonoQ (lane 2), concanavalin A-Sepharose (lane 3), and scrambled peptide-Affigel 10 (lane 4) were separated by SDS-PAGE. The entire material eluted from Ser-69 peptide-Affigel 10 was concentrated and loaded in lane 5. Molecular weight standards are shown at the right. Proteins were stained with SYPRO Ruby. The numbers below lanes 1–5 refer to the percentage of the material loaded onto the gel.

creas or bovine testis converted up to 95% of the peptide into the FGly containing derivative (Figure 1). Under standard conditions, the FGly formation was proportional to the amount of enzyme and time of incubation as long as less than 50% of the substrate was consumed and the incubation period did not exceed 24 hr. The K_M for P23 was 13 nM. The effects of reduced and oxidized glutathione, Ca^{2+}, and pH were comparable to those seen in the assay using ribosome-associated nascent chain complexes as substrate (Fey et al., 2001).

Purification of FGE

For purification of FGE, the soluble fraction (reticuloplasm) of bovine testis microsomes served as the starting material. The specific activity of FGE was 10–20 times higher than that in reticuloplasm from bovine pancreas microsomes (Fey et al., 2001). Purification of FGE was achieved by a four-step chromatography. The first two steps were chromatography on a MonoQ anion exchanger and on concanavalin A-Sepharose. At pH 8, the FGE activity bound to MonoQ and was eluted at 50–165 mM NaCl with 60%–90% recovery. When this fraction was mixed with concanavalin A-Sepharose, FGE was bound. 30%–40% of the starting activity could be eluted with 0.5 M α-methyl mannoside. The two final purification steps were chromatography on affinity matrices derivatized with 16-mer peptides. The first affinity matrix was Affigel 10 substituted with a variant of the ASA65–80 peptide, in which the three residues critical for FGly formation, Cys-69, Pro-71, and Arg-73, were scrambled (scrambled peptide PVSLPTRSCAALLTGR). This peptide did not inhibit FGE activity when added at 10 μM concentration to the in vitro assay and, when immobilized to Affigel 10, did not retain FGE activity. Chromatography on the scrambled peptide affinity matrix removed peptide binding proteins including chaperones of the endoplasmic reticulum. The second affinity matrix was Affigel 10 substituted with a variant of the ASA65–80 peptide, in which the Cys-69 was replaced by a serine (Ser-69 peptide PVSLSTPSRAALLTGR). The Ser-69 peptide affinity matrix efficiently bound FGE. The FGE activity could be eluted with 20%–40% recovery with either 2 M KSCN or 25 μM Ser-69 peptide. Prior to activity determination, the KSCN or Ser-69 peptide had to be removed by dialysis. The substitution of Cys-69 by serine was crucial for the elution of active FGE. Affigel 10 substituted with the wild-type ASA65–80 peptide bound FGE efficiently. However, nearly no activity could be recovered in eluates with chaotropic salts (KSCN, $MgCl_2$), peptides (ASA65–80, or Ser-69 peptide), or buffers with low or high pH.

In Figure 2 the polypeptide pattern of the starting material and of the active fractions obtained after the four chromatographic steps of a typical purification is shown. In the final fraction, 5% of the starting FGE activity and 0.0006% of the starting protein were recovered (8333-fold purification).

The Purified 39.5 and 41.5 kDa Polypeptides Are Encoded by a Single Gene

The 39.5 and 41.5 kDa polypeptides in the final FGE preparation (see Figure 2, lane 5) were subjected to peptide mass fingerprint analysis. The mass spectra of the tryptic peptides of the two polypeptides obtained

```
                                        ▼
  1  MAAPALGLVC GRCPELGLVL LLLLLSLLCG AAGSQEAGTG AGAGSLAGSC   50
 51  GCGTPQRPGA HGSSAAAHRY SREANAPGPV PGERQLAHSK MVPIPAGVFT  100
                                                *
101  MGTDDPQIKQ DGEAPARRVT IDAFYMDAYE VSNTEFEKFV NSTGYLTEAE  150
151  KFGDSFVFEG MLSEQVKTNI QQAVAAAPWW LPVKGANWRH PEGPDSTILH  200
201  RPDHPVLHVS WNDAVAYCTW AGKRLPTEAE WEYSCRGGLH NRLFPWGNKL  250
251  QPKGQHYANI WQGEFPVTNT GEDGFQGTAP VDAFPPNGYG LYNIVGNAWE  300
301  WTSDWWTVHH SVEETLNPKG PPSGKDRVKK GGSYMCHRSY CYRYRCAARS  350
351  QNTPDSSASN LGFRCAADRL PTMD                              374
```

Figure 3. Amino Acid Sequence of Human FGE

The sequence is deduced from the cloned cDNA. The sequence deviates from that encoded by the human cDNA AK075459 at residue 124, F instead of L, and at residue 264, E instead of D. The two peptides identified in bovine FGE are underlined by a stippled bar. The predicted cleavage site for the signal peptidase at Gly-33 is indicated by an arrow head and the N-glycosylation site at Asn-141 by an asterisk. Subdomain one (light gray), subdomain two (gray), and subdomain three (dark gray) are underlined by filled bars.

by MALDI-TOF mass spectrometry were largely overlapping, suggesting that the two proteins originate from the same gene. Among the tryptic peptides of both polypeptides, two abundant peptides (MH^+ 1580.73, SQNTPDSSASNLGFR, and MH^+ 2049.91, MVPIPAGVF TMGTDDPQIK with the two methionine residues oxidized) were found, which matched to the protein encoded by the human cDNA AK075459. The amino acid sequence of the two peptides was confirmed by MALDI-TOF post source decay spectra and by MS/MS analysis using offline nano-electrospray ionisation (ESI) iontrap mass spectrometry. An EST sequence of the bovine ortholog covering the C-terminal part of FGE and matching the sequence of the MH^+ 1580.73 peptide provided additional sequence information for bovine FGE.

Evolutionary Conservation and Domain Structure of FGE

The gene encoding human FGE, designated as *SUMF1* (*su*lfatase *m*odifying *f*actor 1) is located on chromosome 3p26. It spans 105 kb and the coding sequence is distributed over 9 exons. Orthologous genes are found in mouse (87% identity), rat (94% identity), *Drosophila melanogaster* (48% identity), *Anopheles gambiae* (47% identity), and *Fugu rubripes* (63% identity). Orthologous EST sequences are found for a number of further species including cow, pig, chicken, *Xenopus laevis*, *Silurana tropicalis*, zebra fish, other fish species, and sea urchin (for details see J. Landgrebe, T.D., B.S., and K.v.F., unpublished data). The exon-intron structure between the human and the mouse gene is conserved, and the mouse gene on chromosome 6E2 is located within a region syntenic to the human chromosome 3p26. The genomes of *S. cerevisiae* and *C. elegans* lack *SUMF1* homologs. In prokaryotes, 13 orthologs of the *SUMF1* gene were found.

The cDNA for human FGE is predicted to encode a protein of 374 residues (Figure 3). The protein contains a cleavable signal sequence of 33 residues, which indicates translocation of FGE into the endoplasmic reticulum, and contains a single N-glycosylation site at Asn-141. The binding of FGE to concanavalin A suggests that this N-glycosylation site is utilized. Residues 87–367 of FGE are listed in the PFAM protein motif database as a domain of unknown function (PFAM: DUF323). Sequence comparison analysis of human FGE and its eukaryotic orthologs identified in databases indicates that this domain is composed of three distinct subdomains.

The N-terminal subdomain (residues 91–154 in human FGE) has a sequence identity of 39% and a similarity of 85% within the six known full-length eukaryotic FGE orthologs. In human FGE, this domain carries the N-glycosylation site at Asn-141, which is conserved in the other orthologs. The middle part of FGE (residues 179–308 in human FGE) is represented by a tryptophan-rich subdomain (12 tryptophans per 129 residues). The identity of the eukaryotic orthologs within this subdomain is 48%, the similarity 72%. The C-terminal subdomain (residues 327–366 in human FGE) is the most highly conserved sequence within the FGE family. The sequence identity of the human C-terminal subdomain with the eukaryotic orthologs is 85%, the similarity 100%. Within the 40 residues of the subdomain three, four cysteine residues are fully conserved. Three of these cysteines are also conserved in the prokaryotic FGE orthologs.

In vertebrates (man, mouse, rat, and fugu), a paralog of the *SUMF1* gene exists, which is designated as *SUMF2* gene. *SUMF2* originating EST sequences are documented for eight additional vertebrate species. Among the *SUMF2* encoded proteins, the identity is much higher than to the *SUMF1* encoded FGE of the same species. For example, the identity between the human and mouse *SUMF2* encoded protein is 86%, while the identity with the corresponding FGE is 47%–49%. The *SUMF2* encoded proteins share with FGE the subdomain structure. In subdomain three, however, they lack two of the three cysteines conserved among pro- and eukaryotic FGEs.

Expression and Subcellular Localization of FGE

A single transcript of 2.1 kb is detectable by Northern blot analysis of total RNA from skin fibroblasts (data not shown) and poly A^+ RNA from heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas (Figure 4A). Relative to β-actin RNA, the abundance varies by one order of magnitude and is highest in pancreas and kidney and lowest in brain. Various eukaryotic cell lines stably or transiently expressing the cDNA of human FGE or FGE derivatives C-terminally extended by an HA-, Myc-, or His_6-tag were assayed for FGE activity and subcellular localization of FGE. Transient expression of tagged and non-tagged FGE increased the FGE activity

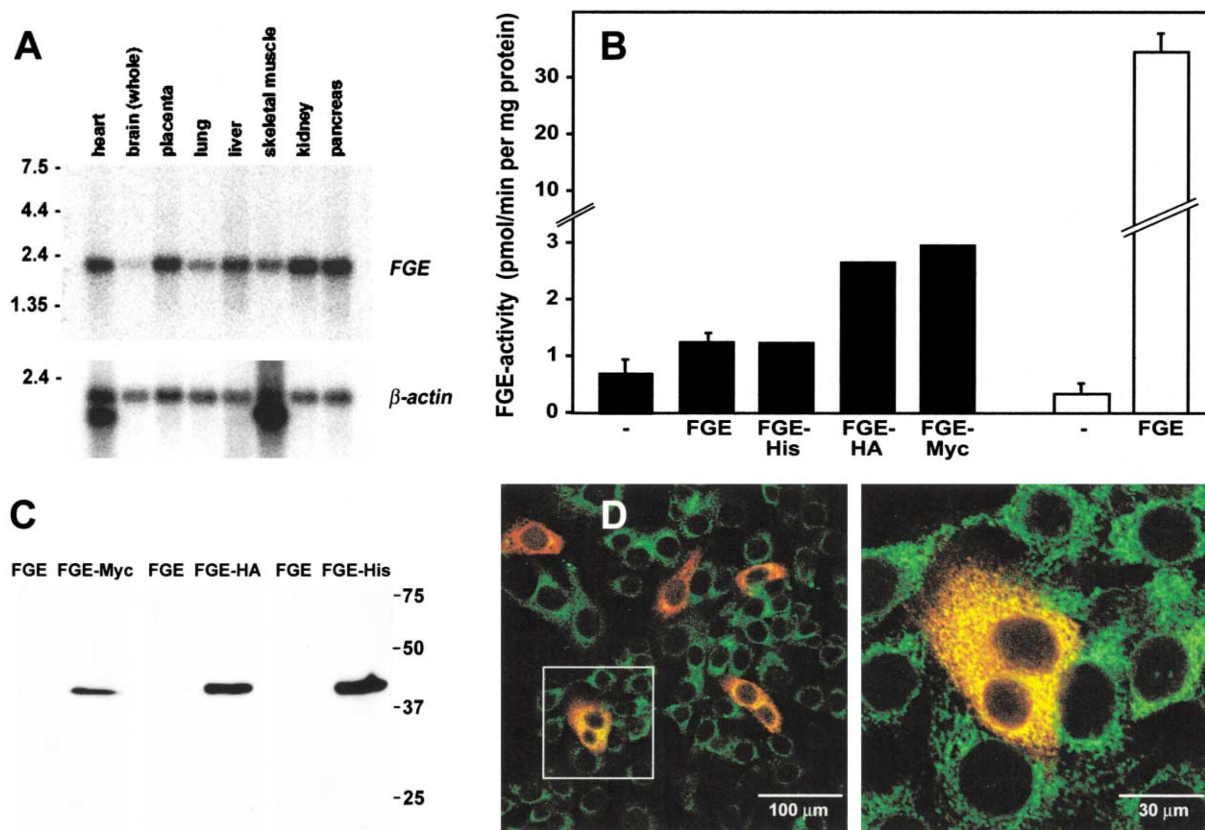

Figure 4. Expression, Molecular Forms, and Subcellular Localization of FGE and C-Terminally Tagged FGE

(A) Northern blot analysis of polyA^{+} RNA from human tissues for FGE (top) and β-actin (bottom).

(B) FGE activity of cells transiently or stably expressing FGE and C-terminally tagged FGE versions. BHK 21 cells (black bars) were transfected with pMPSV vectors containing the cDNA for human FGE or FGE C-terminally tagged with the His$_6$-, HA-, or Myc-tag. Cells were harvested 2 days after transfection. PT67 cells (open bars) were transduced with pLPCX or pLPCX-FGE and selected with puromycin. The range or variation of two to nine independent determinations is indicated.

(C) Western blot analysis of BHK 21 cells transiently transfected with cDNAs encoding FGE or FGE with a C-terminal Myc-, HA-, or His$_6$-tag. Cell extracts were separated by SDS-PAGE, transferred onto PVDF membrane, and probed with monoclonal antibodies against the Myc-, HA-, or His$_6$-epitope.

(D) Subcellular localization of FGE-HA in HT-1080 cells. After fixation and permeabilization, the HA-tag (red) and proteindisulfide isomerase (green) were detected with monoclonal antibodies. In cells expressing FGE-HA, the FGE-HA colocalizes with proteindisulfide isomerase (yellow).

1.6- to 3.9-fold (Figure 4B). Stable expression of FGE in PT67 cells increased the activity of FGE about 100-fold (Figure 4B).

Indirect immunofluorescence showed the colocalization of the variously tagged forms of FGE with proteindisulfide isomerase, a lumenal protein of the endoplasmic reticulum in BHK21, CHO and HT1080 cells (shown in Figure 4D for HA-tagged FGE in HT 1080 cells). Western blot analysis of extracts from BHK 21 cells transiently transfected with cDNA encoding tagged forms of FGE showed a single immunoreactive band with an apparent size between 42 to 44 kDa (Figure 4C).

The FGE Encoding *SUMF1* Gene Carries Mutations in MSD

MSD is caused by a deficiency to generate FGly residues in sulfatases (Schmidt et al., 1995). The FGE encoding *SUMF1* gene is therefore a candidate gene for MSD. We amplified and sequenced the FGE encoding cDNA of seven MSD patients and found nine different mutations that were confirmed by sequencing the genomic DNA (Table 1). The first patient was heterozygous for a 1076C>A substitution, converting the codon for serine 359 into a stop codon (S359X) and a mutation causing the deletion of the 25 residues 149–173 that are encoded by exon 3 and space the first and the second domain of the protein. Genomic sequencing revealed a deletion of nucleotides +5–8 of the third intron (IVS3+5-8 del), thereby destroying the splice donor site of intron 3. The second patient was heterozygous for the mutation causing the loss of exon 3 (IVS3+5-8 del) and a 979C>T substitution, converting the codon for arginine 327 into a stop codon (R327X). The truncated FGE encoded by the 979C>T allele lacks most of subdomain three. The third patient was homozygous for a 1045C>T substitution, replacing the conserved arginine 349 in subdomain three by tryptophan (R349W). The fourth patient was heterozygous for two missense mutations replacing conserved residues in subdomain three: a 1046G>A substitution replacing arginine 349 by glutamine

Table 1. Mutations in MSD Patients

Mutation	Effect on Protein	Remarks	Patient
1076C>A	S359X	truncation of the C-terminal 16 residues	one
IVS3+5-8 del	deletion of residues 149–173	in-frame deletion of exon 3	one, two
979C>T	R327X	loss of subdomain three	two
1045C>T	R349W	substitution of a conserved residue in subdomain three	three, seven
1046G>A	R349Q	substitution of a conserved residue in subdomain three	four
1006T>C	C336R	substitution of a conserved residue in subdomain three	four
836C>T	A279V	substitution of a conserved residue in subdomain two	five
243delC	frameshift and truncation	loss of all three subdomains	five
661delG	frameshift and truncation	loss of the C-terminal third of FGE, including subdomain three	six

Patient one is the MSD patient Mo. in Schmidt et al. (1995) and Rommerskirch and von Figura (1992). Patient six is the MSD patient reported by Burch et al. (1986). The other patients represent unpublished cases.

(R349Q) and a 1006T>C substitution replacing cysteine 336 by arginine (C336R). The fifth patient was heterozygous for an 836 C>T substitution, replacing the conserved alanine 279 by valine (A279V). The second mutation was a single nucleotide deletion (243delC), changing the sequence after proline 81 and causing a translation stop after residue 139. The sixth patient was heterozygous for the deletion of a single nucleotide (661delG), changing the amino acid sequence after residue 220 and introducing a stop codon after residue 266. The second mutation remains to be identified. The seventh patient was homozygous for the same 1045C>T substitution found in the third patient. In addition, we detected two polymorphisms in the coding region of 22 *SUMF1* alleles from controls and MSD patients. 36% carried a 188G>A substitution, replacing serine 63 by asparagine (S63N) and 32% a silent 1116C>T substitution.

Transduction of MSD Fibroblasts with cDNAs Encoding Wild-Type and Mutant FGE

In order to confirm the deficiency of FGE as the cause of the inactivity of sulfatases synthesized in MSD, we expressed the FGE encoding cDNA in MSD fibroblasts utilizing retroviral gene transfer. As a control, we transduced the retroviral vector without cDNA insert. To monitor the complementation of the metabolic defect the activity of ASA, steroid sulfatase (STS) and N-acetylgalactosamine 6-sulfatase (GalNAc6S) were measured in the transduced fibroblasts prior or after selection. Transduction of the cDNA encoding wild-type FGE partially restored the catalytic activity of the three sulfatases in two MSD-cell lines (Table 2) and for STS in a third MSD cell line (data not shown). Selection of transduced fibroblasts increased the activity of ASA and STS 2- to 5-fold, which agrees with the earlier observation that 15%–50% of human diploid fibroblasts are transduced by this procedure (Lübke et al., 2001). It may be noted that the activity of STS reached normal level after selection, while in the same cultures that of ASA and GalNAc6S reached only 20%–50% of control. The sulfatase activities in the MSD fibroblasts transduced with the retroviral vector alone (Table 2) were comparable to those in non-transduced MSD fibroblasts (data not shown). Transduction of FGE encoding cDNA derived from a MSD patient carrying the IVS3+5-8del mutation failed to restore the sulfatase activities (Table 2).

Discussion

FGE Is a Highly Conserved Glycoprotein of the Endoplasmic Reticulum

Purification of FGE from bovine testis yielded two polypeptides of 39.5 and 41.5 kDa that originate from the same gene. We cannot formally exclude that FGE exists in testis in two forms, e.g., due to a limited proteolysis, distinct N-glycosylation, tight binding of a cofactor, or the existence of alternate transcripts. The expression of three differently tagged versions of FGE, each in three different eukaryotic cell lines, as a single form, however, renders it likely that the smaller of the two forms observed in the FGE preparation purified from bovine testis has been generated by limited proteolysis during purification.

The substitution of Cys-69 in the ASA65–80 peptide by serine was critical for the purification of FGE by affinity chromatography. Preliminary experiments had shown that both the Ser-69 peptide and the Cys-69 containing ASA65–80 peptide inhibit FGly formation in an in vitro assay with comparable efficiency. Yet active FGE was only recovered from the affinity matrix substituted with the Ser-69 peptide. The latter cannot be modified by FGE. For yet unknown reasons, active FGE was not recovered after incubation with an affinity matrix substituted with ASA65-80 peptide, which is a substrate of FGE.

FGE has a cleavable signal sequence that mediates translocation across the membrane of the endoplasmic reticulum. The greater part of the mature protein (275 residues out of 340) defines a unique domain, which is likely to be composed of three subdomains (for details see J. Landgrebe, T.D., B.S. and K.v.F., unpublished data). None of the functions to be fulfilled by FGE, such as the recognition of the FGly modification motif in newly synthesized sulfatase polypeptides (Dierks et al., 1999) or the catalysis of the cysteine to FGly conversion can be assigned to one of the subdomains as for none of them homologs exist in proteins with known function.

The catalytic domain in FGE could catalyze the FGly formation in several ways. It has been proposed that FGE abstracts electrons from the thiol group of the cysteine and transfers them to an acceptor. The resulting thioaldehyde would spontaneously hydrolyze to FGly and H_2S (Schmidt et al., 1995). Alternatively, FGE could act as a mixed-function oxygenase (monooxygenase) introducing one atom of O_2 into the cysteine and the

Table 2. Complementation of MSD fibroblasts by Transduction of cDNA Encoding Wild-Type or Mutant FGE

Fibroblasts	FGE Encoding cDNA	Sulfatase Activity ASA[a]	STS[a]	GalNAc6S[a]
MSD 3[b]	–	1.9 ± 0.2	<3	56.7 ± 32
	FGE[c]	7.9	13.5	n.d.
	FGE[d]	12.2 ± 0.2	75.2	283 ± 42
	FGE-IVS3+5-8del[c]	1.8	<3	n.d.
	FGE-IVS3+5-8del[d]	2.1	<3	98.5
MSD 4[b]	–	1.1 ± 0.3	<3	n.d.
	FGE[c]	4.7	17.0	n.d.
Control fibroblasts		58 ± 11	66 ± 31	828 ± 426

[a] The values give the ratio between the sulfatases, ASA (mU/mg cell protein), STS (μU/mg cell protein), GalNAc6S (μU/mg cell protein) and that of the lysosomal reference enzyme ß-hexosaminidase (U/mg cell protein). For control fibroblasts the mean and the variation of 6-11 cell lines is given. Where indicated the range of two cultures transduced in parallel is given for MSD fibroblasts.
[b] The MSD fibroblast number refers to that of the patients in Table 1.
[c] Activity determination prior to selection.
[d] Activity determination after selection.
Abbreviations: n.d., not determined.

other into H_2O with the help of an electron donor such as $FADH_2$. The resulting thioaldehyde hydrate derivative would spontaneously react to FGly and H_2S. Preliminary experiments with a partially purified FGE preparation show a critical dependence of the FGly formation on molecular oxygen. This would suggest that FGE acts as a mixed-function oxygenase. The particular high conservation of subdomain three and the presence of three in all orthologs conserved cysteine residues therein make this subdomain a likely candidate for the catalytic site. It will be of interest to see whether the structural elements mediating the recognition of the FGly motif and the binding of an electron acceptor or electron donor correlate with the domain structure of FGE.

Recombinant FGE is localized in the endoplasmic reticulum. This agrees with the proposed site of its action. FGly residues are generated in newly synthesized sulfatases during or shortly after their translocation into the endoplasmic reticulum (Dierks et al., 1997, 1998a). FGE itself does not contain an ER-retention signal of the KDEL type. Its retention in the endoplasmic reticulum may therefore be mediated by the interaction with other ER proteins. Components of the translocation/N-glycosylation machinery are attractive candidates for such interacting partners.

FGE as a Tool to Improve the Production of Recombinant Sulfatases

Overexpression of sulfatases can be associated with a decrease of the activity of endogenous sulfatases (Anson et al., 1993). Furthermore, only a fraction of the recombinant sulfatases is active. This has been ascribed to the saturation of the mechanism posttranslationally generating the FGly residues in the active site of sulfatases. In accordance with that, a fraction of the recombinant sulfatases retains the unmodified cysteine (Schmidt et al., 1995). Coexpression of FGE and sulfatases may therefore be a means to increase the yield of catalytically active sulfatases. In fact, coexpression of FGE and galactose 6-sulfatase in eukaryotic cells was found to increase the specific activity of galactose 6-sulfatase about 50-fold (M. Heartlein, personal communication). Catalytically active recombinant sulfatases are needed for enzyme replacement therapy of patients suffering from the deficiency of a lysosomal sulfatase, e.g., of arylsulfatase B in Maroteaux-Lamy disease, iduronate sulfatase in Hunter disease, and galactose 6-/N-acetylgalactosamine 6-sulfatase in Morquio A disease (for review see Neufeld and Muenzer, 2001).

Mutations in *SUMF1* Cause MSD

It was the biochemical phenotype of MSD that provided the first hint to a unique posttranslational modification in sulfatases required for their activity. This modification was shown to be the conversion of cysteine to FGly and to be missing in sulfatases from MSD patients (Schmidt et al., 1995). The gene encoding FGE was therefore a strong, but not obligatory, candidate for the site of mutations in MSD. FGE may interact with other components and defects in genes encoding the latter could equally well cause MSD. In seven MSD patients, we indeed found nine different mutations in *SUMF1*, the gene encoding FGE. All mutations have severe effects on the FGE protein by either replacing highly conserved residues in subdomain three (three mutations) or subdomain two (one mutation), or by causing C-terminal truncations of various lengths (four mutations) or large in-frame deletions (one mutation). For three MSD-cell lines, it was shown that transduction of the FGE encoding cDNA partially or fully restores the sulfatase activities. On the contrary, transduction with a FGE encoding cDNA carrying one of the mutations observed in MSD patients did not restore the sulfatase activities. This clearly identifies the *SUMF1* gene as the site of the disease-causing mutations. Using an alternative genetic approach based on the restoration of sulfatase activities by microcell-mediated chromosome transfer, M.P. Cosma et al. (2003 [this issue of *Cell*]) also identified the FGE encoding *SUMF1* gene as the gene carrying the mutations causing MSD.

MSD is both clinically and biochemically heterogenous. A rare neonatal form presenting at birth and developing a hydrocephalus (Vamos et al., 1981; Burch et al., 1986), a common form resembling initially an infantile metachromatic leukodystrophy and subsequently developing ichthyosis- and mucopolysaccharidosis-like

features (Austin, 1973; Rampini et al., 1970; Couchot et al., 1974), and a less-frequent mild form in which the clinical features of a mucopolysaccharidosis prevail (Burk et al., 1984) have been differentiated. Biochemically it is characteristic that a residual activity of sulfatases is detected, which in cultured skin fibroblasts generally is below 10% of controls (Burch et al., 1986; Basner et al., 1979). However, in some MSD cell lines the activity of selected sulfatases can reach the normal range (Yutaka et al., 1981). Furthermore, the residual activity is subject to variations, depending on the cell culture conditions and unknown factors (Fluharty et al., 1978, 1979; Kresse and Holtfrerich, 1980; Chang et al., 1983). Biochemically, MSD has been classified into two groups (Chang et al., 1983; Steckel et al., 1985; Conary et al., 1988). In group I, the residual activity of sulfatases is below 15%, including that of ASB. In group II, the residual activity of sulfatases is higher and particularly that of ASB may reach values of up to 50%–100% of control. All patients reported here fall into group I except patient 5, who falls into group II (ASB activity in the control range) of the biochemical phenotype. Based on clinical criteria, patients 1 and 6 are neonatal cases, while patients 2–5 and 7 have the common form of MSD. The 1045C>T substitution was the only homozygous mutation in MSD patients. Both patient 3 and 7 are of Turkish origin, suggesting that the 1045C>T substitution might be more frequent in Turkey.

The phenotypic heterogeneity suggests that the different mutations in MSD patients are associated with different residual activities of FGE. Preliminary data on PT67 cells stably expressing the FGE mutant with the in-frame deletion of exon 3 indicate that this mutation abolishes FGE activity completely. The characterization of the mutations in MSD, of the biochemical properties of the mutant FGE, and of the residual content of FGly in sulfatases using a recently developed highly sensitive mass spectrometric method (Peng et al., 2003) will provide a better understanding of the genotype-phenotype correlation in MSD.

Is There a Second FGE Encoding Gene in Man and Mouse?

The *SUMF2* genes in vertebrates are predicted to encode proteins with a substantial overall homology to the *SUMF1* encoded FGE. The homology extends over all three subdomains. This raises the possibility that the *SUMF2* encoded proteins have FGly-generating activity. As discussed above, many studies have revealed minor residual sulfatase activities in MSD fibroblasts (Burch et al., 1986; Basner et al., 1979). Even in the most severe form of MSD with neonatal onset residual sulfatase activities were found and the level of residual activity was cell type dependent, being higher in leukocytes than in fibroblasts (Burch et al., 1986). The residual sulfatase activity in MSD may reflect a residual activity of the mutant forms of FGE. However, in patients with severe mutations, such as large deletions, the residual activity of sulfatases may well arise from a FGE activity of the *SUMF2* encoded protein. In fact, a minor FGE activity of the *SUMF2* encoded protein may be essential for the development till birth and the short postnatal life observed for the neonatal form of MSD. In line with a FGE supporting activity of the *SUMF2* encoded proteins are preliminary experiments in which the activity of arylsulfatase A increased when its cDNA was coexpressed with the *SUMF2* cDNA.

Experimental Procedures

In Vitro Assay for FGE

For monitoring the activity of FGE, the N-acetylated and C-amidated 23-mer peptide P23 (MTDFYVPVSLCTPSRAALLTGRS) was used as substrate. The conversion of the cysteine residue in position 11 to FGly was monitored by MALDI-TOF mass spectrometry. A 6 μM stock solution of P23 in 30% acetonitrile and 0.1% trifluoroacetic acid (TFA) was prepared. Under standard conditions, 6 pmol P23 were incubated at 37°C with up to 10 μl enzyme in a final volume of 30 μl 50 mM Tris/HCl (pH 9.0), containing 67 mM NaCl, 15 μM $CaCl_2$, 2 mM DTT, and 0.33 mg/ml bovine serum albumin. To stop the enzyme reaction, 1.5 μl 10% TFA was added. P23 then was bound to ZipTip C18 (Millipore), washed with 0.1% TFA, and eluted in 3 μl 50% acetonitrile, 0.1% TFA. 0.5 μl of the eluate was mixed with 0.5 μl of matrix solution (5 mg/ml α-cyano-4-hydroxy-cinnamic acid (Bruker Daltonics) in 50% acetonitrile, 0.1% TFA) on a stainless steel target. MALDI-TOF mass spectrometry was done with a Reflex III (Bruker Daltonics) using reflectron mode and laser energy just above the desorption/ionization threshold. All spectra were averages of 200–300 shots from several spots on the target. The mass axis was calibrated using peptides of molecular masses ranging from 1000–3000 Da as external standards. Monoisotopic MH^+ of P23 is 2526.28 and of the FGly containing product 2508.29. Activity (pmol product/hr) was calculated on the basis of the peak height of the product divided by the sum of the peak heights of P23 and the product.

Purification of FGE from Bovine Testis

Bovine testes were obtained from the local slaughterhouse and stored for up to 20 hr on ice. The parenchyme was freed from connective tissue and homogenized in a waring blendor and by three rounds of motor pottering. Preparation of rough microsomes (RM) by cell fractionation of the obtained homogenate was performed as described (Meyer et al., 2000) with the following modifications. Three differential centrifugation steps, 20 min each at 4°C, were performed at 500 × g (JA10 rotor), 3000 × g (JA10), and 10000 × g (JA20). From the last supernatant the RM membranes were sedimented (125,000 × g, Ti45 rotor, 45 min, 4°C), homogenized by motor pottering, and layered on a sucrose cushion (50 mM Hepes [pH 7.6], 50 mM KAc, 6 mM $MgAc_2$, 1 mM EDTA, 1.3 M sucrose, and 5 mM β-mercaptoethanol). RMs were recovered from the pellet after spinning for 210 min at 45,000 rpm in a Ti45 rotor at 4°C. Usually 100,000–150,000 equivalents RM, as defined by Walter and Blobel (1983), were obtained from 1 kg of testis tissue. The reticuloplasm, i.e., the luminal content of the RM, was obtained by differential extraction at low concentrations of deoxy Big Chap, as described (Fey et al., 2001).

For FGE purification, 95 ml of reticuloplasm were dialyzed for 20 hr at 4°C against 20 mM Tris/HCl (pH 8.0) and 2.5 mM DTT, and cleared by centrifugation at 125,000 × g for 1 hr. 32 ml aliquots of the cleared reticuloplasm were loaded on a MonoQ HR10/10 column (Amersham Biosciences) at room temperature, and washed and eluted at 2 ml/min with a linear gradient of 0 to 0.75 M NaCl in 80 ml of the Tris buffer. The fractions containing FGE activity, eluting at 50–165 mM NaCl, of three runs were pooled (42 ml) and mixed with 2 ml of concanavalin A-Sepharose (Amersham Biosciences) that had been washed with 50 mM Hepes buffer (pH 7.4), containing 0.5 M KCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM $CaCl_2$, and 2.5 mM DTT. After incubation for 16 hr at 4°C, the concanavalin A-Sepharose was collected in a column and washed with 6 ml of the same Hepes buffer. The bound material was eluted by incubating the column for 1 hr at room temperature with 6 ml 0.5 M α-methylmannoside in 50 mM Hepes (pH 7.4) and 2.5 mM DTT. The elution was repeated with 4 ml of the same eluent. The combined eluates (10 ml) from concanavalin A-Sepharose were adjusted to pH 8.0 with 0.5 M Tris/HCl (pH 9.0) and mixed with 2 ml of Affigel 10 (BioRad) that had been derivatized with 10 mg of the scrambled peptide

(PVSLPTRSCAALLTGR) and washed with buffer A (50 mM Hepes [pH 8.0], containing 0.15 M potassium acetate, 0.125 M sucrose, 1 mM $MgCl_2$, and 2.5 mM DTT). After incubation for 3 hr at 4°C the affinity matrix was collected in a column. The flow through and a wash fraction with 4 ml of buffer A were collected, combined, and mixed with 2 ml of Affigel 10 that had been substituted with 10 mg of the Ser-69 peptide (PVSLSTPSRAALLTGR) and washed with buffer A. After incubation overnight at 4°C, the affinity matrix was collected in a column, washed three times with 6 ml of buffer B (buffer A containing 2 M NaCl and a mixture of the 20 proteinogenic amino acids, each at 50 μg/ml). The bound material was eluted from the affinity matrix by incubating the Affigel twice for 90 min each with 6 ml buffer B containing 25 μM Ser-69 peptide. An aliqout of the eluate was substituted with 1 mg/ml bovine serum albumin, dialyzed against buffer A, and analyzed for activity. The remaining part of the activity (11.8 ml) was concentrated in a Vivaspin 500 concentrator (Vivascience) and solubilized at 95°C in Laemmli SDS sample buffer. The polypeptide composition of the starting material and preparations obtained after the chromatographic steps were monitored by SDS-PAGE (15% acrylamide and 0.16% bisacrylamide) and staining with SYPRO Ruby (BioRad).

Identification of FGE by Mass Spectrometry

For peptide mass fingerprint analysis, the purified polypeptides were in-gel digested with trypsin (Shevchenko et al., 1996), desalted on C18 ZipTip, and analyzed by MALDI-TOF mass spectrometry using dihydrobenzoic acid as matrix and two autolytic peptides from trypsin (m/z 842.51 and 2211.10) as internal standards. For tandem mass spectrometry analysis selected peptides were analyzed by MALDI-TOF post source decay mass spectrometry. Their corresponding doubly charged ions were isolated and fragmented by offline nano-ESI ion trap mass spectrometry (EsquireLC, Bruker Daltonics). The mass spectrometric data were used by Mascot search algorithm for protein identification in the NCBInr protein database and the NCBI EST nucleotide database.

Bioinformatics

Signal peptides and clevage sites were described with the method of von Heijne (1986) implemented in EMBOSS (Rice et al., 2000). N-glycosylation sites were predicted using the algorithm of Brunak (Gupta and Brunak, 2002). Functional domains were detected by searching PFAM-Hidden-Markov-Models (version 7.8) (Sonnhammer et al., 1998). To search for FGE homologs, the databases of the National Center for Biotechnology Information (Wheeler et al., 2002) were queried with BLAST (Altschul et al., 1997). Sequence similarities were computed using standard tools from EMBOSS. Genomic loci organization and synteny were determined using the NCBI's human and mouse genome resources (http://www.ncbi.nlm.nih.gov/genome/guide/) and the human-mouse homology map (http://www.ncbi.nlm.nih.gov/Homology/).

Cloning of Human FGE Encoding cDNA

Total RNA, prepared from human fibroblasts using the RNeasy Mini kit (Qiagen) was reverse transcribed using the Omniscript RT kit (Qiagen) and either an oligo(dT) primer or the *SUMF1*-specific primer 1199nc (CCAATGTAGGTCAGACACG). The first strand cDNA was amplified by PCR using the forward primer 1c (ACATGGCCCGCGG GAC) and, as reverse primer, either 1199nc or 1182nc (CGACTGCT CCTTGGACTGG). The PCR products were cloned directly into the pCR4-TOPO vector (Invitrogen). By sequencing multiples of the cloned PCR products, which had been obtained from various individuals and from independent RT and PCR reactions, the coding sequence of the *SUMF1* cDNA (1125 bp from start to stop codon) was established.

Mutation Detection, Genomic Sequencing, and Northern Blot Analysis

For standard protocols utilized in this study, see Lübke et al. (2001) and Hansske et al. (2002). Northern blots were hybridized with a cDNA probe covering the entire coding region and a β-actin cDNA probe as a control for RNA loading.

Cell Lines and Cell Culture

The fibroblasts from MSD patients 1–7 were obtained from E. Christenson (1, Rigshospitalet Copenhagen), M. Beck (2, Universitätskinderklinik Mainz), A. Kohlschütter (3, Universitätskrankenhaus Eppendorf, Hamburg), E. Zammarchi (4, Meyer Hospital, University of Florence), K. Harzer (5 and 7, Institut für Hirnforschung, Universität Tübingen), and A. Fensom (6, Guy's Hospital, London), respectively. Human skin fibroblasts, HT-1080, BHK21, and CHO cells were maintained at 37°C under 5% CO_2 in Dulbecco's modified Eagle's medium containing 10% fetal calf serum.

Transfection, Indirect Immunofluorescence, Western Blot Analysis, and Detection of FGE Activity

The FGE encoding cDNA was equipped with a 5′ EcoRI site and either a 3′ HA-, c-Myc, or RGS-His_6-tag sequence, followed by a stop-codon and a HindIII site, by add-on PCR using Pfu polymerase (Stratagene) and the following primers: GGAATTCGGGACAACAT GGCTGCG (EcoRI), CCCAAGCTTATGCGTAGTCAGGCACATCATA CGGATAGTCCATGGTGGGCAGGC (HA), CCCAAGCTTACAGGTCT TCTTCAGAAATCAGCTTTTGTTCGTCCATGGTGGGCAGGC (c-Myc), and CCCAAGCTTAGTGATGGTGATGGTGATGCGATCCTCTGTCCA TGGTGGGCAGGC (RGS-His_6). The resulting PCR products were cloned as EcoRI/HindIII fragments into pMPSVEH (Artelt et al., 1988). The plasmids obtained were transiently transfected into HT-1080, BHK21, and CHO cells and grown on cover slips, using Effectene (Qiagen) as transfection reagent. 48 hr after transfection, the cells were analyzed by indirect immunofluorescence as described previously (Lübke et al., 2001; Hansske et al., 2002.), using monoclonal IgG_1 antibodies against HA (BAbCO, Richmond), c-Myc (Santa Cruz), or RGS-His (Qiagen) as primary antibodies. The endoplasmic reticulum marker protein proteindisulfide isomerase (PDI) was detected with a monoclonal antibody of different subtype (IgG_{2A}, Stressgen). The primary antibodies were detected with isotype-specific goat secondary antibodies coupled to CY2 or CY3, respectively (Molecular Probes). Immunofluorescence images were obtained on a Leica TCS Sp2 AOBS laser scan microscope. For Western blot analysis, the same monoclonal antibodies and a HRP-conjugated anti-mouse IgG as secondary antibody were used.

For determination of FGE activity, the trypsinized cells were washed with phosphate-buffered saline containing a mixture of proteinase inhibitors (208 μM 4-(2-aminoethyl)benzene sulfonyl fluoride hydrochloride, 0.16 μM aprotinin, 4.2 μM leupeptin, 7.2 μM bestatin, 3 μM pepstatin A, 2.8 μM E-64); solubilized in 10 mM Tris (pH 8.0), containing 2.5 mM DTT, the proteinase inhibitors, and 1% Triton X-100; and cleared by centrifugation at 125,000 × g for 1 hr. The supernatant was subjected to chromatography on a MonoQ PC 1.6/5 column using a 1 ml gradient of 0 to 375 mM NaCl in the Tris buffer described above and a flow rate of 0.1 ml/min. Fractions eluting at 50–200 mM NaCl were pooled, lyophilized, and reconstituted in one-tenth of the original pool volume prior to determination of FGE activity with peptide P23.

Retroviral Transduction

cDNAs of interest were cloned into the Moloney murine leukemia virus based vector pLPCX and pLNCX2 (Clontech). The transfection of ecotropic ϕNX-Eco cells (ATCC) and the transduction of amphotropic RetroPack PT67 cells (Clontech) and human fibroblasts were done as described (Lübke et al., 2001; Thiel et al., 2002.). For some experiments, pLPCX-transduced PT67 cells were selected with puromycin prior determination of sulfatase activities.

Sulfatase Assays

Activity of ASA, STS, and GalNAc6S were determined as described (Rommerskirch and von Figura, 1992; Glössl and Kresse, 1978).

Acknowledgments

We are gratefully indebted to Martina Balleininger, Frank Kube, Nicole Leister, Klaus Neifer, Petra Schlotterhose, and Annegret Schneemann for their invaluable technical assistence. We thank Drs. S. Höning, J. Landgrebe, T. Lübke, and C. Thiel for their help in immunofluorescence, bioinformatics, and retroviral transduction. This work was supported by the Deutsche Forschungsge-

meinschaft, the Fonds der Chemischen Industrie, and Transkaryotic Therapies, Inc. (Cambridge, MA).

We dedicate this study to M. Karas and F. Hillenkamp to acknowledge the contribution of MALDI-TOF mass spectrometry to the identification of FGE.

Received: February 13, 2003
Revised: March 27, 2003
Accepted: March 31, 2003
Published: May 15, 2003

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Anson, D.S., Muller, V., Bielicki, J., Harper, G.S., and Hopwood, J.J. (1993). Overexpression of N-acetylgalactosamine-4-sulphatase induces a multiple sulphatase deficiency in mucopolysaccharidosis-type-VI fibroblasts. Biochem. J. *294*, 657–662.

Artelt, P., Morelle, C., Ausmeier, M., Fitzek, M., and Hauser, H. (1988). Vectors for efficient expression in mammalian fibroblastoid, myeloid and lymphoid cells via transfection or infection. Gene *68*, 213–219.

Austin, J.H. (1973). Studies in metachromatic leukodystrophy. XII. Multiple sulfatase deficiency. Arch. Neurol. *28*, 258–264.

Basner, R., von Figura, K., Glössl, J., Klein, U., Kresse, H., and Mlekusch, W. (1979). Multiple deficiency of mucopolysaccharide sulfatases in mucosulfatidosis. Pediatr. Res. *13*, 1316–1318.

Boltes, I., Czapinska, H., Kahnert, A., von Bülow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M.A., and Usón, I. (2001). 1.3 Å structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. Structure *9*, 483–491.

Burch, M., Fensom, A.H., Jackson, M., Pitts-Tucker, T., and Congdon, P.J. (1986). Multiple sulphatase deficiency presenting at birth. Clin. Genet. *30*, 409–415.

Burk, R.D., Valle, D., Thomas, G.H., Miller, C., Moser, A., Moser, H., and Rosenbaum, K.N. (1984). Early manifestations of multiple sulfatase deficiency. J. Pediatr. *104*, 574–578.

Chang, P.L., Rosa, N.E., Ballantyne, S.R., and Davidson, R.G. (1983). Biochemical variability of arylsulphatases -A, -B and -C in cultured fibroblasts from patients with multiple sulphatase deficiency. J. Inherit. Metab. Dis. *6*, 167–172.

Conary, J.T., Hasilik, A., and von Figura, K. (1988). Synthesis and stability of steroid sulfatase in fibroblasts from multiple sulfatase deficiency. Biol. Chem. Hoppe Seyler *369*, 297–302.

Cosma, M.P., Pepe, S., Annunziata, I., Newbold, R.F., Grompe, M., Parenti, G., and Ballabio, A. The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. Cell *113*, this issue, 445–456.

Couchot, J., Pluot, M., Schwauch, M.-A., Penaforte, F., and Fandre, M. (1974). La Mucosulfatidose. Etudes de trois cas familiaux. Arch. Fr. Pediatr. *31*, 775–795.

Dierks, T., Schmidt, B., and von Figura, K. (1997). Conversion of cysteine to formylglycine: a protein modification in the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA *94*, 11963–11968.

Dierks, T., Lecca, M.R., Schmidt, B., and von Figura, K. (1998a). Conversion of cysteine to formylglycine in eukaryotic sulfatases occurs by a common mechanism in the endoplasmic reticulum. FEBS Lett. *423*, 61–65.

Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M.A., and von Figura, K. (1998b). Posttranslational formation of formylglycine in prokaryotic sulfatases by modification of either cysteine or serine. J. Biol. Chem. *273*, 25560–25564.

Dierks, T., Lecca, M.R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999). Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. EMBO J. *18*, 2084–2091.

Fey, J., Balleininger, M., Borissenko, L.V., Schmidt, B., von Figura, K., and Dierks, T. (2001). Characterization of posttranslational formylglycine formation by luminal components of the endoplasmic reticulum. J. Biol. Chem. *276*, 47021–47028.

Fluharty, A.L., Stevens, R.L., Davis, L.L., Shapiro, L.J., and Kihara, H. (1978). Presence of arylsulfatase A (ARSA) in multiple sulfatase deficiency disorder fibroblasts. Am. J. Hum. Genet. *30*, 249–255.

Fluharty, A.L., Stevens, R.L., de la Flor, S.D., Shapiro, L.J., and Kihara, H. (1979). Arysulfatase A modulation with pH in multiple sulfatase deficiency disorder fibroblasts. Am. J. Hum. Genet. *31*, 574–580.

Glössl, J., and Kresse, H. (1978). A sensitive procedure for the diagnosis of N-acetyl-galactosamine-6-sulfate sulfatase deficiency in classical Morquio's disease. Clin. Chim. Acta *88*, 111–119.

Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput. 310–22.

Hansske, B., Thiel, C., Lübke, T., Hasilik, M., Höning, S., Peters, V., Heidemann, P.H., Hoffmann, G.F., Berger, E.G., von Figura, K., and Körner, C. (2002). Deficiency of UDP-galactose:N-acetylglucosamine beta-1,4-galactosyltransferase I causes the congenital disorder of glycosylation type IId. J. Clin. Invest. *109*, 725–733.

Hopwood, J.J., and Ballabio, A. (2001). Multiple sulfatase deficiency and the nature of the sulfatase family. In The Metabolic and Molecular Bases of Inherited Disease, C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Childs, K.W. Kinzler, and B. Vogelstein, eds. (New York: McGraw-Hill), pp. 3725–3732.

Kresse, H., and Holtfrerich, D. (1980). Thiosulfate-mediated increase of arylsulfatase activities in multiple sulfatase deficiency disorder fibroblasts. Biochem. Biophys. Res. Commun. *97*, 41–48.

Lübke, T., Marquart, T., Etzioni, A., Hartmann, E., von Figura, K., and Körner, C. (2001). Complementation cloning identifies CDG-IIc, a new type of congenital disorders of glycosylation, as a GDP-fucose transporter deficiency. Nat. Genet. *28*, 73–76.

Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, W. (1998). Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. Biochemistry *37*, 3654–3664.

Meyer, H.-A., Grau, H., Kraft, R., Kostka, S., Prehn, S., Kalies, K.U., and Hartmann, E. (2000). Mammalian Sec61 is associated with Sec62 and Sec63. J. Biol. Chem. *275*, 14550–14557.

Miech, C., Dierks, T., Selmer, T., von Figura, K., and Schmidt, B. (1998). Arylsulfatase from Klebsiella pneumoniae carries a formylglycine generated from a serine. J. Biol. Chem. *273*, 4835–4837.

Neufeld, E., and Muenzer, J. (2001). The mucopolysaccharidosis. In The Metabolic and Molecular Bases of Inherited Disease, C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Childs, K.W. Kinzler, and B. Vogelstein, eds. (New York: McGraw-Hill), pp. 3421–3452.

Peng, J., Schmidt, B., von Figura, K., and Dierks, T. (2003). Identification of formylglycine in sulfatases by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J. Mass Spectrom. *38*, 80–86.

Rampini, S., Isler, W., Baerlocher, K., Bischoff, A., Ulrich, J., and Pluss, H.J. (1970). Die Kombination von metachromatischer leukodystrophie und mukopolysaccharidose als selbstständiges krankheitsbild (mukosulfatidose). Helv. Paediatr. Acta *25*, 436–461.

Rice, P., Longden, I., and Bleasby, A. (2000). The European molecular biology open software suite. Trends Genet. *16*, 276–277.

Rommerskirch, W., and von Figura, K. (1992). Multiple sulfatase deficiency: catalytically inactive sulfatases are expressed from retrovirally introduced sulfatase cDNAs. Proc. Natl. Acad. Sci. USA *89*, 2561–2565.

Schmidt, B., Selmer, T., Ingendoh, A., and von Figura, K. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. Cell *82*, 271–278.

Selmer, T., Hallmann, A., Schmidt, B., Sumper, M., and von Figura, K. (1996). The evolutionary conservation of a novel protein modification: conversion of cysteine to serinesemialdehyde in arylsulfatase from Volvox carteri. Eur. J. Biochem. *238*, 341–345.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. *68*, 850–855.

Sonnhammer, E.L.L., Eddy, S.R., Birney, E., Bateman, A., and Durbin, R. (1998). Multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. *26*, 320–322.

Steckel, F., Hasilik, A., and von Figura, K. (1985). Synthesis and stability of arylsulfatase A and B in fibroblasts from multiple sulfatase deficiency. Eur. J. Biochem. *151*, 141–145.

Szameit, C., Miech, C., Balleininger, M., Schmidt, B., von Figura, K., and Dierks, T. (1999). The iron sulfur protein AtsB is required for posttranslational formation of formylglycine in the *Klebsiella* sulfatase. J. Biol. Chem. *274*, 15375–15381.

Thiel, C., Schwarz, M., Hasilik, M., Grieben, U., Hanefeld, F., Lehle, L., von Figura, K., and Körner, C. (2002). Deficiency of dolichyl-P-Man:Man7GlcNAc2-PP-dolichyl mannosyltransferase causes congenital disorder of glycosylation type Ig. Biochem. J. *376*, 195–201.

Vamos, E., Liebaers, I., Bousard, N., Libert, J., and Perlmutter, N. (1981). Multiple sulphatase deficiency with early onset. J. Inherit. Metab. Dis. *4*, 103–104.

Von Bülow, R., Schmidt, B., Dierks, T., von Figura, K., and Usón, I. (2001). Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis. J. Mol. Biol. *305*, 269–277.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. *14*, 4683–4690.

Walter, P., and Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. Methods Enzymol. *96*, 84–93.

Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Tatusova, T.A., Wagner, L., and Rapp, B. (2002). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *20*, 13–16.

Yutaka, T., Okada, S., Kato, T., Inui, K., and Yabuuchi, H. (1981). Properties of sulfatases in cultured skin fibroblasts of multiple sulfatase deficient patients. Clin. Genet. *20*, 296–303.

Accession Numbers

The GenBank accession number for the *SUMF1* sequence reported in this paper is AY 208752. The detected SNP polymorphisms were also deposited under this accession number.