Cell, Vol. 121, 541–552, May 20, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.03.001

Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme

Thomas Dierks,[1,3] Achim Dickmanns,[2] Andrea Preusser-Kunze,[1] Bernhard Schmidt,[1] Malaiyalam Mariappan,[1] Kurt von Figura,[1,*] Ralf Ficner,[2] and Markus Georg Rudolph[2,*]
[1]Department of Biochemistry II
University of Göttingen
D-37073 Göttingen
Germany
[2]Department of Molecular Structural Biology
University of Göttingen
D-37077 Göttingen
Germany

Summary

Sulfatases are enzymes essential for degradation and remodeling of sulfate esters. Formylglycine (FGly), the key catalytic residue in the active site, is unique to sulfatases. In higher eukaryotes, FGly is generated from a cysteine precursor by the FGly-generating enzyme (FGE). Inactivity of FGE results in multiple sulfatase deficiency (MSD), a fatal autosomal recessive syndrome. Based on the crystal structure, we report that FGE is a single-domain monomer with a surprising paucity of secondary structure and adopts a unique fold. The effect of all 18 missense mutations found in MSD patients is explained by the FGE structure, providing a molecular basis of MSD. The catalytic mechanism of FGly generation was elucidated by six high-resolution structures of FGE in different redox environments. The structures allow formulation of a novel oxygenase mechanism whereby FGE utilizes molecular oxygen to generate FGly via a cysteine sulfenic acid intermediate.

Introduction

Enzymes of the sulfatase family mediate hydrolysis of sulfate esters such as glycosaminoglycans, sulfolipids, and steroid sulfates in eukaryotic cells. Reduced activity of sulfatases is linked to severe lysosomal storage disorders such as mucopolysaccharidoses and metachromatic leukodystrophy and to nonlysosomal disorders such as X-linked ichthyosis and chondrodysplasia punctata (Hopwood and Ballabio, 2001). Causative mechanisms for these diseases include missorting of sulfatases and mutations in sulfatase-encoding genes that lead to inactive enzymes. However, a severe reduction or complete lack of all sulfatase activities, termed multiple sulfatase deficiency (MSD), originates from mutations in the gene coding for FGE, the formylglycine (FGly)-generating enzyme (Cosma et al., 2003; Dierks et al., 2003). Biochemical data (Dierks et al., 1997; Dierks et al., 1998a; Dierks et al., 1998b; Miech et al., 1998; Schmidt et al., 1995; Selmer et al., 1996; Szameit et al., 1999) and crystal structures of human (Bond et al., 1997; Hernandez-Guzman et al., 2003; Lukatela et al., 1998) and bacterial (Boltes et al., 2001) sulfatases revealed a unique FGly residue in the active site to be essential for sulfate ester hydrolysis. Thus, inactive FGE leads to lack of or incomplete generation of the FGly residue in all 13 human sulfatases described to date (Ferrante et al., 2002; Hopwood and Ballabio, 2001; Morimoto-Tomita et al., 2002), with its pathological manifestations.

FGE is localized in the endoplasmic reticulum (ER) and interacts with and modifies the unfolded form of newly synthesized sulfatases (Dierks et al., 1997; Dierks et al., 1998a). In the folded sulfatase, FGly is buried in a deep cleft and is inaccessible to FGE (Boltes et al., 2001; Lukatela et al., 1998). Also present in the ER is the FGE paralog pFGE, which shares 50.6% sequence identity (57.4% homology) with FGE over 283 residues. The function of pFGE is ill defined at present, as it possesses no FGly-generating activity. However, the pFGE gene expression levels are substantial and the pFGE/FGE ratios are approximately equal in all tissue types studied, pointing to a function of pFGE that is related to FGE activity (Mariappan et al., 2005).

The generation of FGly from a cysteine residue is a multistep redox process that involves disulfide bond formation and requires calcium, molecular oxygen, and a reducing agent (Fey et al., 2001; J. Peng, K.v.F., T.D., B.S., A.P.-K., and M.M., unpublished). This process has been reconstituted in vitro by use of sulfatase-derived peptides as substrates and DTT as the reducing agent (Dierks et al., 2003; Preusser-Kunze et al., 2005). Comparison of FGE sequences from different higher eukaryotes and mass spectrometric analyses revealed six conserved cysteine residues that are organized in three disulfide bonds, one of which is partially reduced and may participate in FGly formation. Notably, no redox-active metal ion or any cofactor is required for FGE activity, and molecular oxygen serves as the terminal electron acceptor. The details of how oxygen-dependent cysteine oxidation is mediated in the ER in general remains unclear, despite the crystal structures of two cysteine oxigenases involved in protein disulfide formation, Ero1p and Erv2p (Gross et al., 2002; Gross et al., 2004).

We determined the crystal structure of FGE to delineate the active site and extract a structure-based mechanism for FGly generation. FGE is a single-domain monomer with a surprising paucity of secondary structure and adopts a unique fold that is stabilized by two Ca^{2+} ions. The effect of all mutations found in MSD patients is explained by the FGE structure, providing a molecular basis for MSD. A redox-active disulfide bond is present in the active site of FGE. Most strikingly, an oxidized cysteine residue, possibly cysteine sulfenic acid, was detected that allows formulation of a struc-

*Correspondence: kfigura@gwdg.de (K.v.F.); markus.rudolph@bio.uni-goettingen.de (M.G.R.)
[3]Present address: Department of Biochemistry I, University of Bielefeld, D-33615 Bielefeld, Germany.

Table 1. Data Collection and Refinement Statistics

Data Set #	1: 1Y1E—SS/SH	2: 1Y1I—reduced	3: 1Y1F—SS/Cso	4: 1Y1H—SS/Peo	5: 1Y1J—SO_3^-/Ala	6: 1Y1G—2×SO_3^-
Data Collection	29.8–1.73	31.4–2.58	29.8–1.80	29.9–1.67	41.1–1.55	29.8–1.67
Resolution range (Å)[a]	(1.78–1.73)	(2.67–2.58)	(1.86–1.80)	(1.72–1.66)	(1.61–1.55)	(1.73–1.67)
Measured reflections	89,457 (5,194)	22,390 (156)	108,643 (1,320)	115,895 (1,934)	91,746 (1,024)	240,190 (12,805)
Unique reflections	30,898 (2,243)	7,791 (133)	25,226 (891)	32,268 (1,167)	35,522 (820)	34,756 (3,182)
Multiplicity	2.9 (2.3)	2.9 (1.2)	4.3 (1.5)	3.6 (1.7)	2.6 (1.25)	6.9 (4.0)
Completeness (%)	97.0 (72.2)	79.9 (14.0)	89.7 (32.5)	89.8 (32.9)	80.7 (19.1)	99.3 (93.1)
Mosaicity (°)	1.20	0.86	0.55	0.48	0.63	0.59
R_{sym} (%)[b]	6.3 (26.4)	14.0 (31.1)	7.6 (19.1)	4.9 (13.1)	4.3 (16.2)	5.5 (28.8)
Average I/σ(I)	17.3 (1.6)	6.3 (1.5)	23.2 (3.5)	32.5 (4.5)	26.5 (3.3)	41.1 (4.5)
Refinement	29.8–1.73	31.4–2.61	29.8–1.8	29.9–1.67	41.1–1.55	29.8–1.67
Resolution range (Å)	(1.78–1.73)	(2.67–2.61)	(1.85–1.80)	(1.71–1.67)	(1.59–1.55)	(1.71–1.67)
R_{cryst} (%)[c]	19.4 (27.6)	15.0 (24.5)	14.8 (19.4)	14.6 (19.2)	15.8 (27.6)	14.6 (21.0)
R_{free} (%)[c]	24.2 (34.2)	25.2 (37.6)	19.4 (27.0)	17.6 (26.1)	19.3 (24.3)	17.6 (28.3)
# of residues/waters	272/407	275/255	272/429	275/473	276/471	275/426
Coordinate error (Å)[d]	0.102	0.252	0.071	0.049	0.054	0.052
Rms bonds (Å)/ angles (°)	0.010/1.36	0.017/1.67	0.013/1.41	0.010/1.33	0.010/1.34	0.010/1.27
Ramachandran plot (%)[e]	88.3/10.8/0/0.9	86.2/12.9/0/0.9	89.2/9.9/0/0.9	87.9/11.2/0/0.9	88.4/10.7/0/0.9	87.9/10.7/0.4/0.9
Average B values (Å²)	29.4 ± 7.6	30.2 ± 7.5	16.3 ± 9.2	19.5 ± 9.6	17.5 ± 9.1	25.7 ± 9.4

[a] Values in parentheses correspond to the highest resolution shell.
[b] $R_{sym} = 100 \cdot \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the i^{th} measurement of reflection h and $\langle I(h) \rangle$ is the average value of the reflection intensity.
[c] $R_{cryst} = 100 \cdot \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$, where F_o and F_c are the structure factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} with 5% of test set structure factors.
[d] Based on maximum likelihood.
[e] Calculated using PROCHECK (Laskowski et al., 1993). Numbers reflect the percentage amino acid residues in the core, allowed, generous allowed, and disallowed regions, respectively.

ture-based mechanism for FGly formation from cysteine residues in all sulfatases.

Results and Discussion

Overall FGE Structure

Six independent structures were determined from crystals grown in different redox environments by molecular replacement using pFGE (Dickmanns et al., 2005) as a search model and refined to resolutions between 2.6 Å and 1.55 Å with excellent stereochemistry (Table 1). Main features of the FGE structure include the presence of two Ca^{2+} ions, N-glycosylation at Asn141, two cis-peptide bonds at Pro115 and Pro266, and three disulfide bonds (Figure 1A). FGE is a compact molecule without obvious domain boundaries that have been suggested from sequence comparisons (Landgrebe et al., 2003) and proteolytic experiments (Preusser-Kunze et al., 2005). The protein adopts a novel fold (Figure 1B) that is similar to pFGE (Dickmanns et al., 2005) but also displays important differences. The secondary structure content of FGE is low with only 20% β- and 13% α-structure. 3% of all residues adopt a 3_{10}-helical conformation. β strands β7, β8, β10, β13, and β14 are very short, consisting of only two residues. No structural homologs were found in the DALI database (Holm and Sander, 1993), which puts FGE and pFGE into a new family of proteins adopting an "FGE" fold. The presence of a novel fold testifies to the unique function of FGE and supports the possibility that FGE and pFGE may have evolved from a common ancestor (Landgrebe et al., 2003).

Comparison of FGE and pFGE

FGE and pFGE are the first structurally characterized members of a large protein family registered as DUF323 (domain of unknown function 323) in the PFAM database of protein families (Bateman et al., 2000). This family currently comprises 168 prokaryotic and eukaryotic proteins of very different functions, including xylanase, transcriptional regulators, and nitrate reductors. Based on the average sequence similarity of 25%, it has been concluded that all members of this family adopt an FGE fold (Dickmanns et al., 2005). However, even the close paralogs FGE and pFGE differ significantly already at the sequence level, suggesting extensive structural variation within this protein family. This prompted us to more elaborately compare the FGE and pFGE structures.

FGE and pFGE superimpose closely with a root mean square deviation of 1.07 Å over 240 Cα atoms. There are three larger insertions in the FGE sequence relative to pFGE, which map to two loop regions (Figure 2A). The first insertion of three residues in FGE immediately precedes the N-glycosylation site of pFGE. The second and third insertions mark an increased loop length of 22 residues from Asp304-Lys325 in FGE as opposed to four residues from Ala244-Gln247 in pFGE. Interestingly, a short β strand (β10) is formed by the larger loop in FGE that extends the β sheet β2/β3 present in pFGE (Figures 1 and 2A, right). The insertions are remote from the active site (see below) but conserved among all FGEs, suggesting an important function that is both unrelated to FGly formation and lacking in pFGEs.

Of the complex glycosylation present in FGE (Preus-

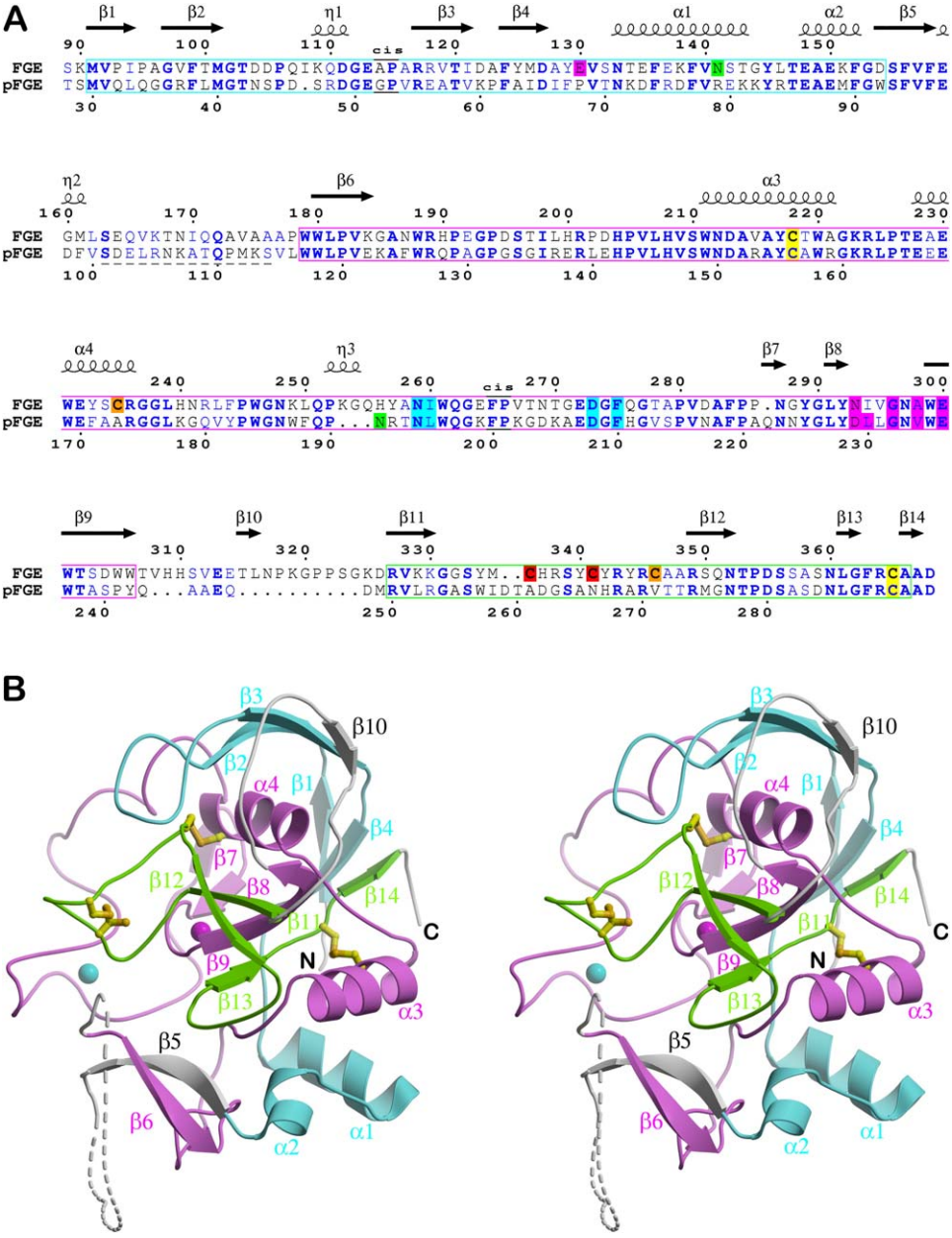


Figure 1. Structure Sequence Relationship of FGE

(A) Structure-based sequence alignment of human FGE and pFGE. The secondary structure elements (η denotes a 3_{10}-helix) for FGE are shown above the sequence. Conserved and homologous residues are colored blue and shown as bold and normal typeface, respectively. The sequence region Ser163-Ala176 (dashed underscored) was removed from FGE by elastase treatment. The disulfide-forming cysteine residues are boxed pairwise in red (active site), orange, and yellow (conserved in pFGE). Green boxes indicate the single N-glycosylation sites in FGE and pFGE. Two cis-prolines are conserved between FGE and pFGE and indicated by black underscores and labeled "cis." Ca^{2+} binding residues share a cyan (site 1) and magenta (site 2) background. The sequence domains assigned on the basis of phylogenetic conservation (Landgrebe et al., 2003) are boxed in cyan, magenta, and green.

(B) Stereo diagram of FGE with the secondary structure elements indicated and the sequence domains colored as in (A). The Ca^{2+} ions are drawn as spheres in cyan (site 1) and magenta (site 2). The missing loop region is indicated as a dashed line.

ser-Kunze et al., 2005), the two proximal N-acetyl-glucosamine residues were identified from the electron density maps at Asn141. While many FGEs, including the enzymes from mouse, rat, chicken, pig, and trout, conserve the glycosylation site, this site is missing in FGE from toad, frog, medaka, and pufferfish. By contrast, all known pFGEs but the one from sea urchin conserve a glycosylation site at Asn191, putting the carbohydrate at the other end of the molecule (distance 35 Å; Figure 2A).

Two metal ions, buried in the hydrophobic core, were clearly identified in the electron density maps of FGE.

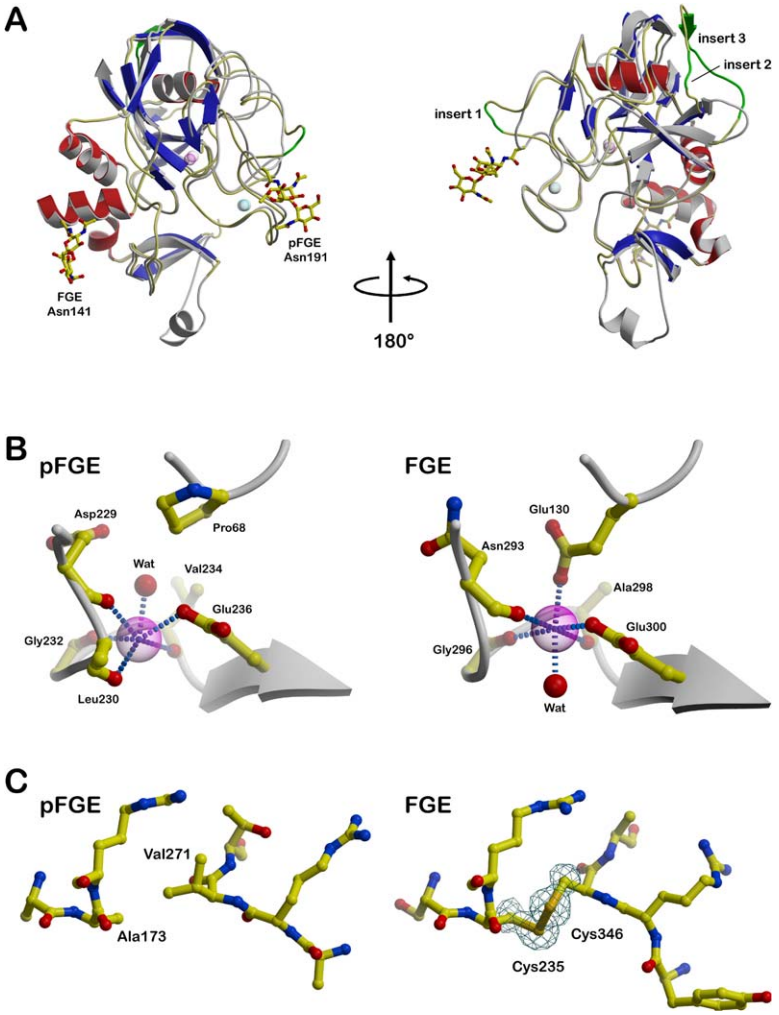


Figure 2. Comparison of FGE and pFGE

(A) Superposition of FGE and pFGE. pFGE is shown as a gray ribbon representation, whereas for FGE α helices are colored red, β strands blue, and loop regions yellow. The two Ca^{2+} ions in pFGE are drawn in gray, and the Ca^{2+} ions in FGE are shown as transparent spheres in cyan (site 1) and magenta (site 2). The three insert regions in FGE relative to pFGE are labeled and shown in green. The carbohydrate moieties are located on opposite sides of the molecules and are drawn as stick models.

(B) Stereochemistry of Ca^{2+} binding site 2 in pFGE (left) and FGE (right). The side chain of pFGE Leu230 is truncated at Cβ for clarity.

(C) Effect of the absence of the second disulfide bond in pFGE (left) compared to FGE (right). The σ_A-weighted $mF_o - DF_c$ omit electron density map around the disulfide bond is contoured at 3σ.

The chemical nature of these ions was determined as Ca^{2+} by analysis of their coordination sphere and by anomalous difference Fourier maps calculated from a data set of high multiplicity (not shown). In addition, the crystallization conditions included >0.2 M $CaCl_2$, ensuring full occupancy of the Ca^{2+} sites. Although K^+ is another candidate ion for the observed binding geometries and the Fourier analyses, given the much higher concentration of Ca^{2+} compared to K^+ in the ER, any significant binding of K^+ to FGE in vivo seems unlikely, and Ca^{2+} will be the natural ligand for FGE. This conclusion is also supported by plasma mass spectroscopic results (see below). Two similarly located Ca^{2+} ions have been identified in pFGE (Dickmanns et al., 2005), which is also retained in the ER. The Ca^{2+} ion at site 1 in FGE is 7-fold coordinated by a bidentate interaction with the Asp273 side chain, the Asn259 side chain, the carbonyl groups of Ile260 and Phe275, and two water molecules. The coordination geometry can best be described as a distorted pentagonal bipyramid and is conserved in pFGE. In pFGE, Ca^{2+} site 2 is coordinated in an irregular fashion by the side chain of Glu236, the carbonyl groups of Asp229, Leu230, Gly232, and Val234, and one water molecule (Figure 2B, left). This ion binding site differs drastically in FGE because of the presence of a new ligand, Glu130, instead of the Pro68 present in pFGE. The Glu130 side chain replaces the water molecule and shifts the Ca^{2+} ion by 0.85 Å relative to its position in pFGE. This abolishes the interaction with the carbonyl group of Leu230, and a water molecule occupies the free coordination site. The net result is an almost perfect octahedral coordination of Ca^{2+} in FGE by the side chains of Glu130 and Glu300, the carbonyl groups of Asn293, Gly296, and Ala298, and a water molecule (Figure 2B, right). As Glu130 is conserved in all FGE sequences and, likewise, Pro68 is conserved in all pFGE sequences, it can be concluded that the Ca^{2+} geometry at site 2 will be the same for all FGEs and pFGEs and exhibit the same differences when FGEs and pFGEs are compared. In summary, the Ca^{2+}-coordinating residues are conserved in most FGE sequences, suggesting that the two Ca^{2+} ions are an integral part of FGE/pFGE protein stability.

The observation that pFGE crystallizes as a dimer in the asymmetric unit (Dickmanns et al., 2005) and, more compellingly, that FGE/pFGE expression levels are similar throughout different tissue types and that pFGE overproduction antagonizes FGE activity (Mariappan et

al., 2005) suggests the possibility that FGE may form heterodimers with pFGE. In solution, no heterodimers were observed by gel permeation chromatography (data not shown), establishing that the affinity of FGE for pFGE, if present, is low. FGE/pFGE dimers are sterically possible, albeit with a small buried surface area and a low surface complementarity (Dickmanns et al., 2005). While these considerations suggest such a dimer to be unstable, they do not rule out a possible interaction of FGE and pFGE in the presence of sulfatases. If FGE/pFGE heterodimerization cannot explain the antagonistic effect of pFGE on FGE activity (Mariappan et al., 2005), another unidentified protein with a function that links pFGE to FGE activity may be involved.

Two out of three disulfide bonds, Cys218-Cys365 and Cys235-Cys346, are permanently present in FGE and contribute to its stability. The Cys218-Cys365 disulfide bond is also present in pFGE and adopts the same structure, establishing its role in protein stabilization. Ala173 and Val271 in pFGE (Figure 2C, left) replace the second disulfide bond in FGE (orange in Figures 1 and 2C, right). Ala173 is conserved among all pFGEs while position 271 can be either valine or isoleucine (sea urchin and trout). The lack of the second disulfide bond in pFGE leads to a rigid body shift of the region Ala269-Thr273 by about 1.8 Å away from Ala173 toward the surface, with a concomitant expansion of the whole pFGE structure. The third disulfide bond in FGE, Cys336-Cys341, has varying occupancies in the FGE crystal structures and participates in catalysis (see below).

Delineation of the Substrate Binding and Active Sites

Cysteine-containing peptides generated from human sulfatases have proven invaluable for an in vitro assay of FGE activity (Dierks et al., 2003). A 23-mer peptide derived from arylsulfatase A, ASA p23-Bpa77, comprising ASA residues 60-TDFYVPVSL**C**TPSRAALBpaTGR-80 (the modified Cys69 residue is bold, the minimal motif for modification is underlined), which carries the photoreactive p-benzoyl-phenylalanine (Bpa) at position Leu77, crosslinks to Pro182 in FGE (Preusser-Kunze et al., 2005).

Mixed disulfide bonds between a substrate peptide and preferentially Cys341, and to a much lesser extent also with Cys336, were observed by mass spectrometry (J. Peng, K.v.F., T.D., B.S., A.P.-K., and M.M., unpublished). In addition, Cys336 and Cys341 are conserved among all FGEs and are in close proximity to each other, facilitating intramolecular disulfide bond formation. The loop between Cys336/Cys341 displays elevated main chain temperature factors, suggesting structural flexibility, which is a hallmark for regions undergoing conformational changes during binding events (Luan et al., 2000). Thus, the redox-active Cys336/Cys341 region in FGE is a prime candidate for a catalytic entity. The active site, therefore, is likely to comprise all three residues: Pro182, Cys336, and Cys341.

The distance between the Cα atoms of Cys341 and Pro182 is $\sim$21 Å. For a fully extended conformation, the distance between Cys69 and Leu77 in the ASA p23 peptide is calculated to 27 Å. Thus, the substrate peptide could easily span the distance between Cys341 and Pro182, and the distances involved suggest that the substrate is bound to FGE in an almost completely extended fashion. No crosslinks were detected with peptides that carried the photoactive Bpa label N-terminal to the modified Cys69 (Preusser-Kunze et al., 2005), further establishing that the orientation of the substrate with respect to FGE is not arbitrary but with the N-terminal part at Cys341 and the C-terminal part located in the vicinity of Pro182.

Indeed, a surface representation of FGE shows an oval-shaped groove of 20 Å length, 10 Å depth, and 12 Å width (yellow in Figure 3) that is bordered by the Cys336/Cys341 pair at one end and Pro182 at the other end. This groove is sufficiently large to host up to six amino acids of a substrate and thus could accommodate the minimal sequence motif of CXPSR (where X is T, S, C, or A) that can be converted to FGly by FGE. Eleven residues (Ala149, Ser155, Trp299, Lys329, Ser333, Met335, Ser356, Ser357, Ala358, Asn360, and Leu361) line the groove. Importantly, the groove is extended in both directions, although not as deep, along the surface of FGE, which enlarges the substrate binding area. Asn107, Arg338, and Tyr342 (magenta in Figure 3) at the N-terminal substrate binding site and Phe152 and Asn154 (blue in Figure 3) at the C-terminal substrate binding site form these areas. In conclusion, the "channel" across the surface of FGE and the groove bordered by the reactive disulfide bond and Pro182 in FGE constitute a functionally bipartite binding/active site for the unfolded sulfatase substrates.

Cys336/Cys341 Is a Redox-Active Disulfide Bond and Essential for Catalysis

We assessed the redox activity of the Cys336/Cys341 pair and its role in the catalytic mechanism of FGE by six independent crystal structures, which were determined from crystals of different age or that were grown under different redox environments. Unlike all other cysteine residues in FGE, Cys336 and Cys341 are extremely sensitive to oxidation. In structure 1 (Figure 4A), obtained from a 3-day-old crystal, the Cys336/Cys341 pair is oxidized and reduced, respectively, in about half of the molecules. A fully reduced FGE molecule was generated by addition of 20 mM cysteine to the crystallization setup (structure 2; Figure 4B). Older FGE crystals of 16 days (structure 3; Figure 4C) and 39 days (structure 4; Figure 4D) clearly show the presence of a partial Cys336-Cys341 disulfide bond. Surprisingly, additional unbiased $F_o - F_c$ electron density is visible in both structures at the part of Cys336 that is not engaged in a disulfide bond. This density is fitted equally well by a cysteine sulfenic acid plus a water molecule that is partially occupied (Cys-S-OH or Cso; Figure 4C) or by a hydroperoxide moiety (Cys-S-O-OH or Peo; Figure 4D). Both of these entities may represent catalytically relevant intermediates, although the peroxide moiety presumably is less stable than the sulfenic acid. The sensitivity of the Cys336/Cys341 pair to oxidation is further borne out by a structure determined from a 22-day-old FGE crystal grown at pH 10.5. This structure shows Cys336 to be oxidized to cysteine sulfonic acid

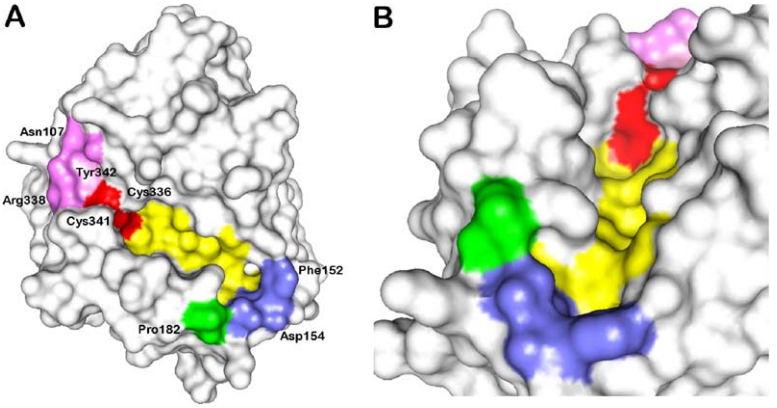


Figure 3. Substrate Binding Groove in FGE

(A) The surface representation of FGE shows a groove stretching from the site of the crosslink at Pro182 to the redox-active cysteine pair Cys336/Cys341. The solvent-accessible surfaces of Pro182 and the cysteine residues are colored green and red, respectively. The surfaces of residues forming the deep groove are colored in yellow. Magenta and blue surfaces mark the areas that bind to parts of the substrate that are N- and C-terminal to the modified cysteine residue.

(B) Side view and close-up of (A).

(structure 5; Figure 4E). No electron density is visible for Cys341 beyond Cβ, suggesting either rotational disorder or desulfuration to alanine. Likewise, addition of 0.5% H_2O_2 to the crystallization setup resulted in complete oxidation of both cysteine residues to the sulfonic acids within $\leq$ 13 days (structure 6; Figure 4F). Oxidation to cysteine sulfonic acid is irreversible, thus eliminating it as an intermediate. But given the fact that the other disulfide bonds do not show any signs of oxidation as opposed to complete oxidation of the Cys336/Cys341 pair, it can be concluded that the redox chemistry during FGly formation must be carried out at the Cys336/Cys341 pair. Oxidation of cysteine to the sulfonic acid is irreversible, leading to a dead-end side product. This raises the problem of FGE inactivation in the ER by reactive oxygen species that are generated during oxidative protein folding (Tu et al., 2000).

The direct implication of Cys336 and Cys341 in FGE catalysis was proven by the inactivity of the serine mutants. The C336S and C341S FGE proteins were purified to homogeneity from stably overproducing HT1080 mutant cell lines and analyzed for FGly-generating activity in vitro. Both mutants are catalytically inactive (<0.2% of wild-type activity). Even extensively prolonged incubation times and using 50-fold more of the mutant proteins as compared to the wild-type control

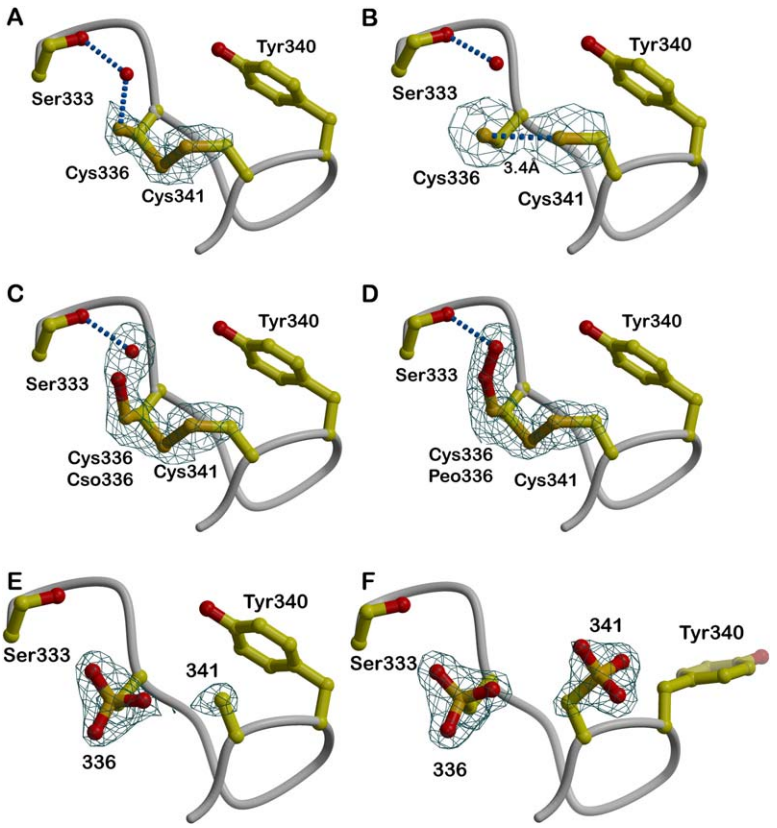


Figure 4. Structural Plasticity of the Active Site Cysteine Residues in FGE Crystal Structures from Different Redox Environments

(A) Cys336 forms a partial disulfide bond with Cys341 in this structure. The other conformation of Cys336 hydrogen bonds to a water molecule, which is also bound by Ser333 (structure 1).

(B) The reduced form of the Cys336/Cys341 pair (structure 2). The distance of the two sulfur atoms is 3.4 Å, which is too far for a disulfide bond (2.03 Å).

(C) Two conformers of the Cys336/Cys341 pair. Cys336 either forms a disulfide bond with Cys341 or is oxidized to the sulfenic acid (structure 3).

(D) Very similar electron density is present at Cys336 in another structure from a crystal grown in the presence of $SrCl_2$ instead of $CaCl_2$ (structure 4). Cysteine peroxide is an alternative description of this electron density.

(E) Electron density of the Cys336/Cys341 region from a structure determined at pH 10.5 (structure 5). Cys336 is oxidized to the sulfonic acid, while Cys341 has no electron density beyond Cβ.

(F) A different oxidation pattern of the same region produced by 0.5% H_2O_2 during crystallization (structure 6). Both cysteine residues are converted to the sulfonic acids. The bulky sulfonic acid at position 341 forces the Tyr340 side chain to reorient toward the solvent. All σ_A-weighted $mF_o - DF_c$ omit electron density maps are contoured at the 4σ level.

did not result in detectable substrate turnover. These data demonstrate that both Cys336 and Cys341 are part of the active site and essential for the FGly-generating activity of FGE.

FGE Employs a Nonclassical Mechanism for Cys Oxidation

Many enzymes catalyzing redox processes use redox-active metals in Fe-S- or Mo-Fe-clusters, or Ni^{2+} or Cu^{2+} to exchange electrons during catalysis. Others rely on nucleotide or protein cofactors such as $NAD(P)^+$, FAD, or the thioredoxins, which employ reversible disulfide bond formation during catalysis. The ER-localized oxidase Ero1p reoxidizes protein disulfide isomerase by a flavin-dependent mechanism with molecular oxygen as the terminal electron acceptor (Gross et al., 2004; Tu et al., 2000). Examples of metal ion-containing oxidases located in the ER include the Cu-dependent monooxigenase X (Xin et al., 2004) and the Fe-dependent dioxigenase prolyl 3-hydroxylase 1 (Vranka et al., 2004). Thus, it seemed possible that FGE could employ either of these or a similar strategy for catalysis.

As organic cofactors neither are needed for catalysis nor are present in the FGE electron density maps, involvement of flavines or nicotinamides, for example, can be ruled out. A thioredoxin-type mechanism seems unlikely because thioredoxins rely on accessory proteins for reoxidation of their disulfide bond, whereas FGE displays multiturnover catalysis independent of other protein factors (J. Peng, K.v.F., T.D., B.S., A.P.-K., and M.M., unpublished). Initially, the finding of two metal ions in the FGE electron density maps raised the question of whether they partake in the mechanism. Metal ion catalysis is essential for FGly generation in prokaryotes such as *K. pneumoniae*, where the FGE analog AtsB uses a Fe-S-cluster to convert a serine residue in the periplasmic sulfatase AtsA to FGly (Fang et al., 2004; Marquordt et al., 2003). However, in FGE several lines of evidence argue against metal ion involvement in catalysis. First, no requirement of a redox-active metal ion was found in in vitro assays. Second, plasma mass spectrometry established molar ratios Ca^{2+}/FGE and K^+/FGE of 1.65 and 0.69, respectively, without significant content of redox-active metals (Preusser-Kunze et al., 2005). Third, redox-active metal ions such as Cu^{2+} in coupredoxins and Fe^{2+} in dioxigenases are often bound to at least one cysteine or histidine residue, respectively (Karlin et al., 1997), which is not the case in FGE. Where Ca^{2+} is present in other redox-active proteins such as horseradish peroxidase (Haschke and Friedhoff, 1978), lignin peroxidase (Poulos et al., 1993), and the di-heme *Paracoccus* cytochrome c peroxidase (Gilmour et al., 1994), it was always found to serve a structural, not a catalytic, purpose. In summary, eukaryotic FGEs must employ a unique mechanism that involves a disulfide bond and that is very different from the redox chemistry seen in conventional oxigenases and dehydrogenases.

The Catalytic Cycle of FGly Generation

Formally, the oxidation of an alcohol or a thiol to the respective aldehyde requires abstraction of a hydride ion from the carbon atom carrying the hydroxyl or thiyl group. This strategy is used by NAD^+-dependent dehydrogenases but cannot be employed by FGE due to the absence of a hydride acceptor in the active site. Another strategy would be the oxidation of the thiol sulfur atom to a sulfenic acid, followed by a β-elimination of hydroxide or water. In the case of a cysteine residue, this elimination requires abstraction of a proton from the Cβ-atom by a catalytic base. This scenario is supported by the cysteine sulfenic acid at position 336 in structure 3 (Figure 4C). Cys336 exhibits two alternate conformations, either forming a disulfide bond with Cys341 or existing as cysteine sulfenic acid. As the redox potential in the ER is more oxidizing than that of the cytosol, both the disulfide-bridged form of FGE and the sulfenic acid are likely reaction intermediates in the catalytic cycle.

Cysteine sulfenic acids are implicated in diverse cellular processes, including signal transduction, transcriptional regulation, oxygen metabolism, and the oxidative stress response (Claiborne et al., 1999). In enzymes, stable cysteine sulfenic acid moieties have been found in NADH peroxidase (Yeh et al., 1996), NADH oxidase (Ahmed and Claiborne, 1989), nitrile hydratase, and the hORF6 (Choi et al., 1998) and AhpC (Ellis and Poole, 1997) peroxiredoxins, where they either are involved in substrate binding or take part in catalysis. The presence of the sulfenic acid in FGE raises the question as to how Cys336 is oxidized. Recently, it has been shown that molecular oxygen is a substrate for FGE and that the stoichiometry of O_2 consumption and FGly formation is close to 1 (J. Peng, K.v.F., T.D., B.S., A.P.-K., and M.M., unpublished). Thus, the cysteine sulfenic acid in FGE could be generated in the ER by reaction of O_2 with Cys336, possibly by a radical mechanism. The putative hydroperoxide at Cys336 in structure 4 may be a product of this reaction, which may then decompose to the sulfenic acid. Apart from the cases where the cysteines were irreversibly oxidized to the sulfonic acids by H_2O_2 (Figure 4E), no oxidation to a sulfenic, sulfinic, or sulfonic acid was ever seen with Cys341. This fact points to Cys336 as the major redox partner during FGly generation. In conjunction with the mass spectrometric result that peptides form mixed disulfide bonds preferentially with Cys341, two specific functions may be assigned to the catalytically active cysteines: Cys341 binds the substrate and Cys336 oxidizes it.

Three conserved residues, Ser333, His337, and Tyr340, are present in the vicinity of the Cys336/Cys341 pair and could serve as catalytic bases for the elimination reaction. Ser333 is the most likely candidate, as it is located only 4.6 Å from the modified Cys336 and 7.7 Å from Cys341. The proximity of Ser333 and Cys336 is congruent with a stabilization of the Cys336 sulfenic acid by Ser333 (3.0 Å distance between the oxygen atoms) and the substrate-oxidizing function of Cys336. Because at some stage the β-elimination of water from the cysteine sulfenic acid requires abstraction of a proton from the Cβ of the substrate cysteine, Ser333 would have to either exhibit a perturbed pKa value or become deprotonated during catalysis. The negative charge of the forming anion would be stabilized by a hydrogen bond to the Nδ atom of the side chain of conserved Asn360, which is only 3.4 Å away from the Ser333 hy-

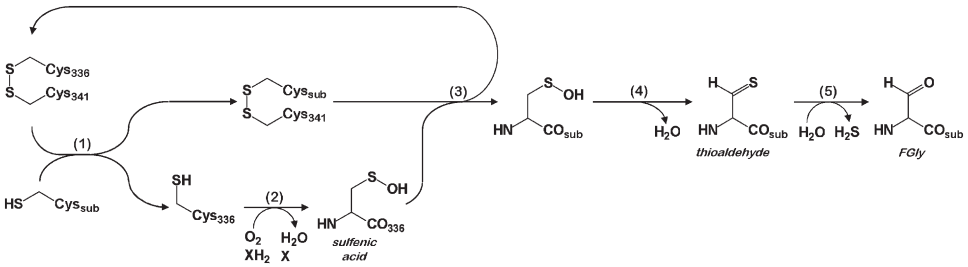


Figure 5. Possible Catalytic Mechanism of FGly Formation that Involves the Cysteine Sulfenic Acid at Position 336 of FGE

(1) Formation of the mixed disulfide intermediate between the substrate and Cys341 of FGE.

(2) Generation of the sulfenic acid at Cys336.

(3) Transfer of the hydroxyl group from Cys336 to the substrate cysteine residue and regeneration of the disulfide form of FGE.

(4) Attack of a catalytic base at Cβ of the substrate cysteine sulfenic acid and elimination of water/hydroxide.

(5) Spontaneous hydrolysis of the thioaldehyde intermediate to FGly. Alternative mechanisms with the cysteine sulfenic acid moiety solely in the peptide are conceivable but do not accommodate the observed modification of Cys336. XH_2 denotes the chemical (DTT) or biological (e.g., glutathione) reducing agent.

droxyl group. Alternatively, a solvent hydroxide ion could deprotonate Ser333 in a concerted fashion during attack of Ser333 on the Cβ-proton of the substrate. This mechanism of activation of the catalytic base is in line with the alkaline (pH 9–10) optimum of FGE activity (Fey et al., 2001) and, more importantly, the presence of electron density for an oxygen atom (water or hydroxide) in structures 1–3, which hydrogen bonds to Ser333 (Figures 4A–4C). The alternative catalytic bases His337 and Tyr340 are ~10 Å and 7 Å away from Cys336, respectively, rendering their involvement in catalysis less likely.

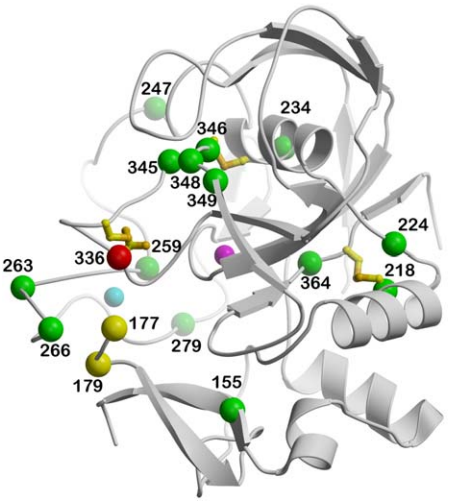
To test the possible roles of residues Ser333, His337, and Tyr340 in the catalytic mechanism of FGE, four mutants (Ser333Ala, Ser333Thr, His337Ala, and Tyr340Phe) were generated and assayed for FGE activity (see Supplemental Table S1 available with this article online). While the Tyr340Phe mutation did not affect FGE activity significantly, the His337Ala mutation resulted in a 5-fold decrease of specific activity. The most dramatic effect is, however, displayed by the Ser333Ala (inactive; <0.5% of wild-type activity) and Ser333Thr (<0.8% of wild-type activity) mutants. Thus, Ser333 must be strongly involved in the FGE catalytic mechanism and is the prime candidate for the catalytic base.

The biochemical and structural data can be integrated into a novel, yet at this stage of the analysis hypothetical, reaction mechanism of FGly formation (Figure 5). In this mechanism, the resting state of FGE is assumed to be the oxidized form with the Cys336-Cys341 disulfide bond. The substrate, i.e., the unfolded sulfatase, undergoes a disulfide exchange reaction to form a mixed disulfide with Cys341. The free Cys336 is oxidized to the sulfenic acid by molecular oxygen in a monooxigenase-like fashion that requires a reducing agent (glutathione or thioredoxin in vivo, or DTT in vitro) and produces a water molecule. The hydroxyl group of the Cys336-OH sulfenic acid is transferred to the substrate cysteine to form a sulfenic acid in the sulfatase. This reaction is similar to a step in the proposed mechanism for the *N. gonorrhoeae pilB* methionine sulfoxide reductase (Lowther et al., 2002) and regenerates the disulfide bonded form of FGE. The catalytic base Ser333 initiates a β-elimination of a good leaving group, the hydroxide ion, by abstraction of a proton from the Cβ-atom of the substrate cysteine sulfenic acid. The product is a thioaldehyde, which is unstable under aqueous conditions and undergoes spontaneous hydrolysis to FGly and H_2S. The net result is the conversion of one molecule O_2 and a thiol into one molecule of water, H_2S, and an aldehyde (Figure 5).

This proposed FGly-generation mechanism raises several testable hypotheses. First, the product H_2S should be identifiable by either radioactive labeling or gas chromatography. Second, the Cys336Ser mutant should trap a covalent FGE-peptide complex. Third, the peptide sulfenic acid should be trapped by a modified substrate peptide containing Cys-CF_2SH at the position of the modified cysteine. Fourth, no oxygen atom from O_2 but from bulk solvent is incorporated into FGly, which could be tested by isotope-labeling experiments. Finally, if molecular oxygen generates the sulfenic acid by a radical mechanism, radical scavengers should abolish FGE activity. This last point has been addressed in initial studies using ibuprofen (1 μM–10 mM) as a radical scavenger. No effect was detected on FGE activity (data not shown), indicating that generation of the cysteine sulfenic acid does not require radical formation unless the lifetime of the radical involved is too short to be efficiently quenched by ibuprofen.

Absence of FGly-Generating Activity in pFGE

Whereas all peptides derived from human sulfatases that contain an FGly modification motif are converted by catalytic amounts of FGE, no such activity is observed with pFGE (Mariappan et al., 2005). The question of why pFGE lacks FGly-generating activity can be directly answered by the absence of the catalytically active disulfide bond (Figure 1A). However, peptides that are substrates for FGE also bind to pFGE (Mariappan et al., 2005), because many of the residues (Ala149, Ser155, Trp299, Ser333, Ser356, Asn360, and Leu361) lining the substrate binding groove in FGE are conserved in pFGE. This points to the possibility that pFGE competes with FGE for substrates. Indeed, an inhibitory effect of pFGE on FGE-mediated activation of



Mutation	Effect on FGE structure
S155P #	Three hydrogen bonds lost; clash with A186
C218Y #	Loss of Cys218/Cys365 SS-bond
R224W *	Loss of five hydrogen bonds
S234R	One hydrogen bond lost; clash with F124, M126
G247R	No space for larger side-chain
N259I *	Ca^{2+}-binding at site 1 abolished
G263V	Larger side-chain ; clash with T263 and T270
P266L *	Cis-peptide bond destabilized
A279V †#	Larger side-chain, clash with A283 and F275
R345C #	Four hydrogen bonds lost
C346W	Loss of Cys235/Cys346 SS-bond
A348P #	Proline side-chain too large for type I β-turn
R349W †#	Five interactions lost
R349Q †#	Four interactions lost
R364C	Four interactions lost
A177P *	Possibly substrate binding compromised
W179S	On top of active site
C336R †#	Catalytically active residue

Figure 6. Molecular Basis of Multiple Sulfatase Deficiency

The known missense multiple sulfatase deficiency (MSD) mutations are mapped onto the FGE structure as spheres and are colored according to the effect they have on the FGE structure. Green, destabilizing; red, catalytically compromising; yellow, inhibition of substrate binding. The Ca^{2+} ions are shown as cyan and magenta spheres. The table compiles a brief structural explanation of the effect the missense mutations have on the FGE structure. The majority of the mutations will prevent folding of FGE into an active conformation by destabilization of the hydrophobic core. Further destabilizing mutations are located at a tight surface loop (263 and 266) and at Ca^{2+}-site 1 (259). Mutations marked with an asterisk, hash mark, and dagger are reported in Cosma et al. (2004), Cosma et al. (2003), and Dierks et al. (2003), respectively. The S234R and C346W mutations are unpublished (S. Marie and M.F. Vincent, personal communication). All other mutations are our own unpublished data. A more detailed analysis of the effect the mutations have on the FGE structure is presented in Supplemental Table S2.

newly synthesized sulfatases has been observed (Mariappan et al., 2005). pFGE may exhibit a chaperone-like function by preserving a pool of unfolded and translationally arrested sulfatases in the ER for FGE in situations of increased sulfatase production.

Molecular Basis for Multiple Sulfatase Deficiency

MSD is a rare (52 alleles in 32 patients) autosomal recessive disease. The almost complete lack of all sulfatase activities leads to severe symptoms such as ataxia; progressive loss of motor abilities, speech, vision and hearing; hepatosplenomegaly; dystosis multiplex; abnormal facies; stiff joints; and ichthyosis, and ultimately results in untimely death. To determine the structural reason for MSD, the missense mutations were mapped onto the FGE crystal structure (Figure 6, left). A total of 18 missense mutations over 17 different residues of mature FGE have been found to date, of which 12 have been published (Cosma et al., 2003; Cosma et al., 2004; Dierks et al., 2003). Fourteen of these mutations lead to substitutions of residues that are strictly conserved among FGE proteins from different organisms. In the remaining three cases (positions 177, 234, and 279), the residues are highly conserved as serine or alanine. The mutations fall into three classes: structure destabilizing, interference with substrate binding, and catalytically inactivating (Supplemental Table S2). Most mutations (15 out of 18) destabilize FGE either by inserting a residue with too large a side chain for efficient packing of the hydrophobic core or by altering the side chain properties such that important interactions are lost (green in Figure 6). Structurally very interesting are the Pro266Leu and Asn259Ile mutations, which affect a conserved cis-peptide bond and Ca^{2+} binding site 1, respectively. The Asn259Ile mutation supports the notion that the Ca^{2+} ions strongly contribute to FGE stability.

Of particular note are the mutations Ala177Pro, Trp179Ser, and Cys336Arg, which have a more drastic effect on substrate binding and catalytic activity of FGE. Ala177 and Trp179 are located close to the active site, i.e., mutation of these residues will most likely compromise substrate binding (yellow in Figure 6). More compellingly, Cys336 is part of the active site and the Cys336Arg mutation will exhibit a strongly decreased, if not complete loss of, catalytic activity (red in Figure 6), similar to the Cys336Ser mutation (see above).

Conclusions

The human FGE forms the bottleneck in the activation of all sulfatases. The structure explains on a molecular basis the effect of several missense mutations that lead to MSD. The severity of the disease and the fact that most of the mutations lead to destabilized, and therefore inactive, FGE allows the conclusion that there is no FGE backup activity in human. This is surprising, as nematodes and fungi lack FGE but still contain FGly-based sulfatases (Landgrebe et al., 2003). Hence, another FGly-generating system must be present in these organisms.

We have determined six individual crystal structures of FGE, which reveal both dead-end products (sulfonic acids) that delineate the active site and putative reaction intermediates (sulfenic acid), which allowed construction of a unique oxigenase reaction mechanism

not hitherto engaged in any other class of enzymes. Based on unbiased high-resolution electron density, we propose a cysteine sulfenic acid as a crucial reaction intermediate in the catalytic cycle of FGly generation. To our knowledge, FGE is the first enzyme to use molecular oxygen as a terminal electron acceptor (J. Peng, K.v.F., T.D., B.S., A.P.-K., and M.M., unpublished) without requirement of an oxygen-activating cofactor such as FAD. The structural dissimilarity of FGE to Ero1p and Erv2p excludes that these ER-localized cysteine oxigenases share an analogous mechanism. Further studies are required to entirely unravel FGly formation by FGE, especially the role of the sulfenic acid. Because FGE is easily inactivated by irreversible oxidation of Cys336 and/or Cys341, the presence of reactive oxygen species in the ER may have prompted evolution of an FGE-protecting system of which the paralog pFGE may be a part.

Experimental Procedures

FGE was cloned and produced from HT1080 fibrosarcoma cells in a glycosylated form as described (Preusser-Kunze et al., 2005). As extensive crystallization trials with the full-length FGE protein were unsuccessful, limited proteolysis with elastase was employed. A mixture of elastase and 9.6 mg/ml FGE (1:1000 w/w) was directly subjected to crystallization trials at 20°C and yielded irregularly shaped crystals overnight from 30% PEG4000, 0.1 M Tris/HCl [pH 8.5], 0.2 M $MgCl_2$. Large single crystals were obtained from 20%–25% PEG4000, 0.1 M Tris/HCl [pH 8.0–9.0], 0.2–0.3 M $CaCl_2$. DTT (2.5 mM) or cysteine (10–20 mM) or H_2O_2 (0.5%) were added, combined with incubation times ranging from 1 to 72 days, in order to generate different redox environments during crystal growth and aging. SDS-PAGE analysis proved the presence of two fragments in the crystals. Blotting and sequencing of these fragments established that the protease had removed the N-terminal residues up to Glu73 and residues Ser163-Ala176. All data were collected in-house at 100K on a MAR345 image plate detector mounted on a MicroMax 007 generator and reduced with the HKL programs (Table 1; Otwinowski and Minor, 1997). This data collection protocol eliminated any possible radiation damage to the active site cysteine residues. The space group is $P2_12_12$ with unit cell dimensions a = 61.8 Å, b = 109.6 Å, c = 43.4 Å. The Matthews coefficient (V_M = 2.3 $Å^3$/Da) suggested one molecule in the asymmetric unit with a solvent content of 46%. All structure determination and refinement calculations were carried out using the CCP4 program suite (CCP4, 1994). The structure was determined from data between 12 Å and 4 Å by molecular replacement with MOLREP using an unrefined initial model of pFGE (Dickmanns et al., 2005). Rigid body refinement and phase extension to the maximum resolution limit resulted in readily interpretable maps that allowed manual rebuilding of the model with O (Jones et al., 1991). All other FGE structures were determined by molecular replacement using the first FGE structure as the starting model. Refinement was performed with REFMAC5 with the same set of 5% of reflections reserved for R_{free} crossvalidation (Brünger, 1992). Water oxygen atoms were assigned automatically with ARP/wARP (Lamzin and Wilson, 1993). In the final models, residues Ser163-Ala176 are absent as a result of proteolysis. Thirteen N- and three C-terminal residues followed by the sequence $RGSH_6$ display poor electron density and are disordered to various extents depending on the actual crystal. Two metal ions are present in all structures (Sr^{2+} in structure 4, Ca^{2+} in all others). Two N-acetyl-glucosamine moieties at Asn141 were visible in the electron density. Two residues, Phe284 and Asn297, have disallowed ϕ/ψ–combinations in the Ramachandran plot, but these residues display excellent electron density. Phe284 is very close to a generously allowed region, and Asn297 binds to the Ca^{2+} ion in site 2. Data collection and refinement statistics are summarized in Table 1. Possible hydrogen bonds, salt bridges, and van der Waals contacts were detected with HBPLUS (McDonald and Thornton, 1994) and CONTACSYM (Sheriff et al., 1987) using default parameters. Secondary structure was assigned with STRIDE (Frishman and Argos, 1995). Figure 1A was generated with ESPript (Gouet et al., 2003). Structure figures were created with Bobscript (Esnouf, 1997) and rendered with Raster3D (Merritt and Murphy, 1994). Surfaces were calculated with PyMol (http://pymol.sourceforge.net/).

Supplemental Data

Supplemental Data include two tables and Supplemental Experimental Procedures and can be found with this article online at http://www.cell.com/cgi/content/full/121/4/541/DC1/.

Acknowledgments

The authors declare that they have no conflicting financial interests. M.G.R. thanks Dagmar Klostermeier for key discussions and helpful comments on the manuscript and Kathrin Gasow for excellent technical assistance. We also thank our colleagues Oliver Einsle and Jianhe Peng for discussions and Lars Schlotawa for generation of the cDNA for the FGE S333A and S333T mutants. This work was supported by grants from the Deutsche Forschungsgemeinschaft, Transkaryotic Therapies, and Fonds der Chemischen Industrie.

Received: December 2, 2004
Revised: February 10, 2005
Accepted: March 2, 2005
Published: May 19, 2005

References

Ahmed, S.A., and Claiborne, A. (1989). The streptococcal flavoprotein NADH oxidase. II. Interactions of pyridine nucleotides with reduced and oxidized enzyme forms. J. Biol. Chem. *264*, 19863–19870.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000). The Pfam protein families database. Nucleic Acids Res. *28*, 263–266.

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

Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., and Guss, J.M. (1997). Structure of a human lysosomal sulfatase. Structure *5*, 277–289.

Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

CCP4 (Collaborative Computational Project, Number 4)(1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.

Choi, H.J., Kang, S.W., Yang, C.H., Rhee, S.G., and Ryu, S.E. (1998). Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution. Nat. Struct. Biol. *5*, 400–406.

Claiborne, A., Yeh, J.I., Mallett, T.C., Luba, J., Crane, E.J., 3rd, Charrier, V., and Parsonage, D. (1999). Protein-sulfenic acids: diverse roles for an unlikely player in enzyme catalysis and redox regulation. Biochemistry *38*, 15407–15416.

Cosma, M.P., Pepe, S., Annunziata, I., Newbold, R.F., Grompe, M., Parenti, G., and Ballabio, A. (2003). The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. Cell *113*, 445–456.

Cosma, M.P., Pepe, S., Parenti, G., Settembre, C., Annunziata, I., Wade-Martins, R., Di Domenico, C., Di Natale, P., Mankad, A., Cox, B., et al. (2004). Molecular and functional analysis of SUMF1 mutations in multiple sulfatase deficiency. Hum. Mutat. *23*, 576–581.

Dickmanns, A., Schmidt, B., Rudolph, M.G., Mariappan, M., Dierks, T., von Figura, K., and Ficner, R. (2005). Crystal structure of human pFGE, the paralog of the Cα-formylglycine generating enzyme. J. Biol. Chem. *280*, 15180–15187.

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., Schmidt, B., Borissenko, L.V., Peng, J., Preusser, A., Mariappan, M., and von Figura, K. (2003). Multiple sulfatase deficiency is caused by mutations in the gene encoding the human Cα-formylglycine generating enzyme. Cell *113*, 435–444.

Ellis, H.R., and Poole, L.B. (1997). Novel application of 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole to identify cysteine sulfenic acid in the AhpC component of alkyl hydroperoxide reductase. Biochemistry *36*, 15013–15018.

Esnouf, R.M. (1997). An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 132–134.

Fang, Q., Peng, J., and Dierks, T. (2004). Post-translational formylglycine modification of bacterial sulfatases by the radical S-adenosylmethionine protein AtsB. J. Biol. Chem. *279*, 14570–14578.

Ferrante, P., Messali, S., Meroni, G., and Ballabio, A. (2002). Molecular and biochemical characterisation of a novel sulphatase gene: Arylsulfatase G (ARSG). Eur. J. Hum. Genet. *10*, 813–818.

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.

Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. Proteins *23*, 566–579.

Gilmour, R., Goodhew, C.F., Pettigrew, G.W., Prazeres, S., Moura, J.J., and Moura, I. (1994). The kinetics of the oxidation of cytochrome c by *Paracoccus* cytochrome c peroxidase. Biochem. J. *300*, 907–914.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Gross, E., Sevier, C.S., Vala, A., Kaiser, C.A., and Fass, D. (2002). A new FAD-binding fold and intersubunit disulfide shuttle in the thiol oxidase Erv2p. Nat. Struct. Biol. *9*, 61–67.

Gross, E., Kastner, D.B., Kaiser, C.A., and Fass, D. (2004). Structure of Ero1p, source of disulfide bonds for oxidative protein folding in the cell. Cell *117*, 601–610.

Haschke, R.H., and Friedhoff, J.M. (1978). Calcium-related properties of horseradish peroxidase. Biochem. Biophys. Res. Commun. *80*, 1039–1042.

Hernandez-Guzman, F.G., Higashiyama, T., Pangborn, W., Osawa, Y., and Ghosh, D. (2003). Structure of human estrone sulfatase suggests functional roles of membrane association. J. Biol. Chem. *278*, 22989–22997.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.

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 Diseases, 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.

Jones, T.A., Cowan, S., Zou, J.Y., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.

Karlin, S., Zhu, Z.Y., and Karlin, K.D. (1997). The extended environment of mononuclear metal centers in protein structures. Proc. Natl. Acad. Sci. USA *94*, 14225–14230.

Lamzin, V.S., and Wilson, K.S. (1993). Automated refinement of protein models. Acta Crystallogr. *D49*, 129–147.

Landgrebe, J., Dierks, T., Schmidt, B., and von Figura, K. (2003). The human SUMF1 gene, required for posttranslational sulfatase modification, defines a new gene family which is conserved from pro- to eukaryotes. Gene *316*, 47–56.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lowther, W.T., Weissbach, H., Etienne, F., Brot, N., and Matthews, B.W. (2002). The mirrored methionine sulfoxide reductases of *Neisseria gonorrhoeae pilB*. Nat. Struct. Biol. *9*, 348–352.

Luan, P., Heine, A., Zeng, K., Moyer, B., Greasely, S.E., Kuhn, P., Balch, W.E., and Wilson, I.A. (2000). A new functional domain of guanine nucleotide dissociation inhibitor (alpha-GDI) involved in Rab recycling. Traffic *1*, 270–281.

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.

Mariappan, M., Preusser-Kunze, A., Balleininger, M., Eiselt, N., Schmidt, B., Gande, S.L., Wenzel, D., Dierks, T., and von Figura, K. (2005). Expression, localization, structural and functional characterization of pFGE, the paralog of the Cα-formylglycine generating enzyme. J. Biol. Chem. *280*, 15173–15179.

Marquordt, C., Fang, Q., Will, E., Peng, J., von Figura, K., and Dierks, T. (2003). Posttranslational modification of serine to formylglycine in bacterial sulfatases. Recognition of the modification motif by the iron-sulfur protein AtsB. J. Biol. Chem. *278*, 2212–2218.

McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. *238*, 777–793.

Merritt, E.A., and Murphy, M.E.P. (1994). Raster3D Version 2.0—a program for photorealistic molecular graphics. Acta Crystallogr. *D50*, 869–873.

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.

Morimoto-Tomita, M., Uchimura, K., Werb, Z., Hemmerich, S., and Rosen, S.D. (2002). Cloning and characterization of two extracellular heparin-degrading endosulfatases in mice and humans. J. Biol. Chem. *277*, 49175–49185.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Poulos, T.L., Edwards, S.L., Wariishi, H., and Gold, M.H. (1993). Crystallographic refinement of lignin peroxidase at 2Å. J. Biol. Chem. *268*, 4429–4440.

Preusser-Kunze, A., Mariappan, M., Schmidt, B., Gande, S.L., Mutenda, K., Wenzel, D., von Figura, K., and Dierks, T. (2005). Molecular characterization of the human Cα-formylglycine generating enzyme. J. Biol. Chem. *280*, 14900–14910.

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, the conversion of cysteine to serinesemialdehyde in arylsulfatase from *Volvox carteri*. Eur. J. Biochem. *238*, 341–345.

Sheriff, S., Hendrickson, W.A., and Smith, J.L. (1987). Structure of myohemerythrin in the azidomet state at 1.7/1.3 Å resolution. J. Mol. Biol. *197*, 273–296.

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.

Tu, B.P., Ho-Schleyer, S.C., Travers, K.J., and Weissman, J.S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. Science *290*, 1571–1574.

Vranka, J.A., Sakai, L.Y., and Bachinger, H.P. (2004). Prolyl 3-hydroxylase *1*, enzyme characterization and identification of a novel family of enzymes. J. Biol. Chem. *279*, 23615–23621.

Xin, X., Mains, R.E., and Eipper, B.A. (2004). Monooxygenase X: a member of the Cu-dependent monooxygenase family localized to the endoplasmic reticulum. J. Biol. Chem. *279*, 48159–48167.

Yeh, J.I., Claiborne, A., and Hol, W.G. (1996). Structure of the native cysteine-sulfenic acid redox center of enterococcal NADH peroxidase refined at 2.8 Å resolution. Biochemistry *35*, 9951–9957.

Accession Numbers

The coordinates and structure factors have been deposited in the Protein Data Bank (accession codes 1Y1E, 1Y1F, 1Y1G, 1Y1H, 1Y1I, and 1Y1J).