



Published in final edited form as:

ACS Chem Biol. 2015 January 16; 10(1): 72–84. doi:10.1021/cb500897w.

## Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications

Mason J. Appel<sup>†</sup> and Carolyn R. Bertozzi<sup>†,‡,§,\*</sup>

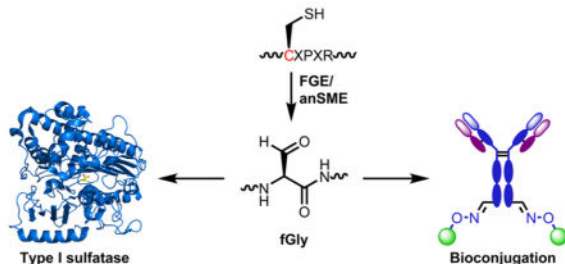
<sup>†</sup>Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, United States

<sup>‡</sup>Department of Chemistry, University of California, Berkeley, Berkeley, California 94720, United States

<sup>§</sup>Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, California 94720, United States

### Abstract

Formylglycine (fGly) is a catalytically essential residue found almost exclusively in the active sites of type I sulfatases. Formed by post-translational oxidation of cysteine or serine side chains, this aldehyde-functionalized residue participates in a unique and highly efficient catalytic mechanism for sulfate ester hydrolysis. The enzymes that produce fGly, formylglycine-generating enzyme (FGE) and anaerobic sulfatase-maturing enzyme (anSME), are as unique and specialized as fGly itself. FGE especially is structurally and mechanistically distinct, and serves the sole function of activating type I sulfatase targets. This review summarizes the current state of knowledge regarding the mechanism by which fGly contributes to sulfate ester hydrolysis, the molecular details of fGly biogenesis by FGE and anSME, and finally, recent biotechnology applications of fGly beyond its natural catalytic function.



### INTRODUCTION

Post-translational modification (PTM) of canonical amino acid side chains is a mechanism for augmenting the chemical diversity of enzymatic catalysis. Many cofactors involved in fundamental metabolic transformations derive from protein backbone or side chain modifications.<sup>1</sup> The novel functionalities created—redox moieties, electrophiles, or metal

\*Corresponding Author: crb@berkeley.edu.

The authors declare no competing financial interest.

chelators, for example—allow for catalytic mechanisms unattainable with canonical protein chemical groups. A relative newcomer to the family of PTM-derived catalytic cofactors is C $\alpha$ -formylglycine (fGly, Figure 1A), a modification of Cys or Ser side chains that is essential for the activity of type I sulfatases.<sup>2</sup> Unlike the complex pathways for generating many of the peptide-derived quinocofactors,<sup>3</sup> the biosynthesis of fGly is accomplished by only one enzymatic step. Astoundingly, type I sulfatases achieve rate enhancements ( $k_{\text{cat}}/k_{\text{uncat}}$ ) of 10<sup>26</sup> for alkylsulfate substrates, making them among the most efficient enzymes ever measured, and this can be attributed in part to the unique catalytic properties of fGly.<sup>4</sup>

Ubiquitous across all domains of life, sulfatases catalyze the hydrolysis of a vast array of natural and synthetic aryl- and alkylsulfate ester substrates. Three divergent classes of sulfatases have been identified, but the type I family members are the most common and the only class found in eukaryotes. In aerobic organisms, type I sulfatases become active when the formylglycine-generating enzyme (FGE) (also referred to as sulfatase-modifying factor 1, or SUMF1)<sup>5,6</sup> catalyzes the oxidation of cysteine to fGly. In humans, 17 sulfatases have been identified, of which 14 have been assigned specific activities in catabolism, signaling, and development.<sup>2,7</sup> Human sulfatases are initially translated into the endoplasmic reticulum (ER); some are retained there, while others are targeted to the lysosome, the Golgi, or the cell surface.<sup>7</sup> Lysosomal sulfatases act on sulfated glycolipids (sulfatides) and glycosaminoglycans, and their activities are necessary for proper degradation of these glycosides. ER-resident sulfatases, most notably steroid sulfatase (STS; arylsulfatase C, ASC), regulate hormone levels by desulfation of inactive precursors such as dehydroepiandrosterone 3-sulfate and iodothyronine sulfate.<sup>2</sup> Secreted sulfatases (Sulf1, Sulf2) modulate the sulfation level of cell-surface heparan sulfate, thereby regulating signaling events critical for development and tumor progression.<sup>8</sup> The disruption of individual sulfatases causes at least eight pathologies in humans, including six lysosomal storage disorders (e.g., mucopolysaccharidoses, metachromatic leukodystrophy), the bone disease chondrodysplasia punctata type 1, and skin disorder X-linked ichthyosis.<sup>9</sup> Deficiency in FGE causes multiple sulfatase deficiency (MSD), a fatal disorder marked by decreased activity of all sulfatases.<sup>10</sup> Microbial sulfatases were historically thought to be utilized for scavenging environmental sulfur, but a growing body of work over the past decade has revealed a much more elaborate role in modulating endosymbiont and host-pathogen interactions by remodeling host sulfation.<sup>11</sup> Given the breadth of research on sulfatase biology, we defer to a number of reviews for a thorough appraisal of sulfatase biochemistry and physiology,<sup>2,12–14</sup> the genetic basis of FGE and sulfatase disorders in humans,<sup>9,15,16</sup> and the pursuit of novel sulfatases for bioengineering applications.<sup>17</sup>

This review will focus the catalytic function of fGly and mechanisms by which enzymes from various organisms are thought to produce this PTM. Finally, we discuss the use of fGly's aldehyde functionality as a chemical handle for site-specific protein chemical modification, a biotechnology application of fGly that has undergone recent commercial translation.

## FGLY IS AN ESSENTIAL POSTTRANSLATION MODIFICATION OF TYPE I SULFATASES

Type I sulfatases are the predominant mediators of sulfate ester hydrolysis in all domains of life. They are abundant, highly conserved, and require the fGly PTM for catalysis.<sup>2</sup> Some sulfatases have been assigned defined biological substrates (e.g., sulfatases that act on the glycosaminoglycans chondroitin and heparan sulfate),<sup>16</sup> while most, particularly from microbial sources, have not been characterized at this level of biochemical detail. Thus, many type I sulfatases are annotated as arylsulfatases, simply because they catalyze the hydrolysis of colorimetric arylsulfate substrates *in vitro*.<sup>2</sup>

fGly is generated from a cysteine or serine residue embedded in a conserved active site-localized sequence motif, (C/ S)XPXRXXXLTGR; cysteine predominates as the modification target in eukaryotes and aerobic microbes whereas cysteine or serine is found in anaerobes.<sup>15,18</sup> The original discovery of fGly was made during biochemical analysis of the fatal congenital illness Multiple Sulfatase Deficiency (MSD).<sup>19</sup> The severe physiological defects associated with this condition reflected a combination of those associated with single sulfatase deficiencies.<sup>10</sup> Analysis of MSD-derived cell lines revealed diminished activity across the type I sulfatase family, despite the fact that protein levels of specific sulfatases appeared unaltered.<sup>10,20</sup> This observation led to speculation that MSD patients lacked a regulatory component or catalytic cofactor required for sulfatase activity. Importantly, cell-fusion complementation of MSD cells with those deficient in a single-sulfatase could restore activity of the deficient sulfatase, confirming the competence of individual MSD sulfatase genes.<sup>21</sup> The inability of MSD cells to produce active recombinant sulfatases likewise suggested a defect in a factor required for activity of several sulfatase gene products.<sup>22</sup>

Direct biochemical characterization of human arylsulfatases A and B (ASA and ASB) revealed a tryptic peptide carrying an unknown modification. At the site where a cysteine residue was predicted, mass spectrometry analysis indicated a novel amino acid that formed an adduct with glycerol and, when treated with NaBH<sub>4</sub>, yielded a serine residue. fGly was thus identified for the first time as a PTM.<sup>19</sup> Subsequently, MSD-derived sulfatases were shown to contain significantly lower amounts of fGly, but the complete loss of sulfatase activity is likely embryonic-lethal in humans.<sup>23</sup>

## FGLY PARTICIPATES DIRECTLY IN SULFATE ESTER HYDROLYSIS

The correlation of fGly deficiency with a loss of sulfatase activity implied a direct role for fGly in enzymatic activity. Absent further information, several modes of reactivity were proposed based on known chemistries of the aldehyde group its electrophilicity as well as the nucleophilicity of the geminal diol formed by aldehyde hydration (Figure 1A). Clarity came from the analysis of several sulfatase crystal structures which support a role for fGly in covalent catalysis.<sup>24</sup> The first sulfatase structure, that of human ASB, was solved by Guss and co-workers in 1997, and one year later the structure of ASA was solved by von Figura and coworkers.<sup>25,26</sup> A third human sulfatase structure, that of steroid sulfatase (ASC, STS), was later reported by Ghosh and coworkers.<sup>27</sup> Notably, despite relatively low sequence

homology, these sulfatases adopt a similar fold to alkaline phosphatases, consistent with their related catalytic functions.

The sulfatase structures resolved several uncertainties regarding fGly. First, the fGly geminal diol (fGly-diol) was detected in ASA, suggesting a role as a catalytic nucleophile. Electron density consistent with a sulfated fGly-diol, a putative intermediate in a covalent catalysis mechanism, was observed in both ASB and ASC.<sup>25,27</sup> However, it is important to note that the presence of an artifactual phosphate ester instead could not be ruled out. Further evidence came from the 1.3 Å resolution structure of the *P. aeruginosa* arylsulfatase (PAS), which clearly showed electron density for the fGly-diol at full occupancy (Figure 1B).<sup>28</sup> An earlier observation that fGly within native folded ASA and ASB is refractory to NaBH<sub>4</sub> reduction is consistent with the presence of the fGly-diol rather than the fGly-aldehyde in the resting enzyme.<sup>19</sup> In addition to fGly, the active sites of type I sulfatases contain a divalent cation, usually Ca<sup>2+</sup> but also Mg<sup>2+</sup>, and a side-chain hydrogen bonding network which coordinates fGly and substrate. The metal helps to bind and polarize the substrate, and the fGly-centered hydrogen bonding network provides for general acid–base catalysis. Three aspartates, one of which H-bonds with an fGly-diol hydroxyl, and a glutamine/asparagine constitute the metal-binding residues, leaving two open coordination sites for binding of sulfate and the second fGly-diol hydroxyl. Several acidic residues bind the substrate, and a histidine base is predicted to assist in catalysis. A concise comparison of the corresponding active site residues for seven (of eight total) fGly enzyme crystal structures is provided by Steinfeld and co-workers (Figure 1C).<sup>24</sup> The first structure of human sulfamidase (*N*-sulfoglucosamine sulfohydrolase, SGSH) was recently reported and overlays closely with previous sulfatases, especially in the active site region, despite low overall sequence identity (between 19 and 25% across fGly enzymes that have been structurally characterized).<sup>24</sup>

The presence of an active site fGly-diol suggested possible catalytic mechanisms (Figure 1D). Guss (ASB)<sup>25</sup> and von Figura (ASA)<sup>26</sup> both proposed that one of the hydroxyls of the fGly-diol acts as a nucleophile for initial attack at the substrate sulfur atom, releasing the alcohol coproduct and forming a sulfoenzyme intermediate. The details of their proposed pathways diverge thereafter, though both emphasize acid-base catalysis. In the latter case, deprotonation of the free remaining fGly-diol hydroxyl group by a conserved histidine would allow E2 elimination of sulfate, generating a transient fGly-aldehyde in a scheme abbreviated “S<sub>N</sub>2-E2.”<sup>26</sup> Guss and co-workers proposed instead that S<sub>N</sub>2 substitution at sulfur by hydroxide or water was responsible for the breakdown of the sulfoenzyme intermediate, which can be designated as the “S<sub>N</sub>2-S<sub>N</sub>2” pathway (Figure 1D).<sup>25</sup> Consistent with the unequal roles of each hydroxyl group in the geminal diol in either proposed mechanism, only the *S* stereoisomer of the transesterified enzyme has been observed in crystal structures of ASB and PAS.<sup>25,29</sup>

Intuitively, one may suppose that the biosynthetic cost of installing the unusual fGly PTM is more than repaid because fGly enables a mechanism unavailable to Ser/Cys/Thr nucleophiles. Replacing cysteine with serine in recombinant sulfatases prevents the conversion to fGly by eukaryotic FGEs and has provided an instructive probe to compare the two proposed mechanisms. In the case of ASA and ASB, the Cys-to-Ser mutant sulfatase

will bind substrate and perform the initial S-O cleavage to form the serine sulfoenzyme and release the alcohol coproduct. But, this adduct is a dead-end product that will not release sulfate even after extended time periods.<sup>30</sup> If the wild-type  $\alpha$ -hydroxy sulfate ester intermediate was resolved by an  $S_N2$ -type attack of water at sulfur, and not by E2 elimination, the serine mutant would be expected to retain at least partial activity. Thus, the E2 pathway is presently favored. Sulfoenzyme formation is detected in only about 20% of the Cys-to-Ser sulfatase mutant, however, and this complex is unable to form in crystals. These facts suggest that the fGly-diol contributes to accelerated enzyme sulfation as well, either by the enhanced nucleophilicity of the diol hydroxyl group compared to a primary alcohol or by specific hydrogen bonding features of the active site that are not replicated in the serine mutant.<sup>31</sup> Since the serine mutant affects both breakdown and formation of the sulfoenzyme intermediate, its behavior alone is not sufficient to prove the elimination mechanism.

Further chemical and theoretical analysis has yielded a framework for understanding the significance of fGly and its likely reaction pathway. First, the uncatalyzed hydrolysis of the S-O bond of sulfate esters proceeds  $10^6$ -fold slower than spontaneous hydrolysis of the P-O bond of phosphate esters.<sup>4</sup> The high barrier for the uncatalyzed reaction, with a  $t_{1/2} = 10^{18}$  years,<sup>4</sup> may explain the evolutionary significance of fGly's novel catalytic strategy. Alkaline phosphatase hydrolyzes phosphate esters through a phosphoserine intermediate that releases inorganic phosphate by attack of a  $Zn^{+}$ -activated hydroxide.<sup>32</sup> This catalytic paradigm may be inadequate for sulfate ester S-O cleavage, or was otherwise impractical or unavailable to sulfatases during evolution, thus requiring the specialized functionality of fGly. Of the two proposed fGly catalyzed pathways, the  $S_N2$ -E2 mechanism requires only one high energy barrier S-O bond cleavage. The rate of the initial  $S_N2$  step may be accelerated by the decreased  $pK_a$  of the geminal diol hydroxyl group (e.g., acetaldehyde hydrate has a  $pK_a$  of 13.57<sup>33</sup>) compared to a conventional hydroxyl group (e.g., 15.93<sup>34</sup>). A recent theoretical study has suggested that the initial nucleophilic attack of fGly-diol on substrate may be rate-determining, and not the subsequent elimination.<sup>35</sup> Perhaps inconsistent with this prediction, a measured  $\beta_{LG}$  of 0 for  $V_{max}$  but  $-0.86$  for  $V_{max}/K_M$  for PAS suggests that the alkoxide leaving group is not involved in the rate-determining step but only the first irreversible chemical step.<sup>36</sup> For the desulfation step, hydrolysis of the sulfoenzyme intermediate by water in the  $S_N2$  pathway may be quite slow compared to E2, again for the reason for S-O bond cleavage. Experimentally, pre-steady-state kinetics of the wild-type enzyme may permit the comparison of the relative sulfation/desulfation rates. Stereochemical characterization of reaction products can also provide valuable clues to the mechanism. In principle, the  $S_N2$ -E2 mechanism would result in net inversion at sulfate while  $S_N2$ - $S_N2$  would result in net retention. In an early study, prior to the discovery of fGly, interception of the sulfoenzyme intermediate generated using a chiral sulfate ester with an exogenous nucleophile produced a new sulfate ester with net retention of configuration.<sup>37</sup> This observation is consistent with an initial  $S_N2$ -like step in the sulfatase mechanism but does not inform the course of desulfation.

Other determinants of reactivity may explain the proposed preference of sulfatases for an E2 release pathway, and the differential specificity of sulfatases and phosphatases in general. A

recent experimental and theoretical study by Krenske and coworkers affirms earlier theoretical studies,<sup>35</sup> which predict a much lower transition state barrier for E2 elimination compared to S<sub>N</sub>2 substitution to release sulfate (estimated  $H^\ddagger$  of 26.5 kcal/mol).<sup>36</sup> This study estimates a  $H^\ddagger$  of only 12 kcal/mol for the corresponding phosphoenzyme intermediate, which may account for why fGly is found mostly in sulfatases, and not in alkaline phosphatase. Wild-type ASA is also inhibited by a prohibitively slow rate of elimination/hydrolysis of the corresponding phosphoenzyme intermediate at fGly.<sup>29</sup> This is not an inherent characteristic of fGly enzymes, however. PAS can catalyze arylphosphate hydrolysis at a lower catalytic efficiency ( $k_{\text{cat}}/K_M$ ) compared to arylsulfates ( $7.9 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$  vs  $4.9 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ ), but with dramatic rate acceleration ( $k_{\text{cat}}/k_{\text{uncat}}$ ) and catalytic proficiency ( $(k_{\text{cat}}/K_M)/k_{\text{uncat}}$ ) compared to the uncatalyzed hydrolysis of arylphosphates; PAS is predicted to release phosphate by elimination as with sulfate.<sup>35,38,39</sup> In the opposing case of promiscuous sulfatase activity in alkaline phosphatase, the source of large catalytic discrimination for phosphate esters over sulfate esters appears to be complex and only partially related to the difference in formal charge between the two substrates.<sup>40</sup> Notably, a family of type I sulfatase/alkaline phosphatase related fGly-dependent enzymes has been recently discovered and structurally characterized. It includes a phosphatase/phosphonatease from *Rhizobium leguminosarum* termed *RIPMH*<sup>41</sup> and an orthologous but highly promiscuous phosphate/phosphonate/sulfate/sulfonate ester hydrolase from *Burkholderia caryophylli* (*BcPMH*).<sup>42</sup> *RIPMH* has a  $>10^3$ -fold preference for phosphonate monoesters/phosphate diesters over sulfate esters and displays burst-phase kinetics, indicating that formation of the phosphoenzyme intermediate is not rate-limiting.<sup>41</sup> This enzyme family reflects the apparent catalytic versatility of fGly as tuned by protein context, as well as the shared ancestry of type I sulfatases and alkaline phosphatases.

It is important to note that two other unrelated sulfatase families exist in some prokaryotes, but unlike type I, they proceed with concomitant oxidation of the alcohol coproduct to an aldehyde (type II), or stereochemical inversion of the alcohol coproduct (type III).<sup>17</sup> It is clear that such changes would not be suitable for the signaling or catabolic roles of eukaryotic sulfatases. Though further experimental evidence is needed to clarify remaining mechanistic uncertainties, it is clear that type I sulfatases are among the most extraordinary biological catalysts known, with rate enhancements ( $k_{\text{cat}}/k_{\text{uncat}}$ ) as high as  $10^{26}$ , and undoubtedly warrant further studies.<sup>4</sup>

## FGE CATALYZES FGLY FORMATION AND IS DYSFUNCTIONAL IN MSD

In 2003, the gene encoding FGE (also termed sulfatase-modifying factor 1 or *SUMF1*) was identified as responsible for MSD by Ballabio et al.<sup>6</sup> and von Figura et al.,<sup>5</sup> using genetic and biochemical approaches, respectively. Von Figura and coworkers exploited FGE's ability to bind tightly but not turn over a peptide containing its recognition motif with serine substituted for cysteine (SXPXR). By affinity enrichment of bovine testis extract, FGE was thus purified for the first time.<sup>5</sup> Investigation of human and bacterial FGEs has revealed a new enzyme family defined by both a novel catalytic mechanism and protein fold. Mapping the mutations in *SUMF1* alleles from MSD patients onto the hFGE crystal structure, as well as *in vivo* and *in vitro* dissection of mutant phenotypes, has provided a rationale for their

deleterious effects. MSD mutations result in structural, catalytic, and intracellular stability defects.<sup>43–45</sup>

Prior to isolation of FGE, fGly formation was studied using intact microsomes and later with reticuloplasm, using an *in vitro* transcription/translation system.<sup>46,47</sup> These studies indicated that fGly generation occurs cotranslationally, and prior to folding. These initial efforts suggested a dependence on thiol reductants, Ca<sup>2+</sup> ions, and an alkaline pH optimum.<sup>48</sup> Subsequently, it was observed that FGE activity is dependent on O<sub>2</sub>, presumably as a terminal electron acceptor (cf. Figure 3C).<sup>49</sup> Interestingly, O<sub>2</sub> availability can limit sulfatase activity *in vivo* by decreasing FGE activity, and the decrease of sulfatase activity levels can mediate hypoxic response.<sup>50</sup>

## MAMMALIAN FGE ACTIVITY IS REGULATED BY PROTEIN PARTNERS

Regulation of mammalian FGE activity occurs spatially and is mediated by protein–protein interactions in the ER and Golgi. Recent work has identified functional interactions with several proteins in the ER and along the secretory pathway including ERp44, PDI, ERGIC-53, and pFGE among others.<sup>51–54</sup> FGE lacks a canonical ER retention sequence; instead, the N-terminal region of FGE (residues 34–88 following signal peptide cleavage) forms both noncovalent interactions and a disulfide linkage with ERp44, a trafficking protein that is the primary means for retrieval of FGE from the Golgi to the ER. In fact, this N-terminal region is also involved in activation of sulfatases by an unknown (ER-retention-independent) mechanism. The region contains a cysteine pair (Cys50 and Cys52) that forms intermolecular disulfide bonds both with Erp44 and with other FGE molecules. Only Cys52 is necessary for FGE activity *in vivo*, but not *in vitro*,<sup>54</sup> while both cysteines are dispensable for retrieval by ERp44 through noncovalent complex formation,<sup>51</sup> though an alternate mutational study disputes this finding (Figure 2).<sup>52</sup>

Protein disulfide isomerase (PDI) is a well-known ER-resident chaperone that allows nascent polypeptides to reach their preferred disulfide configuration by catalyzing fast exchange with its own redox active disulfide.<sup>55</sup> PDI appears to be involved both in retaining FGE in the ER and in promoting activity, and these functions are dependent on catalytically active PDI (Figure 2B).<sup>52</sup> ERGIC-53 is an anterograde cargo binding protein which cycles between the ER and the Golgi.<sup>52</sup> It functions to traffic FGE from the ER to the cis-Golgi and also seems to protect it from proteasomal degradation (Figure 2C). Once in the cis-Golgi, most FGE is retrieved to the ER via ERp44, with the remaining fraction secreted from the cell.<sup>52</sup> It is possible that competition with other ERp44 substrates may regulate FGE secretion by this process. Downstream in the secretory pathway, a fraction of FGE is N-terminally truncated and inactivated by furin and furin-like proteases (Figure 2D). Approximately 20–30% of secreted FGE avoids proteolytic cleavage and remains active.<sup>56</sup> Secreted FGE can traffic back to the ER from the cell surface. Indeed, tissue-specific expression of FGE in FGE-null mice has been shown to activate sulfatases in nontransduced tissues, in this way acting as a paracrine agent. Uptake is dependent in part on the mannose receptor and FGE N-glycosylation (Figure 2D–F).<sup>57</sup> Taken together, a model for the localization and activity of FGE has emerged (Figure 2). However, it remains of great

interest to understand the significance of FGE secretion and reuptake to the ER, as this offers a route for enzyme replacement therapy for treatment of MSD.

Finally, the paralog of FGE (pFGE, encoded by the gene *SUMF2*) was identified by homology shortly after SUMF1 was identified, and its precise role is the subject of ongoing investigation.<sup>6</sup> pFGE displays strong sequence and structural conservation (48% identity) with hFGE, is localized to the ER, and is transcribed at comparable levels to FGE in human tissues, but pFGE lacks FGE activity and is different by several features.<sup>58–60</sup> pFGE lacks the active site cysteine pair and functionally critical N-terminal region of FGE. Instead, pFGE has a C-terminal PGEL sequence, which is a noncanonical variant of the ER retention signal KDEL. Based on coimmunoprecipitation data, pFGE forms heterodimers and higher order oligomers with FGE and also interacts with sulfatases, though direct physical evidence for pFGE interactions has remained elusive (Figure 2G). Curiously, pFGE overexpression decreases sulfatase activation *in vivo* in a dose-dependent manner but does not increase FGE retention in the ER. The meaning of these results is unclear, but it is possible that in appropriate stoichiometry, pFGE may act as a sulfatase-specific chaperone or otherwise mediate the interaction of FGE with sulfatases.<sup>58,60</sup>

## FGE DEFINES A NOVEL PROTEIN FOLD

The first series of crystal structures of human FGE were reported in 2005 by Dierks et al., and FGE was shown to adopt a novel fold, the FGE-fold (Figure 3A).<sup>43,61</sup> Subsequently, the crystal structure of the *S. coelicolor* FGE (scFGE) was solved in our laboratory and overlays closely with hFGE, with an rmsd of 0.65 Å.<sup>49</sup> The FGE fold is defined by one domain of low secondary structure content, approximately 30% combined  $\beta$ -sheet and  $\alpha$ -helix in both orthologs. hFGE binds two Ca<sup>2+</sup> ions and scFGE just one; two structural disulfides and an *N*-glycosite are particular to the human enzyme. A cysteine pair (C336 and C341), shown to be catalytically essential and separated by four residues in a loop, exists in a distribution of the reduced (dithiol, FGE<sub>red</sub>) and oxidized (disulfide, FGE<sub>ox</sub>) forms based on both crystal structures and by mass spectrometry (Figure 3A).<sup>62,43</sup> A cocrystal structure of hFGE C336S with LCTPSRA substrate was obtained by Roeser et al. in 2006 and defined a shallow substrate binding channel which can bind up to six substrate residues in the consensus peptide sequence in an extended conformation (Figure 3B).<sup>63</sup> Additional structures support the existence of a probable O<sub>2</sub> binding pocket adjacent to the active site cysteines. The use of halide bound hFGE crystal structures to define the size and environment of this pocket predict that it could accommodate a negatively charged oxidative intermediate in complex with substrate.<sup>64</sup> Homologues of FGE have been found in a class of bacteria and bacteriophage diversity-generating retroelement (DGR) variable proteins, which generate adaptive protein variation through an RNA intermediate.<sup>65</sup> With an rmsd of 0.98 Å between the catalytic region of FGE and the *Treponema denticola* TvpA variable region, the FGE-fold has since been classified as a subtype of the C-type lectin fold.<sup>66</sup>

## FGE IS A UNIQUE CYSTEINE-DEPENDENT OXIDASE/OXYGENASE

Many biochemical transformations in both primary and secondary metabolism across aerobic organisms harness the reduction potential of O<sub>2</sub>. Reduction of molecular oxygen is



thermodynamically favorable, but there is a high barrier to its reaction with singlet state organic compounds due to its triplet ground state.<sup>67</sup> Most oxidases/oxygenases require reductants of sufficient strength, typically paramagnetic transition metals (mainly Fe<sup>2+</sup> and Cu<sup>2+</sup>), and/or organic cofactors (flavins, pterins, quinones) which generate a stabilized radical following single electron transfer.<sup>68</sup> Subsequent one or two electron transfer events are very favorable and rapid, producing potent O<sub>2</sub>-derived oxidants that may add in to substrate (oxygenase) or become further reduced to H<sub>2</sub>O or H<sub>2</sub>O<sub>2</sub> (oxidase). Cofactorless oxidases/oxygenases have also been identified. They typically promote the direct oxidation of substrates, without the intermediacy of a metal or cofactor, by targeting O<sub>2</sub> to a stabilized carbanion at a specific position of an aryl substrate in a manner comparable to flavin reactivity.<sup>69</sup>

FGE does not possess a canonical metal or organic redox cofactor and thus has been categorized as a cofactorless oxidase/ oxygenase. But, there is no evidence that FGE utilizes any variation of the strategies above to catalyze reaction with O<sub>2</sub>. Instead, mechanistic proposals center on the essential redox-active C336/C341 pair located in the active site. The lability of these cysteines to irreversible oxidation by air or H<sub>2</sub>O<sub>2</sub>, creating cysteine sulfonic acids, suggests that they may mediate reaction of O<sub>2</sub> with the substrate cysteine side chain.<sup>49,43</sup> Mutation of either of these cysteines abrogates activity, and they are positioned closely to the substrate cysteine in a cocrystal structure where Cys336 (hFGE) is mutated to serine (Figure 3B).<sup>63</sup> The isolation of this complex, and the observation of partially oxidized cysteine side chains in various crystal structures, formed the basis of the proposed mechanism shown in Figure 3D. This scheme invokes a disulfide bonded form of FGE (FGE<sub>ox</sub>) as the enzyme's resting state. Substrate binding is then followed by disulfide exchange, forming a covalent mixed disulfide bond between enzyme and substrate. Direct reaction of a free cysteine side chain at C336 with O<sub>2</sub> is proposed to generate a reactive intermediate, a cysteine oxoform presumably, that then targets the substrate cysteine side chain.

In both the hFGE and scFGE apo crystal structures, Cys336 (hFGE numbering) displays additional electron density that can be modeled as a sulfenic acid, peroxy-sulfenic acid, free peroxide, or water. Ser333 is likely critical for catalysis due to its role in coordinating such an intermediate. A substrate cysteine sulfenic acid may be generated through the transfer of oxidative species from one of the above reactive derivatives. Elimination of H<sub>2</sub>O from sulfenic acid would then generate a thioaldehyde intermediate, which would hydrolyze to generate fGly and H<sub>2</sub>S (Figure 3D). Notably, *in vitro* biochemical studies have shown that an external reductant is required to achieve high substrate turnover (e.g., GSH or DTT).<sup>48</sup> Thus, although FGE activates O<sub>2</sub> in a cofactor-independent manner, a reductant may be required to regenerate the enzyme's active state during each catalytic cycle, or at an intermediate stage.<sup>69</sup> It is possible that hFGE relies on PDI as a reducing agent *in vivo*, but there is no direct evidence for this.<sup>52</sup>

## ANSME CATALYZES O<sub>2</sub>-INDEPENDENT fGLY FORMATION ON SER/CYS SULFATASES IN ANAEROBIC MICROBES

Given fGly's impressive catalytic capability, it is not surprising that an O<sub>2</sub>-independent fGly generating system also exists in anaerobic and facultative anaerobic microbes. A family of radical S-adenosylmethionine (RS)-dependent dehydrogenases has been identified to fill this role. AtsB from *Klebsiella pneumoniae* was identified as an fGly generating enzyme by Dierks and coworkers in 1999 and was previously shown to be required for high activity of recombinant *Klebsiella* sulfatase in *E. coli*.<sup>70,71</sup> Curiously, the native substrate of this enzyme contains Ser and not Cys in a recognition motif that is otherwise identical to the FGE-dependent sulfatase sequences. Based on biochemical and taxonomic evidence, the discovery of a Cys-sulfatase-specific ortholog from *Clostridium perf ringens* (cpe) led to the designation of a new enzyme family, the anaerobic sulfatase-maturing enzymes (anSME).<sup>18</sup> anSMEcpe can also oxidize Ser-type substrates,<sup>72</sup> suggesting a similar reaction pathway for both targeted side chains.<sup>73</sup> AtsB, whose native substrates are Ser-type, is 4-fold more active with cysteine-based substrates *in vitro*. This preference for Cys may originate from increased thiol compared to alcohol reactivity at one or more proposed elementary steps. Spectroscopic characterization revealed that anSME, in addition to the canonical RS [4Fe-4S] cluster, contains two auxiliary [4Fe-4S] clusters as well (Aux I and Aux II).<sup>74</sup> Drennan and coworkers recently reported the first crystal structure of anSMEcpe with and without bound substrate (1.6–1.8 Å resolution), showing both Aux clusters in the C-terminal SPASM domain with fully occupied coordination sites, unlike the RS cluster which contains an open Fe-binding site for substrate.<sup>75</sup>

Mechanistic investigation of anSMEs has revealed many details of this reaction (Figure 4). As with all RS enzymes, a canonical [4Fe-4S]<sup>+</sup> cluster reductively cleaves S-adenosylmethionine (SAM) to generate Met and the 5'-deoxyadenosine radical (5'-dA<sup>•</sup>). Initial H<sup>•</sup> abstraction by 5'-dA<sup>•</sup>, as evidenced by substrate deuterium labeling and stereospecific C<sub>β</sub>-methyl substitution, is performed at the pro-S C<sub>β</sub>-H of Cys/Ser in the consensus sequence.<sup>73,76</sup> Deprotonation of the thiol/alcohol C<sub>β</sub>radical by an asparate side chain and charge donation into the radical promotes the second substrate oxidation step. The second one-electron oxidation would yield a thioaldehyde intermediate in the case of cysteine, which would hydrolyze to fGly, or would yield fGly directly in the case of serine. Based on the orientation of Aux I and Aux II in relation to the RS cluster, Aux I is the likely electron acceptor for the second oxidation step. Aux II is proximal to bulk solvent and may receive this electron from Aux I and donate it to a soluble carrier. This pathway accounts for the multiple turnover ability of anSME in the presence of flavodoxin without external reductants, meaning that recycling of the second substrate electron to reduce the RS cluster cannot occur directly from AuxI/II.<sup>75</sup> Further experimentation will clarify the role of each Aux cluster. Despite the strong conservation between substrates of anSME and FGE, the way each enzyme binds substrate is distinct. FGE-substrate interaction hinges on a hydrophobic pocket for Pro binding and several hydrogen bond partners to Arg (in CXPXR), with only 4 of 12 H-bonds to the substrate backbone.<sup>63</sup> However, anSME binds substrate primarily through the substrate backbone, accounting for 14 of 17 H-bonds; this

explains the existence of alternative native anSME substrates containing CXAXR without Pro, though Arg side-chain binding is maintained.<sup>75</sup>

## SOME ORGANISMS HAVE AS YET UNIDENTIFIED FGLY PRODUCING SYSTEMS

Yet a third means for producing fGly has been observed, with the responsible enzyme(s) remaining elusive. Recombinant expression of type I sulfatases in *E. coli* has been shown to produce active enzymes with varying proportions of fGly modification.<sup>77</sup> Berteau and co-workers have shown that a putative *E. coli* anSME ortholog was unnecessary for the conversion of recombinant sulfatases, and that an unidentified, O<sub>2</sub>-dependent, Cys-type fGly-generating enzyme must be responsible for installing the aldehyde. According to sequence data available at the time, as of 2007, approximately 10% of bacterial genomes were predicted to encode arylsulfatase genes without the co-occurrence of obvious FGE or anSME orthologs.<sup>78</sup> This is also true for several fungi, and even for the metazoan *Caenorhabditis elegans*, which contains three sulfatase genes.<sup>12,79</sup> *C. elegans* has been shown to utilize sulfatases for heparan sulfate remodeling, as in humans.<sup>80</sup> Thus, it is likely that these organisms all carry a related system for fGly production with important functional consequences. We believe that identification of the putative third type of fGly-generating enzyme may reveal new aspects of fGly biology.

## FGLY PROVIDES A GENETICALLY ENCODED, SITE-SPECIFIC PROTEIN BIOCONJUGATION HANDLE

During characterization of the sequence determinants of cotranslational FGE activity, it was reported that the minimal five-residue motif, CXPXR, is efficiently converted when embedded in an otherwise nonnative sequence.<sup>81</sup> This observation led us to speculate that the CXPXR sequence could be exploited as a means to genetically encode fGly residues in various proteins of interest. The aldehyde group, in addition to its unique catalytic nature within sulfatases, is also a reactive electrophile with orthogonal reactivity to canonical protein side chain functionalities. We proposed that classical carbonyl-based bioconjugation chemistries, such as oxime and hydrazine formation, could be used to covalently attach various cargo to such “aldehyde tagged” proteins in a site-specific manner. The use of FGE and its five-residue motif to arm recombinant proteins with aldehyde “chemical handles” was first reduced to practice in *E. coli* expression systems.<sup>82</sup> The limited endogenous FGE-like activity in *E. coli* was found to be saturated by overexpressed proteins, necessitating coexpression of ectopic FGE along with the protein of interest in order to achieve high yields of Cys-to-fGly conversion. The *M. tuberculosis* FGE was particularly useful in this regard, as the enzyme was found to have relaxed sequence specificity and tolerated single alanine substitutions within the CXPXR motif.<sup>83</sup> Subsequently, the method was used to produce aldehyde tagged recombinant proteins in mammalian cell expression systems. As eukaryotic FGE is naturally ER localized, membrane-associated and secreted proteins, both of which traffic through the secretory pathway, were found to be viable targets for aldehyde tagging.<sup>84,85</sup>

As a site-specific protein modification method, the aldehyde tag has many advantages compared to other options.<sup>86,87</sup> It is operationally simple, requiring simple cloning of the CXPXR motif at the desired modification site and expression of the protein of interest in FGE-expressing cells. Once installed, the aldehyde can be reacted with aminoxy or hydrazide reagents to form the corresponding oxime and hydrazone conjugates, respectively (Figure 5A). This straightforward process has been used to site-specifically PEGylate proteins,<sup>82</sup> to introduce biophysical probes at discrete locations for single molecule studies of protein dynamics,<sup>88,89</sup> to immobilize proteins and phage on solid materials,<sup>90–92</sup> to construct artificially glycosylated proteins,<sup>93,94</sup> and to generate DNA–protein conjugates.<sup>95</sup> The aldehyde tag method has also been used in conjunction with other bioconjugation reactions to generate chemically fused protein heterodimers (Figure 5B).<sup>96</sup>

The biopharma industry has a growing interest in chemically modified protein therapeutics. Antibody-drug conjugates (ADCs) in particular are highly represented in biopharma pipelines, largely for cancer indications.<sup>97</sup> There is increasing awareness of the benefits of generating such constructs via site-specific attachment of drug molecules to the antibody scaffold.<sup>98,99</sup> This trend presents an opportunity for clinical applications of fGly-functionalized proteins. However, critical to the therapeutic performance of such constructs is stability of the drug–protein linkage. Oximes and hydrazones can undergo reversible hydrolysis at physiologically relevant pHs, an inherent liability to their use in ADCs. To address this issue, we developed a new C–C bond forming reaction of protein-associated fGly residues with alkoxyamine-functionalized indoles (**1**, Figure 5A), an adaptation of the classic Pictet-Spengler reaction that we termed the Pictet-Spengler ligation.<sup>100</sup> A second-generation version, the hydrazino Pictet-Spengler ligation (**2**, Figure 5A) proceeds readily at neutral pH to give stable fGly conjugates.<sup>101</sup> This chemistry has recently been employed to construct homogeneous ADCs comprising the antimitotic natural product maytansine linked to the anti-Her2 antibody trastuzumab (Figure 5C).<sup>102</sup> The use of fGly as a chemical handle for these ADCs enabled precise control over the site and stoichiometry of drug conjugation, thereby facilitating drug optimization.

## CONCLUSION

Many questions remain regarding the molecular details of fGly biogenesis and its role in catalysis. But in the two decades since fGly was first discovered, sulfatases and the enzymes that activate them have become widely appreciated as lessons in enzyme evolution and catalytic strategy. As well, the orthogonal reactivity of the aldehyde group with respect to canonical protein functionality has enabled biotechnology applications of fGly that are approaching clinical translation. The scope of FGE as a tool for protein engineering has yet to be fully explored. Interesting questions abound, such as the amenability of the enzyme to targeted engineering for enhanced performance or altered substrate specificity. As well, the intriguing discovery of fGly outside the type I sulfatase family as required for the phosphatase/phosphonate *R*/PMH and multisubstrate hydro-lase *Bc*PMH hints that fGly could be more widespread across microbial proteomes. And the search for additional FGE-like enzymes is far from over. Organisms as familiar as *E. coli* and *C. elegans* generate fGly on recombinant proteins and produce fGly-dependent sulfatases respectively, but their FGE-

like machineries remain undefined. We anticipate a rich future for fundamental and applied studies of this small but powerful PTM.

## Acknowledgments

We thank J. Klinman, D. Fox, and P. Agarwal for critical reading of the manuscript.

## References

1. Okeley NM, van der Donk WA. Novel cofactors via post-translational modifications of enzyme active sites. *Chem Biol.* 2000; 7:R159–R171. [PubMed: 10903941]
2. Hanson SR, Best MD, Wong CH. Sulfatases: Structure, Mechanism, Biological Activity, Inhibition, and Synthetic Utility. *Angew Chem, Int Ed.* 2004; 43:5736–5763.
3. Klinman JP, Bonnot F. Intrigues and Intricacies of the Biosynthetic Pathways for the Enzymatic Quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ. *Chem Rev.* 2013; 114:4343–4365. [PubMed: 24350630]
4. Edwards DR, Lohman DC, Wolfenden R. Catalytic Proficiency: The Extreme Case of S–O Cleaving Sulfatases. *J Am Chem Soc.* 2011; 134:525–531. [PubMed: 22087808]
5. Dierks T, Schmidt B, Borissenko LV, Peng J, Preusser A, Mariappan M, von Figura K. Multiple Sulfatase Deficiency Is Caused by Mutations in the Gene Encoding the Human C $\alpha$ -Formylglycine Generating Enzyme. *Cell.* 2003; 113:435–444. [PubMed: 12757705]
6. Cosma MP, Pepe S, Annunziata I, Newbold RF, Grompe M, Parenti G, Ballabio A. The Multiple Sulfatase Deficiency Gene Encodes an Essential and Limiting Factor for the Activity of Sulfatases. *Cell.* 2003; 113:445–456. [PubMed: 12757706]
7. Wiegmann EM, Westendorf E, Kalus I, Pringle TH, Lübke T, Dierks T. Arylsulfatase K, a Novel Lysosomal Sulfatase. *J Biol Chem.* 2013; 288:30019–30028. [PubMed: 23986440]
8. Buono M, Visigalli I, Bergamasco R, Biffi A, Cosma MP. Sulfatase modifying factor 1–mediated fibroblast growth factor signaling primes hematopoietic multilineage development. *J Exp Med.* 2010; 207:1647–1660. [PubMed: 20643830]
9. Diez-Roux G, Ballabio A. Sulfatases and Human Disease. *Annu Rev Genomics Hum Genet.* 2005; 6:355–379. [PubMed: 16124866]
10. Austin JH. Studies in metachromatic leukodystrophy: Xii. multiple sulfatase deficiency. *Arch Neurol.* 1973; 28:258–264. [PubMed: 4265903]
11. Ulmer JE, Vilén EM, Namburi RB, Benjdia A, Beneteau J, Malleron A, Bonnaffé D, Driguez PA, Descroix K, Lassalle G, Le Narvor C, Sandström C, Spillmann D, Berteau O. Characterization of Glycosaminoglycan (GAG) Sulfatases from the Human Gut Symbiont *Bacteroides thetaiotaomicron* Reveals the First GAG-specific Bacterial Endosulfatase. *J Biol Chem.* 2014; 289:24289–24303. [PubMed: 25002587]
12. Sardiello M, Annunziata I, Roma G, Ballabio A. Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship. *Hum Mol Genet.* 2005; 14:3203–3217. [PubMed: 16174644]
13. Bojarová P, Williams SJ. Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. *Curr Opin Chem Biol.* 2008; 12:573–581. [PubMed: 18625336]
14. Ghosh D. Human sulfatases: A structural perspective to catalysis. *Cell Mol Life Sci.* 2007; 64:2013–2022. [PubMed: 17558559]
15. von Figura K, Schmidt B, Selmer T, Dierks T. A novel protein modification generating an aldehyde group in sulfatases: its role in catalysis and disease. *BioEssays.* 1998; 20:505–510. [PubMed: 9699462]
16. Dierks T, Schlotawa L, Frese MA, Radhakrishnan K, von Figura K, Schmidt B. Molecular basis of multiple sulfatase deficiency, mucopolipidosis II/III and Niemann-Pick C1 disease -Lysosomal storage disorders caused by defects of non-lysosomal proteins. *Biochim Biophys Acta.* 2009; 1793:710–725. [PubMed: 19124046]

17. Toesch M, Schober M, Faber K. Microbial alkyl- and aryl-sulfatases: mechanism, occurrence, screening and stereo-selectivities. *Appl Microbiol Biotechnol*. 2014; 98:1485–1496. [PubMed: 24352732]
18. Berteau O, Guillot A, Benjdia A, Rabot S. A New Type of Bacterial Sulfatase Reveals a Novel Maturation Pathway in Prokaryotes. *J Biol Chem*. 2006; 281:22464–22470. [PubMed: 16766528]
19. Schmidt B, Selmer T, Ingendoh A, Von Figura K. A novel amino-acid modification in sulfatases that is defective in Multiple Sulfatase Deficiency. *Cell*. 1995; 82:271–278. [PubMed: 7628016]
20. Fluharty AL, Stevens RL, Davis LL, Shapiro LJ, Kihara H. Presence of arylsulfatase A (ARS A) in multiple sulfatase deficiency disorder fibroblasts. *Am J Hum Genet*. 1978; 30:249–255. [PubMed: 27985]
21. Horwitz AL. Genetic complementation studies of multiple sulfatase deficiency. *Proc Natl Acad Sci U S A*. 1979; 76:6496–6499. [PubMed: 118467]
22. Rommerskirch W, von Figura K. Multiple sulfatase deficiency: catalytically inactive sulfatases are expressed from retrovirally introduced sulfatase cDNAs. *Proc Natl Acad Sci U S A*. 1992; 89:2561–2565. [PubMed: 1348358]
23. Schlotawa L, Ennemann EC, Radhakrishnan K, Schmidt B, Chakrapani A, Christen HJ, Moser H, Steinmann B, Dierks T, Gartner J. SUMF1 mutations affecting stability and activity of formylglycine generating enzyme predict clinical outcome in multiple sulfatase deficiency. *Eur J Hum Genet*. 2011; 19:253–261. [PubMed: 21224894]
24. Sidhu NS, Schreiber K, Propper K, Becker S, Uson I, Sheldrick GM, Gartner J, Kratzner R, Steinfeld R. Structure of sulfamidase provides insight into the molecular pathology of mucopolysaccharidosis IIIA. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2014; 70:1321–1335. [PubMed: 24816101]
25. Bond CS, Clements PR, Ashby SJ, Collyer CA, Harrop SJ, Hopwood JJ, Guss JM. Structure of a human lysosomal sulfatase. *Structure*. 1997; 5:277–289. [PubMed: 9032078]
26. Lukatela G, Krauss N, Theis K, Selmer T, Gieselmann V, von Figura K, Saenger W. 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*. 1998; 37:3654–3664. [PubMed: 9521684]
27. Hernandez-Guzman FG, Higashiyama T, Pangborn W, Osawa Y, Ghosh D. Structure of Human Estrone Sulfatase Suggests Functional Roles of Membrane Association. *J Biol Chem*. 2003; 278:22989–22997. [PubMed: 12657638]
28. Boltes I, Czapinska H, Kahnert A, von Bülow R, Dierks T, Schmidt B, von Figura K, Kertesz MA, Usón I. 1.3 Å Structure of Arylsulfatase from *Pseudomonas aeruginosa* Establishes the Catalytic Mechanism of Sulfate Ester Cleavage in the Sulfatase Family. *Structure*. 2001; 9:483–491. [PubMed: 11435113]
29. Chruszcz M, Laidler P, Monkiewicz M, Ortlund E, Lebioda L, Lewinski K. Crystal structure of a covalent intermediate of endogenous human arylsulfatase A. *J Inorg Biochem*. 2003; 96:386–392. [PubMed: 12888274]
30. Recksiek M, Selmer T, Dierks T, Schmidt B, von Figura K. Sulfatases, Trapping of the Sulfated Enzyme Intermediate by Substituting the Active Site Formylglycine. *J Biol Chem*. 1998; 273:6096–6103. [PubMed: 9497327]
31. von Bulow R, Schmidt B, Dierks T, von Figura K, Uson I. Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis. *J Mol Biol*. 2001; 305:269–277. [PubMed: 11124905]
32. Cleland WW, Hengge AC. Enzymatic Mechanisms of Phosphate and Sulfate Transfer. *Chem Rev*. 2006; 106:3252–3278. [PubMed: 16895327]
33. Bell RP, Onwood DP. Acid strengths of the hydrates of formaldehyde, acetaldehyde and chloral. *Trans Faraday Soc*. 1962; 58:1557–1561.
34. Murto J. Nucleophilic Reactivity of Alkoxide Ions toward 2,4-Dinitro-fluorobenzene and the Acidity of Alcohols. *Acta Chem Scand*. 1964; 18:1043–1053.
35. Marino T, Russo N, Toscano M. Catalytic Mechanism of the Arylsulfatase Promiscuous Enzyme from *Pseudomonas Aeruginosa*. *Chem—Eur J*. 2013; 19:2185–2192. [PubMed: 23280779]

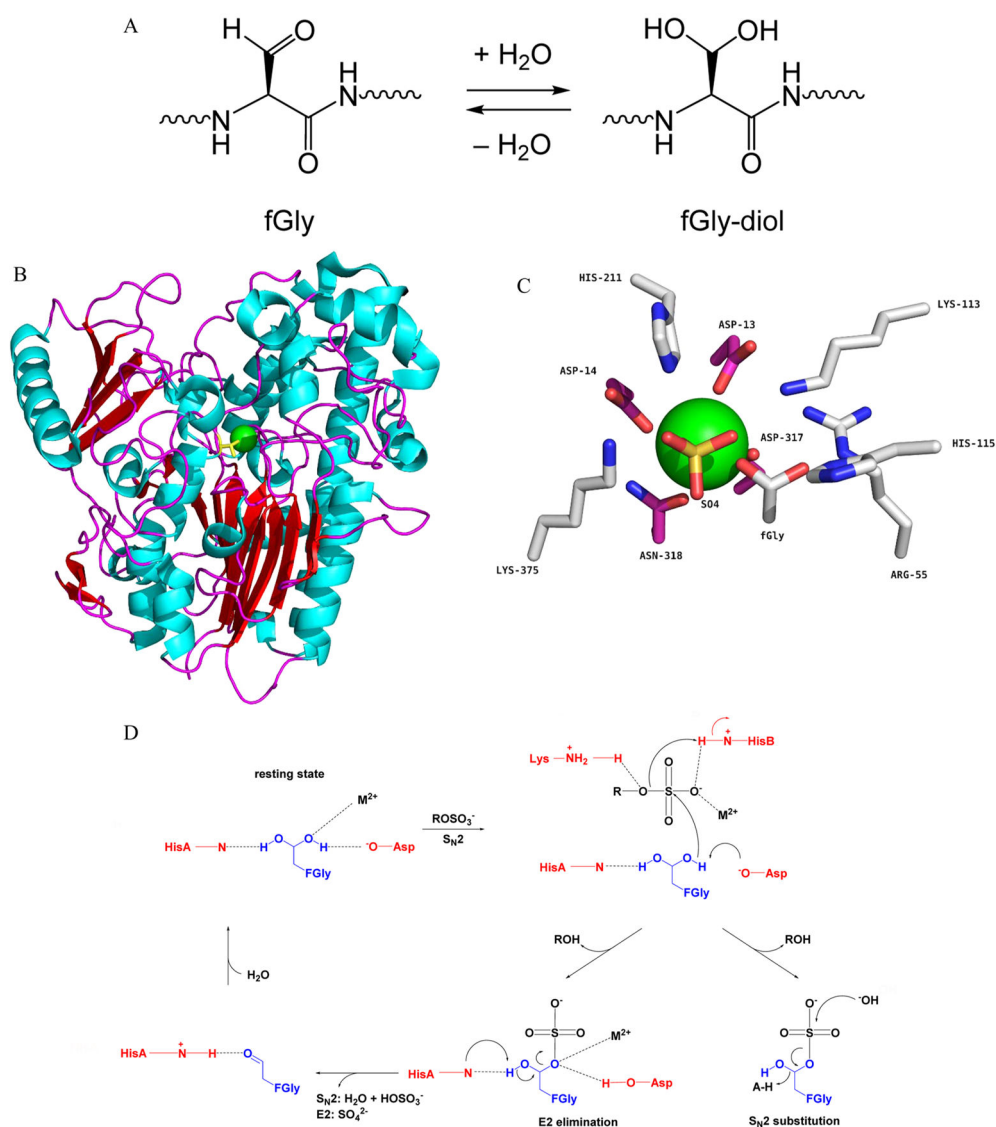
36. Williams SJ, Denehy E, Krenske EH. Experimental and Theoretical Insights into the Mechanisms of Sulfate and Sulfamate Ester Hydrolysis and the End Products of Type I Sulfatase Inactivation by Aryl Sulfamates. *J Org Chem.* 2014; 79:1995–2005. [PubMed: 24555731]
37. Chai CL, Loughlin WA, Lowe G. The stereochemical course of sulphuryl transfer catalysed by arylsulphatase II from *Aspergillus oryzae*. *Biochem J.* 1992; 287:805–812. [PubMed: 1445242]
38. Olguin LF, Askew SE, O'Donoghue AC, Hollfelder F. Efficient Catalytic Promiscuity in an Enzyme Superfamily: An Arylsulfatase Shows a Rate Acceleration of 1013 for Phosphate Monoester Hydrolysis. *J Am Chem Soc.* 2008; 130:16547–16555. [PubMed: 19554727]
39. Luo J, van Loo B, Kamerlin SC. Examining the promiscuous phosphatase activity of *Pseudomonas aeruginosa* arylsulfatase: a comparison to analogous phosphatases. *Proteins.* 2012; 80:1211–1226. [PubMed: 22275090]
40. Andrews LD, Zalatan JG, Herschlag D. Probing the Origins of Catalytic Discrimination between Phosphate and Sulfate Monoester Hydrolysis: Comparative Analysis of Alkaline Phosphatase and Protein Tyrosine Phosphatases. *Biochemistry.* 2014; 53:6811–6819. [PubMed: 25299936]
41. Jonas S, van Loo B, Hyvönen M, Hollfelder F. A New Member of the Alkaline Phosphatase Superfamily with a Formylglycine Nucleophile: Structural and Kinetic Characterisation of a Phosphonate Monoester Hydrolase/Phosphodiesterase from *Rhizobium leguminosarum*. *J Mol Biol.* 2008; 384:120–136. [PubMed: 18793651]
42. van Loo B, Jonas S, Babbie AC, Benjdia A, Berteau O, Hyvönen M, Hollfelder F. An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proc Natl Acad Sci U S A.* 2010; 107:2740–2745. [PubMed: 20133613]
43. Dierks T, Dickmanns A, Preusser-Kunze A, Schmidt B, Mariappan M, von Figura K, Ficner R, Rudolph MG. Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme. *Cell.* 2005; 121:541–552. [PubMed: 15907468]
44. Annunziata I, Bouche V, Lombardi A, Settembre C, Ballabio A. Multiple sulfatase deficiency is due to hypomorphic mutations of the SUMF1 gene. *Hum Mutat.* 2007; 28:928. [PubMed: 17657823]
45. Schlotawa L, Radhakrishnan K, Baumgartner M, Schmid R, Schmidt B, Dierks T, Gartner J. Rapid degradation of an active formylglycine generating enzyme variant leads to a late infantile severe form of multiple sulfatase deficiency. *Eur J Hum Genet.* 2013; 21:1020–1023. [PubMed: 23321616]
46. Dierks T, Schmidt B, von Figura K. Conversion of cysteine to formylglycine: a protein modification in the endoplasmic reticulum. *Proc Natl Acad Sci U S A.* 1997; 94:11963–11968. [PubMed: 9342345]
47. Dierks T, Lecca MR, Schmidt B, von Figura K. Conversion of cysteine to formylglycine in eukaryotic sulfatases occurs by a common mechanism in the endoplasmic reticulum. *FEBS Lett.* 1998; 423:61–65. [PubMed: 9506842]
48. Fey J, Balleininger M, Borissenko LV, Schmidt B, von Figura K, Dierks T. Characterization of post-translational formylglycine formation by luminal components of the endoplasmic reticulum. *J Biol Chem.* 2001; 276:47021–47028. [PubMed: 11600503]
49. Carlson BL, Ballister ER, Skordalakes E, King DS, Breidenbach MA, Gilmore SA, Berger JM, Bertozzi CR. Function and Structure of a Prokaryotic Formylglycine-generating Enzyme. *J Biol Chem.* 2008; 283:20117–20125. [PubMed: 18390551]
50. Bhattacharyya S, Tobacman JK. Hypoxia reduces arylsulfatase B activity and silencing arylsulfatase B replicates and mediates the effects of hypoxia. *PLoS One.* 2012; 7:e33250. [PubMed: 22428001]
51. Mariappan M, Radhakrishnan K, Dierks T, Schmidt B, von Figura K. ERp44 mediates a thiol-independent retention of formylglycine-generating enzyme in the endoplasmic reticulum. *J Biol Chem.* 2008; 283:6375–6383. [PubMed: 18178549]
52. Fraldi A, Zito E, Annunziata F, Lombardi A, Cozzolino M, Monti M, Spampanato C, Ballabio A, Pucci P, Sitia R, Cosma MP. Multistep, sequential control of the trafficking and function of the multiple sulfatase deficiency gene product, SUMF1 by PDI, ERGIC-53 and ERp44. *Hum Mol Genet.* 2008; 17:2610–2621. [PubMed: 18508857]

53. Gande SL, Mariappan M, Schmidt B, Pringle TH, von Figura K, Dierks T. Paralog of the formylglycine-generating enzyme – retention in the endoplasmic reticulum by canonical and noncanonical signals. *FEBS J.* 2008; 275:1118–1130. [PubMed: 18266766]
54. Mariappan M, Gande SL, Radhakrishnan K, Schmidt B, Dierks T, von Figura K. The Non-catalytic N-terminal Extension of Formylglycine-generating Enzyme Is Required for Its Biological Activity and Retention in the Endoplasmic Reticulum. *J Biol Chem.* 2008; 283:11556–11564. [PubMed: 18305113]
55. Wilkinson B, Gilbert HF. Protein disulfide isomerase. *Biochim Biophys Acta.* 2004; 1699:35–44. [PubMed: 15158710]
56. Ennemann EC, Radhakrishnan K, Mariappan M, Wachs M, Pringle TH, Schmidt B, Dierks T. Proprotein Convertases Process and Thereby Inactivate Formylglycine-generating Enzyme. *J Biol Chem.* 2013; 288:5828–5839. [PubMed: 23288839]
57. Zito E, Buono M, Pepe S, Settembre C, Annunziata I, Surace EM, Dierks T, Monti M, Cozzolino M, Pucci P, Ballabio A, Cosma MP. Sulfatase modifying factor 1 trafficking through the cells: from endoplasmic reticulum to the endoplasmic reticulum. *EMBO J.* 2007; 26:2443–2453. [PubMed: 17446859]
58. Mariappan M, Preusser-Kunze A, Balleininger M, Eiselt N, Schmidt B, Gande SL, Wenzel D, Dierks T, von Figura K. Expression, localization, structural, and functional characterization of pFGE, the paralog of the *Ca*-formylglycine-generating enzyme. *J Biol Chem.* 2005; 280:15173–15179. [PubMed: 15708861]
59. Dickmanns A, Schmidt B, Rudolph MG, Mariappan M, Dierks T, von Figura K, Ficner R. Crystal structure of human pFGE, the paralog of the *Ca*-formylglycine-generating enzyme. *J Biol Chem.* 2005; 280:15180–15187. [PubMed: 15687489]
60. Zito E, Fraldi A, Pepe S, Annunziata I, Kobinger G, Di Natale P, Ballabio A, Cosma MP. Sulphatase activities are regulated by the interaction of sulphatase-modifying factor 1 with SUMF2. *EMBO Rep.* 2005; 6:655–660. [PubMed: 15962010]
61. Roeser D, Dickmanns A, Gasow K, Rudolph MG. De novo calcium/sulfur SAD phasing of the human formylglycine-generating enzyme using in-house data. *Acta Crystallogr, Sect D: Biol Crystallogr.* 2005; 61:1057–1066. [PubMed: 16041070]
62. Preusser-Kunze A, Mariappan M, Schmidt B, Gande SL, Mutenda K, Wenzel D, von Figura K, Dierks T. Molecular Characterization of the Human *Ca*-formylglycine-generating Enzyme. *J Biol Chem.* 2005; 280:14900–14910. [PubMed: 15657036]
63. Roeser D, Preusser-Kunze A, Schmidt B, Gasow K, Wittmann JG, Dierks T, von Figura K, Rudolph MG. A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme. *Proc Natl Acad Sci U S A.* 2006; 103:81–86. [PubMed: 16368756]
64. Roeser D, Schmidt B, Preusser-Kunze A, Rudolph MG. Probing the oxygen-binding site of the human formylglycine-generating enzyme using halide ions. *Acta Crystallogr, Sect D: Biol Crystallogr.* 2007; 63:621–627. [PubMed: 17452787]
65. Alayoubi M, Guo H, Dey S, Golnazarian T, Brooks GA, Rong A, Miller JF, Ghosh P. Structure of the Essential Diversity-Generating Retroelement Protein bAvd and Its Functionally Important Interaction with Reverse Transcriptase. *Structure.* 2013; 21:266–276. [PubMed: 23273427]
66. Le Coq J, Ghosh P. Conservation of the C-type lectin fold for massive sequence variation in a *Treponema* diversity-generating retroelement. *Proc Natl Acad Sci U S A.* 2011; 108:14649–14653. [PubMed: 21873231]
67. Ho, RN.; Liebman, J.; Valentine, J. Overview of the Energetics and Reactivity of Oxygen. In: Foote, C.; Valentine, J.; Greenberg, A.; Liebman, J., editors. *Active Oxygen in Chemistry.* Springer; Netherlands: 1995. p. 1-23.
68. Klinman JP. How Do Enzymes Activate Oxygen without Inactivating Themselves? *Acc Chem Res.* 2007; 40:325–333. [PubMed: 17474709]
69. Fetzner S, Steiner R. Cofactor-independent oxidases and oxygenases. *Appl Microbiol Biotechnol.* 2010; 86:791–804. [PubMed: 20157809]
70. Murooka Y, Ishibashi K, Yasumoto M, Sasaki M, Sugino H, Azakami H, Yamashita M. A sulfur- and tyramine-regulated *Klebsiella aerogenes* operon containing the arylsulfatase (*atsA*) gene and the *atsB* gene. *J Bacteriol.* 1990; 172:2131–2140. [PubMed: 2180918]

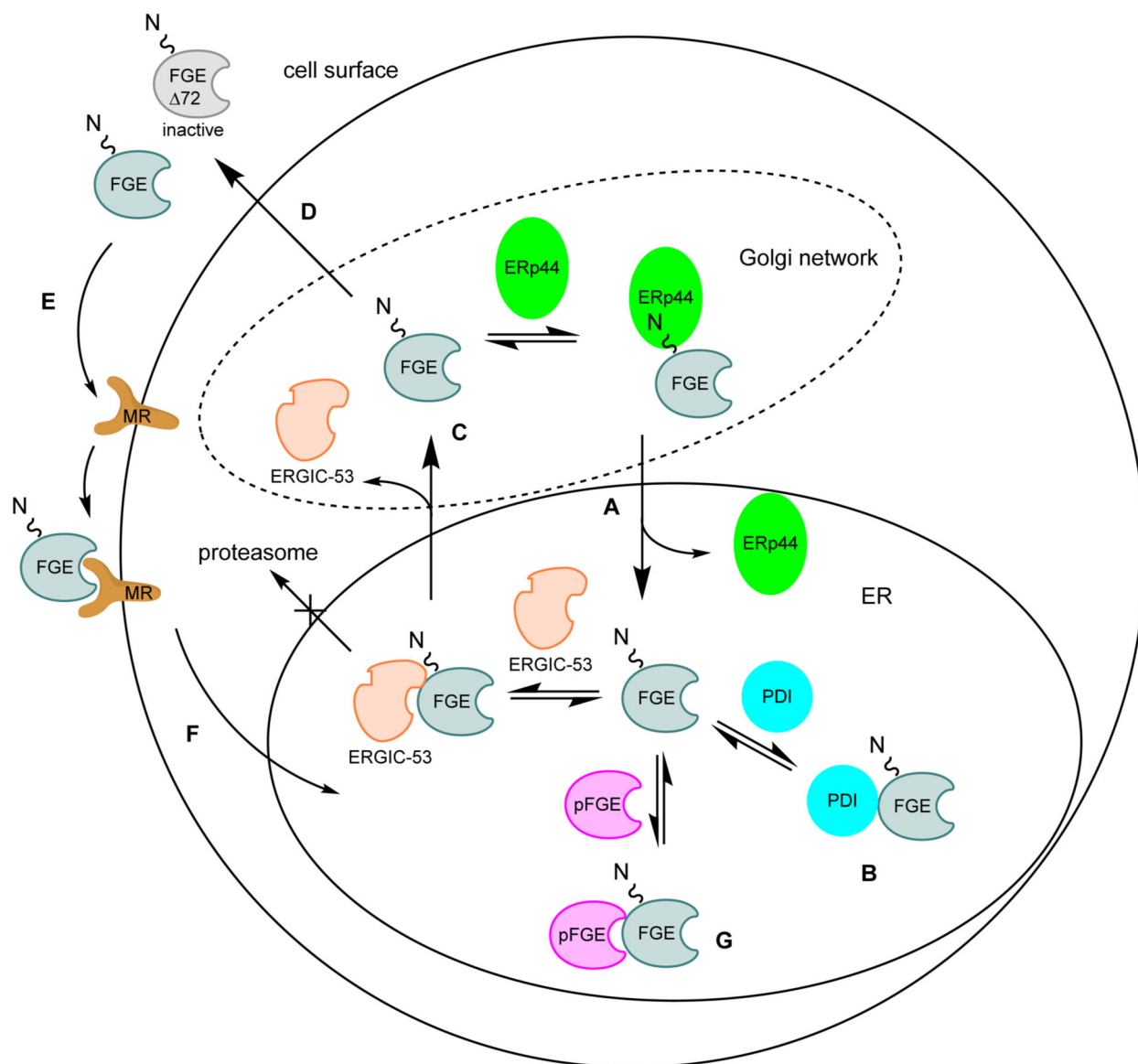


71. Szameit C, Miech C, Balleininger M, Schmidt B, von Figura K, Dierks T. The iron sulfur protein AtsB is required for post-translational formation of formylglycine in the *Klebsiella sulfatase*. *J Biol Chem*. 1999; 274:15375–15381. [PubMed: 10336424]
72. Benjdia A, Subramanian S, Leprince J, Vaudry H, Johnson MK, Berteau O. Anaerobic sulfatase-maturing enzymes, first dual substrate radical S-adenosylmethionine enzymes. *J Biol Chem*. 2008; 283:17815–17826. [PubMed: 18408004]
73. Grove TL, Ahlum JH, Qin RM, Lanz ND, Radle MI, Krebs C, Booker SJ. Further Characterization of Cys-Type and Ser-Type Anaerobic Sulfatase Maturing Enzymes Suggests a Commonality in the Mechanism of Catalysis. *Biochemistry*. 2013; 52:2874–2887. [PubMed: 23477283]
74. Grove TL, Lee KH, St Clair J, Krebs C, Booker SJ. In vitro characterization of AtsB, a radical SAM formylglycine-generating enzyme that contains three [4Fe-4S] clusters. *Biochemistry*. 2008; 47:7523–7538. [PubMed: 18558715]
75. Goldman PJ, Grove TL, Sites LA, McLaughlin MI, Booker SJ, Drennan CL. X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine post-translational modification. *Proc Natl Acad Sci U S A*. 2013; 110:8519–8524. [PubMed: 23650368]
76. Benjdia A, Leprince J, Sandström C, Vaudry H, Berteau O. Mechanistic Investigations of Anaerobic Sulfatase-Maturing Enzyme: Direct C $\beta$ H-Atom Abstraction Catalyzed by a Radical AdoMet Enzyme. *J Am Chem Soc*. 2009; 131:8348–8349. [PubMed: 19489556]
77. Dierks T, Miech C, Hummerjohann J, Schmidt B, Kertesz MA, von Figura K. Post-translational formation of formylglycine in prokaryotic sulfatases by modification of either cysteine or serine. *J Biol Chem*. 1998; 273:25560–25564. [PubMed: 9748219]
78. Benjdia A, Dehò G, Rabot S, Berteau O. First evidences for a third sulfatase maturation system in prokaryotes from *E. coli* aslB and ydeM deletion mutants. *FEBS Lett*. 2007; 581:1009–1014. [PubMed: 17303125]
79. Landgrebe J, Dierks T, Schmidt B, von Figura K. The human SUMF1 gene, required for post-translational sulfatase modification, defines a new gene family which is conserved from pro- to eukaryotes. *Gene*. 2003; 316:47–56. [PubMed: 14563551]
80. Townley RA, Bülow HE. Genetic Analysis of the Heparan Modification Network in *Caenorhabditis elegans*. *J Biol Chem*. 2011; 286:16824–16831. [PubMed: 21454666]
81. Frese MA, Dierks T. Formylglycine aldehyde Tag–protein engineering through a novel post-translational modification. *ChemBioChem*. 2009; 10:425–427. [PubMed: 19130455]
82. Carrico IS, Carlson BL, Bertozzi CR. Introducing genetically encoded aldehydes into proteins. *Nat Chem Biol*. 2007; 3:321–322. [PubMed: 17450134]
83. Rush JS, Bertozzi CR. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J Am Chem Soc*. 2008; 130:12240–12241. [PubMed: 18722427]
84. Wu P, Shui W, Carlson BL, Hu N, Rabuka D, Lee J, Bertozzi CR. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc Natl Acad Sci U S A*. 2009; 106:3000–3005. [PubMed: 19202059]
85. Rabuka D, Rush JS, deHart GW, Wu P, Bertozzi CR. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nat Protoc*. 2012; 7:1052–1067. [PubMed: 22576105]
86. Rabuka D. Chemoenzymatic methods for site-specific protein modification. *Curr Opin Chem Biol*. 2010; 14:790–796. [PubMed: 21030291]
87. Rashidian M, Dozier JK, Distefano MD. Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjugate Chem*. 2013; 24:1277–1294.
88. Shi X, Jung Y, Lin LJ, Liu C, Wu C, Cann IKO, Ha T. Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat Methods*. 2012; 9:499–503. [PubMed: 22466795]
89. Ghoneim MK, Spies M. Direct Correlation of DNA Binding and Single Protein Domain Motion via Dual Illumination Fluorescence Microscopy. *Nano Lett*. 2014; 14:5920–5931. [PubMed: 25204359]
90. Cho H, Jaworski J. A portable and chromogenic enzyme-based sensor for detection of abrin poisoning. *Biosens Bioelectron*. 2014; 54:667–673. [PubMed: 24334282]

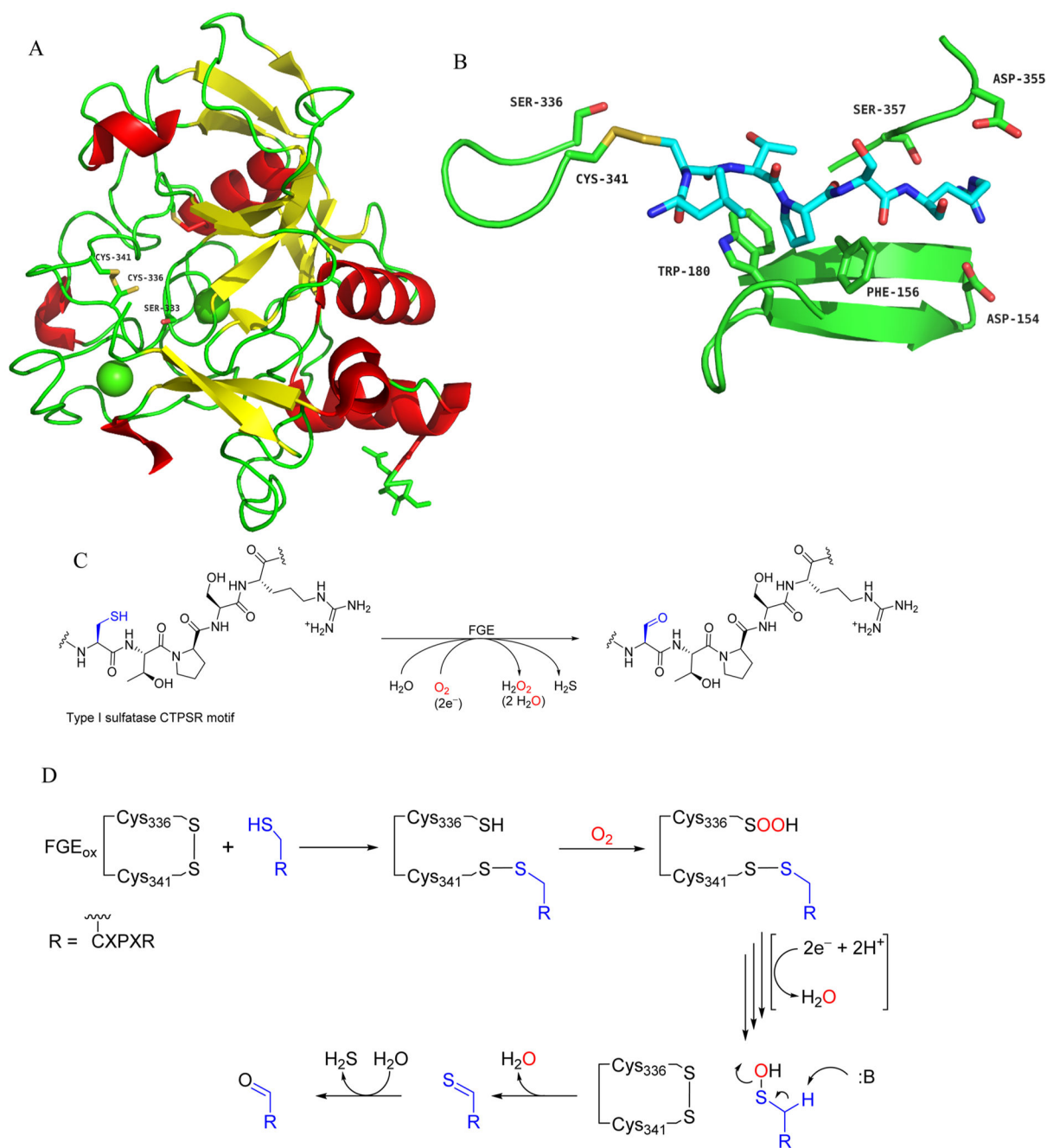
91. Kwak EA, Jaworski J. Controlled surface immobilization of viruses via site-specific enzymatic modification. *J Mater Chem B*. 2013; 1:3486–3493.
92. Cho H, Jaworski J. Enzyme directed formation of un-natural side-chains for covalent surface attachment of proteins. *Colloids Surf, B*. 2014; 122:846–850.
93. Hudak JE, Yu HH, Bertozzi CR. Protein glycoengineering enabled by the versatile synthesis of aminoxy glycans and the genetically encoded aldehyde tag. *J Am Chem Soc*. 2011; 133:16127–16135. [PubMed: 21866965]
94. Smith EL, Giddens JP, Iavarone AT, Godula K, Wang LX, Bertozzi CR. Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjugate Chem*. 2014; 25:788–795.
95. Liang SI, McFarland JM, Rabuka D, Gartner ZJ. A Modular Approach for Assembling Aldehyde-Tagged Proteins on DNA Scaffolds. *J Am Chem Soc*. 2014; 136:10850–10853. [PubMed: 25029632]
96. Hudak JE, Barfield RM, de Hart GW, Grob P, Nogales E, Bertozzi CR, Rabuka D. Synthesis of Heterobifunctional Protein Fusions Using Copper-Free Click Chemistry and the Aldehyde Tag. *Angew Chem, Int Ed*. 2012; 51:4161–4165.
97. Chari RVJ, Miller ML, Widdison WC. Antibody–Drug Conjugates: An Emerging Concept in Cancer Therapy. *Angew Chem, Int Ed*. 2014; 53:3796–3827.
98. Junutula JR, Raab H, Clark S, Bhakta S, Leipold DD, Weir S, Chen Y, Simpson M, Tsai SP, Dennis MS, Lu Y, Meng YG, Ng C, Yang J, Lee CC, Duenas E, Gorrell J, Katta V, Kim A, McDorman K, Flagella K, Venook R, Ross S, Spencer SD, Lee Wong W, Lowman HB, Vandlen R, Sliwkowski MX, Scheller RH, Polakis P, Mallet W. Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat Biotechnol*. 2008; 26:925–932. [PubMed: 18641636]
99. Agarwal P, Bertozzi CR. Site-specific antibody-drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjugate Chem*. 2014 Online.
100. Agarwal P, van der Weijden J, Sletten EM, Rabuka D, Bertozzi CR. A Pictet-Spengler ligation for protein chemical modification. *Proc Natl Acad Sci U S A*. 2013; 110:46–51. [PubMed: 23237853]
101. Agarwal P, Kudirka R, Albers AE, Barfield RM, de Hart GW, Drake PM, Jones LC, Rabuka D. Hydrazino-Pictet-Spengler Ligation as a Biocompatible Method for the Generation of Stable Protein Conjugates. *Bioconjugate Chem*. 2013; 24:846–851.
102. Drake PM, Albers AE, Baker J, Banas S, Barfield RM, Bhat AS, de Hart GW, Garofalo AW, Holder P, Jones LC, Kudirka R, McFarland J, Zmolek W, Rabuka D. Aldehyde Tag Coupled with HIPS Chemistry Enables the Production of ADCs Conjugated Site-Specifically to Different Antibody Regions with Distinct in Vivo Efficacy and PK Outcomes. *Bioconjugate Chem*. 2014; 25:1331–1341.



**Figure 1.** Hydration of fGly to fGly-diol, *Pseudomonas aeruginosa* arylsulfatase structure and active site architecture (PAS, PDB: 1HDD), and proposed catalytic mechanisms of type I sulfatase. (A) fGly is rapidly hydrated to a geminal diol, fGly-diol, and this form predominates in the sulfatase resting state. (B) fGly diol, yellow;  $Ca^{2+}$ , green. (C) Highlighted features:  $Ca^{2+}$ , green; metal binding, magenta; substrate binding and catalytic, gray. (D) Following the binding of sulfate ester substrate, fGly-diol is activated for nucleophilic attack at sulfur by aspartate (Asp317). The sulfoenzyme is formed, and desulfation may proceed by one of two pathways, hydrolysis by sulfur ( $S_N2$ ) or, more probable, elimination from the remaining fGly-diol hydroxyl (E2), catalyzed by a histidine base (His115). Catalytic mechanism scheme adapted with alteration from ref 28.

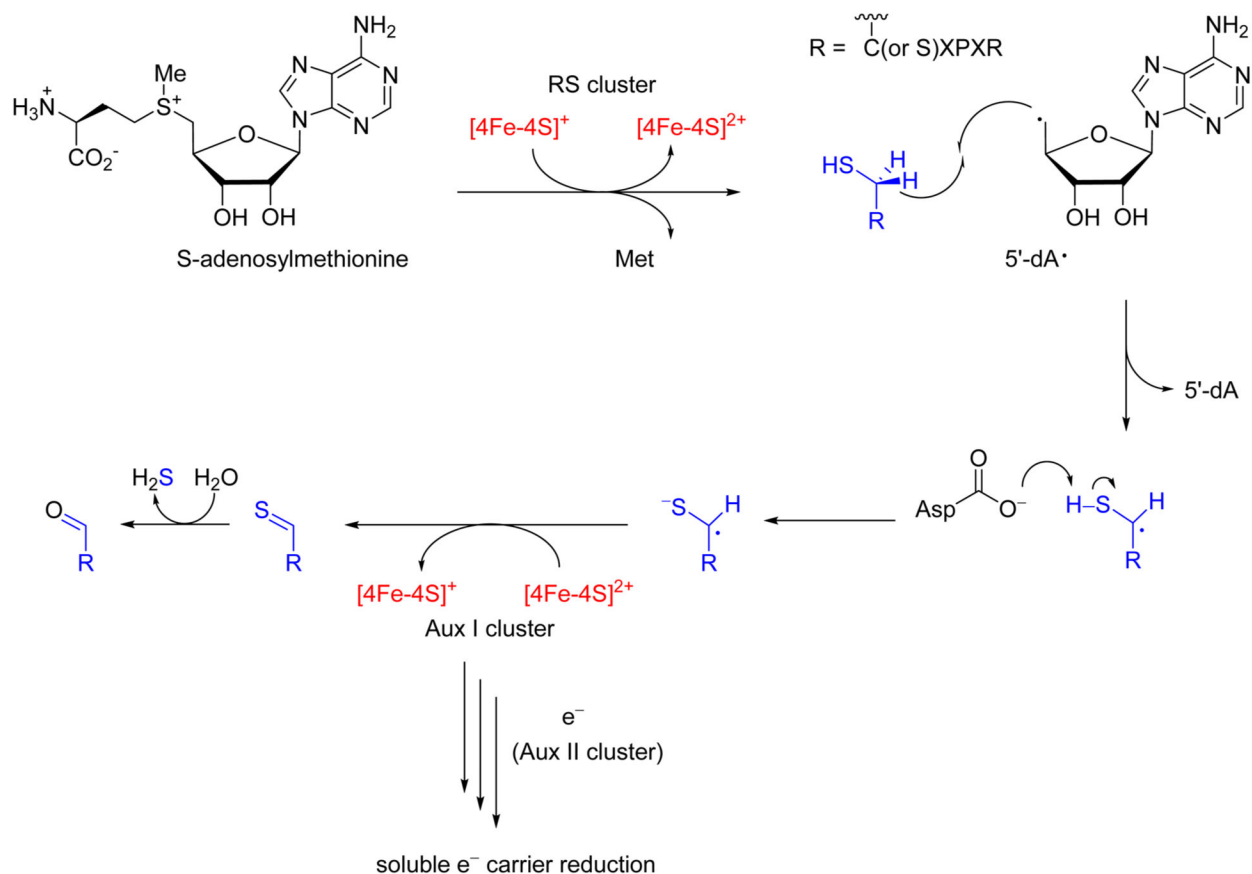


**Figure 2.** Regulation and trafficking of mammalian FGE by protein partners. (A) ERp44 binds the N-terminal extension of FGE and delivers it back to the ER. (B) Protein disulfide isomerase (PDI) retains FGE in the ER through binding and increases catalytic activity, perhaps by acting as a disulfide reductant. (C) ERGIC-53 transports FGE to the Golgi and prevents degradation by the proteasome. (D) A portion of FGE is trafficked to the cell surface and secreted through the secretory pathway; furin and other proteases cleave and thereby inactivate 70–80% of secreted FGE. (E) Secreted FGE binds mannose receptor (MR) through its N-glycan. (F) Extracellular FGE can be trafficked back to the ER in an active state. (G) pFGE interacts with FGE for an unknown purpose, but may chaperone unfolded sulfatases or mediate their interaction with FGE.

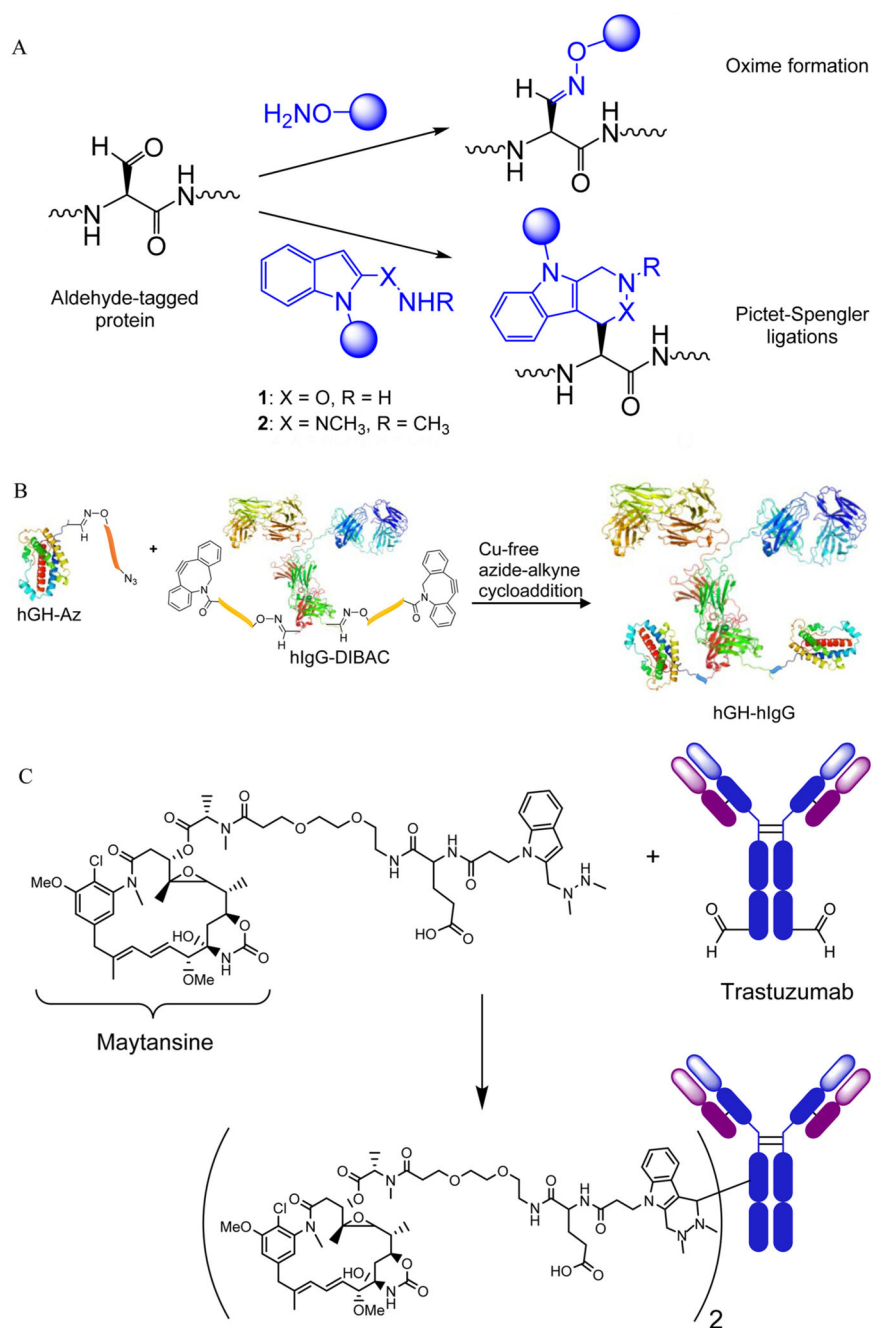
**Figure 3.**

FGE structure, overall reaction, and *speculative* mechanistic proposal. (A) Crystal structure of human FGE (PDB: 1Y1E), with active site cysteines (Cys336, Cys341; engaged in partial disulfide bond), the essential serine (Ser333), and  $\text{Ca}^{2+}$  ions (green) highlighted. (B) Close-up of human FGE Cys336S (PDB: 2AFY) in complex with substrate as Cys341-LCTPSRA mixed disulfide, with substrate binding residues highlighted. (C) Overall reaction catalyzed by FGE on generic CTPSR substrate. Given the  $2e^-$  oxidation of cysteine to formylglycine,  $\text{O}_2$  is reduced to  $\text{H}_2\text{O}_2$ . However, as has been proposed, the concomitant input of an

additional  $2e^-$  and protons would result in the production of 2  $H_2O$  instead. (D) Abbreviated summary of proposed mechanism provided by Roeser et al.<sup>63</sup>  $FGE_{ox}$  binds substrate through thiol–disulfide exchange with substrate at Cys341. Cys336 then reacts with  $O_2$  to generate a peroxysulfenic acid. Decomposition of this intermediate, and perhaps the input of two external electrons, would regenerate  $FGE_{ox}$ , release 1  $H_2O$ , and form a substrate cysteine sulfenic acid. Base-catalyzed elimination of  $H_2O$  from substrate sulfenic acid produces a transient thioaldehyde, which gives fGly upon hydrolysis.

**Figure 4.**

Mechanism of fGly biosynthesis by the anaerobic sulfatase maturing enzyme (anSME). The reduced radical S-adenosylmethionine (RS) [4Fe-4S]<sup>+</sup> cluster first binds and reductively cleaves SAM to produce methionine (Met) and 5'-deoxyadenosyl radical (5'-dA<sup>•</sup>). The 5'-dA<sup>•</sup> abstracts the *pro-S* C<sub>β</sub> hydrogen from the conserved substrate cysteine or serine residue. Thiol deprotonation by an active site aspartyl side chain generates a transient radical anion, which is thought to be oxidized by auxiliary [4Fe-4S]<sup>2+</sup> cluster I (Aux I). Auxiliary cluster II may oxidize Aux I and transfer the second substrate electron to a soluble carrier for regeneration of the reduced RS cluster.



**Figure 5.** Principle and application of aldehyde tag engineering. (A) fGly is installed on a protein target and used as a bioconjugation handle with aminooxy, or modified Pictet-Spengler reagent functionalized molecules. (B) Bifunctional protein fusion of human growth hormone (hGH) and human IgG (hIgG) synthesized by aldehyde tagging each followed by Cu-free azide-alkyne cycloaddition (reproduced with permission from ref 96, copyright Wiley-VCH, 2012). (C) ADC produced site-specifically and with precise drug/antibody ratio, shown with



hydrazino Pictet-Spengler ligation product, for recent example of trastuzumab and maytansine.<sup>102</sup>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript