Computational analysis of bacterial sulfatases and their modifying enzymes

Andreas Schirmer and Roberto Kolter

The sequence analysis of enzymes that might modify bacterial sulfatases should be useful in the task of identifying the human sulfatase-modifying homologs – enzymes that are defective in the rare inherited disease multi-sulfatase deficiency.

Address: Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA.

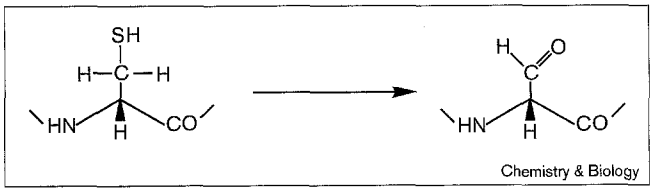
Correspondence: Roberto Kolter
E-mail: kolter@mbcrr.harvard.edu

Chemistry & Biology August 1998, **5**:R181–R186
http://biomednet.com/elecref/1074552100 5R0181

Sulfatases are evolutionarily conserved enzymes, found in diverse bacteria and eucarya, that catalyze the hydrolytic cleavage of sulfate from organic O-sulfate ester [1–4]. Human sulfatases are involved in the turnover of mucopolysaccharides and sulfolipids and in the synthesis of steroid hormones [4,5], whereas bacterial sulfatases appear to be used primarily in sulfur scavenging [1]. Eucaryotic sulfatases contain a C_α-formylglycine (FGly) in their active site that is essential for catalytic activity. This unusual residue is derived from a cysteine, but the exact mechanism of this post-translational conversion remains unknown [6,7] (Figure 1).

A clue regarding the enzymology of this modification comes from studies of multi-sulfatase deficiency (MSD), a rare inherited human disease [8]. Individuals with MSD produce inactive sulfatase proteins that contain the unmodified cysteine, suggesting an enzymatic activity is responsible for the modification and this activity is lacking in MSD [7]. Interestingly, all the information necessary for the modification to occur in humans is contained in a highly conserved 15 amino acid sequence surrounding the cysteine residue [9]. A very similar sequence is also present

Figure 1



The conversion of cysteine to C_α-formylglycine.

in bacterial sulfatases. In the primary sequence of some bacterial sulfatases, however, there is a serine residue instead of a cysteine in the position that specifies FGly. Miech *et al.* [10] recently demonstrated that, in fact, in the arylsulfatase of *Klebsiella aerogenes* FGly was generated from a serine residue. Yet, when the active-site cysteine in the human sulfatase ArsA was replaced with serine, the *in vitro* modification to FGly was completely abolished [9]. These observations suggest the existence of both cysteine-specific and serine-specific sulfatase-modifying enzymes. The computational analysis presented here reveals likely candidates for the respective sulfatase-modifying enzymes and proposes a molecular mechanism for the unique conversion of the active-site cysteine to FGly.

Bacterial sulfatases

The arylsulfatases of *Pseudomonas aeruginosa*, $AtsA_{Pae}$ [11], and *K. aerogenes*, $AtsA_{Kae}$ [1], are known bacterial sulfatases similar to eucaryotic sulfatases. In $AtsA_{Pae}$ the active-site cysteine is conserved in the primary sequence, whereas $AtsA_{Kae}$ has a serine residue at this position, leading us to propose that bacterial sulfatases can be classified into cysteine (cys)-type and serine (ser)-type sulfatases with respect to the amino acid that is converted to FGly in the active enzymes. Similarity searches of all annotated bacterial sequences revealed additional sulfatase homologs in *Escherichia coli*, *Mycobacterium tuberculosis* and *Sinorhizobium meliloti*, confirming this classification (Table 1).

E. coli encodes three putative sulfatases (F571, YidJ and AslA), but none of them has any assigned function or substrate (Table 1). Their primary sequences show all the features expected for members of the sulfatase family: a length of 500–600 amino acids, 20–30% sequence identity to other sulfatases, and a highly conserved region around the active site [4]. Both F571 and AslA are ser-type sulfatases whereas YidJ belongs to the cys-type sulfatases (Table 1). *Mycobacterium tuberculosis* encodes five sulfatase homologs, four of which are larger than the known sulfatases. Again, there are no experimental data available on their function, but all five are cys-type sulfatases (Table 1). BetC, a cys-type sulfatase homolog of *Sinorhizobium meliloti*, has been identified as part of a gene cluster encoding enzymes that convert choline to glycine betaine [12].

It is noteworthy that, in terms of sequence similarity, there is no difference between the two types, that is there is no greater degree of similarity between ser-type or cys-type sulfatases when compared to each other than there is between ser-type and cys-type sulfatases. This is true for

Table 1

Classification of known and hypothetical bacterial sulfatases with similarities to eucaryotic sulfatases.

Organism	Protein/ORF	Substrate	Length (amino acids)	Region surrounding the active-site residue		Signal sequence	Reference/ Accession No.
*Homo sapiens**	ArsA (ASA)	Cerebroside sulfate	507	Pos. 65–80	PVSL**C**T**P**S**R**AALLT**G**R	+	[7]
Cys-type sulfatases							
P. aeruginosa	$AtsA_{Pae}$	Arylsulfates	533	Pos. 47–62	TAST**C**S**P**T**R**SMLLT**G**T	–	[11]
S. meliloti	BetC	Choline sulfate	512	Pos. 50–65	SSPL**C**A**P**A**R**ASFMA**G**Q	–	U39940
E. coli	YidJ (F497)	Unknown	497	Pos. 48–63	CSPV**C**T**P**A**R**AGLFT**G**I	–	[23]
M. tuberculosis	MTCY210.30	Unknown	787	Pos. 79–94	TTAL**C**S**P**T**R**ASLLT**G**R	–	Z84395
M. tuberculosis	MTCI376.13c	Unknown	787	Pos. 85–100	TTAL**C**S**P**T**R**QALLT**G**R	–	Z95972
M. tuberculosis	MTCI418A.01c	Unknown	970	Pos. 254–269	VTAV**C**S**P**T**R**AALLT**G**R	–	Z96070
M. tuberculosis	MTCY22D7.04c	Unknown	603	Pos. 53–68	GSLA**C**V**P**S**R**PTIFT**G**Q	–	Z83866
M. tuberculosis†	MTCY63.01c/ MTV035.24c	Unknown	465	Pos. 54–69	TAPL**C**T**P**S**R**GSLFT**G**R	–	Z96800/ AL021930
Ser-type sulfatases							
K. aerogenes	$AtsA_{Kae}$	Arylsulfates	464	Pos. 68–83	TSPM**S**A**P**A**R**SMLLT**G**N	+	[1]
E. coli	AslA	Unknown	551	Pos. 68–83	SQPS**S**S**P**T**R**ATILT**G**Q	+	[28]
E. coli	F571	Unknown	560	Pos. 128–143	AHGV**S**G**P**S**R**AAIMT**G**R	+	U00096

Hypothetical bacterial sulfatases have been identified by performing basic BlastP2 searches [29] of all nonredundant GenBank and EMBL CDS translations (as of 12 May 1998) with $AtsA_{Kae}$ or Ats_{Pae} as query sequences. *The residues 65–80 of ArsA contain all the information necessary for the modification of the active site Cys69 to C_α-formylglycine [7]. †The sequence has been assembled from MTCY63.01c (amino terminus) and MTV035.24c (carboxyl terminus). *P, Pseudomonas; S, Sinorhizobium; E, Escherichia; M, Mycobacterium; K, Klebsiella.*

both the complete protein sequence and the highly conserved region around the active site (Table 1). All ser-type sulfatases contain typical amino-terminal signal sequences, however, suggesting a periplasmic localization, whereas the cys-type sulfatases are predicted to be cytoplasmic.

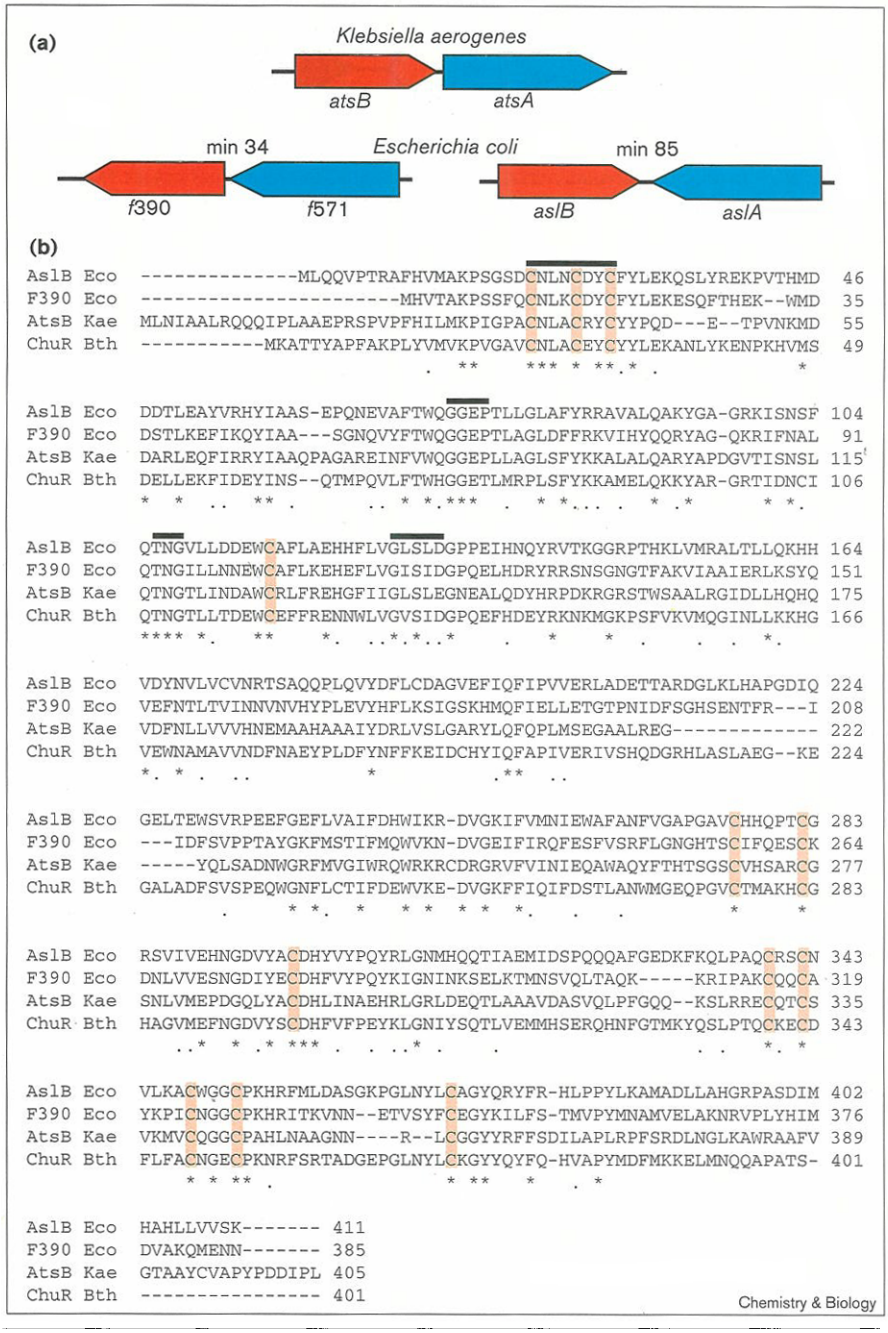
Putative serine-specific sulfatase-modifying enzymes

Each of the three ser-type sulfatases has an adjacent gene (*atsB*, *aslB* and *f*390) encoding a protein of about 400 amino acids (Figure 2a) and these proteins are ~ 40% identical (Figure 2b). When $AtsA_{Kae}$ was expressed in *E. coli*, AtsB was required to obtain sulfatase activity [1]. Although the initial observation was interpreted as meaning simply that AtsB has a positive role in expression of sulfatase activity [1], subsequent publications suggested that AtsB was a transcriptional activator for $AtsA_{Kae}$ [13,14]. Evidence suggesting that AtsB and its homologs are not transcription factors but rather sulfatase-modifying enzymes, however, comes from scrutiny of their primary sequence — they do not contain an apparent helix–turn–helix motif or any other DNA-binding motif expected for transcriptional regulators. The most striking feature of the amino acid sequence of AtsB and its homologs is the presence of 12 conserved cysteine residues (Figure 2b), which are not reminiscent of cysteine-containing zinc finger motifs [15], arranged in three clusters. Motif 1 consists of three highly conserved cysteines (Cys–X_3–Cys–X_2–Cys) near the amino terminus followed by another conserved cysteine residue at a distance of 83–87 amino acids. Motif 2, with three cysteine residues (Cys–X_5–Cys–X_{14}–Cys), and motif 3, with five cysteine residues (Cys–X_2–Cys–X_5–Cys–X_3–Cys–X_{11-17}–Cys), are located in the carboxy-terminal region (Figure 2b). A high number of conserved and clustered cysteine residues in a protein is a strong indicator for the presence of iron–sulfur [Fe–S] center. Most [Fe–S] proteins contain more than one cluster, and they are predominantly respiratory enzymes in which the [Fe–S] clusters are involved in electron transfer during redox reactions [16]. The arrangement of the conserved cysteine residues within the primary structure of AtsB and its homologs strongly suggests that these proteins contain [Fe–S] clusters, implying that they function as oxidoreductases. We therefore hypothesize that AtsB, AslB and F390 are not transcription factors but, instead, they are modifying proteins that oxidize the active-site serine of their respective accompanying sulfatases to FGly.

There is a fourth gene, *chuR* from *Bacteroides thetaiotaomicron*, homologous to *atsB*, *aslB* and *f*390 [13] (Figure 2b). *B. thetaiotaomicron* can utilize sulfated mucopolysaccharides such as chondroitin sulfate and heparin [17]. A *ChuR* mutant was found to be unable to grow either on chondroitin sulfate or on heparin [13]. The finding that a single gene affected both chondroitin sulfate and heparin utilization was unexpected, because there are two different,

Figure 2

Serine-specific sulfatase-modifying enzymes. **(a)** Organization of genes encoding ser-type sulfatases (blue boxes) and their accompanying modifying proteins (red boxes). **(b)** Alignment of ser-specific sulfatase-modifying enzymes generated by using the program CLUSTAL [30]. Identical and similar amino acids are indicated in the bottom line by asterisks and points, respectively. The conserved cysteine residues of the sulfatase-modifying enzymes are orange and regions conserved in all members of a larger group of proteins have a bar on top. Eco, *Escherichia coli*; Kae, *Klebsiella aerogenes*; Bth, *Bacteroides thetaiotaomicron*. References: AtsA [1], AslA [29], ChuR [13], F390 (GenBank accession number U00096).

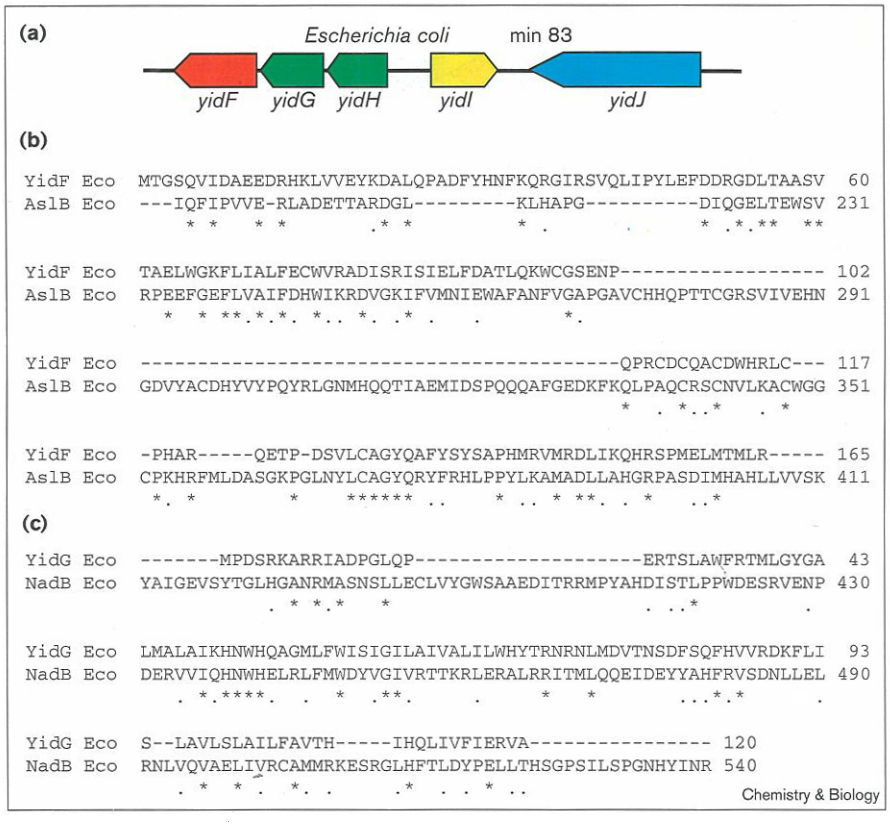


independently regulated pathways for the degradation of these compounds in *B. thetaiotaomicron* [18]. Interestingly, of the chondroitin sulfate utilization genes, only chondroitin-6-sulfatase activity was missing in the *chuR* mutant. Cheng *et al.* [13] concluded that *chuR* is a regulatory gene that controls the transcription of chondroitin sulfate and heparin utilization genes. If, instead, ChuR is a sulfatase-modifying enzyme, ChuR would modify at least two different sulfatases that desulfate either chondroitin-6-sulfate or one of the sulfated compounds of heparin. If this is so, these sulfatases are probably ser-type sulfatases and the reported phenotype of the *chuR* mutant is a form of bacterial MSD.

Putative cysteine-specific sulfatase-modifying enzymes

Are the enzymes that modify the cys-type sulfatases likely to be similar to the ser-specific sulfatase-modifying enzymes described above? Because the 15 amino acid residues essential for the modification are highly conserved in both types of sulfatases (Table 1), it is likely

Figure 3



Cysteine-specific sulfatase-modifying enzymes. **(a)** Organization of the genes of a cys-type sulfatase cluster [23]. YidF is similar to the carboxy-terminal region of a ser-specific sulfatase-modifying enzyme. YidG and YidH are similar to oxidoreductases. YidI does not display any similarity to known proteins and YidJ is a cys-type sulfatase. Alignment of **(b)** YidF with the carboxy-terminal region of a ser-specific sulfatase-modifying enzyme, AslB [28], and **(c)** YidG with the carboxy-terminal region of L-aspartate oxidase, NadB [24], generated using the program CLUSTAL [30]. Identical and similar amino acids are indicated in the bottom line by asterisks and points, respectively.

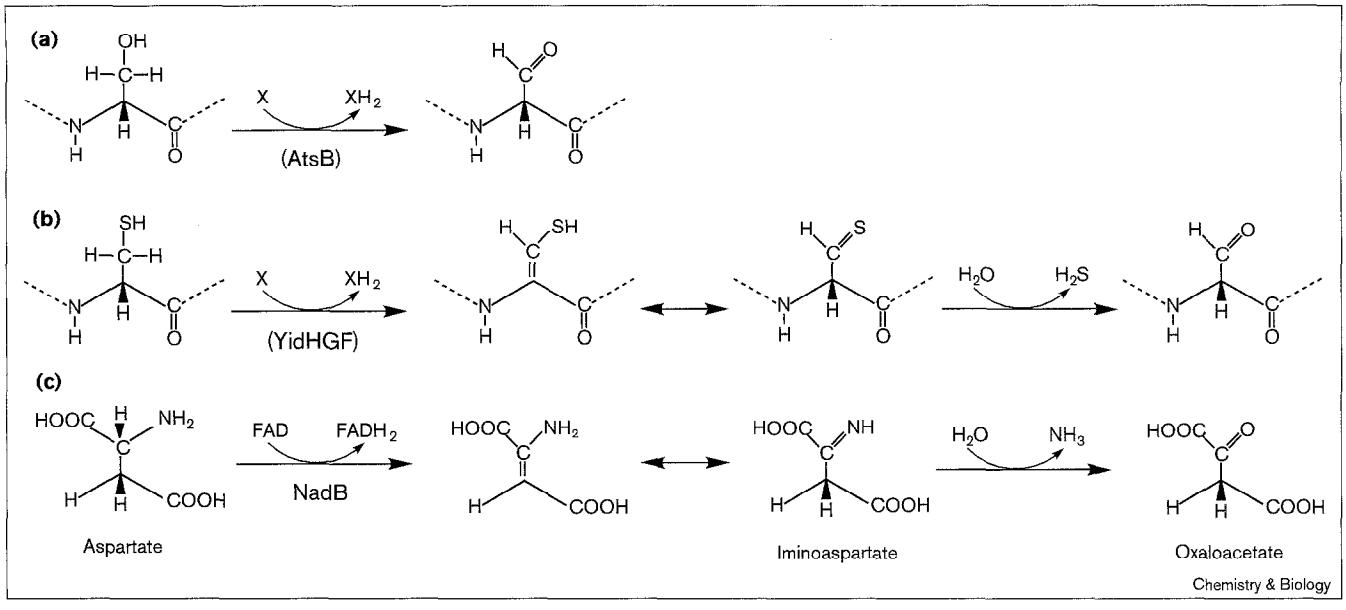
that the substrate-binding domains are similar. Catalytic domains, on the other hand, could be very different and sequence analysis suggests that this is the case.

The four putative ser-type-specific sulfatase-modifying proteins have several short sequences in their amino-terminal region that are also present in a much larger group of proteins (sequences with a bar on top in Figure 2b). The function of all of the proteins in this larger group are unknown but some might be involved in modification steps during the biosynthesis of several cofactors, such as molybdopterin, heme d_1 and PQQ [19–22]. If the limited similarity is indicative of conservation of a basic catalytic mechanism this would be an indication that the amino-terminal region of the ser-specific sulfatase-modifying enzymes contains the catalytic domain. Thus, a substrate-binding domain could be found in the carboxy-terminal region. Interestingly, in the vicinity of the cys-type sulfatase gene *yidJ* of *E. coli*, there is a gene, *yidF* [23], encoding a protein of 165 amino acids that is about 20% identical to the carboxy-terminal region of the ser-specific modifying enzymes (Figure 3b). YidF is unlikely to be an catalytically active sulfatase-modifying enzyme because it lacks their entire amino-terminal region. YidF could contain the information to bind to the conserved 15 amino acid region of a sulfatase, however.

The *yidF* gene (*f*165) is part of a gene cluster containing two more open reading frames (ORFs), *yidG* (*f*120) and *yidH* (*f*115; Figure 3a) [23] that could encode the catalytic subunits of a cys-specific sulfatase-modifying enzyme. YidG and YidH share some similarity with each other. They appear to be integral membrane proteins and both display limited similarities to oxidoreductases: YidG is similar to the carboxy-terminal region of L-aspartate oxidase, the product of *nadB*, of *E. coli* (Figure 3c) [24] and YidH is similar to a subunit of a membrane-bound methane monoxygenase of *Methylococcus capsulatus* [25].

NadB is a flavin adenine dinucleotide (FAD)-dependent flavoprotein that, *in vitro*, catalyzes the oxidative deamination of L-aspartate to oxaloacetate [24]. By sequence homology, NadB belongs to the succinate dehydrogenase/fumarate reductase family [26]. It is therefore assumed that the initial reaction carried out by NadB is the removal of two reducing equivalents from L-aspartate to form a -C=C- double bond. This intermediate is in equilibrium with iminoaspartate, which spontaneously hydrolyzes to yield oxaloacetate and ammonia (Figure 4c). The reaction mechanism of NadB is very reminiscent of the proposed mechanism by which the active-site cysteine in human sulfatases is converted to C_α-formylglycine [7]: the oxidation of an amino group to an imino group by NadB corresponds to the

Figure 4



Mechanisms of C_α-formylglycine formation in the active site of **(a)** ser-type sulfatases and **(b)** cys-type sulfatases compared to **(c)** the *in vitro* reaction of L-aspartate oxidase, NadB. It is proposed here that reaction (a) is carried out by AtsB and its homologs and reaction (b) by YidHGF. The reaction mechanisms shown in (b) and (c) are similar: the oxidation of a sulfhydryl group to a thioaldehyde (b) is analogous to the oxidation of an amino to an imino group (c), which subsequently hydrolyzes to yield the aldehyde and hydrogen sulfide or ammonia , respectively. (b) Has been modified from Schmidt *et al.* [7] and (c) from Flachmann *et al.* [24].

oxidation of a sulfhydryl group to a thioaldehyde by a sulfatase-modifying enzyme and these intermediates are subsequently hydolyzed to yield the aldehyde and either ammonia or hydrogen sulfide, respectively (Figure 4b,c). It further implies that the initial step of the oxidative desulfhydrylation of the active-site cysteine could be an elimination reaction catalyzed by YidG. A similar enzymatic reaction, that is an oxidative decarboxylation of the carboxy-terminal cysteine of a lantibiotic precursor peptide, has been reported as part of the biosynthesis of epidermin [27]. This reaction is catalyzed by the flavin mononucleotide (FMN)-dependent flavoprotein EpiD. YidG and EpiD share little sequence similarity, however.

The reaction catalyzed by NadB comprises the reoxidation of the reduced coenzyme $FADH_2$ by molecular oxygen, which regenerates the electron acceptor FAD and yields H_2O_2 [24]. An analogous reoxidation of the electron acceptor during the oxidative desulfhydrylation of the active-site cysteine is possible and could be catalyzed by YidH which is similar to a monooxygenase [25]. As there is no apparent FAD- or FMN-binding site by sequence scrutiny in any of the hypothetical proteins encoded by *yidHGF*, however, the nature of this electron acceptor remains unknown.

Taken together, we hypothesize that the enzyme complex encoded by the *yidHGF* genes modifies YidJ, the cys-type sulfatase of *E. coli*. Because only cys-type sulfatases have been found in humans, the YidHGF complex could be the functional homolog of the human sulfatase-modifying enzyme. We believe that our identification of the bacterial homolog will point the way to the identification of the analogous human genes, which are probably mutated in individuals with an inherited MSD.

References

1. Murooka, Y., *et al.*, & Yamashita, M. (1990). A sulfur- and tyramine-regulated *Klebsiella aerogenes* operon containing the arylsulfatase (*atsA*) gene and the *atsB* gene. *J. Bacteriol.* **172**, 2131-2140.
2. Yang, Q., Angerer, L.M. & Angerer, R.C. (1989). Structure and tissue-specific developmental expression of a sea urchin arylsulfatase gene. *Dev. Biol.* **135**, 53-65.
3. Hallmann, A. & Sumper, M. (1994). An inducible arylsulfatase of *Volvox carteri* with properties suitable for a reporter-gene system. *Eur. J. Biochem.* **221**, 143-150.
4. Parenti, G., Meroni, G. & Ballabio, A. (1997). The sulfatase gene family. *Curr. Opin. Gen. Dev.* **7**, 386-391.
5. Bond, C.S., *et al.*, & Guss, J.M. (1996). Structure of a human lysosomal sulfatase. *Structure* **5**, 277-289.
6. Selmer, T., Hallmann, A., Schmidt, B., Sumper, M. & 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.
7. Schmidt, B., Selmer, T., Ingendoh, A. & von Figura, K. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **82**, 271-278.
8. Kolodny, E.H. & Fluharty, A.L. (1995). Metachromatic leukodystrophy and multiple sulfatase deficiency: sulfatide lipidosis. In *The metabolic and molecular bases of inherited disease.* (R. Scriver, A.L. Beaudet, W.S. Sly, & Valle, D. eds), pp. 2693-2741, McGraw-Hill, New York.
9. Dierks, T., Schmidt, B. & von Figura, K. (1997). Conversion of cysteine to formylglycine: A protein modification in the endoplasmic reticulum. *Proc. Natl Acad. Sci.* **94**, 11963-11968.

10. Miech, C., Dierks, T., Selmer, T., von Figura, K. & Schmidt, B. (1998). Arylsulfatase from *Klebsiella pneumoniae* carries a formylglycine generated from serine. *J. Biol. Chem.* **273**, 4835-4387.
11. Beil, S., *et al.*, & Kertesz, M.A. (1995). Purification and characterization of arylsulfatase synthesized by *Pseudomonas aeruginosa* PAO during growth in sulfate-free medium and cloning of the arylsulfatase gene (*atsA*). *Eur. J. Biochem.* **229**, 385-394.
12. Pocard, J.A., *et al.*, & Le Rudelier, D. (1997). Molecular characterization of the *bet* genes encoding glycine betaine synthesis in *Sinorhizobium meliloti* 102F34. *Microbiology* **143**, 1369-1379.
13. Cheng, Q., Hwa, V. & Salyers, A.A. (1992). A locus that contributes to colonization of the intestinal tract by *Bacteroides thetaiotaomicron* contains a single regulatory gene (*chuR*) that links two polysaccharide utilization pathways. *J. Bacteriol.* **174**, 7185-7193.
14. Murooka, Y., Azakami, H. & Yamashita, M. (1996). The momoamine regulon including synthesis of arylsulfatase and monoamine oxidase. *Biosci. Biotech. Biochem* **60**, 935-941.
15. Coleman, J.E. (1992). Zinc proteins: Enzymes, storage proteins, transcription factors, and replication proteins. *Annu. Rev. Biochem.* **61**, 897-946.
16. Beinert, H. (1990). Recent development in the field of iron-sulfur proteins. *FASEB J.* **4**, 2483-2491.
17. Salyers, A.A. & O'Brien, M. (1980). Cellular location of enzymes involved in chondroitin sulfate breakdown by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **143**, 772-780.
18. Hwa, V. & Salyers, A.A. (1992). Evidence for differential regulation of genes in the chondroitin sulfate utilization pathway of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **174**, 342-344.
19. Rivers, S.L., McNairn, E., Blasco, F., Giordano, G. & Boxer, D.H. (1993). Molecular genetic analysis of the *moa* operon of *Escherichia coli* K-12 required for molybdenum cofactor biosynthesis. *Mol. Microbiol.* **8**, 1071-1081.
20. Kletzin, A., Mukund, S., Kelly-Crouse, T.L., Chan, M.K., Rees, D.C. & Adams, M.W.W. (1995). Molecular characterization of genes encoding the tungsten-containing aldehyde oxidoreductase from *Pyrococcus furiosus* and formaldehyde ferredoxin oxidoreductase from *Thermococcus litoralis*. *J. Bacteriol.* **177**, 4817-4819.
21. Kawasaki, S., Arai, H., Kodama, T. & Igarashi, Y. (1997). Gene cluster for dissimilatory nitrite reductase (*nir*) from *Pseudomonas aeruginosa*: Sequence and identification of a locus for heme d_1 biosynthesis. *J. Bacteriol.* **179**, 235-242.
22. Meulenberg, J.J., Sellink, E., Riegman, N.H. & Postman, P.W. (1992). Nucleotide sequence and structure of the *klebsiella pneumoniae pqq* operon. *Mol. Gen. Genet.* **232**, 284-294.
23. Burland, V., Plunket III, G. & Blattner, F.R. (1993). DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: Organizational symmetry around the origin of replication. *Genomics* **16**, 551-561.
24. Flachmann, R., *et al.*, & Gassen, H.G. (1988). Molecular biology of pyridine nucleotide biosynthesis in *Escherichia coli*: Cloning and characterization of quinolinate synthesis genes *nadA* and *nadB*. *Eur. J. Biochem.* **175**, 221-228.
25. Semrau, J.D., *et al.*, & Lidstrom, M.E. (1995). Particulate methane monooxygenase genes in methanotrophs. *J. Bacteriol.* **177**, 3071-3079.
26. Mortarino, M., *et al.*, & Ronchi, S. (1996). L-aspartate oxidase from *Escherichia coli*. I. Characterization of coenzyme binding and product inhibition. *Eur. J. Biochem.* **139**, 418-426.
27. Kupke, T., Kemptner, C., Jung, G. & Goetz, F. (1995). Oxidative decarboxylation of peptides catalyzed by flavoprotein EpiD. *J. Biol. Chem.* **270**, 11282-11289.
28. Daniels, D.L., PLunkett III, G., Burland, V. & Blattner, F.R. (1992). Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* **257**, 771-778.
29. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
30. Thompson, J.D., Higgins, D.G. & Gibbon, T.J. (1994). ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.