How Does Nature Cleave Sulfuric Acid Esters? A Novel Posttranslational Modification of Sulfatases

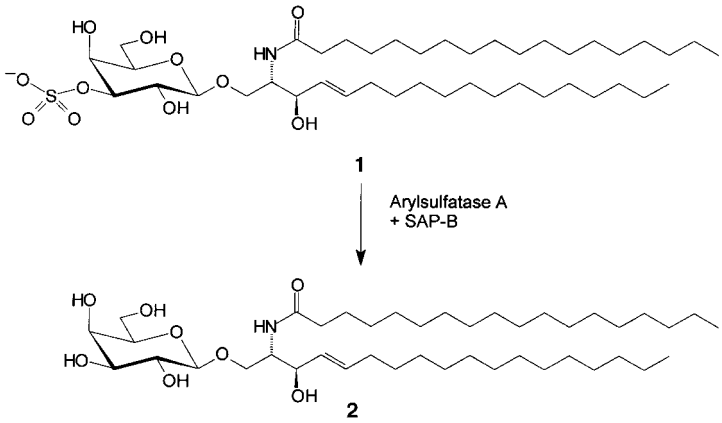
Gunther Uhlhorn-Dierks, Thomas Kolter, and Konrad Sandhoff*

Alkyl- and arylsulfates are relatively stable compounds that can only be cleaved under drastic conditions, for example by alkaline hydrolysis. They are widely used as anionic detergents. Sulfuric acid esters of alcohols exist as naturally occurring substances, for example in heparin, in the mucopolysaccharides chondroitin sulfate and keratan sulfate, in sulfated steroids, and as tyrosine sulfate in the peptide hormones gastrin and cholecystokinin. Conjugation with sulfate is also of relevance for drug metabolism. The analysis of an inherited disease by the group of Kurt von Figura has led to the basic understanding of the enzymatic reactions involved in sulfuric acid ester hydrolysis. A new posttranslational protein modification required for the catalytic mechanism has been identified.[1]

Nine different sulfatases have been identified in humans that are necessary for the enzymatic degradation of organic sulfuric acid esters.[2, 3] With the exception of steroid sulfatase they are localized in the lysosomes of the cell. Different inherited disorders are known that are caused by the deficiency of distinct sulfatases leading to a tissue-specific accumulation of the nondegradable sulfuric acid ester. Metachromatic leukodystrophy (MLD), mucopolysaccharidosis, and ichthyosis can be distinguished.

It was demonstrated in patients suffering from MLD that arylsulfatase A (ASA) is defect.[4, 5] The name arylsulfatase is based on their ability to cleave artificial aromatic sulfuric acid esters. Arylsulfatase A cleaves sulfatides in the presence of a cofactor, a sphingolipid activator protein (SAP-B) (Scheme 1).[6] MLD leads to the accumulation of sulfatide (galactocerebroside sulfate) in different organs, a loss of the cerebral white matter of the patients, and to death during childhood. The name "metachromatic leukodystrophy" results from the metachromatic stain of the storage material in histological sections, such that on staining with cresyl violet, the absorption maximum is changed so that the deposits appear brown.[4]

In contrast to MLD, the activities of all known sulfatases are severely decreased in a rare disorder called Austin's disease or multiple sulfatase deficiency (MSD). The clinical manifestation is a combination of symptoms of individual sulfatase deficiencies. The genetic basis of MSD is still unknown. Complementation studies have shown that the genes for single sulfatases are not affected and that it is not a fortuitous common failure of two or more enzymes.[7, 8] After expression of the complementary DNA (cDNA) from an intact sulfatase in MSD patient cells the resulting enzymatic activity appears to be severely reduced.[9] These experimental results led to the proposal of a secondary defect such as an incomplete co- or posttranslational modification that is necessary for the catalytic activity of all sulfatases.

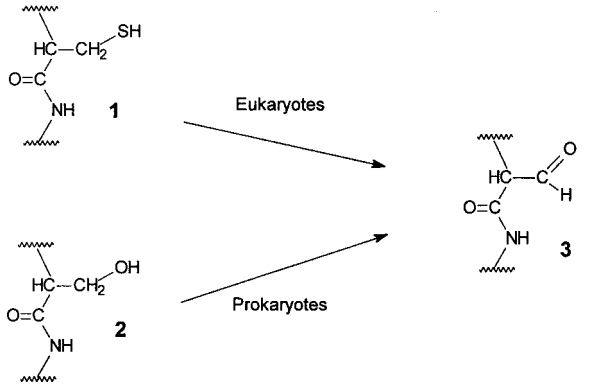


Scheme 1. Hydrolysis of sulfatide **1** to give galactosylceramide **2**. The reaction is catalyzed by arylsulfatase A. Sphingolipid activator protein B (SAP-B) is a necessary cofactor. The deficiency of arylsulfatase A leads to metachromatic leucodystrophy.

With this in mind the enzymatic active arylsufatase A was purified and analyzed for the presence of such a posttranslational modification.[1] After digestion with the protease trypsin and enzymatic removal of N-glycosydically linked oligosaccharides a peptide was isolated with a mass different from that predicted by the DNA sequence. The mass of this peptide, which comprises the amino acids 59 to 73 of ASA, was amazingly 74 Da less than expected after reductive carboxymethylation. In general posttranslational modifications result in mass increases—apart from proteolytic processing or glycane trimming. By sequencing, metabolic labeling with [^{35}S]methionine, and MALDI-MS it could be shown that this peptide (ASA 59 to 73) contains no cysteine residue at position 69 as predicted by the ASA cDNA, but a formylglycine residue (2-amino-3-oxopropanoic acid, Scheme 2). After reduction with $NaBH_4$ the peptide mass increased by two units and serine was identified at position 69 through sequencing.

[*] Prof. Dr. K. Sandhoff, G. Uhlhorn-Dierks, Dr. T. Kolter
Kekulé-Institut für Organische Chemie und Biochemie der Universität
Gerhard-Domagk-Strasse 1, D-53121 Bonn (Germany)
Fax: (+49) 228-73-7778
E-mail: sandhoff@uni-bonn.de

  1433-7851/98/3718-2453 $ 17.50+.50/0



Scheme 2. Posttranslational modification of the amino acid cysteine **1** (in eukaryotes) and serine **2** (in specific prokaryotes: *Klebsiella pneumoniae* and *Escherichia coli*) to give formylglycine **3**.

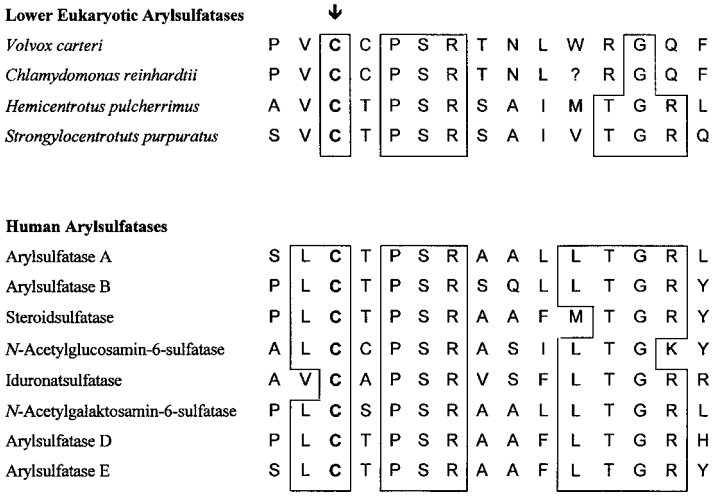


Figure 1. Comparison of the sequences from different sulfatases. The cysteine residue which is converted into formylglycine is marked by an arrow.

The same modification was also detected in the human N-acetylgalactosamine-4-sulfatase (arylsulfatase B, ASB)[1] and in the arylsulfatase from the green alga *Volvox carteri*.[10] This indicates the evolutionary conservation of this protein modification in eukaryotes.

Which amino acids contain the sulfatases from patients with MSD, the starting point for this investigation? In arylsulfatase A and B overexpressed in MSD cells, mainly the unmodified cysteine residue instead of formylglycine was found. However up to 20% of the investigated sulfatases were modified.[1] Therefore, MSD has to be regarded as a partial defect of a posttranslational modification resulting in an incomplete loss of catalytically active sulfatases.

The identity of the modifying enzyme that catalyzes the conversion of protein-bound cysteine to formylglycine is not yet known. Since it has been shown that the synthesis of the cysteine-containing precursors is not impaired, a modification on the level of a putative formylglycine containing aminoacyl tRNA can be excluded. The transformation of the cysteine residue occurs after synthesis of the peptide backbone and its translocation into the lumen of the endoplasmic reticulum, where N-glycosylation also takes place. With the aid of peptide constructs it could be demonstrated that formylglycine is formed after N-glycosylation.[11]

The signal for posttranslational modification is often coded within the amino acid sequence. In the case of arylsulfatase A the trigger for modification was localized by construction of chimeric proteins from ASA sequences and non-ASA sequences. A linear sequence of sixteen amino acids (four in front of and eleven after the cysteine) is sufficient to initiate the oxidation of the cysteine residue to the formylglycine residue.[11] The amino acid sequence in this region is strictly conserved among all eukaryotic sulfatases (Figure 1).

It is not directly evident, how a sulfuric acid ester can be cleaved with the assistance of an aldehyde. Two different reaction mechanisms are conceivable: first the nucleophilic attack of a sulfate oxygen at the aldehyde; second the nucleophilic substitution at the sulfur atom by an enol or hydrate of the aldehyde. The crystal structure of arylsulfatase A[12] and B[13] as well as a biologically verified site-directed mutated sulfatase[14] provide information on this problem. The recently published crystal structure shows that arylsulfatase A exists as a homo-octamer composed of a tetramer of dimers $(\alpha_2)_4$.[12] The structure revealed great similarity to alkaline phosphatase although the sequence homology is rather low. In its assumed active site a magnesium ion is found which is coordinated to the oxygen atom of formylglycine (Figure 2). The observed

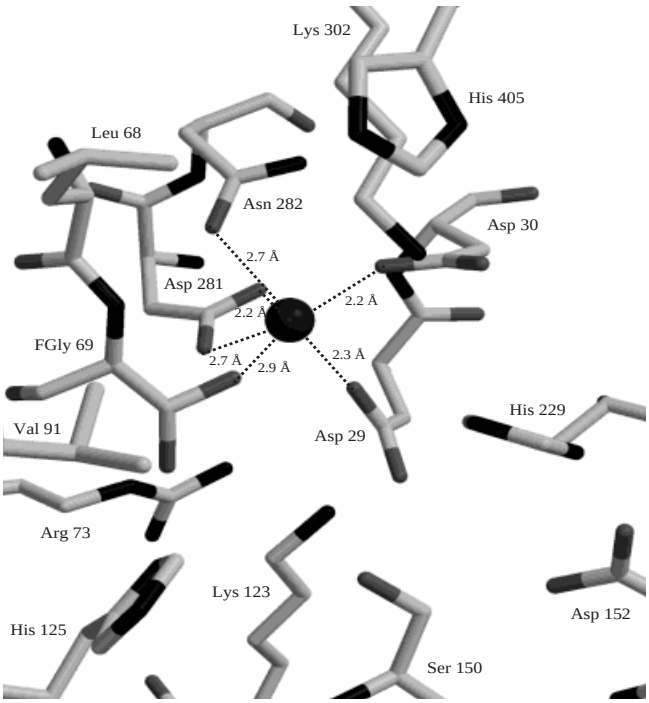
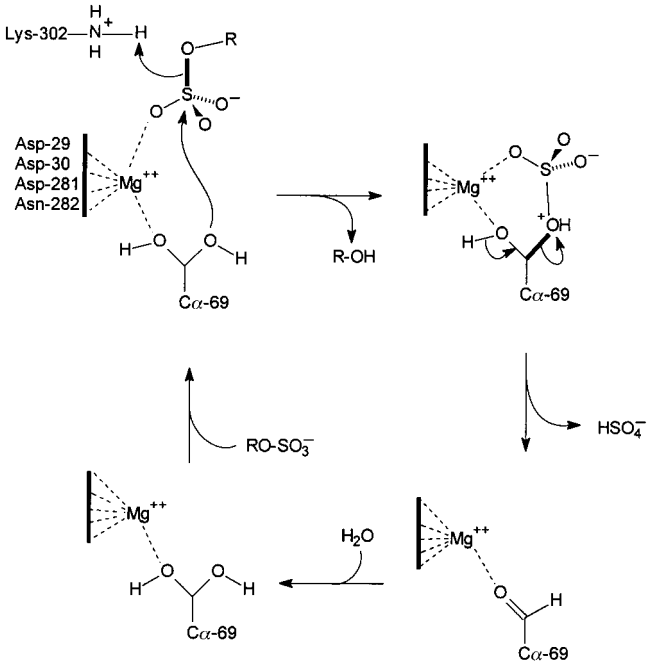


Figure 2. Part of the three-dimensional structure of human arylsulfatase A.[12] The distances between selected side chain atoms and the central magnesium ion are given.

electron density was interpreted as an aldehyde hydrate by the authors. In alkaline phosphatase two zinc ions are in the active site and coordinate the phosphate; the catalytic active amino acid is a serine.

A hypothetical mechanism of sulfate cleavage is shown in Scheme 3. One of the geminal oxygen atoms of the aldehyde

  1433-7851/98/3718-2454 $ 17.50+.50/0



Scheme 3. Hypothetical mechanism of sulfuric acid ester hydrolysis (modified according to ref. [12]).

performs a nucleophilic attack on the sulfur atom of the sulfate; the galactosyl ceramide residue serves as the leaving group. The magnesium ion increases the electrophilicity of the sulfur atom and facilitates the nucleophilic attack of the hydrate oxygen atom leading to the formation of a sulfated enzyme intermediate. Such an intermediate was postulated earlier on the grounds of the anomalous kinetics of arylsulfatase A.[15] With [^{35}S]-labeled substrate it could be shown that ASA binds sulfate. Further characterization of the modified enzyme failed since the label was lost after denaturation. It was considered that the sulfate group is not present in an ester linkage which would be stable under denaturing conditions.[16, 17]

Taking the recent findings into account an explanation is as follows: the driving force for the fission of the sulfated enzyme is the reformation of the carbonyl bond of formylglycine. No hydrolysis of a stable S–O bond is required, instead the less stable C–O bond is broken. The conclusions derived from the crystal structure concerning the catalytic mechanism are supported by experiments with a modified enzyme. The cysteine residue in the precursor protein of ASA and ASB was replaced by serine by site-directed mutagenesis. This serine residue is not oxidized to formylglycine in the cell lines used for the expression. The generated mutants are able to cleave [^{35}S]-labeled substrate, however the sulfate remains tightly bound to the enzyme.[14] The serine oxygen atom is sufficiently nucleophilic to cleave the sulfuric acid ester but the assistance of a second oxygen atom to release the enzyme-bound sulfate is lacking. Only in alkaline environment and after the addition of galactose as acceptor for the released sulfate a desulfatation of the enzyme takes place and a catalytic cycle starts. The proposed mechanism is comparable to that of alkaline phosphatase. First a serine residue attacks the phosphoric acid ester of the substrate. In the second step the resulting phosphoserine is hydrolyzed by a water molecule. In sulfatases, a second hydroxyl group of the hydrated formylglycine is required for sulfate release.

The cysteine residue, which is posttranslationally converted to formylglycine, is highly conserved in all known eukaryotic sulfatases (see Figure 1).[10] However, among the five prokaryotic sulfatase sequences known to date there are two exceptions. In *Klebsiella pneumoniae* and *Escherichia coli,* the DNA codes for a serine instead of cysteine. The characterization of *Klebsiella pneumoniae* sulfatase on the protein level showed that the initially incorporated serine is converted to a formylglycine residue (see Scheme 2).[18] The incomplete conversion is explained with saturation of the enzymes necessary for the transformation. In contrast to eukaryotic cells, which cannot convert a serine residue, this amino acid serves as a precursor for the catalytically active amino acid in prokaryotic cells. So far it is not known whether *Klebsiella pneumoniae* can also modify a cysteine residue into a formylglycine residue.

The recent results should enable the identification of the enzymes that catalyze the oxidation of cysteine residues or serine residues by suitable methods such as that of the two-hybrid system. This would also lead to an improved understanding of the molecular basis of multiple sulfatase deficiency which is assumed to be caused by a defective posttranslational modification. The quoted articles illustrate how the analysis of inherited diseases can lead to the discovery of new chemical and biochemical principles.

German version: *Angew. Chem.* **1998**, *110*, 2591 – 2594

Keywords: enzyme catalysis · glycolipids · proteins · sulfatases

[1] B. Schmidt, T. Selmer, A. Ingendoh, K. von Figura, *Cell* **1995**, *82*, 271 – 278.
[2] E. F. Neufeld, J. Muenzer in *The Metabolic Bases of Inherited Disease* (Eds.: C. Scriver, A. L. Beaudet, W. S. Sly, D. Valle), McGraw-Hill, New York, **1995**, S. 2465 – 2494.
[3] A. Ballabio, L. J. Shapiro in ref. [2], pp. 2999 – 3022.
[4] E. H. Kolodny, A. L. Fluharty in ref. [2], pp. 2693 – 2741.
[5] J. H. Austin, A. S. Balasubramanian, T. N. Paattabiraman, S. Saraswathi, D. K. Basu, B. K. Bachhawat, *J. Neurochem.* **1963**, *10*, 805 – 816.
[6] E. Mehl, H. Jatzkewitz, *Hoppe-Seyler's Z. Physiol. Chem.* **1964**, *339*, 260 – 276.
[7] K. Fedde, A. L. Horwitz, *Am. J. Hum. Genet.* **1984**, *36*, 623 – 633.
[8] A. Ballabio, G. Parenti, E. Napolitano, P. Di Natale, G. Andria, *Hum. Genet.* **1985**, *70*, 315 – 317.
[9] W. Rommerskirch, K. von Figura, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2561 – 2565.
[10] T. Selmer, A. Hallmann, B. Schmidt, M. Sumper, K. von Figura, *Eur. J. Biochem.* **1996**, *238*, 341 – 345.
[11] T. Dierks, B. Schmidt, K. von Figura, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11 963 – 11 968.
[12] G. Lukatela, N. Krauss, K. Theis, T. Selmer, V. Gieselmann, K. von Figura, W. Saenger, *Biochemistry* **1998**, *37*, 3654 – 3664.
[13] C. S. Bond, P. R. Clements, S. J. Ashby, C. A. Collyer, S. J. Harrop, J. J. Hopwood, J. M. Guss, *Structure* **1997**, *5*, 277 – 289.
[14] M. Recksiek, T. Selmer, T. Dierks, B. Schmidt, K. von Figura, *J. Biol. Chem.* **1998**, *273*, 6096 – 6103.
[15] C. I. Prosser, A. B. Roy, *Biochim. Biophys. Acta* **1980**, *613*, 130 – 139.
[16] A. Waheed, R. L. Van Etten, *Arch. Biochem. Biophys.* **1980**, *203*, 11 – 24.
[17] A. B. Roy, T. J. Mantle, *Biochem. J.* **1989**, *261*, 689 – 697.
[18] C. Miech, T. Dierks, T. Selmer, K. von Figura, B. Schmidt, *J. Biol. Chem.* **1998**, *273*, 4835 – 4837.

  1433-7851/98/3718-2455 $ 17.50+.50/0