Post-translational generation of Cα-Formylglycine in Prokaryotic Sulfatases by Radical SAM-Proteins

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

Vorgelegt von
Qinghua FANG
aus Shanghai, CHINA

Göttingen 2003

D7

Referent: Prof. Dr. Dr. h. c. K. von Figura

Korreferent: Prof. Dr. H.-J. Fritz

Tag der mündlichen Prüfung:

Abbreviations

AHT	anhydrotetracycline
ASA	Human arylsulfatase A
AtsA	*K pneumoniae* arylsulfatase
AtsB	*K pneumoniae* arylsufatase activation protein
BPB	Brom phenol blue
CCS	putative *C. crescentus* sulfatase
DTT	Dithiothreitol
DVS	putative *D. vulgaris* sulfatase
D. vulgaris	*Desulfavibrio vulgaris*
DHB	2,5-Dihydroxybenzonic acid
DNPH	2,4-Dinitrophenylhydrasine
FGE	FGly generation enzyme
FGly	Cα-formylglycine
GSH	glutathione
GST	glutathione S-transferase
HPLC	High performance liquid chromatography
IPTG	isopropyl ß-D-thiogalactopyranoside
K. pneumoniae	*Klebsiella pneumoniae*
MADLI-MS	matrix-assisted laser desorptional ionization time-of-flight mass spectrometry
M. mazei	*Methanosarcina mazei*
MMS	Putative *M. mazei* sulfatase
MSD	Multiple sulfatase deficiency
NTA	nitrilotriacetic acid
PBS	phosphat-buffered solution
P. aeruginosa	*Pseudomonas aeruginosa*
PAS	*P. aeruginosa* arylsulfatase
pNCS	*p*-nitrocatechol sulfate
P2	cysteine/serine containing form of peptide 2

P2*	FGly-containing form of peptide 2
RT	room temperature
SAHC	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SUMF	Sulfatase Modifying Factor
TBS	Tris-buffered solution
X-gal	5-bromo-4-chloro-2-indolyl-ß-D-galactopyranosida

Contents

1 Introduction **1**
  1.1 The sulfatases 1
    1.1.1 Bacterial sulfatases 1
    1.1.2 Human sulfatases and associated diseases 3
    1.1.3 Formylglycine in eu- and prokaryotic sulfatases 6
    1.1.4 Formylglycine-hydrate in the active site of sulfatases 8
    1.1.5 Mechanism of formylglycine-hydrate mediated sulfate ester hydrolysis 8
    1.1.6 Sequence motif directing formylglycine formation 10
  1.2 Formylglycine generating machineries 11
    1.2.1 AtsB: the putative serine-modifying system 11
    1.2.2 FGE: a cysteine-modifying system of pro- and eukaryotes 14
    1.2.3 X: unknown cysteine-modifying system 16
  1.3 Goals and experimental approach 17

2. Materials **20**
  2.1 Appliances 20
  2.2 Chemicals 20
  2.3 Radioactively labelled substances 20
  2.4 Enzymes and standards 20
  2.5 Protein determination, translation and transcription Kits, DNA Kits 20
  2.6 Antibody 21
    2.6.1 Primary antibodies 21
    2.6.2 Secondary antibodies 21
  2.7 Bacterial cell culture media 21
  2.8 Frequently used buffers and solutions 22
  2.9 Plasmid vectors 22
  2.10 *Escherichia coli* strains 23

3. Methods **24**
  3.1 *E. coli* culture 24
  3.2 Preparation of DNA 24
    3.2.1 Genomic DNA preparation from *Caulobacter crescentus* and *Desulfovibrio vulgaris* 24
    3.2.2 Analytical isolation of plasmid DNA from *E. coli* cells (Qiagen handbook) 25
    3.2.3 Preparative isolation of DNA from *E. coli* cells (Qiagen handbook) 25
  3.3 Purification of DNA 26
    3.3.1 Ethanol precipitation 26
    3.3.2 Phenol/Chloroform-extraction 26
    3.3.3 Purification of DNA-fragments by gel-extraction 26
    3.3.4 PCR purification 27
  3.4 Determination of concentration and the purity of DNA 27
  3.5 Electrophoretic separation of DNA fragments in agarose gels 27
  3.6 Digestion of DNA with restriction nucleases 28
  3.7 Dephosphorylation of DNA 28

3.8 Filling up overhanging DNA ending 29
3.9 Cloning of recombined DNA in *E. coli* 29
    3.9.1 Ligation of vector and insert DNA 29
    3.9.2 Preparation of competent cells 29
    3.9.3 Transformation by electroporation 29
3.10 PCR amplification of DNA 30
    3.10.1 Mutagenesis 31
    3.10.2 Colony PCR to identify positive clone 31
    3.10.3 Sequencing of DNA 31
3.11 Protein expression 32
    3.11.1 Induction of *E. coli* by IPTG 32
    3.11.2 Induction of *E. col*i by AHT 32
3.12 Cell lysis 32
    3.12.1 Lysis by Ultra sonificaiton 32
    3.12.2 Lysis by "French Press" 33
3.13. Protein purification 33
    3.13.1 6xHis-fusion protein purification 33
    3.13.2 GST-fusion protein purification 34
    3.13.3 Strep-fusion protein purification 34
3.14 Protein analysis and detection 35
    3.14.1 SDS-polyacrylamide gel electrophoresis 36
    3.14.2 Detection of protein in polyacrylamide gels 36
        3.14.2.1 Staining with Roti-Blue colloidal Coomassie 36
        3.14.2.2 Staining with silver (Bassam *et al.*, 1991) 37
        3.14.2.3 Western blot on PVDF membrane 37
        3.14.2.4 Determination of protein concentration 38
    3.14.3 Arylsulfatase activity assay (Beil *et al.*, 1995) 38
3.15 Antibody production 39
3.16 Pull-down assay 39
3.17 *In vitro* formylglycine generation assay 40
3.18 Identification of proteins by matrix-assisted laser desorption ionisation / time-of-flight (MALDI-TOF) mass spectrometry 41
    3.18.1 In-gel digestion of protein separated by SDS-PAGE 41
    3.18.2 Extraction of tryptic peptides from the gel pieces 41
    3.18.3 ZipTip purification of the trypic peptides before analysis by MALDI-TOF 41
    3.18.4 Sample/matrix preparation of MALDI-MS (drying droplet method) 42
    3.18.5 Obtaining the mass spectra on MALDI-TOF 43

4. Results 44
4.1 *In vivo* formylglycine modification of *Klebsiella* sulfatase AtsA promoted by AtsB and its orthologs 44
    4.1.1 Coexpression of AtsA in *E. coli* together with AtsB and its orthologs 44
    4.1.2 formylglycine generation in AtsA mediated by AtsB and its orthologs 46
4.2 Purification of AtsB 48
    4.2.1 Purification of GST-AtsB 48
        4.2.1.1 Cloning of GST-AtsB 48

4.2.1.2 GST-AtsB expression, *in vivo* activity and purification 48
4.2.2 Purification of Strep-AtsB and Strep-AtsB-His_6 53
4.3 Physical interaction of AtsB with AtsA 55
4.3.1 Cloning of GST-AtsA-(21-112) 56
4.3.2 Interaction of AtsB with the formylglycine-modification motif of GST-AtsA-(21-112) 56
4.4 Substrate specificity of AtsB 58
4.4.1 Interaction of AtsB with putative serine-type sulfatases of *E. coli* 58
4.4.2 AtsB can modify serine- and cysteine-type sulfatases 60
4.4.2.1 AtsB and MM-AtsB can modify the cysteine-type sulfatase MMS 60
4.4.2.2 AtsB can not modify the threonine-type sulfatase DVS 63
4.5 Requirement of AtsA's signal peptide for AtsB-mediated formylglycine modification 65
4.5.1 AtsB coexpression does not lead to activation of PAS-C51S 65
4.5.2 AtsB-dependent formylglycine formation in AtsA is strongly reduced after removal of AtsA's signal peptide 67
4.5.3 AtsB-dependent expression of active signal peptide-containing PAS-C51S 69
4.6 Requirement of S-adenosylmethionine for formation of a functional AtsB/AtsA complex 70
4.6.1 SAM improves binding of AtsB to the formylglycine-modification motif of AtsA 70
4.6.2 Formation of a ternary AtsB/SAM/AtsA-complex 72
4.7 SAM-dependent *in vitro* formylglycine-formation by AtsB 73
4.8 ^{83}GGE^{85} of AtsB is required for SAM-binding, AtsA-interaction and formylglycine-formation 77
4.9 Requirement of conserved cysteine residues in AtsB for formylglycine-generation *in vivo* and *in vitro* 80

5. Discussions 83
5.1 AtsB is a prokaryotic formylglycine-generation enzyme 83
5.2 AtsB-dependent and AtsB-independent formylglycine-modification 84
5.3 Cell-biological aspects of AtsB-mediated formylglycine-formation 85
5.4 Mechanism of AtsB-mediated formylglycine-modification 88

6. Summary **93**

References **95**

1. Introduction

1.1 The sulfatases

A large proportion of the sulfur in various soil and sediment environments is present as sulfate esters. In addition, sulfate ester also can be found in a variety of biological compounds, such as glycosaminoglycans (dermatan sulfate, chondroitin sulfate, keratan sulfate, and heparan sulfate), sulfated glycolipids (cerebroside sulfate), tyrosine-sulfate-carrying proteins, and sulfated hydroxysteroids (cholesterol sulfate and dehydroepiandrosterone sulfate).

These sulfate esters are hydrolysed by sulfatases, which are found in both prokaryotes and eukaryotes. Mammalian sulfatases are involved in the turnover of endogenous sulfated substrates (von Figura *et al.*, 1998). On the other hand, in lower eukaryotes and prokaryotes, sulfatases are expressed under conditions of sulfur starvation and function in sulfate scavenging from exogenous substrates (Dodgson *et al.*, 1982).

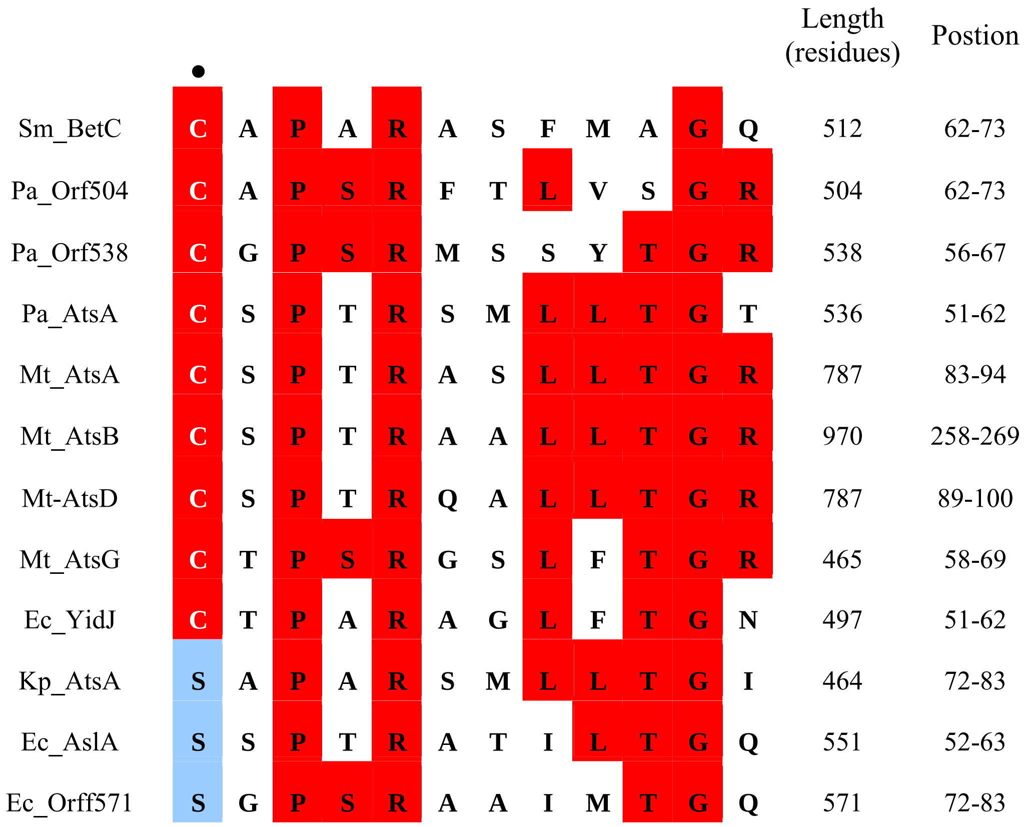
The sulfatases catalyze the hydrolysis of sulfate esters and sulfamides (O-sulfates and N-sulfates).

$R\text{-}OSO_3^- + H_2O \rightarrow R\text{-}OH + H^+ + SO_4^{2-}$

$R\text{-}NHSO_3^- + H_2O \rightarrow R\text{-}NH_3^+ + SO_4^{2-}$

1.1.1 Bacterial sulfatases

Bacterial sulfatases have been identified and studied in a variety of species (Figure 1.1), including enterobacteria (*Klebsiella, Salmonella*, *Enterobacter*, *Proteus*, *Serratia*), and pseudomonads (*Pseudomonas*, *Comamonas*, *Mycobacteria*, and Cyanobacteria). In recent years, genetic and biochemical studies have concentrated primarily on the arylsulfatase enzymes AtsA from *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*, at least in part because these enzymes are closely related to the human arylsulfatases.



	●												Length (residues)	Postion
Sm_BetC	C	A	P	A	R	A	S	F	M	A	G	Q	512	62-73
Pa_Orf504	C	A	P	S	R	F	T	L	V	S	G	R	504	62-73
Pa_Orf538	C	G	P	S	R	M	S	S	Y	T	G	R	538	56-67
Pa_AtsA	C	S	P	T	R	S	M	L	L	T	G	T	536	51-62
Mt_AtsA	C	S	P	T	R	A	S	L	L	T	G	R	787	83-94
Mt_AtsB	C	S	P	T	R	A	A	L	L	T	G	R	970	258-269
Mt-AtsD	C	S	P	T	R	Q	A	L	L	T	G	R	787	89-100
Mt_AtsG	C	T	P	S	R	G	S	L	F	T	G	R	465	58-69
Ec_YidJ	C	T	P	A	R	A	G	L	F	T	G	N	497	51-62
Kp_AtsA	S	A	P	A	R	S	M	L	L	T	G	I	464	72-83
Ec_AslA	S	S	P	T	R	A	T	I	L	T	G	Q	551	52-63
Ec_Orff571	S	G	P	S	R	A	A	I	M	T	G	Q	571	72-83

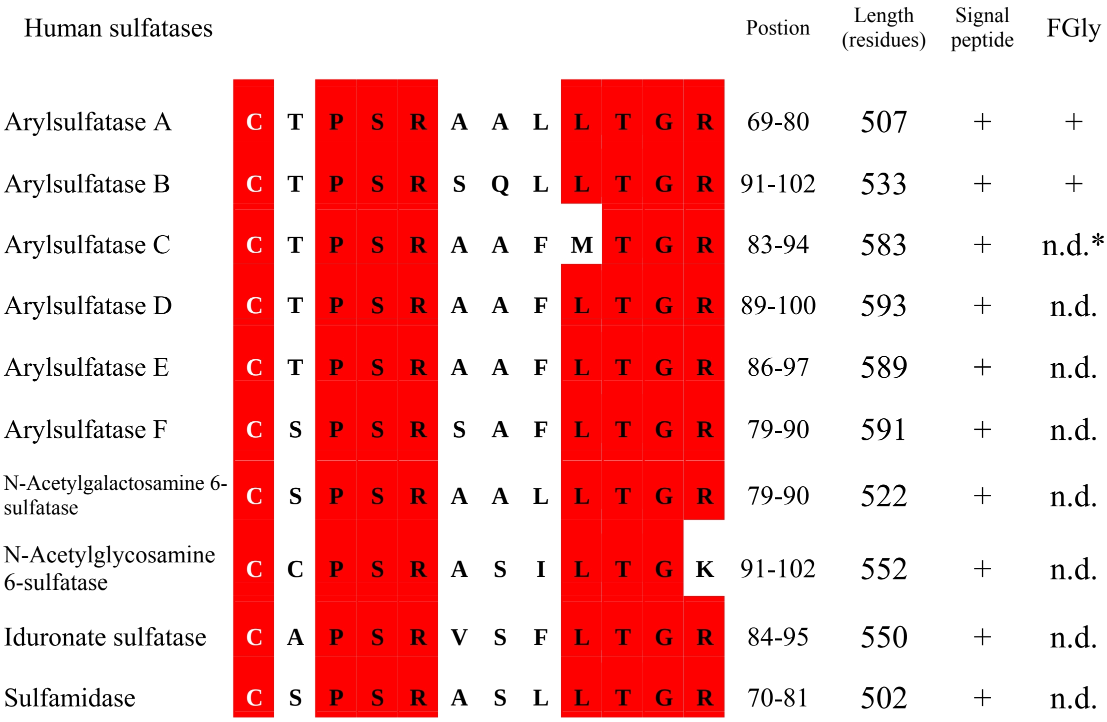
Figure 1.1 Sequence similarities among members of bacterial sulfatases and related proteins representing putative sulfatases. The proteins shown are: *S. Meliloti* BetC, *P. aeruginosa* Orf504, Orf538 and AtsA, *M. tuberculosis* AtsA, AtsB, AtsD, AtsG, *K. pneumoniae* AtsA, *E. coli*YidJ, AslA and Orf f571. The cysteine/serine that is or may be modified to formylglycine is indicated with a dot. The sequence shown represents the region homologous to the motif directing FGly formation in human sulfatases, thereby serving as a sulfatase signature. Residues boxed in grey are identical to the human semiconsensus sequence CxPSRxxLLTGR (see Figure 1.2).

These two proteins are relatively similar at the protein sequence level (37 % amino acid identity), and in size (60 kDa for the *Pseudomonas* enzyme, and 62 KDa for the mature *Klebsiella* sulfatase). However, there are also significant differences. The *Klebsiella* sulfatase is a periplasmic enzyme, and carries a typical signal peptide, whereas the *Pseudomonas* enzyme is an intracellular protein, since the N-terminal amino acid sequence corresponds to that encoded by the 5'-end of the gene (Kertesz *et al.*, 1999). The pH optimum for the *Klebsiella* enzyme is 7.4 (Okamura *et al.*, 1976), whereas the *Pseudomonas* sulfatase shows highest activity at a more alkaline pH value of 8.9 (Beil *et al.*, 1995).

Arylsulfatases are also commonly found in multiple forms in the mycobacteria, but there has been no attempt to characterize the enzymes in detail. However, arylsulfatase activity patterns have been used as a means of identifying and distinguishing different Mycobacterium species, and the presence of arylsulfatase is routinely used as phenotypic test for the genus Mycobacterium.

1.1.2 Human sulfatases and associated diseases

Twelve sulfatases have been characterized biochemically in human (Table 1.1), ten of which have been identified at the molecular level (Figure 1.2). The biological importance of the human sulfatases is highlighted by the manifestation of eight known inherited metabolic diseases that are caused by the specific deficiency of single sulfatases. Each of these disorders is associated with the impaired desulfation of specific sulfated metabolites (Table 1.1). For example, Arysulfatase A, localized in the lysosomes, catalyzes the initial step in the catabolism of cerebroside 3-sulfate and some other related sulfolipids that share a galactose 3-sulfate residue. Deficiency of this sulfatase leads to metachromatic leukodystrophy, a lysosomal storage disorder that results in severe neurological symptoms because of extensive demyelination in the central and peripheral nervous system (von Figura *et al.*, 2001). Arylsulfatase B desulfates N-acetylgalactosamine 4-sulfate residues in the degradative pathways of dermatan sulfate and chondroitin 4-sulfate (Neufeld *et al.*, 2001). Arylsulfatase C is a membrane-bound enzyme of the endoplasmic reticulum and the plasma membrane and is involved in cholesterol metabolism and the biosynthesis of steroid hormones. Deficiency of arylsulfatase C leads to X-linked ichthyosis, a relatively mild disease of the skin (Ballabio *et al.*, 2001). Arylsulfatases D, E, and F are non-lysosomal membrane proteins, which have been located to the endoplasmic reticulum (arylsulfatases D and F) or the Golgi apparatus (arylsulfatase E) (Parenti *et al.*, 1997). The natural substrates of arylsulfatases D, E, and F are not known, but deficiency of arylsulfatase E leads to X-linked chondrodysplasia punctata, a disorder of bone and cartilage (Franco *et al.*, 1995; Daniele *et al.*, 1998). A newly identified member of the sulfatase family, Sulf1 (KIAA 1077), is localized on the cell surface and was shown to be involved in regulation of Wnt signalling during embryo patterning (Dhoot *et al.*, 2001; Ai *et al.*, 2003). By heparan sulfate desulfation this sulfatase triggers Wnt release from heparansulfate proteoglycans at the cell surface, a step required for embryonic muscle specification.



Human sulfatases													Postion	Length (residues)	Signal peptide	FGly
Arylsulfatase A	C	T	P	S	R	A	A	L	L	T	G	R	69-80	507	+	+
Arylsulfatase B	C	T	P	S	R	S	Q	L	L	T	G	R	91-102	533	+	+
Arylsulfatase C	C	T	P	S	R	A	A	F	M	T	G	R	83-94	583	+	n.d.*
Arylsulfatase D	C	T	P	S	R	A	A	F	L	T	G	R	89-100	593	+	n.d.
Arylsulfatase E	C	T	P	S	R	A	A	F	L	T	G	R	86-97	589	+	n.d.
Arylsulfatase F	C	S	P	S	R	S	A	F	L	T	G	R	79-90	591	+	n.d.
N-Acetylgalactosamine 6-sulfatase	C	S	P	S	R	A	A	L	L	T	G	R	79-90	522	+	n.d.
N-Acetylglycosamine 6-sulfatase	C	C	P	S	R	A	S	I	L	T	G	K	91-102	552	+	n.d.
Iduronate sulfatase	C	A	P	S	R	V	S	F	L	T	G	R	84-95	550	+	n.d.
Sulfamidase	C	S	P	S	R	A	S	L	L	T	G	R	70-81	502	+	n.d.

Figure 1.2 Sequence similarities among members of the human sulfatase protein family downstream of the cysteine to be converted to FGly. Amino

acid residues, conserved in more than 60 % of human sulfatases are indicted by blocks. * n.d.- not determined

In addition to single sulfatase deficiencies, a rare inherited disorder named multiple sulfatase deficiency (MSD) is known, in which the activities of all sulfatases are severely reduced. The single sulfatase deficiency syndroms are caused by mutations in the corresponding sulfatase genes. However, the molecular basis for MSD was identified as a deficiency of a novel posttranslational protein modification that is essential for all sulfatases to gain their enzymatic activities. This modification leads to the formation of an active site Cα-formylglycine (FGly) residue and is catalyzed by a cysteine-modifying enzyme located in the endoplasmic reticulum.

Recently, the FGE (FGly generating enzyme) was identified biochemically (Dierks *et al.*, 2003) and also by functional complementation of MSD cells using microcell-mediated chromosome transfer (Cosma *et al.*, 2003). In all MSD patients analyzed, mutations were found in the *SUMF1* gene (Sulfatase Modifying Factor 1), which encodes FGE. All mutations have severe effects on the FGE protein by either replacing highly conserved residues or by causing C-terminal truncations of various lengths or large in-frame deletions (Dierks *et al.*, 2003). Expression of FGE encoding cDNA in MSD fibroblasts partly (*e.g.* arylsulfatase A) or fully (steroid sulfatase) restored sulfatase activity (Dierks *et al.*, 2003; Cosma *et al.*, 2003). These data indicate that the FGE encoded by the *SUMF* 1 gene is responsible for the posttranslational modification required for sulfatase activity. And losing functional FGE leads to MSD.

Table 1.1 Human sulfatases, their subcelluar localization, natural substrates and metabolic disorders caused by their deficiencies [a]

Sulfatase	Subcellular localization	Natural substrate	Disease
Arylsulfatase A	Lysosomes	Cerebroside 3-sulfate	Metachromatic leukodystrophy
Arylsulfatase B	Lysosomes	Dermatan sulfate, Chondroitin 4-sulfate	MPS VI (MaroteauxLamy)
N-Acetylgalactosamine 6-sulfatase (Galactose 6-sulfatase)	Lysosomes	Keratan sulfate, Chondroitin 6-sulfate	MPS IV A (Morquio A)
Exo-N-Acetylglucosamine 6-sulfatase	Lysosomes	Heparan sulfate, Keratan sulfate	MPS III D (Sanfilippo D)
Iduronate 2-sulfatase	Lysosomes	Dermatan sulfate, Heparan sulfate	MPSII (Hunter)
Sulfamidase (Heparan N-Sulfatase)	Lysosomes	Heparan sulfate	MPS IIIA (Sanfilippo A)
Glucuronate 2-Sulfatase [b]	Lysosomes	Heparan sulfate	not established
N-Acetylglucosamine 3-sulfatase [b]	Lysosomes	Heparan sulfate	not established
Arylsulfatase C (Steroid sulfatase)	Endolasmic reticulum, plasma membrane	3-β-Hydroxysteroid sulfates	X-linked ichthyosis
Aryslulfatase D	Endoplasmic reticulum	Unknown	Not established
Arylsulfatase E	Golgi apparatus	Unknown	Chondrodysplasia punctata
Arylsulfatase F	Endoplasmic reticulum	Unknown	Not established

[a] The eponyms of the various mucopolysaccaridoses (MPS) are given in parentheses.

[b] Sulfatase not yet identified (only characterised biochemically)

[c] Endo-N-Acetylglucosamine 6-O-sulfatase

1.1.3 Formylglycine in eu-and prokaryotic sulfatases

Bacterial arylsulfatases are similar in sequence to eukaryotic sulfatases, and the highest similarity is observed in the region of the cysteine that is converted to FGly (compare figure 1.1 and figure 1.2). In all eukaryotic and most prokaryotic sulfatases, this cysteine is conserved in the primary sequence. However, in some bacterial sulfatases, there is a serine residue instead of the cysteine (Figure 1.1). Both the cysteine and the serine are converted to FGly as shown for *Pseudomonas* sulfatase (Dierks, *et al.,* 1998) and *Klebsiella* sulfatase (Miech *et al*., 1998). This leads us to propose that bacterial sulfatases can be classified into cysteine-type and serine-type sulfatases with respect to the amino acid that is converted to FGly in the active enzymes. 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 among the members of the serine-type or cysteine-type sulfatases family than there is between serine-type and cysteine-type sulfatases (Schirmer and Kolter, 1998). This is true for both the complete protein sequence and the highly conserved region around the active site. However, as mentioned above, a distinction of the two types of sulfatases can be made with respect to their localization. All serine-type sulfatases contain typical amino-terminal signal sequences, suggesting a periplasmic localization, whereas the cysteine-type sulfatases are predicted to be cytoplasmic enzymes. This difference in subcellular localization was experimentally verified for the cysteine-type sulfatase of *P. aeruginosa* (Kertesz *et al.,* 1999) and the serine-type sulfatase of *K. pneumoniae* (Marquordt *et al.*, 2003).

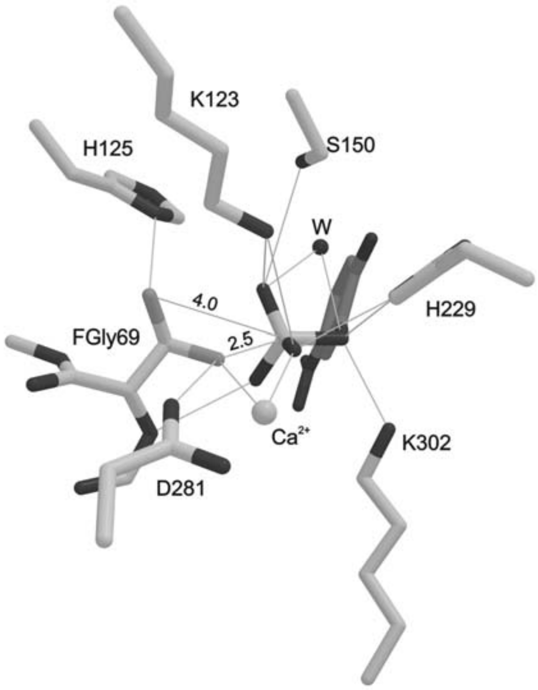


Figure 1.3 Active site of Arylsulfatase A. The sulfate group of the substrate is positioned in the active site by coordination to lysine^{123}, serine^{150}, histidine^{229}, lysine^{302}, a water molecule (W) and the Ca^{2+} ion. The nucleophilic attack on the sulfur by the juxtaposed Oγ1 hydroxyl of formylglycine^{69} hydrate (distance 2.5 Å) is possible because of proton abstraction by aspartate^{281}.

1.1.4 Formylglycine-hydrate in the active site of sulfatases

The three-dimensional structures of three human sulfatases, Arylsulfatase A, B and C and of AtsA of *P. aeruginosa* have been determined showing a very similar α/β fold (Bond *et al.*, 1997; Lukatela *et al.*, 1998; Hernandez-Guzman *et al.,* 2003**;** Boltes *et al.*, 2001). The formylglycine as the catalytic residue is buried in a cavity formed by charged and highly conserved amino acids (Fig 1.3). An enzyme intermediate could be trapped that upon incubation with sulfate ester was covalently sulfated at the position of the formylglycine (Recksiek *et al.*, 1998). The two-fold disordered electron density of the formylglycine side chain reflects an aldehyde-hydrate ($-CH(OH)_2$) with two geminal hydroxyls protruding into the cavity. This hydrate is stabilized by a Ca^{2+} ion (Fig 1.3), that furthermore is held in place by three aspartates and one asparagine. Two lysines, two histidines, and a serine complete the arrangement of residues potentially involved in binding and cleavage of the sulfate group within the active site.

1.1.5 Mechanism of formylglycine-hydrate mediated sulfate ester hydrolysis

The catalytic mechanism of sulfate ester cleavage proposed initially for arylsulfatase A includes two half-cycles (Fig 1.4). In the first half cycle, one of the two geminal oxygens of the aldehyde hydrate attacks the sulfur of the sulfate group and a covalently sulfated enzyme intermediate is formed. Simultaneously the substrate alcohol is released. In the second half cycle, sulfate is eliminated from the enzyme-sulfate intermediate by an intramolecular rearrangement induced by the second oxygen. By this “intramolecular hydrolysis” the aldehyde group is regenerated. The essential role of FGly-hydrate in the formation and break-down of the transient enzyme-sulfate ester has been shown indirectly (Recksiek *et al.*, 1998). Mutants of Arylsufatase A and Arylsufatase B were expressed in which FGly had been replaced by a serine. These mutants were able to cleave sulfate ester and to generate a sulfated enzyme intermediate with a serine sulfate ester in position 69 (Arylsufatase A) or 91 (Arylsufatase B). However, due to the lack of a second hydroxyl group, the release of the sulfate group regenerating the aldehyde did not occur. Further molecular details of the FGly-hydrate mediated sulfate ester hydrolysis are given in Fig.1.4.

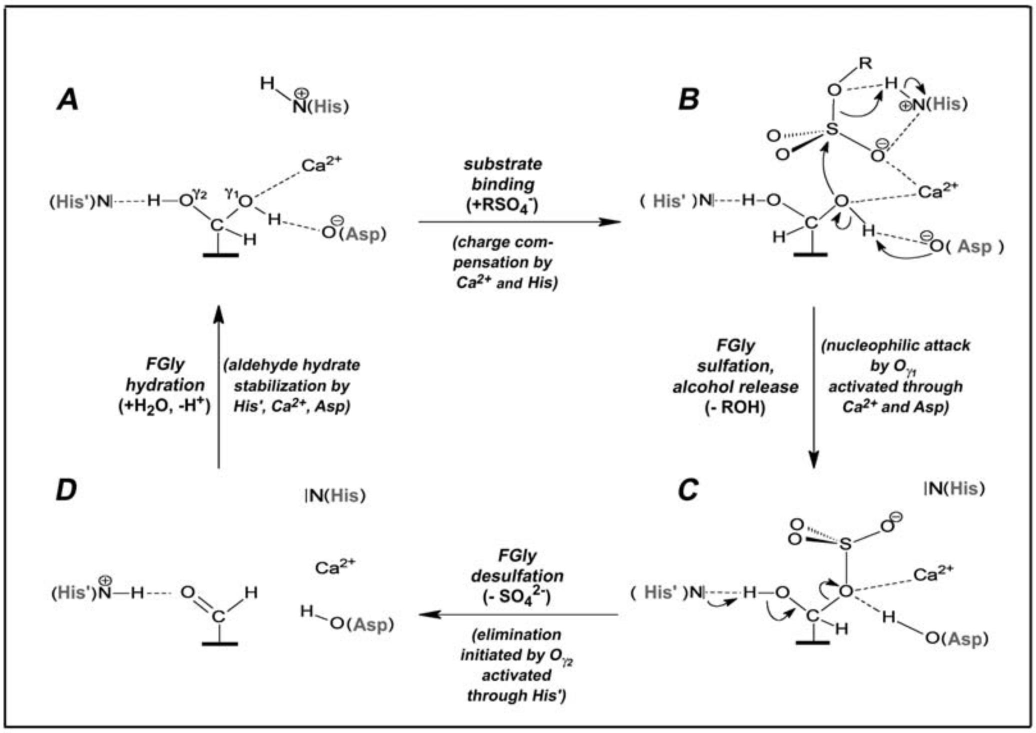

A
B
C
D
N(His)
Ca2+
(His')N----H—Oγ2 γ1O
O(Asp)
(His')N----H—O
O(Asp)
R
substrate
binding
(+RSO4-)
(charge compensation by
Ca2+ and His)
FGly
sulfation,
alcohol release
(- ROH)
(nucleophilic attack
by Oγ1
activated through
Ca2+ and Asp)
FGly
desulfation
(- SO42-)
(elimination
initiated by Oγ2
activated
through His')
FGly
hydration
(+H2O, -H+)
(aldehyde hydrate
stabilization by
His', Ca2+, Asp)
(His')N—H---- O
O(Asp)

Figure 1.4 Scheme of the proposed catalytic mechanism for sulfate ester hydrolysis (Boltes *et al.*, 2001), as shown for the active site of *Pseudomonas* sulfatase. In the geminal diol of FGly-hydrate, one of the alcohol groups, $O^{\gamma 2}$, is oriented towards the core of the protein and is stabilized by the neighbouring His^{115} residue (A). The other alcohol group $O^{\gamma 1}$ is very close (2.5 Å) to the sulfate sulfur atom to start a nucleophilic attack on the sulfur, and its nucleophilicity is enhanced both by its coordination to the calcium atom and the possibility of a proton transfer to the carboxyl group of the $aspartate^{317}$, which would be stabilized in turn by the divalent cation (B). The orientation of the sulfate with the side of the sulfate tetrahedron facing FGly-$O^{\gamma 1}$ group is stabilized by the charged side chains of $histidine^{211}$ and $lysine^{375}$ residues. After substrate desulfation, the alcohol is released and diffuses out of the catalytic pocket. The sulfate is now covalently bound to the enzyme (C) and additionally coordinated to the calcium atom through two of its oxygen atoms. The C-O bond in the FGly-sulfate ester is polarized through all the contacts of the sulfate

group to positively charged residues, favouring its elimination. The final step of the reaction is elimination of sulfate and regeneration of the aldehyde (D). The aldehyde hydrate is regenerated by addition of a water molecule and is stabilized by hydrogen bonds to histidine^{115}, arginine (not shown), and the calcium atom. Equivalent positions of active site residues in arylsulfatase A (see Fig. 1.3) are FGly^{69}(FGly^{51}), His^{125}(His^{115}), Lys^{302}(Lys^{375}), His^{229}(His^{211}), and Asp^{317}(Asp^{281}).

1.1.6 Sequence motif directing formylglycine formation

Members of the sulfatase protein family, from bacteria to man, show highest sequence homology in the N-terminal region including the FGly residue that has been found in sulfatases of prokaryotic (Fig 1.1 and 1.4), lower eukaryotic (Selmer *et al.*, 1996) and human origin (Fig 1.2 and 1.3). The sequence information directing conversion of cysteine to FGly in human arylsulfatase A is confined to residues –1 to +11 with respect to the position of the cysteine to be modified (Dierks *et al.*, 1999). Within this region, two sequence motifs directing FGly formation were identified by a mutational analysis. The core motif, which is absolutely required, comprises the pentapeptide CTPSR starting with cysteine to be modified. The most crucial residues, besides the cysteine, are Pro(in position +2) and Arg (in position +4). A second 'auxiliary' motif is located directly C-terminal of the core CTPSR motif and comprises a stretch of seven amino acids (AALLTGR), which include the highly conserved tetrapeptide LTGR. This second motif is likely to assist in the presentation of the cysteine to the modifying enzyme, as has been concluded from peptide inhibition experiments (Dierks *et al.*, 1999). In experiments with *de novo* synthesized ASA fragments, it was shown that a 16mer sequence (-4 to +11), comprising both the core and the auxiliary motif, is essential and sufficient for the formation of FGly (Dierks *et al.*, 1997).

As outlined above (see 1.1.3), prokaryotic sulfatases can be classified into serine-type and cysteine-type sulfatase with respect to the amino acid that is converted to FGly. Nevertheless, the FGly generation motifs in these two kinds of sulfatases both are very similar to that of eukaryotic sulfatases (see Figure 1.1). All known sulfatases share the key motif (C/S-X-P-X-R). Because of this similarity of the FGly generation motif, the mechanism of cysteine-type and serine-type sulfatase modification likewise was expected to be very similar.

1.2 Formylglycine generating machineries

In chemical respect, the proposed classification of sulfatase-modifying enzymes into cysteine-specific and serine-specific systems has mechanistic implications (Fig 1.5). Conversion of a serine residue to FGly requires merely an oxidation, whereas the modification of cysteine to FGly proceeds formally via oxidation to the thio-aldehyde, followed by a hydrolytic step. It turned out that different modifying systems exist for cysteine-sulfatase modification and serine-sulfatase modification. The conversion of a thiol into an aldehyde group in eukaryotic sulfatases is catalyzed by FGE in the ER (Dierks *et al.*, 2003). During protein synthesis a cysteine is incorporated into the sulfatase, which is converted into FGly upon translocation into the ER (Dierks *et al.*, 1997). Serine-type sulfatase AtsA of *Klebsiella* is periplasmically located, and the serine-modifying enzyme AtsB which is described in this thesis, is cytosolic. Also AtsB function seems to be linked to transport of AtsA through the membrane (see Result).

1.2.1 AtsB: the putative serine-modifying system

An obvious candidate for the serine-modification enzyme was found immediately when the genetic structure of the arylsulfatase gene in *K. pneumoniae* was examined. This is the *atsB* gene product, which is encoded in a bicistronic operon with the *atsA* gene, encoding the arylsulfatase (Szameit *et al.*, 1999).

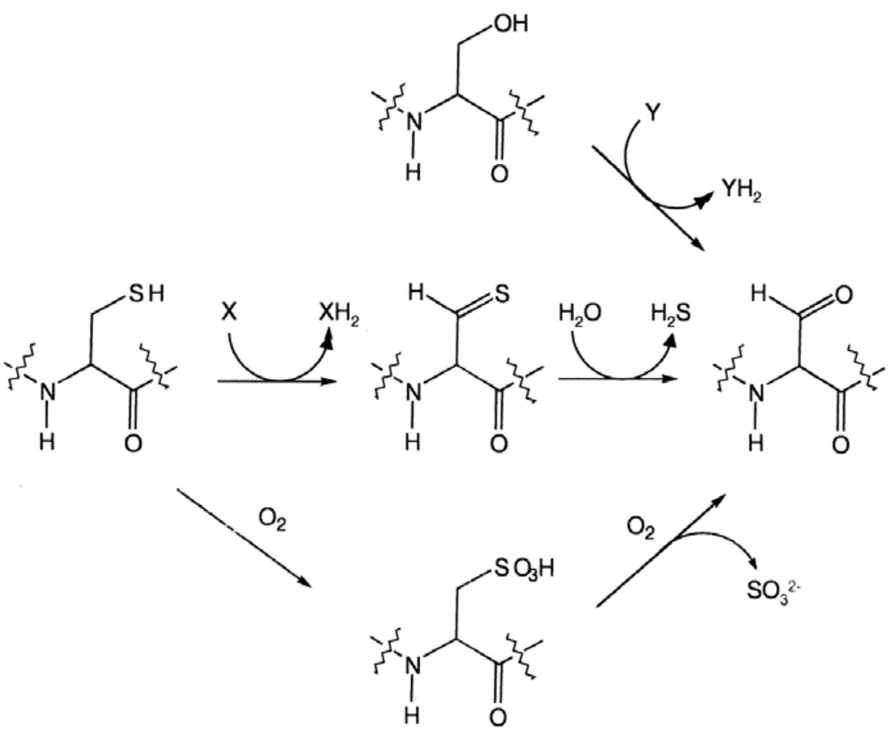


Figure 1.5 Mechanisms for FGly formation in *Cysteine-type* and *Serine-type* arylsulfatases. Conversion of serine requires only an oxidative step, with an unknown electron acceptor, whereas cysteine conversion requires an additional hydrolytic step. Alternatively, cysteine oxidation to cysteate and subsequent oxygenolytic cleavage would also lead to FGly. (Kertesz 1999)

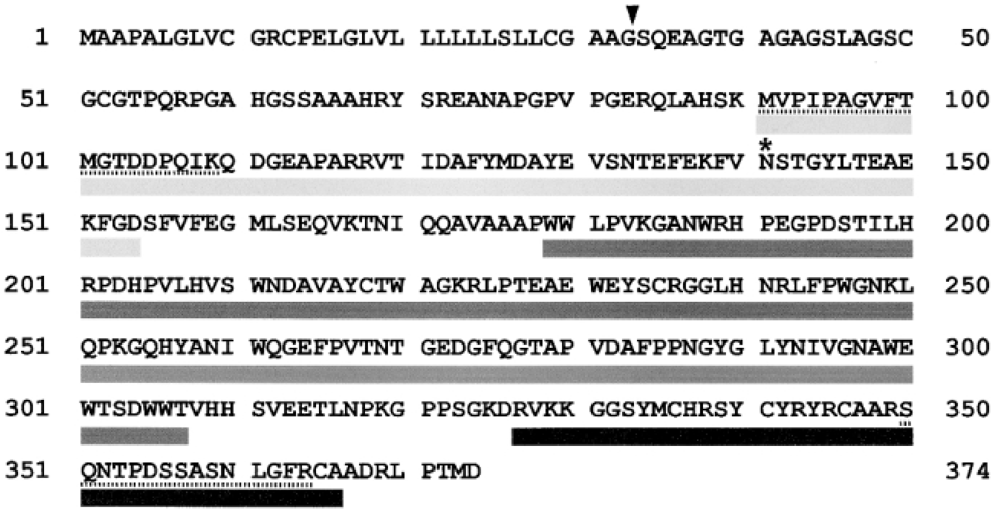
The *atsB* gene product is required for synthesis of active serine-arylsulfatase, and was originally proposed to be a transcriptional activator for the *atsBA* operon (Murooka *et al.*, 1990). However, the 44-kDa AtsB protein does not show any of the motifs expected for a transcriptional regulator (*e.g.* helix-turn-helix DNA binding regions), and expression of the *atsA* gene, driven by the strong *lac* promoter, in *E. coli* in the absence of *atsB* yielded only inactive arylsulfatase (Szameit *et al.*, 1999). In the inactive arylsulfatase, it was found that the serine at position 72 was completely unmodified. However, co-expression of the *atsA* and *atsB* genes on separate plasmids led to synthesis of an active sulfatase, in which the arylsulfatase serine-72 was 50% modified to FGly (Szameit *et al.*, 1999). The AtsB contains 12 cysteine residues in three clusters whose arrangement (cluster 1: C-X3-C-X2-C; cluster 2: C-X5-C-X14-C; cluster 3: C-X2-C-X5-C-X3-C) is reminiscent of [Fe-S] iron sulfur centers (Schirmer and Kolter, 1998). The cluster 1, with two tyrosines flanking the last cysteine thereby lowering the midpoint potential of the cluster (Johnson, 1998) and a predicted S-adenosylmethionine (SAM) binding site (GGEPLL) in AtsB, implicate that this protein belongs to the Radical SAM protein superfamily (Sofia *et al.*, 2001). This structure is consistent with the putative function of the modification enzyme as an oxidoreductase. At least one further member of this protein family, namely the oxygen-independent coproporphyrinogen-III oxidase HemN, has been shown to function as an oxidizing enzyme (Layer *et al.*, 2002). And SAM may be the essential co-factor of this oxidation. As FGly formation form a serine most likely is a single enzymatic reaction, AtsB may be sufficient for FGly formation in serine-type sulfatases.

Although no endogenous sulfatase activity has ever been measured in *E. coli*, this species carries two genes encoding putative serine-type sulfatases termed *aslA* and *f571*. Like the *Klebsiella atsA* gene, *aslA* and *f571* have an adjacent gene in the same operon (*aslB* and *f390*, respectively) encoding an AtsB homolog of about 400 amino acids (Schirmer and Kolter, 1998). These homologs did not take over function in the absence of AtsB (Szameit *et al.*, 1999), which most likely is due to repression of their operons. All three AtsB homologs are 34-41% identical and represent members of the radical SAM iron sulfur proteins (Sofia *et al.*, 2001). There is a further gene, *chuR* from *Bacteroides thetaiotaomicron*, homolgous to *atsB*, *aslB* and *f390* that has been implicated in sulfatase activation (Szameit *et al.*, 1999). *B. thetaiotaomicron* can utilize sulfated mucopolysaccharides such as chondroitin sulfate and heparin. A *chuR* mutant was found to be unable to grow either on chondroitin sulfate or on heparin (Cheng *et al.*, 1992). The finding that a single gene affected both chondroitin sulfate and heparin utilization was unexpected, because there are two different, independently regulated pathways for the degradation of these compounds in *B. thetaiotaomicron*. Interestingly, of the chondroitin sulfate utilization genes, only chondroitin-6-sulfatase activity was missing in the *chuR* mutant. Cheng *et al.* 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 (Schirmer and Kolter, 1998). If this is so, these sulfatases are probably serine-type sulfatases and the reported phenotype of the *chuR* mutant is a form of bacterial MSD.

1.2.2 FGE: a cysteine-modifying system of pro- and eukaryotes

Recently, the human Cα-formylglycine generating enzyme (FGE), encoded by the *SUMF1* gene, has been identified independently after biochemical purification of FGE from bovine testis (Dierks *et al.*, 2003) or by functional complementation of MSD cells using microcell-mediated chromosome transfer (Cosma *et al.*, 2003).

SUMF1 gene orthologs were found in numerous species, including *Homo sapiens, Mus musculus, Fugu rubripes, Drosophila melanogaster, Anopheles gambiae*, and also in a number of bacterial species. A high degree of sequence similarity was observed along the entire length of the protein sequence, with the exception of the hydrophobic N-terminal signal peptide region. The greater part of the mature protein (275 residues out of 340) defines a unique domain, which is likely to be composed of three subdomains (Fig 1.6 , Dierks *et al.*, 2003).



1 MAAPALGLVC GRCPELGLVL LLLLLSLLCG AAGSQEAGTG AGAGSLAGSC 50

51 GCGTPQRPGA HGSSAAAHRY SREANAPGPV PGERQLAHSK MVPIPAGVFT 100

101 MGTDDPQIKQ DGEAPARRVT IDAFYMDAYE VSNTEFEKFV NSTGYLTEAE 150

151 KFGDSFVFEG MLSEQVKTNI QQAVAAAPWW LPVKGANWRH PEGPDSTILH 200

201 RPDHPVLHVS WNDAVAYCTW AGKRLPTEAE WEYSCRGGLH NRLFPWGNKL 250

251 QPKGQHYANI WQGEFPVTNT GEDGFQGTAP VDAFPPNGYG LYNIVGNAWE 300

301 WTSDWWTVHH SVEETLNPKG PPSGKDRVKK GGSYMCHRSY CYRYRCAARS 350

351 QNTPDSSASN LGFRCAADRL PTMD 374

Figure 1.6 Amino Acid Sequence of Human FGE (Dierks *et al.,* 2003). The sequence is deduced from the cloned cDNA. The two peptides identified in bovine FGE are underlined by a stippled bar. The predicted cleavage site for the signal peptidase at Gly-33 is indicated by an arrow head and the N-glycosylation site at Asn-141 by an asterisk. Subdomain one (light gray), subdomain two (gray), and subdomain three (dark gray) are underlined by filled bars.

So far none of these subdomains could be associated with functional aspects such as substrate binding and catalysis. The catalytic domain in FGE could catalyze the FGly formation in several ways. It has been proposed that FGE abstracts electrons from the thiol group of the cysteine and transfers them to an acceptor. The resulting thioaldehyde would spontaneously hydrolyze to FGly and H_2S (Schmidt *et al.*, 1995). Alternatively, FGE could act as a mixed-function oxygenase (monooxygenase) introducing one atom of O_2 into the cysteine and the other into H_2O with the help of an electron donor such as $FADH_2$. The resulting thioaldehyde hydrate derivative would spontaneously react to FGly and H_2S. FGly formation is critically depended on molecular oxygen (J. Peng, unpublished). This would suggest that FGE acts as a mixed-function oxygenase or, alternatively, as an H_2O_2-producing oxygenase. Also oxidation to cysteate followed by oxygenolytic cleavage could explain the oxygen dependence (see Fig 1.5). The particular high conservation of subdomain three (Fig 1.6) and the presence of three in all orthologs conserved cysteine residues therein make this subdomain a likely candidate for the catalytic site. It will be of interest to see whether the structural elements mediating the recognition of the FGly motif and the binding of an electron acceptor or electron donor correlate with the domain structure of FGE.

Among the species with prokaryotic *SUMF1* genes, 7 have fully sequenced genomes allowing a neighborhood analysis of the FGE loci (Landgrebe *et al.*, 2003). In four of the 7 genomes a cysteine-type sulfatase is found in direct vicinity of FGE. In all four cases, FGE and the sulfatase are predicted to be cotranscribed from a common promoter. The genomic neighborhood of FGE and sulfatase genes in prokaryotes provides additional evidence for the assumption that the bacterial FGEs are functional orthologs of human FGE (Landgrebe *et al.*, 2003).

1.2.3 X: unknown cysteine-modifying system

The genomes of *E. coli, S. cervisiae*, *N. Crassa* and *C. elegans* lack *SUMF1* genes indicating a phylogenetic gap and the existence of an alternative FGly-generating system for cysteine-sulfatases. *E. coli* has the ability to modify some cysteine-arylsulfatases. This was shown experimentally for the arylsulfatase of *Pseudomonas aeruginosa* (PAS). Even after strong overexpression in *E. coli*, this cytosolic sulfatase was quantitatively converted to the active FGly-bearing enzyme (Dierks *et al.*, 1998). Hence, the *E. coli* cytosol contains the modifying machinery. This machinery is expressed even under excessive supply with inorganic sulfate. Thus, expression of *E. coli*’s cysteine-modifying system is independent of the sulfur status of the cells, in contrast to expression of the sulfatase structural genes, as studied in *P. aeruginosa* and *Klebsiella pneumoniae*. The presence of FGly modification systems in *E. coli*, which lacks an active sulfatase gene, and the difference in the regulatory pattern between bacterial sulfatases and the enzyme system that modifies them suggests that other as yet unidentified bacterial proteins also undergo a similar FGly modification.

E. coli's cysteine-modification system (system X) acts on highly similar FGly-modification motifs as the FGE system. However, X represents a distinct system, because *E. coli* does not contain FGE homologs. It should be noted that the human Arysulfatase G after expression in *E. coli* was found to retain the critical cysteine in unmodified form (S. Schulz, unpublished), which suggests that system X and FGE in fact show different requirements with respect to their substrates.

1.3 Goals and experimental approach

In this thesis, the posttranslational generation of FGly in prokaryotic sulfatases was studied. It was known that AtsB, a predicted radical SAM iron sulfur protein of *Klebsiella pneumoniae*, is required for FGly formation and that AtsB obviously is specific for FGly generation in serine-type sulfatases. *In vivo* expression experiments had established that coexpression of AtsB is required to express active, FGly-modified *Klebsiella* sulfatase AtsA in *E. coli*. Moreover, pilot *in vitro* experiments, using GST-pull-down technology, indicated that AtsB interacts specifically with the serine to be modified. However, to establish AtsB as the key serine-oxidizing enzyme and to elucidate the mechanism of its action, an *in vitro* FGly modification system was required. The knowledge on AtsB's enzymatic mechanism was restricted to bioinformatic prediction of three FeS centers, coordinated by up to 12 conserved cysteine residues, and of a hypothetical SAM-binding site. Experimental evidence, however, supporting their functional importance was lacking.

Concerning the specificity of AtsB it is of interest whether *K. pneumoniae* AtsB can only modify *K. pneumoniae* AtsA or serine-type sulfatases in general. A new AtsB homolog was found in the archaebacterium *Methanosarcina mazei* that was designated MM-AtsB. But MM-AtsB is encoded in an operon with the *mms* gene, encoding a putative cysteine-sulfatase. The following questions arise: Can MM-AtsB, and possibly also AtsB modify cysteine-type sulfatases? Can MM-AtsB modify serine-type sulfatases? What determines serine- and/or cysteine-specificity? What determines whether a cysteine-type sulfatase is modified by system X or by AtsB?

The fact that *K. pneumoniae* sulfatase AtsA is a periplasmic protein and AtsB is a cytosolic protein, this implies that modification of the serine-72 of AtsA occurs prior to translocation into the periplasm. AtsB does not have transmembrane domains and is recovered from cell lysates as a soluble protein. AtsA shows a modification degree of 50% when purified from the periplasm, whereas the modification degree is lower than 10% when purified from total cell lysates. This implies that FGly modification is related to the sulfatase translocation process and may occur at the inner surface of the plasma membrane. This issue needed further investigation.

To summarize, the main goal of this study was to characterize the system and mechanism underlying FGly-formation in serine-type sulfatases. Towards this goal the following steps had to be addressed:

1 To establish an *in vitro* FGly generation assay for the serine-type sulfatase AtsA;

2 To find out the cofactor of the serine-type sulfatase modification system;

3 To validate the functional importance of the predicted FeS-centers and the predicted SAM-binding site of AtsB for FGly generation in order to establish the modification reaction mechanism;

4. To find out the substrate specificity of AtsB;
5. To find out as to whether sulfatase modification is temporally/spatially related to sulfatase export to the periplasm.
6. To purify AtsB for structural analysis and functional characterization under defined *in vitro* conditions.

2. Materials

2.1 Appliances:

Analytic and preparative HPLC	
SMART-system with following columns:	Amersham Biosciences, Freiburg
μ peakC2C18PC3.2/3 (C2C18 2.1 X 30 mm)	Amersham Biosciences, Freiburg
Mono Q ® PC 1.6/5	Amersham Biosciences, Freiburg
μ Peak Detector	Amersham Biosciences, Freiburg
Intelligent Dark Box II, Las-1000+	Fuji, Japan
Automatic Protein sequencer: Type 447A	PE Biosystems
With PTH-Amino acid analyzer, Type 120A	PE Biosystems
DNA-Sequencer: Type 310	ABI, PE Biosystems
Electroporator	Stratagene
Liquid scintillation counter 1900TR	Packard, Frankfurt/Main
French ® Pressure Cell Press, SLM-Aminco®	Spctronic Instruments
Vol: big cell: 37 ml; mini cell: 3.7 ml	
Gene Amp PCR system 2400	Perkin Elmer Cetus, USA
MALDI-TOF Mass Spectrometer, REFLEX III	Bruker Daltonics, Bremen

2.2 Chemicals

All chemicals used were of analytical grade and were purchased from Boehringer, Merck, Roth, Gibco, Difco, Sigma or Serva, unless otherwise stated.

2.3 Radioactively labelled substances

S-Adenosyl-L-[methyl-^{3}H]methionine 74.0 Ci/mmol	Amersham Biosciences, UK

2.4 Enzymes and Standards

Klenow-polimerase	New England Biolabs, Bad Schwalbach
Restriction endonuleases	New England Biolabs, Bad Schwalbach
akaline phosphatase	Boehringer, Mannheim

λ/Hind III- Standard	Gibco BRL, Eggenstein
1-kb-DNA-Ladder	Gibco BRL, Eggenstein
Precision Protein standards for PAGE	BioRad
Trypsin modified, sequencing grade	Roche
Oligonucleotides	IBA, Göttingen
Pfu-Ultra-DNA-Polymerase	Pharmacia, Freiburg
Proteinase K	Qiagen, Hilden
Taq-DNA-Polymerase	Pharmacia, Freiburg
Ultrapure dNTP Set	Pharmacia, Freiburg
Lysozyme from egg	Merck

2.5 Protein determination, Translation and Transcription Kits, DNA Kits

BioRad Protein Assay	BioRad, München
Qiagen Plasmid Miniprep Kit	Qiagen, Hilden
Qiagen Plasmid Midiprep Kit	Qiagen, Hilden
Qiagen DNeasy Tissue Kit	Qiagen, Hilden
QIAquick PCR-purification Kit	Qiagen, Hilden
Qiagen Gel Extraction Kit	Qiagen, Hilden
Quick Change TM	Stratagene, USA
Ni-NTA Agarose	Qiagen, Hiden
GSH- Agarose	Sigma
Strep-tag II/Tactin	IBA, Göttingen
SuperSignal ® Chemiluminescent Substrat	Pierce, USA

2.6 Antibody

2.6.1 Primary antibodies

Name	Antigen	Immunised species	purification	Referencd
αDVS	DVS	Rabbit, polyclone	Serum	This work
αAtsA	AtsA	Rabbit, polyclone	Serum	Miech *et al.*, 1998
αAtsB	AtsB	Rabbit, polyclone	Serum	Marquordt *et al.*, 2003
αPAS	PAS	Rabbit polyclone	Serum	Marquordt et al., 2003
αGST	GST-TAG	Goat, Polyclone	Serum	Amersham Pharmacia, Freiburg
αStrep	Strep-TAG	Mouse, monoclone	-	IBA, Göttingen
αGroEL	GroEL	Mouse, monoclone	-	
αHA	HA-TAG	Mouse, monoclone	-	Boehringer, Mannheim
αHis	His_6-TAG	Mouse, monoclone	-	Qiagen, Hilden

2.6.2 Secondary antibodies

Goat anti rabbit, HRP conjugated	Dianova, Hamburg
Goat anti mouse, HRP conjugated	Dianova, Hamburg
Rabbit anti goat, HRP conjugated	Dianova, Hamburg

2.7 Bacterial cell culture media

All media were autoclaved for 20 min at 121 ^{o}C abd stired at 4 ^{o}C. Solid media were obatined adding bacto-agar at the final concentration of 1.5 % (w/v)

Lurta-Bertami (LB)-Medium: 10 g/l Bacto Typton, 5 g/l Bacto Yeast Extract 5 g/l NaCl 1ml/l 1M NaOH
SOC Medium: 20 g/l Bacto-peptone, 5g/l Yeast extract 0.5 g/l NaCl
Additives: 0.1-1 mM IPTG
0.2 mg/l AHT
Antibiotics: 100 μg/ml Ampicilline
34 μg/ml Chloramphenicol in Methanol

2.8 Frequently used buffers and solutions

para-Nitrocatecholsulfate	20 mM	pNCS
	150 mM	NaCl
	100 mM	Tris-HCl, pH 7.4 or 8.9
50 X TAE solution	2 M	Tris-HCl
	0.1 M	EDTA
		(pH adusted to 8.0 with acetic acid)
TBS buffer	150 mM	NaCl
	10 mM	Tris-HCl, pH 7.4
PBS buffer	137 mM	NaCl
	3 mM	KCl
	8 mM	Na_2HPO_4
PBS buffer	0.2 g/L	KCl
(for GST purification)	0.25 g/L	KH_2PO_4
	1.8 g/L	Na_2HPO_4
	8.2 g/L	NaCl
10 X DNA loading buffer	30% (w/v)	Ficoll
	0.25% (w/v)	Brophenolblue
	0.25% (w/v)	Xylen cyanol FF
	0.5 M	EDTA, pH 8.0
Protease- Inhibitor-Cocktail	2 mM	AEBSF (inhibits serine proteases)
	1 mM	EDTA
	130 μM	Bestatin (inhibits aminopeptieases)
	14 μM	E-64 (inhibits cysteine proteases)
	1 μM	Leupeptin (inhibits both serine and cysteine proteases)
	0.3 μM	Aprotinin (inhibits serine proteases)
2 X Laemmli loading buffer	0.1 M	Tris-HCl. pH 6.8
	2% (w/v)	SDS
	2% (v/v)	β-mercaptoethanol
	20% (v/v)	Glycerol
	0.002 % (w/v)	Bromphenolblue

2.9 Plasmid vectors:

Plasmid vector	Antibiotics	Firm
pBlueScript II KS	ampicillin	Stratagene
pBBRM1CS	chloramphenicol	

pGEX-KG	ampicillin	Pharmacia Amersham
pETBlue-1	amplicillin	Novagen
pASK-IBA7	amplicillin	IBA, Goettingen
PT-Groel	chloramphenicol	Yasukawa *et al.,* 1995
PT-Trx	chloramphenicol	Yasukawa *et al*., 1995

2.10 *Escherichia coli* strains:

Strani	**Genotype**	**Firm**
DH5α	supE44, thi-1, recA1, relA1, hsdR17($r_K^-m_K^+$), thi-1, ΔlacU169 (Φ80 lacZΔM15), endA1, gyrA (Nal^r)	GibcoBRL, Eggenstein
BL21-(DE3)	*E. coli* B F$^-$ ompT hsdS(r_B^- m_B^-) dcm$^+$ Tet^r gal λ(DE3) endA Hte [argU ileY leuW Cam^r]	Stratagene, Heidelberg
NovaBlue Singles ™	*endA1hsdR17*($R_{k12}^-m_{k12}^-$)*supE4thi-1recA1gyrA96 relA1lac*F'[*proA*$^+$*B*$^+$*lacI*^{q}*ZΔM15*::Tn10(Tc^R)]	Novagen
Tuner TM (DE3) pLacI	F- ompT hsdSB (rB -mB-) gal dcm lacY1 (DE3) pLacI (CmR)	Novagen

3. Methods

Molecular biological and biochemical methods used in this work are in more detail described in standard manuals like: “Molecular Cloning” (Sambrook *et al.,* 2001), “Current Protocols in Molecular Biology” (Ausubel *et al.*, 1997), “Current Protocols in Protein Science” (Coligan *et al.*, 1997).

3.1 *E. coli* culture

E. coli strains were grown in the LB media with appropriate antibiotic at 37 °C or at lower temperatures with 260 rpm agitation. For short-time storage, bacteria plates were sealed with Parafilm and stored at 4 °C. For long time storage, 0.75 ml of overnight cultures were mixed with 25 ml 80% sterile glycerol, rapidly frozen in liquid nitrogen or dry ice and then kept at –80 °C.

3.2 Preparation of DNA

3.2.1 Genomic DNA preparation from *Caulobacter crescentus* and *Desulfovibrio vulgaris*

QIAGEN DNeasy® Tissue Kit was used for Genomic DNA preparation from *Caulobacte crescentus* and *Desulfovibrio vulgaris*. About $2X10^9$ cells is harvested by centrifugation. The cell pellet was resuspended in 180 µl Buffer ATL. Add 20 µl proteinase K, mix by vortexing, and incubate at 55 °C with shaking until the pellet is completely lysed. Then 200 µl Buffer AL was added to the sample, mixed thoroughly by vortexing and incubated at 70 °C for 10 min. 200 µl ethanol was added to precipitate the DNA. The sample was transferred to DNeasy spin column, and centrifuge at 6,000 g for 1 min. After discard flow-though, 500 µl Buffer AW1 and 500 µl Buffer AW2 were used to wash the column. In order to dry the spin column membrane, the column was centrifuged at 13000 rpm for 3 min. Finally genomic DNA was eluted by 200 µl Buffer AE. This genomic DNA can be used directly as PCR template. Buffer AW1, AW2, ATL and AE were included in DNeasy Tissue Kit.

3.2.2 Analytical isolation of plasmid DNA from *E. coli* cells (Qiagen handbook)

Small amounts of plasmid DNA were isolated according to the protocol provided by producer of the buffer system used for preparation.

100 µl/ml of Ampeicilline were added to 10 ml LB-medium and cells were incubated overnight by 37 °C with shaking at 240 rpm. 10 ml of this bacterial culture was centrifuged for 5 min by 5000 rpm and sediment was resuspended in 500 µl of ice-cold P1 buffer. After addition of 500 µl of P2 buffer sample was incubated for 5 min by RT. 700 µl buffer N3 were added, mixed and sample was precipitated on ice for 5 min. After sedimentation for 10 min by 14000 rpm at RT the clear supernatant was transferred to the mini-column and centrifuged for 1 min by 14000 rpm. The flow through was discarded and bound to the filter DNA was centrifuged for 1 min at 14000 rpm with 750 µl of PE buffer. The flow through was again discarded and sample was centrifuged for another 1 min without addition of buffers to remove remaining ethanol. The elution of DNA was performed with 100 µl distilled H_2O.

3.2.3 Preparative isolation of DNA from *E. coli* cells (Qiagen handbook)

Larger amounts of plasmid DNA were isolated according to the protocol of the producer of the buffer system used.

200 ml of the overnight grown bacterial culture was centrifuged (JA-10 rotor, 7000 rpm, 15 min, 4 °C). Bacterial sediment was resuspended in 4 ml P1 buffer and transferred in JA-20 tubes. After addition of 4 ml P2 the tubes were several times turned upside down to allow lysate of the cells and incubated for no more than 5 min at room temperature. After addition of 4 ml of buffer P3 samples were immediately mixed, neutralised for 10 min on ice and centrifuged for 30 min by 12000 rpm at 4 °C in JA-20 rotor. The supernatant was transferred on the equilibrated with 10 ml QBT buffer Qiagen column and bound DNA was washed twice with 10 ml QC buffer. The elution of DNA was performed with 5 ml QF

buffer in 25 ml Correx-tubes and eluted DNA was precipitated with 0.7 ml isopropanol (SS34 rotor, 14000 rpm, 4 °C). The DNA pellet was dissolved in 500 µl distilled H_2O, transferred in a reaction tube and after addition of 0.3 vol 7 M NH_4Ac and 1 ml 100 % ethanol precipitated for the second time for 30 min at –80 °C. After centrifugation in eppendorf centrifuge for 20 min by 14000 rpm at 4 °C DNA was washed with 70 % ethanol, dried in the air and dissolved in the appropriate volume of distilled H_2O. After determination of concentration and purity concentration of DNA was adjusted to 1 µg/µl and DNA was kept at –20 °C.

3.3 Purification of DNA

3.3.1 Ethanol precipitation

Precipitation and concentration of DNA was performed by adding to extracted DNA 2-2.5 volumes of ethanol after addition of 0.3 volumes of 7 M NH_4Ac. Sample was incubated for 30 min at –80 °C. After precipitated by centrifugation for 20 min at 14000 rpm at 4 °C, DNA was washed with 70 % ethanol and the precipitate was dried in the air at room temperature.

3.3.2 Phenol/Chloroform-extraction

Extraction of DNA with a mixture of phenol/chloroform/isoamylalcohol (25:24:1) and chloroform/isoamylalcohol (24:1) is a standard method for removal of proteins from DNA preparations. The purification was performed as follows: 1 volume of phenol mixture was added to DNA solution, well mixed and centrifuged for 5 min at 14000 rpm. The upper DNA containing phase was carefully taken out and in order to remove the rest of phenol was mixed with 1 volume of chloroform/isoamylalcohol mixture. After mixing and centrifugation at 14000 rpm for 5 min the upper phase was taken out.

3.3.3 Purification of DNA-fragments by gel-extraction

The DNA-fragments were excised from the gel with a scalpel and transferred into a 1.5 ml tube. 3 folds volume Buffer QG was added to the tube and the tube was incubated at 50 °C until the gel was disappeared. 1 fold volume isopropanol was added to the tube and the sample was transferred to Qiagenquick-column and centrifuged for 1 min by 14000 rpm. The flow through was discarded and bound to the filter DNA was centrifuged for 1 min at 14000 rpm with 750 µl of PE buffer. The flow through was again discarded and sample was centrifuged for another 1 min without addition of buffers to remove remaining ethanol. The elution of DNA was performed with 50 µl distilled H_2O.

3.3.4 PCR purification

5 fold volume Buffer PB was added to PCR products. The sample was vortexed and transferred to the Qiagenquick-column and centrifuged for 1 min by 14000 rpm. The flow through was discarded and bound to the filter DNA was centrifuged for 1 min at 14000 rpm with 750 µl of PE buffer. The flow through was again discarded and sample was centrifuged for another 1 min without addition of buffers to remove remaining ethanol. The elution of DNA was performed with 50 µl distilled H_2O

3.4 Determination of concentration and the purity of DNA

As described in “Molecular Cloning” (Sambrook *et al.,* 2001), it is possible to quantify nucleic acids and to evaluate their purity by spectrophotometric analysis. DNA and RNA absorb light of 260 nm wavelength, proteins (aromatic amino acids) absorb light of 260 nm wavelength too, but absorption is much stronger at 280 nm.

The ratio A260/A280 gives an estimation of DNA purity. For pure DNA, A260/A280 ratio is about 1.8.

Spectrophotometric conversion: 1A260 of double-stranded DNA = 50 µg/ml
1 A260 of single-stranded DNA = 33 µg/ml
1 A260 of single-stranded RNA = 40 µg/ml

3.5 Electrophoretic separation of DNA fragments in agarose gels

For analysis and purification of DNA fragments of different size horizontally oriented agarose gels were used (Sambrook *et al.,* 2001). Agarose concentration of the gels was dependent on the size of DNA fragments to be separated:

For preparation of the gel the appropriate agarose amount dissolved in 300 ml 1x TAE buffer was cooked in the microwave oven and after cooling it down to 60 °C ethidium bromide was added to reach the end concentration of 0.5 μg/ml. Agarose solution was applied into a gel chamber and combs with appropriate pocket size were introduced. After the gel was ready it was transferred into an electrophoresis chamber filled with 1xTAE buffer. The samples were mixed with 0.2 volume of 6x DNA-loading buffer and transferred into gel pockets. Electrophoresis was performed by 4-5 V/cm. Etidium bromide is binding to DNA fragments and that makes possible visualisation of DNA by UV light. The detection limit of ethidium bromide stained gels is approximately 5 ng of DNA per band. The agarose gels were developed on the UV-transilluminator and documented with the help of a video system.

3.6 Digestion of DNA with restriction nucleases

Restriction enzyme digestion was carried out according to standard procedures (Sambook *et al.*, 2001). Depending on the enzymes used and their cutting sites, sticky-ended (5'- or 3'- protruding single strand DNA) or blunt-ended DNA fragments can be generated. Restricted DNA fragments were purified either by gel electrophoresis and extraction using a QIAGEN gel extraction kit, or by using a QIAGEN PCR and nucleotides purification kit.

3.7 Dephosphorylation of DNA

To avoid undesired self ligation of plasmid DNA, the 5' phosphate groups of the DNA were dephosphorylated using calf intestinal phosphatase (CIP, Boehringer), employer at a concentration of 1U/1μg DNA at 37 °C for 30 min using the appropriate buffer delivered with the enzyme. Dephosphorylated DNA was purified by gel electrophoresis.

3.8 Filling up overhanging DNA ending

For filling up overhanging DNA ending, blunt-ended DNA was first prepared using the Klenow fragment of DNA polymerase according to standard procedure, followed by purification with the QIAGEN PCR purification kit.

3.9 Cloning of recombined DNA in *E. coli*

3.9.1 Ligation of vector and insert DNA

T4 DNA ligase catalyses the ATP-dependent ligation of blunt or complementary sticky ends of DNA. The enzyme was used according to the supplier’s instructions. Sticky-end ligations were carried out at 16 °C for 2-4 hr, typically using a 1:3 vector: insert molar ratio and roughly 40 ng vector DNA.

3.9.2 Preparation of competent cells

10 ml of overnight culture of E. coli were used to inoculate 1L fresh LB medium. The culture was grown at 37 °C with 260 rpm agitation until an OD600 of 0.5 was achieved. Cells were harvested by centrifugation (15 min at 4,000 g, 4 °C). The cell pellet was washed 3 times with 200 ml sterile cold water and 1 time with 20 ml sterile cold 10% (v/v) glycerol. Finally, cells were resuspended in 1 volume sterile cold 10 % glycerol, dispensed in 100 µl aliquots and frozen in liquid nitrogen. The frozen cells were stored at – 80 °C.

3.9.3 Transformation by electroporation

50 µl electro-competent cells were thawed on ice and transferred to a chilled 0.2 cm electroportion cuvette. 4 µl ligation mixture was added and the sample was kept on ice for 5 min. Thereafter, the cuvetter was transferred to a Gene Pulser electroporation chamber and pulsed once with 25 µF, 2500 V, 200 Ohms. 0.45 ml LB medium was added immediately after the pulse and the sample were plated on LB-agar plates containing the appropriate antibiotic and incubated

overnight at 37 °C. The efficiency of electroporation can be 10^7 –10^9 transformants per µg of DNA.

3.10 PCR amplification of DNA

The polymerase chain reaction (PCR) is a very useful technique that allows to produce high yields of specific DNA target sequences. PCR was used in this work:

To isolate specific genes from bacterial genomic DNA

To create appropriate restriction sites at the terminal of DNA fragments to be cloned into different vectors.

To create appropriate tag at either the beginning or the terminal of DNA fragments.

To make mutagenesis

To check the correct clones by colony PCR

To sequence the DNA

Most PCR protocols are performed at the 20 µl – 100 µl scale, larger volumes are not recommended.

A typical 50 µl reaction mixture consist of:

1-10 ng plasmid DNA or 50-100 ng genomic DNA
20 pmol forward primer
20 pmol reverse primer
1x nucleotide mix (200 µM of each dNTP)
1x PCR buffer with MgCl2 (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$)
1 U DNA polymerase (Taq or Pfu polymerase)
HPLC-H_2O

The reaction is incubated in a thermocycler device where the temperature can be changed rapidly. Usually there is a preheating step of 2 min at 95 °C to denature the template DNA. Then another 30-35 cycles is following:

denaturing: 95 °C (30 –60 sec)

annealing 45 – 60 °C (30-60 sec)

elongation 72 °C (2 min for 2kb plasmid DNA or 1 kb genomic DNA)

The last cycle is followed by an extra elongation step of 5 – 10 min at 72 °C.

The annealing temperature is dependent on the primers composion, on their Tm (melting temperature) and on their homology with the template. The primers may have modifications such as extensions at their 5’ ends or point mutations.

3.10.1 Mutagenesis

The Quick Change site-directed mutagenesis kit from STRATAGENE was used to create point mutations in plasmid DNA containing cloned genes. The basic procedure starts with a supercoiled, ds DNA vector, with an insert of interest and two long oligonucleotide primers (30-50 bases) containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by Pfu-Ultra DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid. After temperature cycling, the products is treated by DpnI at 37 °C for 1-2 hr. The DpnI is used to digest the methylated non mutated parental DNA template since DpnI is specific for methylated and hemimethylated DNA. The nicked vector DNA incorporating the desired mutations is then transformed in *E. coli.*

3.10.2 Colony PCR to identify positive clone

Bacteria were taken with a tip from agar plates, dissolved in 60 µl HPLC water and boiled at 95 °C for 10 min, after that 10 µl of the mixture were used as template for PCR reaction. Alternatively, 1 µl Bacteria culture medium was taken as template PCR reaction.

3.10.3 Sequencing of DNA

DNA fragments were sequenced using the thermo Sequenas dye terminator cycle sequencing kit from Amersham on an ABI373A sequencing device according to the manufacture's instructions. For the sequence PCR, 200 –500 ng dsDNA is used as template DNA. 0.04 µl 0.1 nM primer and 2 µl sequence premix are added. There is 25 thermocycles is used as following:

95 °C 30 sec; 50 °C 30 sec; 60 °C 4 min

After thermocycles, the DNA in the mixture is precipitated by adding 50 µl pure ethanol and 2 µl 3M NaAc. the precipitated DNA was once washed by 250 µl 70% ethanol and resuspended by adding 25 µl water. Automated sequencing was performed by K. Neifer in this department.

3.11 Protein expression

3.11.1 Induction of *E. coli* by IPTG

For pGEX-KG, pBSKII vector, IPTG was used to induct the protein expression. 1 ml of overnight cultures, grown in the presence of 100 µg/ml ampicillin, were used to inoculate 100 ml fresh liquid media containing the same antibiotic. The turbidity of the samples was monitored and IPTG added to 0.2-1.0 mM upon reaching an A600 value of 0.3-0.5. The cells was grown for another 3 hr at 37 °C to expression the protein.

3.11.2 Induction of *E. coli* by AHT

For Strep-tag protein, AHT was used to induct the protein expression. 1 ml of overnight cultures, grown in the presence of 100 µg/ml ampicillin, were used to inoculate 100 ml fresh liquid media containing the same antibiotic. The turbidity of the samples was monitored and AHT added to 0.2-1.0 mM upon reaching an A600 value of 0.3-0.5. The cells was grown for another 3 hr at 37 °C to expression the protein.

3.12 Cell lysis

3.12.1 Lysis by ultrasonification

To the crash bacteria cells, Heat system Ultrasonics, USA (type 220-F) was used. The cells was resuspended in 0.1 M Tris-HCl, pH7.4 or pH8.0 and kept into an ice bath. The subsequent was set at step 8 and sample was sonificated for 10 min.

3.12.2 Lysis by "French Press"

"French Press" lysate was more efficient than ultra sonficiation especially for large volume cell lysate. The Mini Cell is fit for up to 3.5 ml cell lysate and the 40K Cell is fit for up to 35 ml cell lysate. 850 bar (for Mini Cell) and 1200 bar (for 40K Cell) was used to crash the bacteria cells.

3.13. Protein purification

3.13.1 6xHis-fusion protein purification

6xHis-fusion proteins can be purified on Ni-NTA metal affinity chromatography matrices (Janknecht *et al.,* 1991). Purification can be performed under native or denaturing conditions.

Purification under native conditions: 5 ml cell pellet from 1 l IPTG-induced bacteria were dissolved in 20 ml lysis buffer and lysed by French Press. Cell debris was eliminated by centrifugation for 10 min at 10000g (4 °C). The 10.000g supernatant was transferred to a new tube containing 0.5-1 ml of a 50% slurry of Ni-NTA resin (prewashed with lysis buffer), and incubated at 4 °C for 1 hr with end-over-end rotation. After that the resin was loaded onto a column and washed with 5 ml buffer-1 and 5 ml buffer-2. Finally, the protein was eluted with elution buffer.

Buffer-1 Tris-HCl pH 7.5 20 mM, 300 mM NaCl, 10 mM Imidazole (pH 7.5)
Buffer-2 Tris-HCl pH 7.5 20 mM, 300 mM NaCl, 20 mM Imidazole (pH 7.5)
Elution buffer Tris-HCl pH 7.5 20 mM, 300 mM NaCl, 250 mM Imidazole (pH 7.5)

Purification under denaturing conditions: The cell pellet from 100 ml culture was dissolved in 10 ml buffer-B and lysed by French Press. The lysate was centrifuged for 30 min at 10000g. The supernatant was transferred to a new tube containing 0.5 –1 ml of 50% slurry of Ni-NTA resin and incubated at 4 °C for 1-2 h with rotation.

After that the resin was loaded onto a column and washed twice with 10 ml buffer-C. Finally the protein was eluted with Buffer D and E.

Buffer-B	100 mM NaH_2PO_4, 10 mM Tris-HCl, 8M urea (pH 8.0)
Buffer-C	100 mM NaH_2PO_4, 10 mM Tris-HCl, 8M urea (pH 6.3)
Buffer-D	100 mM NaH_2PO_4, 10 mM Tris-HCl, 8M urea (pH 5.9)
Buffer-E	100 mM NaH_2PO_4, 10 mM Tris-HCl, 8M urea (pH 4.5)

3.13.2 GST-fusion protein purification

The cell pellet from 100 ml culture was dissolved in 10 ml PBS and lysed by French Press. The lysate was centrifuged for 30 min at 10,000g. The supernatant was transferred to a new tube containing 0.5 –1 ml of 50% slurry of glutathione Sepharose (prebalanced with lysis buffer) and incubated at 4 °C for 1-2 h with rotation. After that the Sepharose was loaded onto a column and washed twice with 10 ml PBS. Finally the protein was eluted with 3 times 1 bad volume elution Buffer.

Elution Buffer: 10 mM glutathione; 50 mM Tris-HCl

3.13.3 Strep-fusion protein purification

The cell pellet from 100 ml culture was in 5 ml Buffer W and lysed by French Press. The lysate was centrifuged for 30 min at 10,000 g (4 °C). The supernatant was transferred to the Strep-Tactin affinity column (1 ml Strep-Tactin ® Sepharose prebalanced by 2 ml Buffer W). The protein was binded to the column by flowing though under gravity. After the sample has completely entered the column, wash the column 5 times with 1 ml of Buffer W. Finally the protein was eluted with 6 times 0.5 ml Buffer E. The column can be regenerated by washing the column 3 times with 5 ml Buffer R.

BufferW:	100 mM Tris-HCl pH 8.0, 150 mM NaCl
Buffer E:	100 mM Tris-HCl pH8.0, 150 mM NaCl, 2.5 mM desthiobiotin
Buffer R:	100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM HABA

3.14 Protein analysis and detection

3.14.1 SDS-polyacrylamide gel electrophoresis

The electrophoretic separation of the proteins was performed through "high-Tris" discontinuous SDS polyacrylamide gel electrophoresis. A system with vertical oriented glass plates (16 X 16 cm; 1mm spacer) was used. The gels were prepared as following: first the separating gel (12.5 % acrylamide) was put between glass plates and covered with a layer of butanol. After polymerization of the gel was completed, butanol was removed and the spacebetween glass plateds was dried with Watman paper. The concentrating gel was put on top of the polymerized separating gel and the sample combs were introduced immediately. After approximately 10 min polymerization was completed and the sample combs were removed. The glass plates with the gel between them were fixed inside the electrophoresis chamber and covered with electrophoresis buffer. Protein samples were mixed with 2X Laemmli loading buffer 1:1, denatured by 95 °C 7 min and centrifuged with 14000 rpm 3 min. The supernatant was introduced into the concentrating gel pockets. Electrophoresis was performed for 2-3 hours by the constant current of 50 mA.

Eletrophoresis buffer:

SDS (w/v) 1 g/l Glycine (w/v); 14.4 g/l Tris (w/v); 6.05 g/l

Stacking gel	
Ingredients	
30 % Acrylamide solution (ml)	0.4
1% bisacrylamide solution (ml)	0.275
dd H_2O (ml)	1.175

0.5 M Tris-HCl, pH 6.8 (ml)	0.625
10% (w/v) SDS (µl)	25
TEMED (µl)	2.5
20% Ammoniumperoxiddisulfate (µl)	10
Separating gel	
Ingredients	12.5
30 % Acrylamid solution (ml)	7.3
1% bisacrylamid solution (ml)	2.9
dd H_2O (ml)	2.7
1.5 M Tris-HCl, pH 8.8 (ml)	4.3
10% (w/v) SDS (µl)	175
TEMED (µl)	14.5
20% Ammoniumperoxiddisulfate (µl)	58

3.14.2 Detection of protein in polyacrylamide gels

3.14.2.1 Staining with Roti-Blue colloidal Coomassie

Colloidal Coomassie staining is one of the most sensitive staining protocols. Due to its colloidal properties the dye binds with high specificity to proteins and only minimal to the gel matrix. This allows visualization of proteins separated by SDS-PAGE with sensitivity as high as 30 ng of protein. Roti-Blue colloidal coomassie was purchased from Roth and staining was performed according to the protocol as follows. Immediately after completion of electrophoresis gels were incubated in the fixing solution for 60 min with shaking. The staining solution was applied on to the gels for 2-3 hours. After completion of staining gels were incubated in the washing solution for 5 min and then kept in the stabilizing solution or prepared for drying in the drying solution.

Fixing solution: 20% (v/v) methanol, 0.85% (v/v) o-phosphoric acid,

Staining solution: 20% (v/v) methanol, 20 % (v/v) 5 x concentrated Roti-Blue colloidal Coomassie

Washing solution: 25% (v/v) methanol

Stablising solution: 20% (w/v) ammonium sulfate

Drying solution: 10% (v/v) glycerol, 20% (v/v) ethanol

3.14.2.2 Staining with silver (Bassam *et al.*, 1991)

Staining of proteins separated on SDS-gel with silver allows quick and effective visualization of even small amounts of protein. After completion of electrophoresis gels were incubated with shaking in fixing solution for 1 hour. After that gels were washed 2 times for 20 min in 30% ethanol and equilibrated for 20 min in water. Equilibrated gels were washed for 1 min in 0.03 % $Na_2S_2O_3$, washed with water and incubated in staining solution for 20 min at 4 °C. Then the gels were washed in water twice for 30 seconds and developed in the developing solution until desired intensity of the bands was obtained. Developing was stopped by transferring the gels into a new chamber with 5 % acetic acid. Gels were kept in 1 % acetic acid or washed with water to prepare them for drying.

Fixing solution: 10 % (v/v) acetic acid, 40% (v/v) ethanol

Staining solution: 0.2 % (w/v) silver nitrat, 0.008% (w/v) formic aldehyde

Developing solution: 3 % (w/v) sodium bicarbonate, 0.018 % (v/v) formic aldehyde

3.14.2.3 Western blot on PVDF membrane

Proteins separated by SDS-PAGE were transferred onto Nitrocellulose or PVDF membranes for western blot analysis by chemiluminescence. 6 sheets of whatman paper were cut according to the size of the gel. Three sheets were immersed in the semidry anode buffer and placed in the chamber. A glass pipette was rolled on it to remove any air bubbles. Then the membrane dipped in the cathode buffer was placed on top of it and again the air bubbles were removed. The gel was also soaked in the cathode buffer before placing on the membrane. Another three whatman paper sheets were dipped in cathode buffer and placed on top of the gel. The glass pipette was again rolled on the set up to remove air bubbles. The proteins were transferred onto the membrane using current at 1 mA/cm^2 for 60 - 90 min.

After the transfer, membrane was incubated in PBS containing 5 % milk powder for 1 hour at room temperature to block all non-specific interaction sites on the membrane. After blocking, primary antibody suitably diluted in blocking buffer, was added onto the membrane and incubated overnight in the cold room on a rocker. The non-specifically bound antibody was washed off by incubating the blot on a rocker with blocking buffer, changing the buffer once every 10 min for three times. The membrane was then incubated with the secondary antibody for 1 hour at room temperature and the membrane was washed 4 times, 10 min each, with PBS. The proteins were detected by chemiluminescence.

Chemiluminescent Substrate from PIERCE

Enhancer and peroxide solutions were mixed in a 1:1 ratio and incubated at room temperature for 5 min. The substrate was washed off and the membrane was wrapped in a polythene sheet and signals were detected using a CCD camera or an X-ray film.

3.14.2.4 Determination of protein concentration

Measuring of the protein concentration by Bradford method is based on the ability of Coomassie-staining to change its absorption maximum after binding to proteins from 465 nm to 595 nm. Quantitative measurement was performed with staining solution and according to the protocol “BioRad Protein Assay” (Richmond, USA) spectrometrically by 595 nm.

3.14.3 Arylsulfatase activity assay (Beil *et al.,* 1995)

Arylsulfatase was assayed as 4-nitrocatechol release from 4-nitrocatechol sulfate at 37 °C. The assay mixture (0.5 ml) contained 100 mM Tris/HCl (pH8.9 for PAS and pH 7.4 for AtsA) and 10 mM 4-nitrocatechol sulfate, and the reaction was started by addition of the enzyme or cell lysate containing enzyme. After suitable times, the reaction was stopped by diluting portions tenfold into 1 M NaOH, and the 4-nitrocatechol produced was quantified spectrophotometrically (λ=515nm, ε=12.4 $mM^{-1}cm^{-1}$). The arylsulfatase activity can be calculated by following equation.

$$Activity = \Delta E \times \frac{V \times 1000}{\varepsilon \times v \times t} \times F$$

Whereas

Activity: (mU/ml)

ΔE: increased absorption at 515nm

ε: 12.4

v: volumn of added enzyme or cell lysate (ml)

V: assay volumn (ml)

F: dilution factor (10)

t: reaction time (min)

3.15 Antibody production

DVS-His$_6$ purified by Ni-NTA-agrose affinity chromatography was used as antigen to raise polyclonal antibody in New Zealand white rabbit. 500 µg of DVS-His$_6$ was made up to 400 µl with PBS and mixed with 400 µl of Specol. A homogenous emulsion was prepared by vigorously mixing antigen and specol solution in a leuer lock syringes connected by a capillary tube. This emulsion was injected into the rabbit; pre-immune serum was collected before injection to use as a negative control in western blots and immunoprecipitations. 250 µg of the antigen was used for further booster injections. Booster injections were given once in two weeks and bleeds were collected two weeks after booster injections and analysed by western blot analysis. The rabbit was sacrificed two weeks after the fourth booster injection.

After each bleed the blood was incubated at room temperature for 4 hours to allow for clot formation and centrifuged at 13,000 g for 15 min at 4°C. The supernatant was carefully transferred into new tubes, prepared small aliquots and frozen at -20°C or at -80°C for long-term storage.

3.16 Pull-down assay

Pull down assay in this work was used for detect the binding between sulfatase fragment and sulfatase modification enzyme. The cell pellet, which contains GST-sulfatase fragment was resuspended in 8M urea/PBS and cracked by French Press. The cell lysate was centrifuged at 13,000 g for 15 min at 4 °C. The supernatant was dialyzed by PBS at 4 °C 1 hour 2 times. After dialyze, the supernatant was centrifuged again at 13,000 g for 15 min at 4 °C and 0.7 ml supernatant from 28 ml initial cell medium was transferred to the column containing 0.4 ml of 50% slurry of glutathione sepharose (prebalanced with PBS) and the lysate was flow though the column by gravity force. The sepharose was washed 3 times with 1.5 ml PBS. Then 0.7 ml cell lysate from 28 ml cell medium, which contains sulfatase modification enzymem was transferred to the column and flow though the column by gravity force without incubation. The sepharose was washed again 5 times 1.5 ml PBS. Finally the protein was eluted with 2 times 0.3 ml elution Buffer. The elution samples was loaded to SDS-PAGE and check the GST-sulfatase fragment and sulfatase modification enzyme by western blot.

3.17 *In vitro* formylglycine generation assay

In vitro FGly generation assay was established to determine the key factor of the sulfatase modification.

The cell lysate of GST-AtsA-(21-112) and AtsB was prepared as mentioned in 3.16. The reaction mixture, which contains 0.7 ml GST-AtsA-(21-112) cell lysate (or purified GST-AtsA-(21-112) protein), 0.7 ml AtsB cell lysate and certain amount SAM, was incubated at 37 °C for 15 to 3 hours. In some cases, SAM was added again to the reaction mixture. After incubation, the reaction mixture was transferred to 0.4 ml of 50% slurry of glutathione sepharose (prebalanced with PBS) and the lysate was flow though the column by gravity force. The sepharose was washed by 5 times 1.5 ml PBS and finally eluted by 1 time 0.4 ml elution buffer. 30 µl of elution sample was loaded to the SDS-PAGE. After coomassie staining, the stained band containing GST-AtsA-(21-112) protein was cut and digested by trypsin. After digestion, the peptide was detected by MALDI-MS to see whether or not the FGly was generated.

3.18 Identification of proteins by matrix-assisted laser desorption ionisation / time-of-flight (MALDI-TOF) mass spectrometry

3.18.1 In-gel digestion of protein separated by SDS-PAGE

The stained bands were excised from the gel with a scalpel and transferred into a 0.2 ml tube. The excised bands could have been kept at 4 °C for several days or frozen at –20 °C.

The analysed bands were cut into approximately 1x1 mm pieces and the staining was washed away by the following procedure:

- adding of 100 µl of HPLC-grade water; shaking the tube at 37 °C for 30min; discarding supernatant.
- adding of 100 µl of 25 mM of NH_4HCO_3; shaking the tube at 37 °C for 30 min; discarding supernatant.
- adding of 100 µl of 50% acetonitril, 25 mM NH_4HCO_3; shaking the tube at 37 °C for 30 min; discarding supernatant.
- adding of 100 µl of 100% acetonitril; shaking the tube at 37 °C for 10 min; discarding supernatant.

The washed gel pieces were dried at RT for 30 min. For in gel digesting 20 µl ice cold trypsin solution were added. Trypsin was purchased specially from Promega: V5111, Seq. Grade Modified. Gel pieces covered with trypsin solution were kept on ice for 15 min. For digestion the samples were incubated overnight in 37 °C chamber.

3.18.2 Extraction of tryptic peptides from the gel pieces

During overnight incubation with trypsin the proteins in the gel pieces were cleaved by trypsin with the formation of tryptic peptides. For further analysis the peptides should have been extracted from the gel matrix.

- centrifuge the tube at 13000 rpm for 30 sec; take the supernatant;
- add 20 µl 0.1 % TFA; shaking the tube at 37 °C for 15 min; take the supernatant;
- add 20 µl 50 % acetonitril/0.5 % TFA; shaking the tube at 37 °C for 15 min; take the supernatant.

The supernatant, containing tryptic peptides was dried in a vacuum concentrator (Speed Vac).To resolve dried peptides 10 µl of 0.1 % TFA was added to the tubes, the tubes were vortexed. Tryptic peptide solution could be kept at – 20 °C.

3.18.3 ZipTip purification of the trypic peptides before analysis by MALDI-MS

Purification of the tryptic peptides with ZipTip is used to remove salts and other contaminants that could disturb the following mass spectromectical analysis. Before purification the ZipTip was equilibrated with 0.1 % TFA/50 % acetonitril by pipetting the solution in and out of the tip 1 time with following washing with 0.1 % TFA.

Binding of the peptides to ZipTip matrix was performed by pipetting the solution through the tip for 4 times. The unbound components were washed away by 4 times 10 µl 0.1 % TFA. The peptides were eluted with 3 µl 0.1 % TFA/ 50 % acetonitril solution. The eluate was used for MALDI-MS analysis.

3.18.4 Sample/matrix preparation of MALDI-MS (drying droplet method)

For preparation of the DHB matrix, 1000 µl of H_2O was added to 5 mg of 2,5-dihydroxybenzonic acid with following vortexing for 1 min at RT and ultrasound sonification for 5 min. The solution was centrifuged in an Eppendorf centrifuge at 13000 rpm for 5 min. The supernatant was used for cocrystallization with the sample.

For preparation of DNPH matrix, 1 mg of 2,4-dinitrophenylhydrazone (DNPH) was added to 100 µl 50% acetonitril/0.5% TFA with following vortexing for 1 min at RT. The solution was centrifuged in an Eppendorf centrifuge at 13000 rpm for 1 min. The supernatant was used for cocrystallization with the sample. For best results this saturated matrix solution had to be made freshly.

For cocrystallization of the sample and matrix 0.5 µl of either DHB or DNPH matrix solution was carefully mixed with 0.5 µl of the sample on the MALDI-target and dried in the air at the room temperature.

3.18.5 Obtaining the mass spectra on MALDI-MS

MALDI positive ion mass spectra were obtained with a Bruker Daltonik Reflex III, using 337 nm nitrogen laser, with 200 ns extraction delay. Spectra were obtained as averages of 100 laser shots.

4 Results

4.1 *In vivo* formylglycine modification of *Klebsiella* sulfatase AtsA promoted by AtsB and its orthologs

The arylsulfatase AtsA of *K. pneumoniae* is a serine-type sulfatase that carries an FGly residue generated by oxidation of serine 72. In *E. coli*, expression of active, FGly containing AtsA essentially requires coexpression of the *K. pneumoniae atsB* gene that is adjacent to *atsA* on the same operon. AtsB is predicted to be a 44-kDa radical SAM (S-adenosylmethionine) protein with three FeS centers. Also in some other prokaryotes, there are *atsB* orthologs adjacent to putative sulfatase genes, such as *aslB* and *f390* in *E. coli* or *mm-atsB* in the archaebacterium *Methanosarcina mazei*. It should be noted that *mm-atsB* is the only *atsB* ortholog that is located adjacent to a cysteine-type sulfatase gene. Cysteine-type sulfatases, however, so far were believed to be modified exclusively by non-AtsB proteins (see chapter 1.3). Here, *atsB* and two of its orthologs, namely *aslB* and *mm-atsB*, were cloned together with *atsA* into artificial operons and expressed in *E. coli* to check whether or not AtsA can be activated not only by AtsB but also by its orthologs.

4.1.1 Coexpression of AtsA in *E. coli* together with AtsB and its orthologs

In order to study the role of *atsB* and its orthologs on expression of active *Klebsiella* arylsulfatase, encoded by the *atsA* gene, *atsB/atsA*, *aslB/atsA* and *mm-atsB-His$_6$/atsA* operons were cloned. Since in prokaryotes arylsulfatase expression is repressed during growth in the presence of sulfate, we cloned those operons without their endogenous promoters into an expression vector downstream of the *lac* promoter, allowing controlled expression at logarithmic growth. The proteins were expressed in the presence of isopropyl thiogalactopyranoside (IPTG) under aerobic conditions. The recombinant AtsA protein was detected by western blotting using antibodies that had been raised against purified *Klebsiella* arylsulfatase (Fig. 4.1). Its electrophoretic mobility was in agreement with the predicted mass of 62 kDa. The AtsB and MM-AtsB-His$_6$-proteins also were detected by western blotting using antibodies against purified AtsB or against the His$_6$-tag (Fig. 4.1).

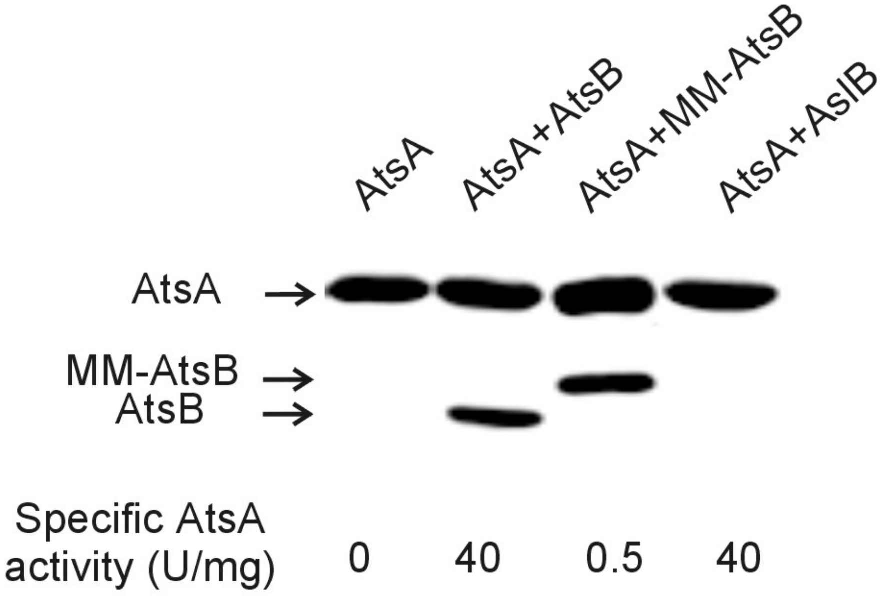


Figure 4.1 Active expression of *Klebsiella* sulfatase AtsA depends on coexpression of AtsB or its orthologs. The *atsA* gene or *atsB/atsA*, *aslB/atsA* or *mm-atsB-His$_6$/atsA* operons were expressed in *E. coli* DH5α at logarithmic growth in the presence of isopropyl thiogalactopyranoside. After expression, the soluble protein from the cell lysate was assayed for sulfatase activity and content of the recombinant proteins. The expressed sulfatase protein (AtsA) and the two orthologous proteins AtsB and MM-AtsB were detected by western blotting using anti-AtsA, anti-AtsB or anti-His$_6$-antibodies. Equal volumes of cell lysate were loaded on the SDS gel. The specific sulfatase activity as determined for the four different AtsA expressions is given below each lane.

Expression of AtsA alone did not lead to any detectable sulfatase activity, although the AtsA protein was produced at normal levels (Fig. 4.1 lane 1). Coexpression of AtsA with AtsB (Fig. 4.1 lane 2) or AslB (Fig. 4.1 lane 4) resulted in active AtsA expression, and the specific activity was around 40 U/mg AtsA protein. When MM-AtsB was coexpressed with AtsA, the expressed AtsA showed a low but significant sulfatase activity of 0.5 U/mg AtsA protein (Fig. 4.1 lane 3), *i.e.* 80-fold lower as compared to that observed after coexpression with AtsB or AslB.

Therefore, both AtsB and its orthologs can act on AtsA to convert it into its active form. The *E. coli* ortholog AslB can fully substitute for AtsB, whereas the archaebacterial MM-AtsB has a relatively poor activation ability.

4.1.2 Formylglycine generation in AtsA mediated by AtsB and its orthologs

In order to verify that the activation of AtsA was due to its FGly-modification, the AtsA protein was analyzed for the presence of FGly in position 72. The cell lysate containing recombinant AtsA was separated by SDS-PAGE and stained with Coomassie Blue. The stained bands containing AtsA were excised and subjected to in-gel digestion with trypsin and extraction of tryptic peptides. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) was used to identify the FGly-containing peptide. Residue 72 is part of the tryptic peptide 2 comprising AtsA residues 63-76 (M-S-Q-Y-Y-T-S-P-M-S/FGly-A-P-A-R). The expected mass for the serine 72-containing P2 is 1589 Da and that for the FGly 72-containing P2* is 1587 Da. Since the difference between modified and unmodified AtsA peptide is just 2 Da, DNPH (2,4-dinitrophenyl hydrazine) was used as a matrix for MALDI-MS. The FGly containing peptide P2* reacts with DNPH to form a hydrazone derivative (theoretical mass 1767 Da). Upon absorbance of laser energy this hydrazone is efficiently desorbed leading to a signal that is more abundant than that in other matrices. Due to a mass increment of 180 Da (Peng *et al.*, 2003), it can easily be discriminated from that of the serine 72-containing P2.

Fig. 4.2A shows that no P2* DNP-hydrazone could be detected by MALDI-MS for AtsA expressed in the absence of AtsB or its orthologs. This result is consistent with the lack of sulfatase activity (Fig. 4.1). When AtsA is expressed alone, the serine-72 is not modified to FGly and the protein is fully inactive. On the contrary, a clear 1767 Da-peak was detected by MALDI-MS when analyzing peptides from AtsA that had been expressed in the presence of AtsB (Fig. 4.2B) or AslB (Fig. 4.2C). The unmodified serine72-containing peptide 2 (1589 Da) was also detected (not shown). Therefore, when expressed in the presence of AtsB or AslB, serine72 of AtsA in both cases was partially modified to FGly. It was found out earlier that even when purified from *K. pneumoniae*, only about 60 % of AtsA polypeptides show FGly-modification (Miech *et al.,* 1998). Also after expression in *E. coli*, the FGly-modifying activity appears to be limiting. AtsB and AslB stimulate FGly-modification with comparable efficiency leading to similar sulfatase activities of AtsA (Figs. 4.1 and 4.2).

For coexpression of AtsA and MM-AtsB, one weak peak at 1586 Da was detected (data not shown) using Dihydroxybenzonic acid (DHB) as a matrix. Because of the contaminated tryptic peptide in 1766 Da, DNPH can not be used as a matrix to confirm the FGly in this peptide.

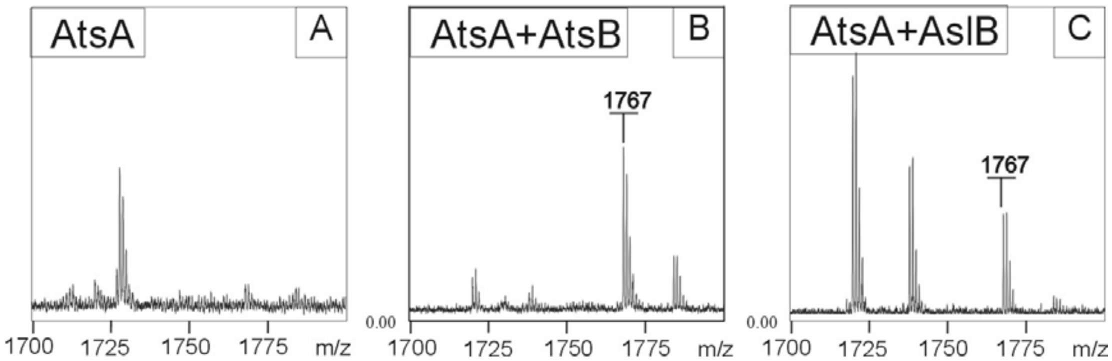


Figure 4.2 AtsB and its orthologs promote FGly-modification of *Klebsiella* sulfatase. About 20 µl soluble *E. coli* lysate containing AtsA, which had been expressed alone (A) or in the presence of AtsB (B), or AslB (C), were analysed by SDS-PAGE and Coomassie Blue staining. The band containing AtsA was subjected to in-gel digestion with trypsin. The tryptic peptides recovered from the gel slice and subsequent Ziptip-C18 purification were analysed by MALDI-MS using DNPH as a matrix. The m/z 1767 peak is the FGly72-containing peptide forming the FGly-DNP hydrazone upon reaction with DNPH. The m/z 1589 peak is the serine72-containing peptide.

In general, it may be concluded that the AtsB-orthologs, though being under control of specific sulfatase operon promoters are not specific for the respective sulfatase. Not only *Klebsiella* AtsB but also its homologs can activate *Klebsiella* sulfatase. Despite being much less efficient, even the archaebacterial

MM-AtsB, that naturally is coexpressed with a cysteine-sulfatase from the same transcript, can activate the serine-sulfatase AtsA.

4.2 Purification of AtsB

In order to solve the 3D-structure, analyse its FeS centers (EPR spectroscopy), and investigate the reaction mechanism of AtsB, we tried to purify it by affinity chromatography. For this purpose, GST-AtsB, AtsB-His$_6$, Strep-AtsB and Strep-AtsB-His$_6$ fusion proteins were expressed in *E. coli*. AtsB-His$_6$ protein turned out to be expressed in largely insoluble form and all efforts for solubilization and renaturation in the absence of agents like urea failed (data not shown).

4.2.1 Purification of GST-AtsB

4.2.1.1 Cloning of GST-AtsB

The *atsB* ORF was fused in frame to the 3' end of the GST-encoding sequence of the pGEX-KG vector. To facilitate insertion into the multi cloning site of this vector, we introduced an EcoRI site 5'of codon 2 and a HindIII site 3' of the stop codon of *atsB*, using add-on PCR. The *atsB* gene was subcloned as a 1.2-kb EcoRI/HindIII fragment into the pGEX-KG vector. For overexpression the pGEX-atsB plasmid was transformed to *E. coli* BL21 cells. In order to test whether GST-AtsB can activate AtsA *in vivo*, the pGEX-atsB plasmid was also transformed to *E. coli* DH5α containing the plasmid pBBR-atsA. Double transformants carrying both plasmids were selected due to their ampicillin and chloramphenicol resistance. The presence of the two genes was verified by PCR analysis.

4.2.1.2 GST-AtsB expression, *in vivo* activity and purification

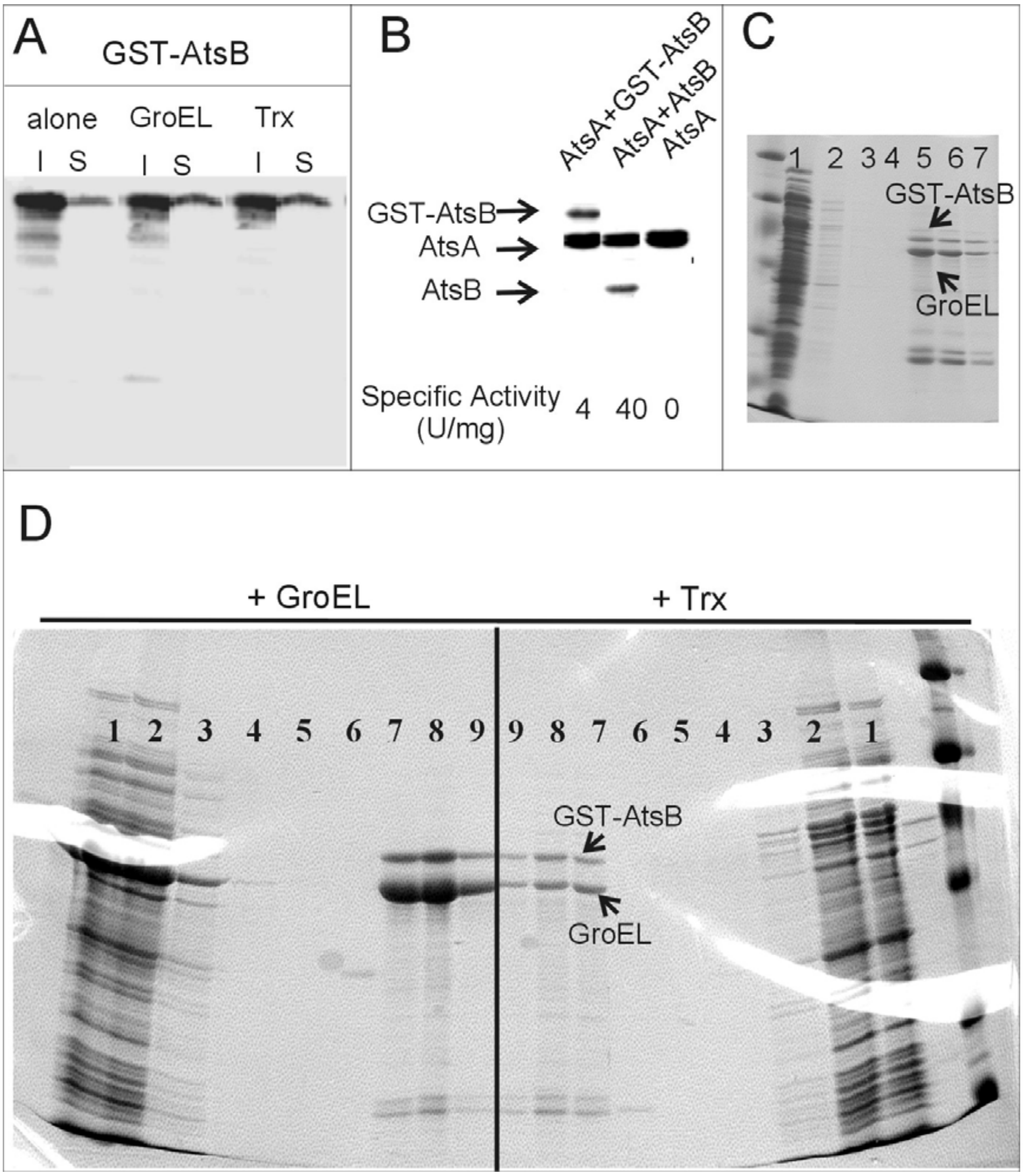

A
GST-AtsB
alone
GroEL
Trx
I
S
I
S
I
S
B
AtsA+GST-AtsB
AtsA+AtsB
AtsA
GST-AtsB
AtsA
AtsB
Specific Activity (U/mg)
4
40
0
C
1 2 3 4 5 6 7
GST-AtsB
GroEL
D
+ GroEL
+ Trx
1 2 3 4 5 6 7 8 9
9 8 7 6 5 4 3 2 1
GST-AtsB
GroEL

Figure 4.3 Expression and purification of GST-AtsB fusion protein. A) Cells expressing GST-AtsB with and without coexpression of GroEL or thioredoxin (Trx) were cracked using a "French Press". Aliquot samples of the soluble(S), and insoluble(I) protein fractions were prepared and analyzed on SDS-PAGE followed by western blotting using anti-AtsB antibodies. B) The *atsA* gene was expressed in the absence or presence of *atsB* or *gst-atsB* in *E. coli* DH5α at logarithmic growth in the presence of isopropyl thiogalactopyranoside. After expression, the soluble protein of the cell lysate was assayed for sulfatase activity and protein content. The expressed sulfatase protein (AtsA) and AtsB (AtsB and GST-AtsB fusion protein) were detected by western blotting after loading equal volumes of cell lysate on the SDS gel. The specific sulfatase activity as determined for the three different AtsA expressions is given below each lane. C) Following expression in *E. coli* BL21 (40ml culture), GST-AtsB was purified on GSH-agarose (200µl). The 60 KDa GroEL was copurified as identified by mass

spectrometric fingerprint analysis of its tryptic peptides. 20µl aliquots of each 1 ml wash or 200µl eluate were loaded. Proteins were visualized by Coomassie staining. 1: flow-through; 2-4; washes; 5-7; eluates. D) GST-AtsB (40ml culture), co-expressed with GroEL/GroES or thioredoxin, was purified on GSH-agarose (200µl). The 60 KDa GroEL is copurified. 20µl aliquots of each 1ml wash or 200µl eluate were loaded. Proteins were visualized by Coomassie staining. 1: cell lysate; 2: flow-though; 3-6; washes; 7-9; eluates.

Fig. 4.3A shows the expression of GST-AtsB protein in *E. coli* BL21. The GST-AtsB protein was detected by western blotting using anti-AtsB antibodies. The electrophoretic mobility of the fusion protein was in agreement with the predicted mass of 71 kDa. Only around 5-10% of GST-AtsB was soluble, and most GST-AtsB was found in inclusion bodies (Fig. 4.3A).

We tested whether GST-AtsB can activate AtsA *in vivo* by coexpression of GST-AtsB and AtsA. As shown in Fig. 4.3B, expression of GST-AtsB and AtsA resulted in production of active arylsulfatase, which means that at least some serine-72 in AtsA was modified to FGly. Therefore, GST-AtsB is functional *in vivo*. However, the specific sulfatase activity of AtsA, when coexpressed with GST-AtsB, is only 4 U/mg, *i.e.* 10% of that activity resulting from coexpression with wild-type AtsB (40 U/mg). Hence, the GST-AtsB fusion protein has a much lower activation ability than AtsB.

GST-AtsB was purified on GSH-agarose. The purification is documented in Fig. 4.3C. After SDS-PAGE two major proteins were detected in the elutate (Fig. 4.3C). The upper band, corresponding to a molecular weight of around 71 kDa, was GST-AtsB. The lower band, showing a molecular weight of around 60 kDa, was the *E. coli* chaperone GroEL, as determined by MALDI-MS fingerprint analysis of its tryptic peptides (data not shown). Because GroEL did not bind to GSH-agrose in the absence of GST-AtsB (not shown), the GroEL had been associated with the GST-AtsB fusion protein and therefore was co-eluted by glutathione (GSH). GroEL catalyzes correct folding of newly synthesized polypeptides, a process requiring ATP. Since GroEL bound to GST-AtsB, GST-AtsB may not be folded correctly, which could explain its low modification ability. In fact, association of GroEL with the flexible linker region downstream of the GST domain has been observed for many GST-fusion proteins. Treatment of GST-AtsB with ATP did not remove GroEL (data not shown).

Because most of the GST-AtsB expressed in *E. coli* was not soluble and even the soluble part showed GroEL-association, we coexpressed GST-AtsB with GroEL and its co-chaperone GroES or with thioredoxin to improve the solubility and folding of the GST-AtsB fusion protein. Fig. 4.3A shows that the solubility of GST-AtsB was increased to 10-15% after coexpression with GroEL/ES or thioredoxin. Nevertheless, in chromotographies stating from the soluble fraction of these coexpressing cells, GroEL still copurified with GST-AtsB (Fig. 4.3D).

We tried to remove GroEL together with the GST-domain by cleaving the fusion protein with thrombin at a thrombin cleavage site that had been inserted between the GST and the AtsB domain (Fig. 4.4). Two bands of AtsB were observed upon SDS-PAGE and western blot analysis of thrombin cleavage products (Fig. 4.4 A and C). The smaller product was due to cleavage C-terminal of Arg19 of AtsB, as determined by N-terminal sequencing. This position represents an endogenous thrombin cleavage site. After thrombin cleavage of GSH-agarose-bound GST-AtsB, some AtsB was eluted by PBS (Fig. 4.4A lanes 6-8), and some was eluted together with GST by GSH (Fig. 4.4A and B lanes 9-11). However, most of AtsB still stuck to the GSH-beads and was eluted only by SDS (Fig. 4.4A and C lane 12). Since GroEL-elution did not require GSH and therefore was not removed from AtsB in the PBS fractions (Fig. 4.4C). The yield of AtsB eluting with PBS or GSH was too low for the intended experiments.

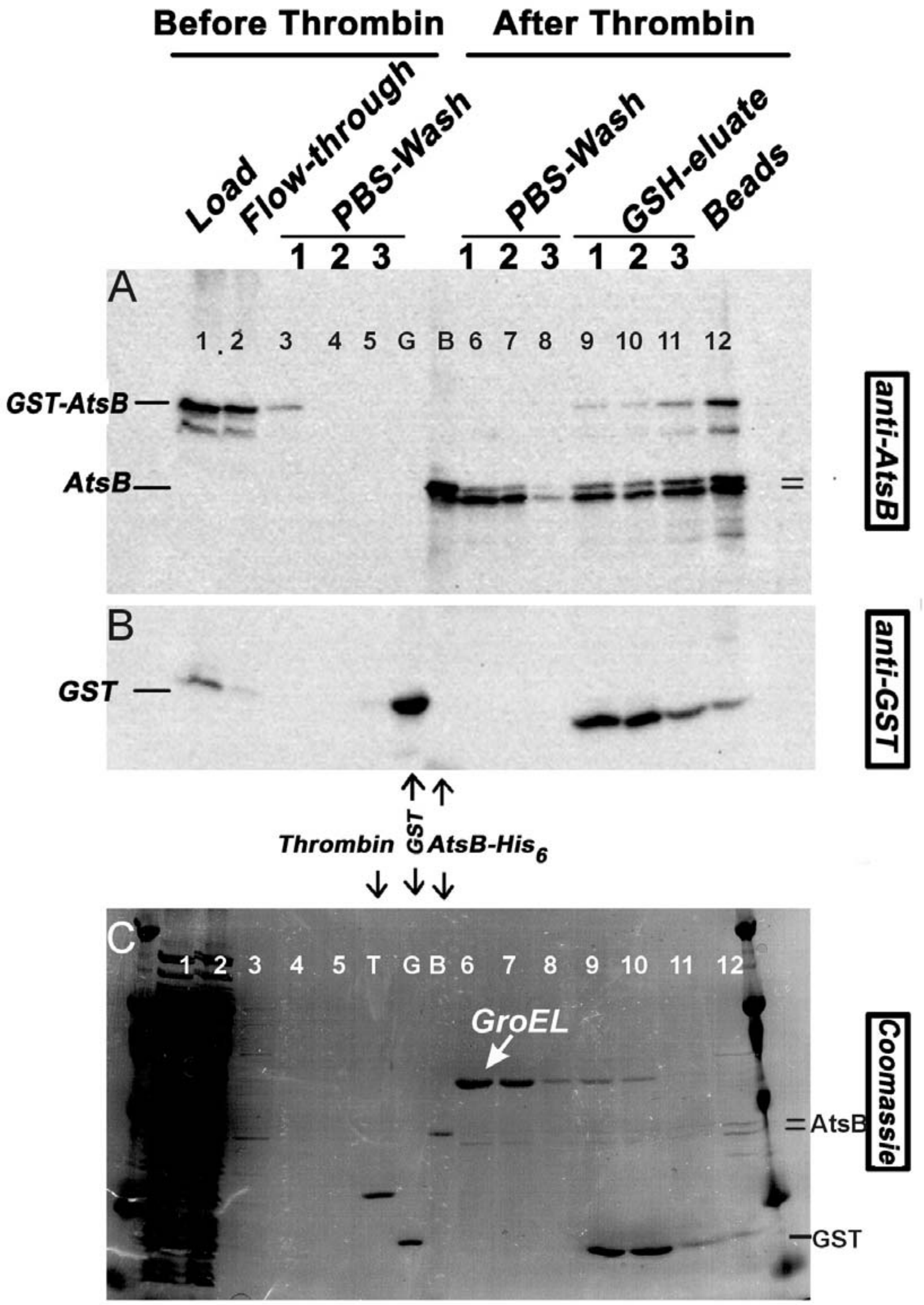

Before Thrombin
After Thrombin
Load
Flow-through
PBS-Wash
1 2 3
PBS-Wash
1 2 3
GSH-eluate
1 2 3
Beads
A
1 2 3 4 5 G B 6 7 8 9 10 11 12
GST-AtsB
AtsB
anti-AtsB
B
GST
anti-GST
Thrombin
GST
AtsB-His6
C
1 2 3 4 5 T G B 6 7 8 9 10 11 12
GroEL
AtsB
GST
Coomassie

Figure 4.4 On column cleavage of GST-AtsB fusion protein by thrombin. Following expression in *E. coli* BL21 (40ml), GST-AtsB was loaded on a GSH-column (200µl) and the column was washed with PBS. 10U thrombin was added to the column and the column was incubated at RT for 5h. After incubation, and washing by PBS (3x2ml), GSH-agarose was eluted with 3x200µl of 10mM

glutathione pH8.0. The GSH-agarose beads were boiled with SDS-laemmli buffer (200µl). 20µl of each wash and eluate were loaded on SDS-PAGE, Proteins were detected by western blotting using anti-AtsB (A) or anti-GST antibodies (B) or by Commassie staining (C). 1: cell lysate; 2: flow-though; 3-5: PBS washes; 6-8: PBS washes after thrombin cleavage; 9-11: GSH eluates; 12 Beads eluted with SDS; B: AtsB standard; G: GST-standard; T: Thrombin standard.

4.2.2 Purification of Strep-AtsB and Strep-AtsB-His$_6$

AtsB was equipped with an N-terminal StrepII-tag (NH_2-WSHPQFEK-COOH) followed by a factor Xa cleavage site (IEGR). For expression of Strep-AtsB, its ORF was placed under the transcriptional control of the *tetA* promoter of the pASK-IBA7 vector. To facilitate insertion into the multi cloning site of this vector, we introduced a BsaI site 5' of the *atsB*, using PCR methods. The PCR product was cloned as a BsaI/HindIII fragment into the pASK-IBA7 vector. The resulting pASK-IBA7-atsB plasmid was transformed to *E. coli* BL21 cells. The Strep-atsB plasmid was also transformed to *E. coli* DH5α containing pBBR-atsA plasmid, to test whether Strep-AtsB can active AtsA *in vivo*. These double transformants were selected due to their ampicillin and chloramphenicol resistance. The presence of the genes was verified by PCR analysis.

The Strep-AtsB was expressed in the presence of AHT as an inducer of the *tetA* promoter under aerobic conditions, whereas double transformants containing Strep-AtsB and AtsA were induced by both AHT and IPTG. Fig. 4.5A shows that 30-35% of Strep-AtsB was expressed as soluble protein. In order to check its activation ability, Strep-AtsB was coexpressed with AtsA. The specific sulfatase activity of AtsA, coexpressed with Strep-AtsB, was 40 U/mg, which was comparable to that of AtsA coexpressed with wild-type AtsB. Therefore, AtsB and Strep-AtsB, promoted FGly generation with similar efficiency and the StrepII-tag did not influence AtsB function.

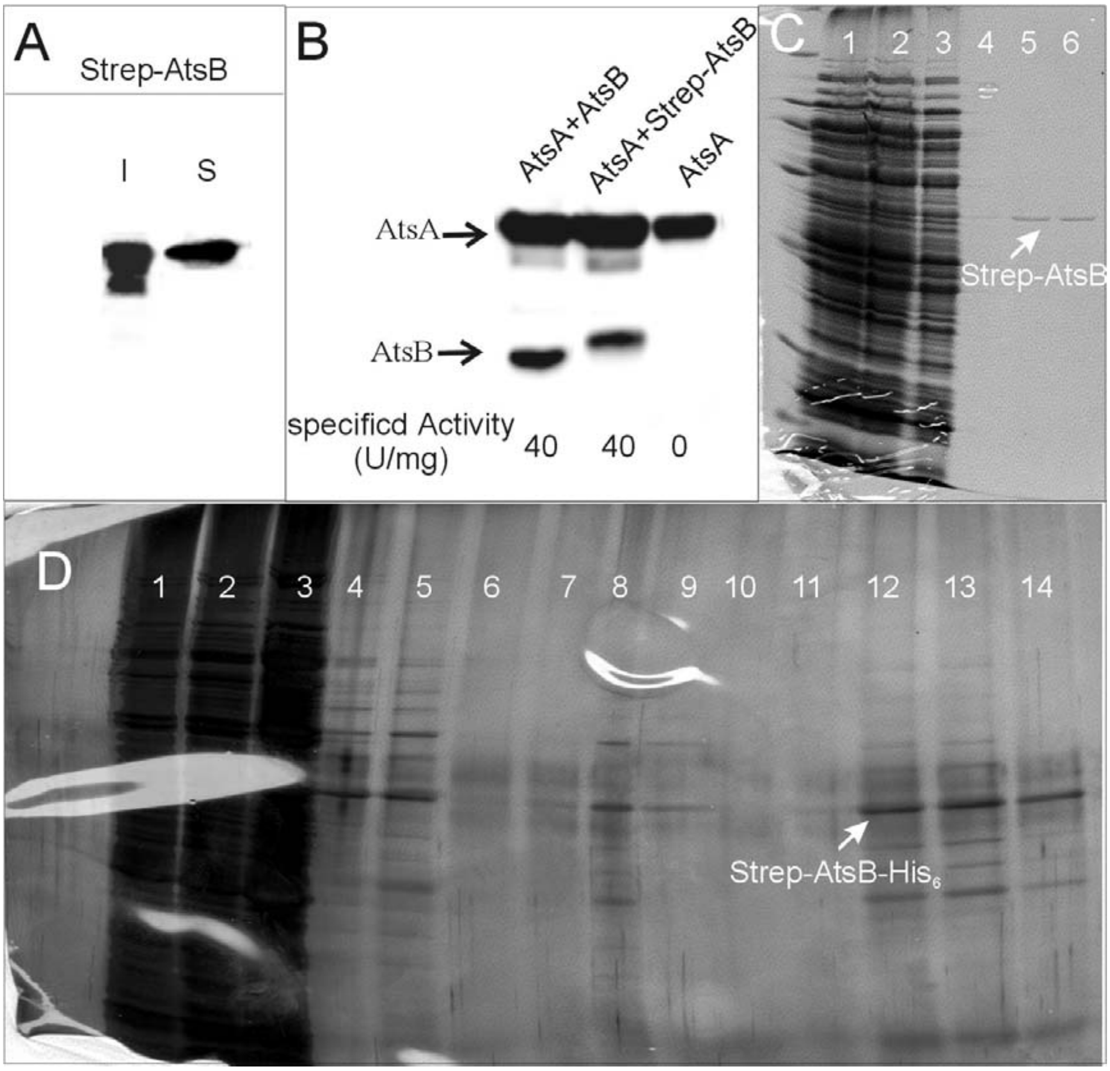


Figure 4.5 Expression and purification of Strep-AtsB fusion protein. A) Cells expressing Strep-AtsB were cracked using a "French Press". Aliquot samples of the soluble(S), and insoluble(I) protein fractions were prepared and analysed on SDS-PAGE followed by western blotting using anti-AtsB antibodies. B) The *atsA* gene was expressed in the absence or presence of *atsB* or *strep-atsB* in *E. coli*. After expression, the soluble protein of the cell lysate was assayed for sulfatase activity and protein content. The expressed sulfatase protein (AtsA) and AtsB (AtsB or Strep-AtsB fusion protein) were detected by western blotting after loading equal volumes of cell lysate on the SDS gel. The specific sulfatase activity as determined for the three different AtsA expressions is given below each lane. C) Following expression in *E. coli* (100ml), Strep-AtsB was purified using Strep-Tactin affinity column (200µl). 20µl of each 2 ml wash or 400µl eluate were loaded. Proteins were visualized by Coomassie staining. 1: cell lysate; 2: flow-through; 3-4; washes; 5-6; eluates. D) Strep-AtsB-His$_6$ from 100ml *E. coli* culture, was purified in sequence first on Ni-NTA agarose (200µl) and then on Strep-Tactin affinity column (200µl) with intermittent dialysis. 20µl of each 2 ml wash or 400µl eluate were loaded on an SDS-gel. Proteins were visualized by silver staining. 1: cell lysate; 2: flow-though of Ni-NTA column; 3: wash of Ni-NTA column; 4-5: eluates of Ni-NTA column; 6-7: dialysis buffer; 8: loading sample of Strep-Tactin column; 9-11; washes of Strep-Tactin column; 12-14: eluates of Strep-Tactin column.

Strep-AtsB was purified via Strep-Tactin affinity column under native conditions. In the eluates, Strep-AtsB was the major protein detectable by Coomassie blue staining (Fig. 4.5C lane 5-6). The purity of Strep-AtsB was higher than 80 %.

Strep-AtsB-His$_6$ fusion protein was used to further improve the purity of AtsB. The RGS-His$_6$ tag was subcloned as a 737 bp EagI/EcoRV fragment from a *atsB-His$_6$* construct (Szmeit *et al.*, 1999) into the Strep-AtsB plasmid (thereby replacing the corresponding EagI/EcoRV fragment). Fig. 4.5D shows the Strep-AtsB-His$_6$ purification. Strep-AtsB-His$_6$ was first purified on nickel-NTA-agarose under native conditions. The purity of eluting protein (Fig. 4.5D, lane 5) was around 40-50 %. After removal of imidazol by dialysis, the eluates (Fig. 4.5D lane 8) were subjected to chromatography on a Strep-Tactin affinity column. The final purity of Strep-AtsB-His$_6$ was around 60-70 % (Fig. 4.5D, lane 12-14). Compared to the purity of Strep-AtsB purified only on Strep-Tactin (Fig. 4.5C), this two-step affinity chromatography purification did not improve AtsB's purity.

At present Strep-AtsB is the best construct for structural analysis, because it has the highest purity after one step affinity purification, and the Strep-tag does not change AtsB's activity. However, the expression of Strep-AtsB is just 3 mg/L in *E. coli* culture, and after purification only one third of this protein is recovered (data not shown). Since AtsB is an FeS protein, which is sensitive to oxygen, aerobic growth may be inappropriate for its expression. The expression and purification of AtsB under anaerobic conditions is presently being studied in cooperation with Dr. Selmer (University of Marburg). Purified Strep-AtsB was used for some functional *in vitro* experiments described below (see Chapter 4.7).

4.3 Physical interaction of AtsB with AtsA

AtsB is essential for FGly-modification of AtsA. To find out whether AtsB plays a direct role in this posttranslational protein modification, it was tested whether AtsB makes contact to AtsA and whether this interaction is dependent on the FGly modification motif of AtsA. To address this question, we performed biochemical *in vitro* interaction experiments using a glutathione *S*-transferase (GST) pull-down assay.

4.3.1 Cloning of GST-AtsA-(21-112)

For construction of a GST-fusion protein, codons 21-112 of *atsA*, *atsA-S72C*, *atsA-S72A/P74A/R76A* was amplified by PCR using primers that add a 5' *Eco*RI site and a 3' *Sac*I site. The PCR products were cloned as *Eco*RI/*Sac*I fragments into the pGEX-KG vector in frame with its glutathione *S*-transferase encoding sequence. The plasmids were transformed to *E. coli* BL21 cells.

4.3.2 Interaction of AtsB with the FGly-modification motif of GST-AtsA-(21-112)

AtsA-(21–112) and its S72C and S72A/P74A/R76A mutant forms were expressed in *E. coli* as C-terminal appendices of GST. The fusion proteins were recovered from inclusion bodies. After solubilization in 8 M urea, dialysis against PBS and centrifugation, the soluble material was bound to glutathione-agarose columns. After washing, the soluble fraction of a lysate of *E. coli* expressing AtsB was applied. After further washing steps, the GSH-agarose column was eluted with glutathione. The eluate was analyzed on a western blot using anti-GST and anti-AtsB antibodies (Fig. 4.6). GST-AtsA fusion proteins were recovered in the glutathione eluate as full-length proteins and, in part, as C-terminally truncated forms (Fig. 4.6A). Significant amounts of AtsB (about 600 ng of AtsB/µg of nontruncated GST-AtsA-(21–112) (*i.e.* 0.54 mol/ mol)) were detected in the eluate of the GST-AtsA-(21–112) column (Fig. 4.6B), whereas a much lower amount of AtsB was present in the eluate of the GST-AtsA-(21–112)-S72C column (about 0.04 mol of AtsB/mol of GST-AtsA-(21–112)-S72C (Fig. 4.6B). No AtsB was detected in the eluates of columns loaded with GST, PBS (Fig. 4.6) or GST-AtsA-S72A/P74A/R76A (data not shown). In conclusion, AtsB firmly interacts with AtsA-(21–112) in a Serine 72-dependent manner. This finding strongly suggests that AtsB is the FGly-generating enzyme for the serine-type sulfatase AtsA of *K. pneumoniae*. It should be noted, however, that upon FGly-analysis of GST-AtsA, after incubation with AtsB, no modification could be detected by MALDI-MS (not shown).

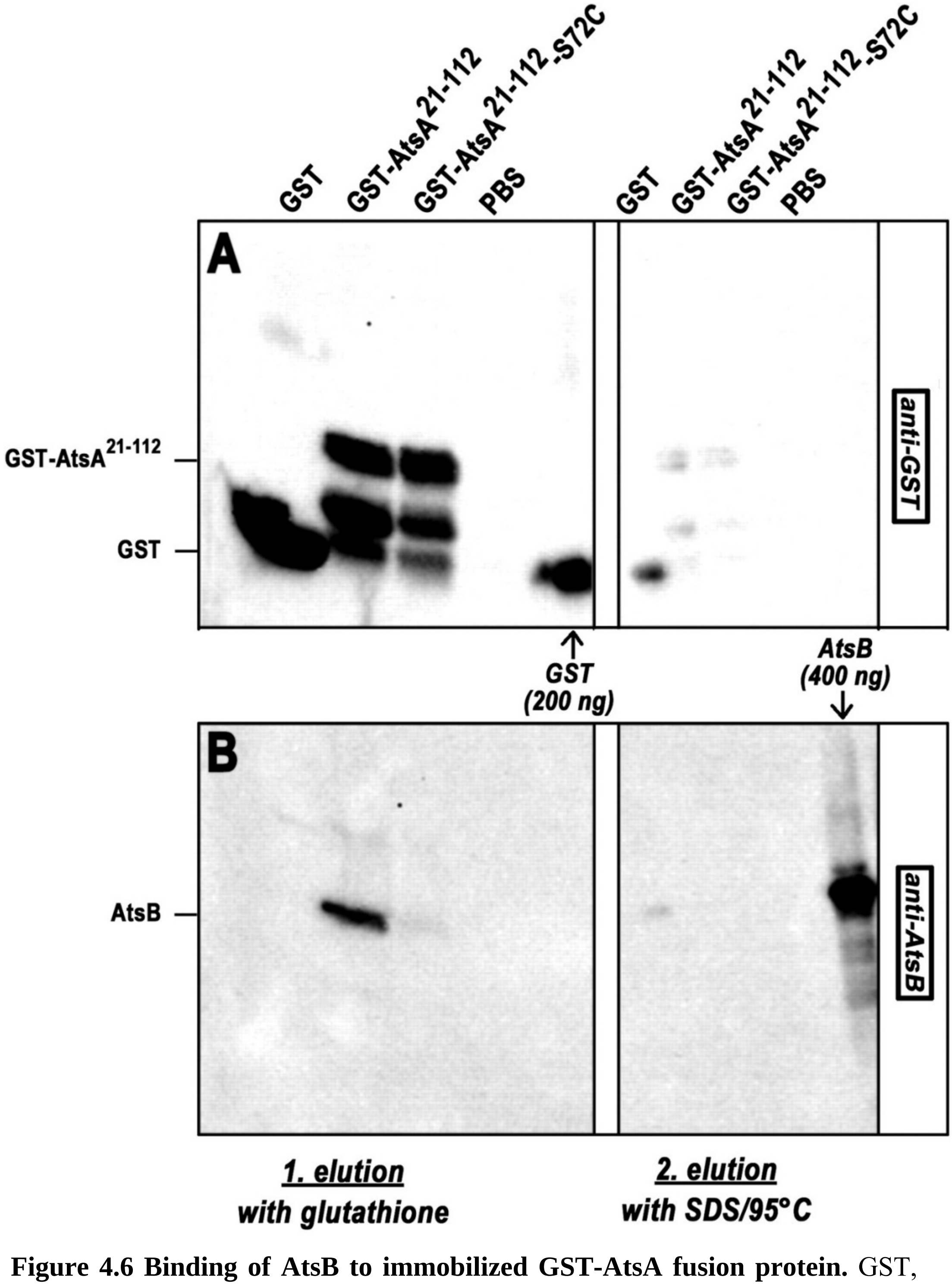

GST
GST-AtsA21-112
GST-AtsA21-112 S72C
PBS
GST
GST-AtsA21-112
GST-AtsA21-112 S72C
PBS
A
B
GST-AtsA21-112
GST
AtsB
anti-GST
anti-AtsB
GST (200 ng)
AtsB (400 ng)
1. elution with glutathione
2. elution with SDS/95°C

Figure 4.6 Binding of AtsB to immobilized GST-AtsA fusion protein. GST, GST-AtsA-(21-112) and GST-AtsA-(21-112)-S72C were solubilized in 5 M urea. After dialysis against PBS and centrifugation, the soluble material or, as a control, PBS buffer were loaded on glutathione-agarose, which then was washed thoroughly. All four columns were then loaded with the soluble fraction of an *E. coli* lysate (in PBS) containing AtsB protein. After another three washing steps, the columns were eluted with glutathione and then boiled with SDS-PAGE sample buffer. The eluates obtained with glutathione and SDS, were analyzed on a

western blot using goat anti-GST (A) or rabbit anti-AtsB antibodies (B). The positions of GST (27 kDa), GST-AtsA-(21-112) (39 kDa), and AtsB (44 kDa) are indicated. In addition to full-length GST-AtsA-(21-112) fusion proteins, the anti-GST antibody detected C-terminally truncated forms of about 33 and 29 kDa. 200 ng of purified GST (*A*) and 400 ng of purified AtsB (*B*) were used as standards, which allowed us to quantitate the detected western blot signals. Controls demonstrated that the washings efficiently removed unbound GST fusion proteins and AtsB (not shown). Elution with glutathione was nearly quantitative, since only minute amounts of GST fusion proteins were extracted from the glutathione-agarose beads by SDS.

4.4 Substrate specificity of AtsB

4.4.1 Interaction of AtsB with putative serine-type sulfatases of *E. coli*

There are two putative serine-type sulfatases in *E. coli*, termed F571 and AslA. And, there are also *atsB* orthologs, *f390* and *aslB*, adjacent to the *f571* and *aslA* ORF respectively. As described above (Chapter 4.1), we found that AslB of *E. coli* modifies the serine-type sulfatase AtsA of *Klebsiella*. This raises the question whether vice versa *K. pneumoniae* AtsB can modify other serine-type sulfatases, such as F571 and AslA of *E. coli*. The substrate of F571 and AslA is unknown; therefore we could not detect the sulfatase activity of these two enzymes after expression with and without AtsB. As an alternative approach, we used the *in vitro* interaction assay (GST-pull-down assay), to study whether AtsB can bind to AslA and F571.

In this pull-down experiment, codons 25-176 of *aslA* and 41-185 of *f571*, encoding the critical serines in position 136 and 143, respectively, were amplified by PCR using primers that added a 5'EcoRI site and a 3'SacI site. The PCR products were cloned as EcoRI/SacI fragments into the pGEX-KG vector in frame with its glutathione S-transferase encoding sequence.

Similar to the experiments with GST-AtsA(21-112) (Fig. 4.6), GST-AslA(25-176) and GST-F571(41-185) were expressed in *E. coli* BL21, recovered from inclusion bodies, and the renatured protein was bound to glutathione-agarose columns. After washing, the soluble fraction of an *E. coli* lysate expressing AtsB was applied. The GSH-agarose columns, after extensive washing, were eluted with glutathione. The eluates were analyzed on a western blot using anti-GST and anti-AtsB antibodies (Fig. 4.7). GST-AslA(25-176) and GST-F571(41-185) were detected in the eluates (Fig. 4.7A lane 1 and 2). However, no coelution of AtsB was observed. Here GST-AtsA(21-112) and GST-AtsA(21-112)S72A were used as positive and negative controls (Fig. 4.7 lane 3 and 4). Thus, no *in vitro* interaction between AtsB and AslA or F571 was detected. This does not necessarily mean that AtsB can not act on the *E. coli* sulfatases *in vivo*. Under the applied *in vitro* conditions the long N-terminal sequence preceding the FGly-motif (residues 25-135 and 41-142 in GST-AslA and GST-F571, respectively) may fold into a structure that impairs proper interaction with AtsB. Extending the corresponding sequence (residues 21-71) in GST-AtsA also impaired the interaction with AtsB (data not shown). As will be shown in Chapter 4.5.3, AtsB *in vivo* activates a C51S-mutant form of the arylsulfatase PAS of *P. aeruginosa*. This indicates that AtsB activity is not restricted to the sulfatase of its operon.

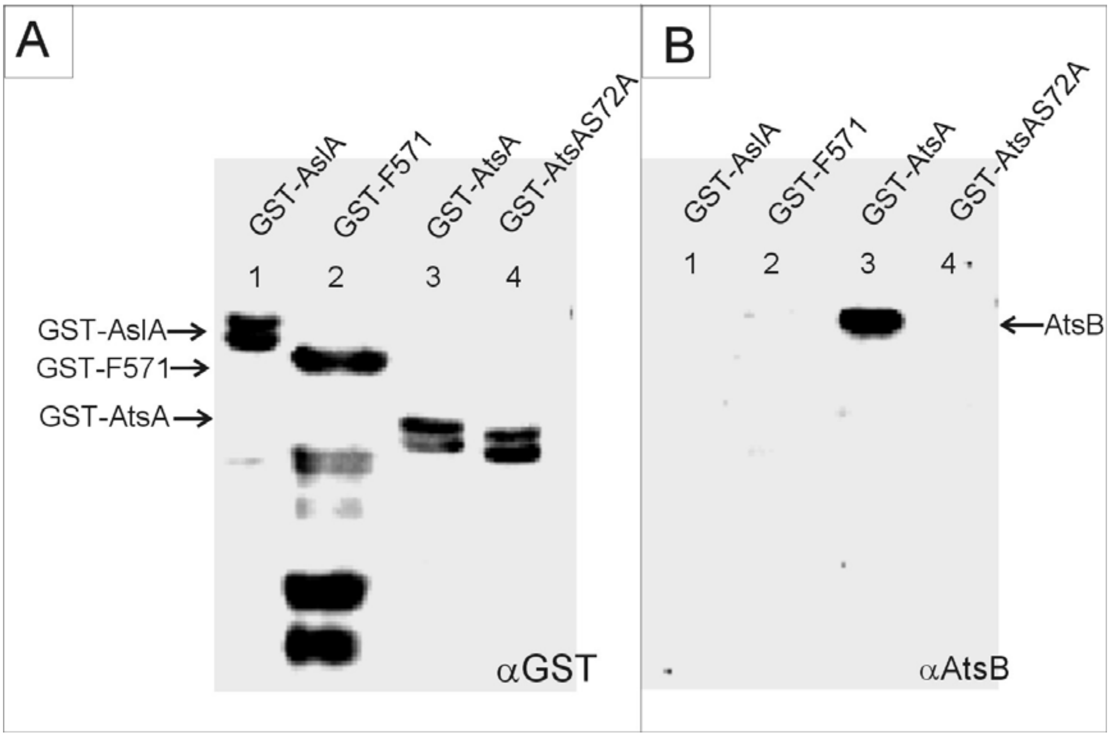


Figure 4.7 *In vitro* interaction of AtsB with AtsA orthologs (F571, AslA). GST-AslA-(25-176), GST-F571-(41-185), GST-AtsA-(21-112) and GST-AtsA-(21-112)S72A were solubilized in 8 M urea. After dialysis against PBS and centrifugation, the soluble material was loaded on GSH-agarose, which then was washed thoroughly. All four columns were then loaded with the soluble fraction of an *E. coli* lysate (in PBS) containing AtsB protein. After another three washing steps, the columns were eluted with glutathione. The eluates were analysed on a

western blot using goat anti-GST (A) or rabbit anti-AtsB antibodies (B). The positions of GST-AslA-(25-176) (44KDa), GST-F571-(41-185) (42KDa), GST-AtsA-(21-112) (39KDa), and AtsB (44KDa) are indicated. The signals detected for GST-AtsA in lanes 3 and 4 suggest slightly different electrophoretic mobilities. It was verified on other gels that the two proteins have identical size (not shown).

4.4.2 AtsB can modify serine- and cysteine-type sulfatases

4.4.2.1 AtsB and MM-AtsB can modify the cysteine-type sulfatase MMS

Two kinds of sulfatases can be discriminated with respect to the amino acid residue that is converted to FGly, namely serine-type and cysteine-type sulfatases (see Chapter 1.1.3). Normally *atsB* and its orthologs are adjacent to serine-type sulfatase genes, *e.g.* *aslB* is adjacent to *aslA*, *f390* to *f571*, and *atsB* to *atsA*. It was interesting to find that in the archaebacterium *Methanosarcina mazei*, an *atsB* ortholog is adjacent to a putative cysteine-type sulfatase ORF. We refer to these orthologous genes as *mm-atsB* and *mms*, respectively. In view of this sulfatase operon structure the question arises as to whether MM-AtsB and/or AtsB can modify MMS.

The *mms* and *mm-atsB* genes were amplified from *Methanosarcina mazei* genomic DNA by PCR. To facilitate insertion into the multi cloning site of pBSK KS II vector, we introduced suitable restriction sites 5’ and 3’ of the *mms* and *mm-atsB* gene, respectively, using PCR methods. A C-terminal Arg-Gly-Ser-$(His)_6$ tag was added to both MM-AtsB and MMS to allow for purification of these proteins and detection of their expression. In order to study the role of AtsB and MM-AtsB on expression of active MMS, $\textit{mm-atsB-His}_6\textit{/mms-His}_6$ and $\textit{atsB/mms-His}_6$ operons were cloned. All the plasmids were transformed to *E. coli* DH5α, and the proteins were expressed in the presence of isopropyl thiogalactopyranoside. The expressed MMS did not show activity towards the synthetic arylsulfate *p*-nitrocatechol sulfate. Because the natural substrate of MMS is unknown, we could not detect the sulfatase activity of MMS. Therefore, we directly analyzed whether FGly was generated in MMS by MALDI-MS methods.

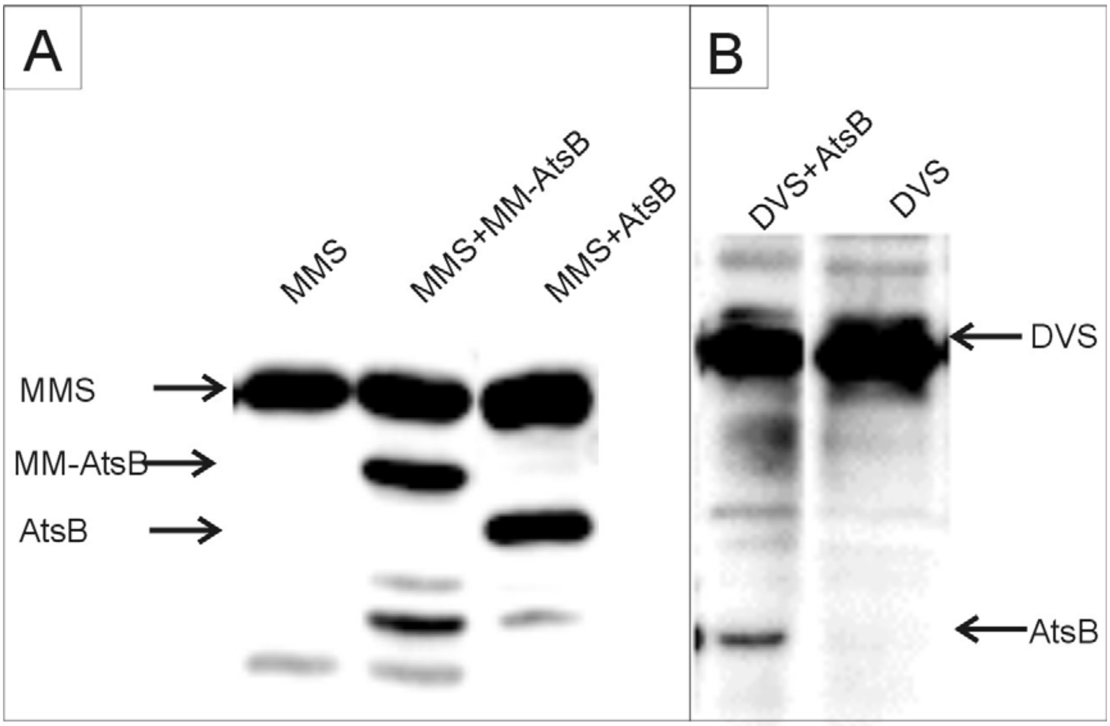


Figure 4.8 Expression of putative sulfatases MMS from *Methanosarcina mazei* (A) and DVS from *Desulfavibrio vulgaris* (B) in the absence and presence of AtsB. The *dvs*, *atsB/dvs*, *mms-His*$_6$, *atsB/mms-His*$_6$, *mm-atsB-His*$_6$*/mms-His*$_6$ operons were expressed in *E. coli* DH5α as indicated. After expression, the soluble protein from the cell lysate was assayed for protein content. The expressed DVS, MMS-His$_6$, AtsB, MM-AtsB-His$_6$ were detected by western blotting, after loading equal volumes of cell lysate (corresponding to 0.8ml culture) on the SDS gel, by anti-His$_6$ antibodies or by anti-DVS and anti-AtsB antibodies .

The expression of MMS-His$_6$ and the coexpression of AtsB or MM-AtsB-His$_6$ was detected by western blotting using antibodies directed against AtsB or against the His$_6$-tag (Fig. 4.8 A). Coexpression of AtsB or MM-AtsB-His$_6$ did not influence the MMS-His$_6$ expression level. MMS-His$_6$ protein was purified on nickel-NTA-agarose, separated by SDS-PAGE and stained with Coomassie Blue. The stained bands containing MMS-His$_6$ were excised and subjected to in-gel digestion with trypsin and extraction of tryptic peptides. MALDI-MS was used to identify the FGly modification. Residue 55 is part of the tryptic peptide 2 comprising MMS residues 48-59 (T-D-S-Y-G-E-Q-S-C/FGly-T-A-G-R). The expected mass for the cysteine 55-contaning P2 is 1520 Da and that for the FGly 55-containing P2* is 1502 Da. After hydrazone derivatization by DNPH, the mass for P2* would shift to 1682 Da, whereas the mass for P2 should not change (1520 Da).

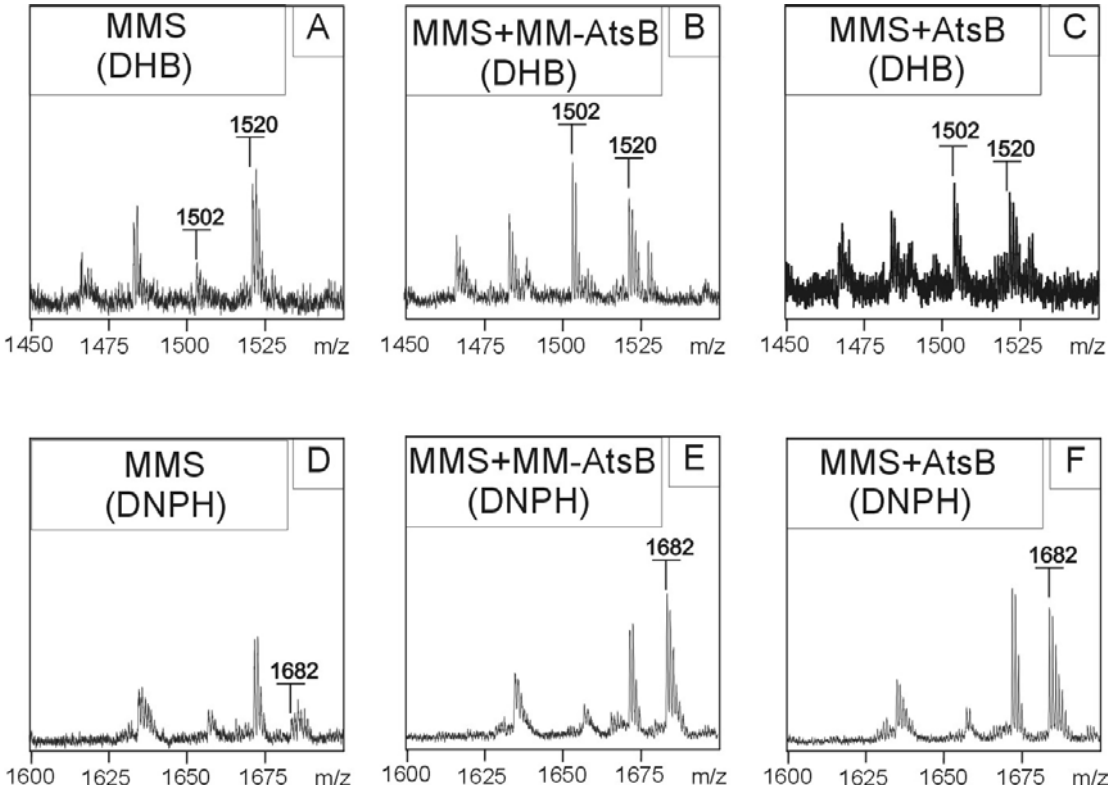


Figure 4.9 FGly generation in *Methanosarcina mazei* sulfatase (MMS) depends on coexpression of AtsB or MM-AtsB. About 1µg purified MMS-His_6 protein, which had been expressed with or without AtsB or MM-AtsB, were analysed by SDS-PAGE and Coomassie Blue staining. The bands containing MMS were subjected to in-gel digestion. The tryptic peptides recovered from the gel slices were purified on Ziptip-C18 and analysed by MALDI-MS using DHB (A, MMS-His_6; B, MMS-His_6 coexpressed with AtsB; C, MMS-His_6 coexpressed with MM-AtsB) and DNPH (D, MMS-His_6; E, MMS-His_6 coexpressed with AtsB; F, MMS-His_6 coexpressed with MM-AtsB) as matrix. The m/z 1502 peak represents the FGly55-containing peptide; the m/z 1520 peak the cysteine55-containing peptide; and the m/z 1682 peak corresponds to the FGly55-containing peptide after reaction with DNPH to the FGly-DNP hydrazone derivative.

Fig. 4.9A shows that no or only traces of FGly-containing P2* could be detected by MALDI-MS for the MMS-His_6 protein expressed in the absence of AtsB or MM-AtsB. The signal/noise ratio of the m/z 1502 Da peak was around 2. The unmodified peptide P2 containing cysteine was detected very clearly (mass 1520 Da). As calculated from the height of the m/z 1520 and m/z 1502 peaks, the FGly modification degree of the analyzed MMS protein was lower than 10%. Using DNPH as a matrix, only a very weak m/z 1682 peak was detected (signal/ noise ratio around 2). Therefore, when MMS was expressed alone, only traces of cysteine-55 were modified to FGly. *E. coli*, has an unknown FGly generation system (see Introduction), which can modify the cysteine-type sulfatase PAS from *P. aeruginosa*. When PAS is expressed alone in *E. coli*, the cysteine-51 is quantitatively modified to FGly (Dierks *et al.,* 1998). But in MMS, only traces of cysteine were modified to FGly by *E. coli*'s unknown FGly generation system. In conclusion, this FGly generation system shows a restricted substrate specificity towards cysteine-type sulfatases.

When MMS-His$_6$ was expressed together with MM-AtsB or AtsB, FGly was generated with much higher efficiency, as shown in Fig. 4.9B and Fig. 4.9C, respectively. The modification efficiencies of MMS-His$_6$ in both cases were around 60% as estimated from the m/z 1502 and m/z 1520 signal intensities in Fig. 4.9B and C. The generated FGly was confirmed by efficient DNPH derivatization (mass 1682 Da in Fig. 4.9E and F). Thus, coexpression of MM-AtsB or AtsB greatly improved the FGly generation in MMS, indicating that i) MM-AtsB is the FGly generation enzyme of MMS in *M. mazei* and ii) AtsB can modify both serine-type sulfatases (such as AtsA (Chapter 4.1)) and cysteine-type sulfatases (such as MMS). The posttranslational oxidation of a conserved serine or cysteine residue to FGly by the same AtsB protein indicates that these processes are mechanistically very similar despite the different chemical nature of the oxidation reaction.

4.4.2.2 AtsB can not modify the threonine-type sulfatase DVS

The *dvs* gene encodes a putative sulfatase in *Desulfavibrio vulgaris*. Normally the amino acid residue in sulfatases, which is modified to FGly, is either a serine or a cysteine. However, in *dvs*, a threonine is encoded at the critical position, as we could verify by sequencing of corresponding PCR products generated from a genomic template of *D. vulgaris*. In the *D. vulgaris* genome, no *atsB* or *fge* orthologs can be found. Here we studied whether AtsB and/or *E. coli*'s unknown FGly generation system can modify this threonine to FGly or to the related formylalanine (FAla).

The *dvs* gene was amplified from *Desulfavibrio vulgaris'* genomic DNA by PCR and cloned into the multi cloning site of pBSK KS II vector using a KpnI and a HindIII site added 5' and 3' of the *dvs* gene, respectively. In order to study the role of AtsB on expression of active DVS, a *dvs-atsB* operon was constructed. The plasmids were transformed to *E. coli* DH5α, and the proteins were expressed in the presence of isopropyl thiogalactopyranoside. The expression of DVS and AtsB was detected by western blotting using antibodies that had been raised against purified DVS or AtsB (Fig. 4.8B). In the presence and absence of AtsB, DVS was expressed at the same level (Fig. 4.8B). However, no arylsulfatase activity was detected when using *p*-nitrocatechol sulfate as a substrate. Since this substrate is only turned over by a subset of arylsulfatases, the observed catalytic inactivity of DVS, expressed in *E. coli* in the presence and absence of AtsB does not necessarily mean that the threonine was not modified.

In order to directly test for threonine modification in DVS, the soluble sulfatase protein was separated by SDS-PAGE and stained with Coomassie Blue. The stained bands containing DVS were excised and subjected to in-gel digestion with trypsin. MALDI-MS was used to identify the critical tryptic peptides. Threonine56 is part of the tryptic peptide 2 comprising DVS residues 39-60 (F-A-Q-N-G-F-L-F-E-N-A-Y-S-E-G-L-P-T/FGly/FAla-I-P-V-R). The mass of the threonine 56-containing P2 peptide is 2469 Da, that of the FGly56-containing P2* 2453 Da, and that of the FAla P2** 2467 Da. After hydrazone derivatization by DNPH, the mass of P2* could change to 2633 Da and that of P2** to 2647 Da.

Fig. 4.10A and B show that only the unmodified Thr-56 containing peptide P2 was detected (mass 2469 Da) in DVS, independent of its coexpression with AtsB. Therefore, no FGly or FAla was generated in the DVS protein. This result was confirmed when using DNPH as a matrix for MALDI-MS (Fig. 4.12 C and D). In conclusion, neither *E. coli*'s unknown modification system nor AtsB can generate FGly from threonine in DVS. Thus, DVS may either not be modified in the heterologous expression system, may require threonine-specific factors or may function in *D. vulgaris* without modification despite having an almost perfect modification motif.

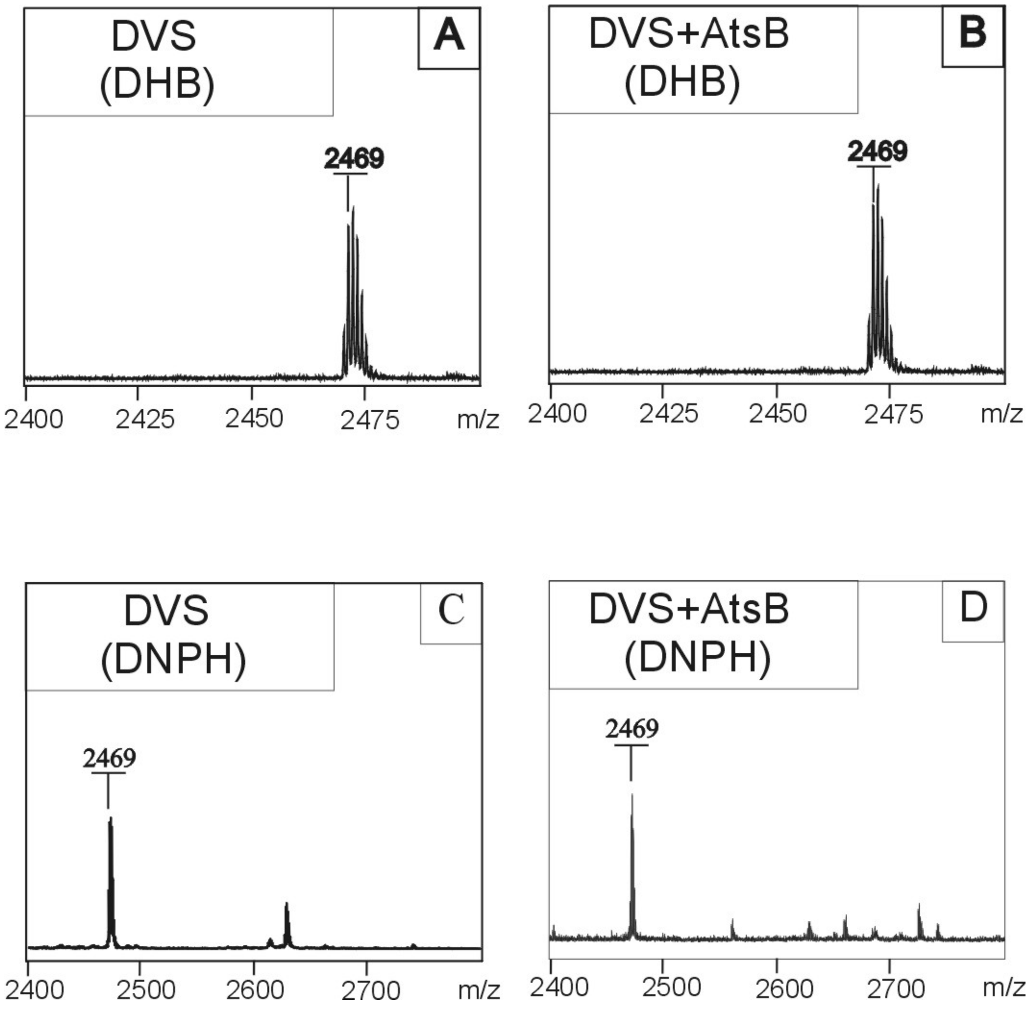


Figure 4.10 The putative sulfatase DVS of *Desulfavibrio vulgaris* remains unmodified after expression in *E. coli* with or without AtsB. About 20µl soluble cell lysate containing DVS (see Fig. 4.8B), which had been expressed alone or in the presence of AtsB, were analysed by SDS-PAGE and Coomassie Blue staining. The bands containing DVS were subjected to in-gel digestion. The tryptic peptides recovered from the gel slice were purified on Ziptip-C18 and analysed by MALDI-MS using DHB (A, DVS; B, DVS coexpressed with AtsB) and DNPH (C, DVS; D, DVS coexpressed with AtsB) as matrix. The m/z 2469 peak represents the Thr 56-containg peptide. No formyl or oxo-group containing peptide was observed, as would have been detected by a corresponding mass increment after reaction with DNPH.

4.5 Requirement of AtsA's signal peptide for AtsB-mediated formylglycine modification

4.5.1 AtsB Coexpression does not lead to activation of PAS-C51S

The arylsulfatase PAS of *Pseudomonas aeruginosa* is a cysteine-type sulfatase that is quantitatively modified upon expression of its structural gene in *E. coli* (Dierks *et al.,* 1998). When the critical cysteine residue 51 was substituted by a serine (PAS-C51S), no FGly formation was observed (Dierks *et al.,* 1998). Since the sequence motifs (SXPXR and LTG) determining AtsB-dependent FGly formation in AtsA are fully conserved also in PAS-C51S, the C51S form of PAS was expressed in *E. coli* with or without AtsB. As shown in Fig. 4.11A, PAS-C51S was expressed as an inactive polypeptide both in the absence and presence of AtsB. Control experiments demonstrated that expression of AtsB does not affect the expression of active wild type PAS (Fig. 4.11A). Therefore, we conclude that PAS, when converted to a serine-type sulfatase, is not a substrate for the AtsB-dependent FGly-generating machinery. Apart from the FGly-modification motif additional parts of the substrate protein must be recognized by this machinery.

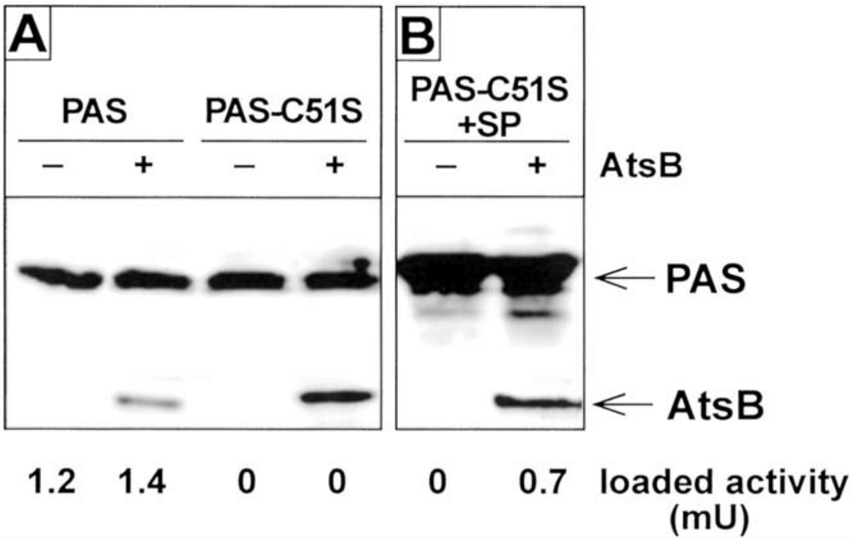


Figure 4.11 Expression of catalytically active PAS-C51S requires AtsB and the presence of a signal peptide. The wild-type and C51S forms of *Pseudomonas* sulfatase PAS were expressed in *E. coli* in the absence or presence of AtsB, as indicated. The western blot shows the PAS and AtsB polypeptides, as recovered from the soluble fraction of total cell lysates. The sulfatase activities present in the samples loaded for SDS-PAGE are given below each lane. PAS and PAS-C51S were expressed as cytosolic (A) or as secretory proteins (B) (*i.e.* without or with the signal peptide of AtsA, respectively, engineered at the N-terminus of PAS (PAS-C51S+SP)). PAS-C51S+SP was catalytically active when coexpressed with AtsB. Its specific activity, as calculated after densitometric

quantification of PAS protein on the western blot, corresponded to about 10% of wild type PAS activity.

4.5.2 AtsB-dependent formylglycine formation in AtsA is strongly reduced after removal of AtsA's signal peptide

One of the major differences between AtsA and PAS-C51S is the presence or absence of a signal peptide, respectively. Therefore, we deleted the signal peptide of AtsA and investigated whether a cytosolic version of AtsA (AtsAΔSP) is synthesized in active form when coexpressed with AtsB. It turned out that AtsAΔSP showed a very low, albeit significant, catalytic activity of 0.84±0,14 units/mg (*i.e.* about 1% of wild type AtsA activity) (Fig. 4.12A). In the absence of AtsB, AtsAΔSP was expressed as a completely inactive protein (data not shown). To examine for the presence of FGly, hexa-Histidine-tagged versions of wild type AtsA and AtsAΔSP were coexpressed with AtsB. The sulfatases were purified on Ni^{2+}-NTA-agarose and analyzed for FGly modification. For this purpose, tryptic peptides were generated and subjected to HPLC on a reversed-phase column, which allowed us to separate unmodified and modified peptide 2 (P2 and P2*) comprising serine or FGly at position 72, respectively. P2 and P2* eluted in adjacent fractions. By amino acid sequencing of these fractions (Fig. 4.12, B and C), we found that 56% of the wild-type AtsA polypeptides carried the FGly.

In the cytosolic AtsAΔSP, traces of FGly could be detected. By mass spectrometry of the HPLC fraction that should contain P2*, we clearly could detect the FGly-containing peptide. As shown in Fig. 4.12E, some P2* (1587 Da) was detected that specifically reacted with DNPH, used as a matrix for MALDI-MS, to the corresponding hydrazone (1767 Da; Fig. 4.12F). This hydrazone formation clearly indicates the presence of a formyl group. The majority of the peptide in the analyzed HPLC fraction, however, corresponded to the Ser72-containing P2 (1589 Da, Fig. 4.12E), indicating incomplete separation of P2* and P2 by HPLC. This was confirmed by amino acid sequencing. A serine residue was identified at the position of the expected FGly (Fig. 4.12D). Moreover, no clear decrease of sequencing efficiency at the position of the FGly and consecutive residues, an obligatory effect of FGly-containing peptides (Fig. 4.12, compare B and D), was observed. As estimated from the yield of amino acids in the sequencing cycles before and after the position of the FGly in the P2* fraction (Fig. 4.12D) and, in comparison, in the P2-fraction (not shown), P2* contributes less than 10% to the P2/P2* mixture in the P2*-fraction and less than 2% to total P2/P2* of AtsAΔSP. In conclusion, expression of a signal peptide-deleted AtsA results in a sulfatase with a 100-fold lower specific enzymatic activity. This agrees with a similarly reduced FGly content.

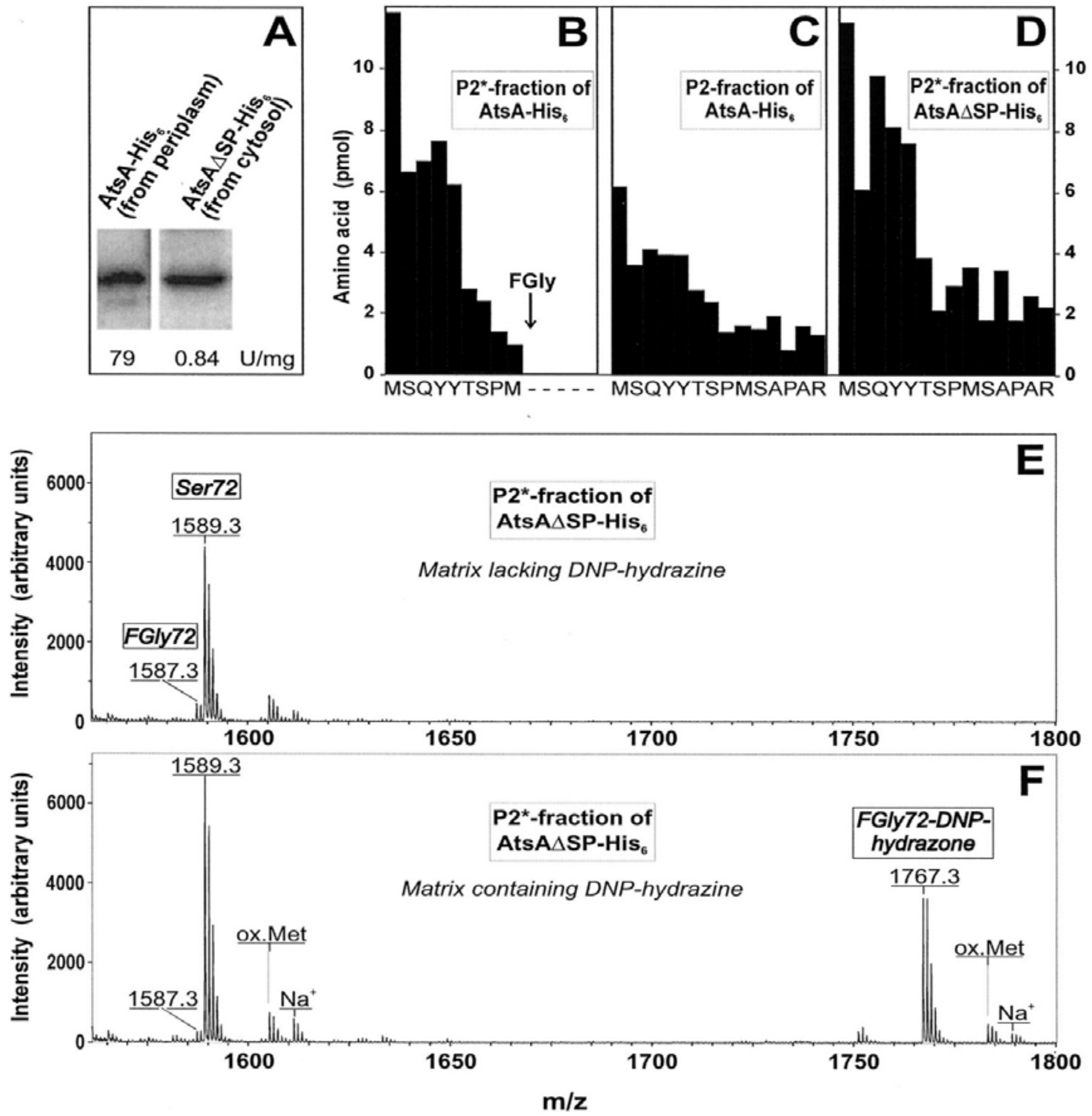


Figure 4.12 The signal peptide is required for efficient FGly modification of AtsA. AtsB and AtsA-His$_6$ were coexpressed in *E. coli*, the latter as periplasmic wild type, or as cytosolic (*i.e.* signal peptide-deleted (AtsAΔSP-His$_6$)) proteins. The AtsA proteins, recovered from the periplasm (AtsA-His$_6$) or from spheroplasts (AtsAΔSP-His$_6$), were purified on Ni-NTA-agarose, and their

specific catalytic activities (units/mg of AtsA protein) were determined. A, the western blot shows 0.1 and 0.2 µg of periplasmic and cytosolic AtsA-His6, respectively, having a specific activity of 79 and 0.84±0.14 units/mg (n=5), respectively. B-F, the purified proteins were subjected to tryptic digestion, and their tryptic peptides were separated by reversed-phase HPLC. The modified and the unmodified forms of peptide 2 (P2* and P2, respectively) eluted in adjacent fractions, as detected by mass spectrometry. B-C, sequencing of the P2* and P2 fraction demonstrated that 56.2±7.5% of the periplasmic wild type AtsA containing FGly. For P2*, Edman degradation is blocked at the position of the FGly (B, cycle 10) and reduced in the preceding cycle. For FGly quantitation, only the amino acid yields in cycles 3-8 (QYYTSP) were considered, since the first two cycles showed some background. D, for the cytosolic AtsAΔSP-His$_6$, the HPLC fraction corresponding to P2* mainly contained unmodified P2, as evidenced by sequencing of the entire peptide showing a serine in position 72 and signals in the following cycles. E, mass spectrometry detected clearly low signals for P2*(1587 Da) and high signals for P2 (1589 Da). F, using DNPH as a matrix for MALDI-MS, P2*, but not P2, was converted into the corresponding hydrazone (1767 Da). Parts of P2, P2*, and the P2*-hydrazone derivative contained an oxidized methionine or were desorbed as Na^+ adducts, as indicated.

4.5.3 AtsB-dependent expression of active signal peptide-containing PAS-C51S

To find out whether the presence of a signal peptide would allow for AtsB-dependent FGly formation in PAS-C51S, PAS-C51S+SP was constructed by fusing PAS-C51S to the signal peptide of AtsA. When expressed in the absence of AtsB, PAS-C51S+SP was catalytically inactive. However, when coexpressed with AtsB a significant sulfatase activity was measured (Fig. 4.11B), which corresponded to up to 10% of that of wild type PAS (Fig. 4.11A). It should be noted that the signal peptide of most of the expressed PAS-C51S+SP remained unprocessed, as evidenced by a slightly lower electrophoretic mobility (Fig. 4.11B). In fact, less than 5% of PAS-C51S protein and less than 10% of arylsulfatase activity was recovered in the periplasm of cells coexpressing PAS-C51S+SP and AtsB (data not shown). Independent of this obvious translocation deficiency, the catalytic activity of PAS-C51S+SP in contrast to the inactivity of PAS-C51S supports the observation made for AtsAΔSP, namely that AtsB-dependent FGly formation in serine-type sulfatases requires the presence of a signal peptide. Furthermore, it shows that AtsB can act on a non-*Klebsiella* sulfatase that carries a serine-type FGly modification motif and a signal peptide (see also Chapter 4.4).

4.6 Requirement of S-adenosylmethionine for formation of a functional AtsB/AtsA complex

AtsB is predicted to be an FeS protein containing 3 FeS centers that are coordinated by the following cysteine clusters: C35-x_3-C39-x-Y-C42-Y, C270-x_5-C276-x_{14}-C291, and C331-x_2-C334-x_5-C340-x_3-C344. There are two additional conserved but more isolated cysteines in positions 127 and 357 that may contribute to FeS center formation. The first cysteine cluster, Cx_3CxYCY, with two aromatic residues flanking the third cysteine, is characteristic for a protein family including biotin synthase, lipoate synthase, and oxygen-independent coproporphyrinogen-III oxidase HemN. Moreover, all these proteins have in common that they use S-adenosylmethionine (SAM) as a cofactor to form a deoxyadenosyl radical involved in catalysis. Even more, AtsB shares variants of a sequence motif (^{83}GGEPLL^{88}) with these proteins that may be part of an SAM-binding site (see below). Therefore AtsB is predicted to be a radical SAM protein with SAM being the essential factor for its activity. Here, we investigated this hypothesis by testing the importance of SAM for AtsB-AtsA interaction and AtsB's activity.

4.6.1 SAM improves binding of AtsB to the formylglycine-modification motif of AtsA

To test for a possible involvement of SAM in AtsB-mediated FGly-modification of AtsA we performed *in vitro* AtsB-AtsA interaction experiments in the absence or presence of SAM and/or its demethylated analog S-adenosylhomocysteine (SAHC). GST-AtsA(21-112) and an AtsB-containing *E. coli* cell lysate were mixed and incubated for 15 min before being loaded to GSH-agarose. We observed an association of AtsB with GST-AtsA that both bound to and co-eluted from GSH agarose (Fig. 4.13A). It turned out that addition of 500µM SAM significantly improved AtsB binding, whereas 500 µM SAHC completely inhibited AtsB binding (Fig. 4.13A). When both SAHC and SAM were added to the interaction mixture, the AtsA/AtsB interaction was improved showing a dominant effect of SAM. The association of GST-AtsA and AtsB, and its stimulation by SAM required incubation of all interaction partners at 25-37^{o}C for at least 5 min; at 0°C interaction was rather inefficient (Fig. 4.13B). This suggests that formation of a stable complex, which can be pulled down by GSH-agarose, involves a rate-limiting step that is strictly temperature-dependent. In the presence of 500µM SAM, formation of this complex increased with time over a 60-minute period (Fig. 4.13C). Without addition of SAM, the complex formed initially (up to 30 min) was lost upon longer incubation. The latter is attributed to the presence of some endogenous SAM in the *E. coli* lysate which initially promotes complex formation and decomposes during the 37°C incubation, thus destabilizing the complex. Control experiments showed that GST-AtsA/AtsB complex formation in the presence of only 50 µM SAM reached a maximum after ≈15 min at 37°C (data not shown). At ≥300 µM SAM the complex was stable for more than 60 min (Fig. 4.13C and data not shown).

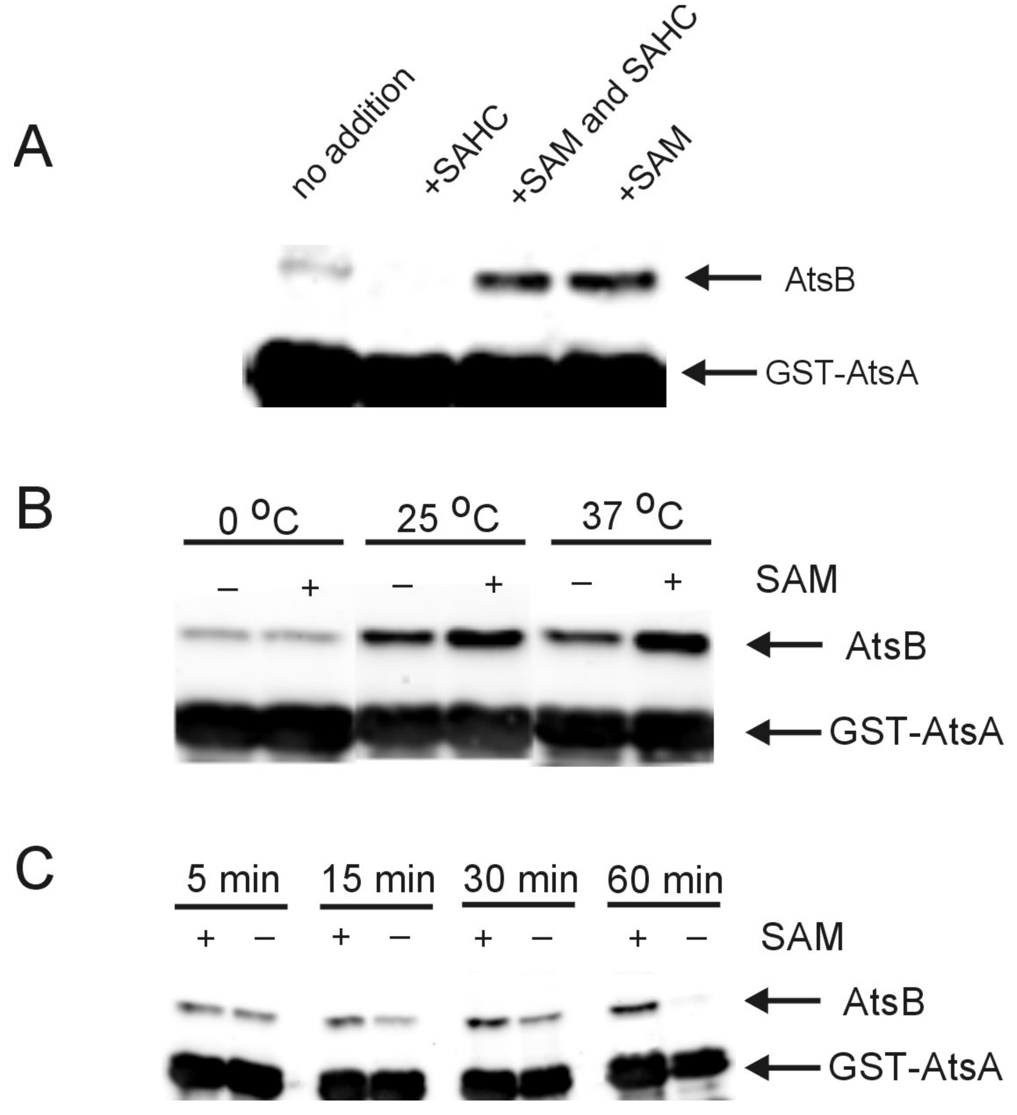

A
no addition
+SAHC
+SAM and SAHC
+SAM
AtsB
GST-AtsA
B
0 °C
25 °C
37 °C
– + – + – +
SAM
AtsB
GST-AtsA
C
5 min
15 min
30 min
60 min
+ – + – + – + –
SAM
AtsB
GST-AtsA

Fig. 4.13 SAM-dependent AtsB/AtsA association. (**A**) *In vitro* interaction reactions of GST-AtsA fusion protein and AtsB in the presence of *E. coli* protein were set up with or without supplementation of 500 µM SAM and/or SAHC, as described in Methods. After incubation for 15 min at 37°C the interaction mixtures were loaded on GSH-agarose columns. The GSH-eluates of these columns, obtained after washing thoroughly (see Methods), were analyzed by western blotting using anti-GST and anti-AtsB antibodies. The positions of the detected proteins are indicated. (**B**) western analysis of *in vitro* interaction reactions (see A) that had been incubated in the absence or presence of 500 µM SAM for 15 min at 0°C, 25°C or 37°C, as indicated. (**C**) western analysis of *in vitro* interaction reactions (see A) that had been incubated in the absence or presence of 500 µM SAM for various times at 37°C, as indicated.

4.6.2 Formation of a ternary AtsB/SAM/AtsA-complex

To find out whether SAM was present in the interaction complex of AtsB/AtsA, we added ^{3}H-labeled SAM as a tracer to the incubation mixture containing 50 µM SAM. After loading this mixture to GSH-agarose and extensive washing of the column, the GSH-eluate was collected and analyzed for AtsB and GST-AtsA by western blotting and for SAM by liquid scintillation counting. It turned out that significant amounts of radioactivity coeluted with AtsB, whereas no radioactivity was recovered in the eluate of GSH-agarose loaded with an incubation mixture lacking AtsB (Fig. 4.14A compare lanes 1 and 4). Two further controls included GST-AtsA-S72C or GST-AtsA-S72A/P74A/R76A (GST-AtsA-S72A) as baits that showed clearly reduced (40 % of wild-type control) or no interaction with AtsB, respectively. In the corresponding eluates also the SAM–associated radioactivity was reduced to around 50 % or was completely absent, respectively (Fig. 4.14A, lanes 2 and 3). Hence, it maybe concluded that the amount of bound [^{3}H]SAM parallels the amount of AtsB associated with GST-AtsA. It should be noted that the required preincubation for 15 min at 37^{o}C led to clearly enhanced binding of AtsB to the GST-AtsA-S72C mutant, which was <10% of wild-type GST-AtsA when omitting the incubation (see Chapter 4.3 and Fig. 4.6).

SAM stimulated GST-AtsA/AtsB interaction in a concentration-dependent manner. Under standard conditions (incubation for 15 min at 37^{o}C), half-maximum stimulation was observed at about 20-40 µM SAM and saturation at about 150-300 µM SAM (Fig. 4.14B). Also the amount of [^{3}H]SAM coeluting from GSH-agarose together with GST-AtsA and AtsB showed a clear dependence on the applied SAM concentration. At ≥300 µM SAM nearly equal amounts (pmoles) of AtsB and SAM were recovered in the eluate (Fig. 4.14B); from 3 experiments a SAM:AtsB ratio of 0.69±0.06 was calculated. At non-saturating SAM concentrations the amount of bound SAM was clearly lower than the amount of bound AtsB (Fig. 4.14B). This difference may reflect a cooperative stimulation of AtsB binding by SAM in order to avoid deoxyadenosyl radical formation in the absence of substrate (see Discussion). However, we can not exclude that some SAM is lost from the AtsB/GST-AtsA complex due to catalytic turnover (see below) and due to the non-steady-state conditions on the affinity column during washing and elution. Taken together the data shown here, in combination with previous results indicating an AtsB/GST-AtsA ratio of 0.54 (see Chapter 4.3), clearly suggest that SAM-binding to AtsB efficiently promotes formation of a ternary AtsA/AtsB/SAM complex. In this complex the three interaction partners may be present in a 1:1:1 ratio. This conclusion is also supported by control experiments testing the time-course of ternary complex formation under various conditions, which showed that at saturating SAM concentrations the amount of bound [^{3}H]SAM always paralleled the amount of AtsB associated with GST-AtsA (see also Fig. 4.14A).

4.7 SAM-dependent *in vitro* formylglycine-formation by AtsB

The data shown above suggest that FGly-modification of serine depends on SAM. To directly measure the effect of SAM on AtsB's FGly-generating activity, an *in vitro* system with limiting SAM-levels had to be established. So far we failed to observe FGly-formation *in vitro*, when GST-AtsA was incubated with a soluble extract of *E. coli* overexpressing AtsB (see Chapter 4.3). Here we used essentially the same components, this time supplementing the modification reaction with 500 µM SAM. After incubation at 37°C, GST-AtsA was isolated from the mixture by GSH-agarose, subjected to digestion with trypsin and MALDI-MS analysis using DNP-hydrazine as a matrix. This method efficiently converts the FGly-containing tryptic peptide of AtsA (1587 Da) into the corresponding hydrazone (1767 Da) upon laser-assisted desorption from a dinitrophenyl hydrazine (DNP-hydrazine) matrix, as described above (Figs. 4.2 and 4.9). Formation of the FGly-modified peptide of AtsA, detected as the DNP-hydrazone (1767 Da), was indeed observed when the incubation was performed in the presence of SAM (Fig. 4.15A), but not in its absence (data not shown). The molecular identity of the DNP-hydrazone peptide was confirmed by the indicative 2-oxodihydroimidazole-derivatized fragment ion (523 Da, Fig. 4.15C) desorbing with high efficiency upon MALDI-PSD analysis of the 1767 Da parent ion (for generation of this fragment from FGly-containing peptides see Peng *et al.*, 2003). The formation of FGly was dependent on AtsB (Fig. 4.15D) and on the FGly-modification motif of AtsA (Fig. 4.15E). Substitution of Serine 72 by cysteine in AtsA prevented FGly-formation by AtsB (Fig. 4.15E).

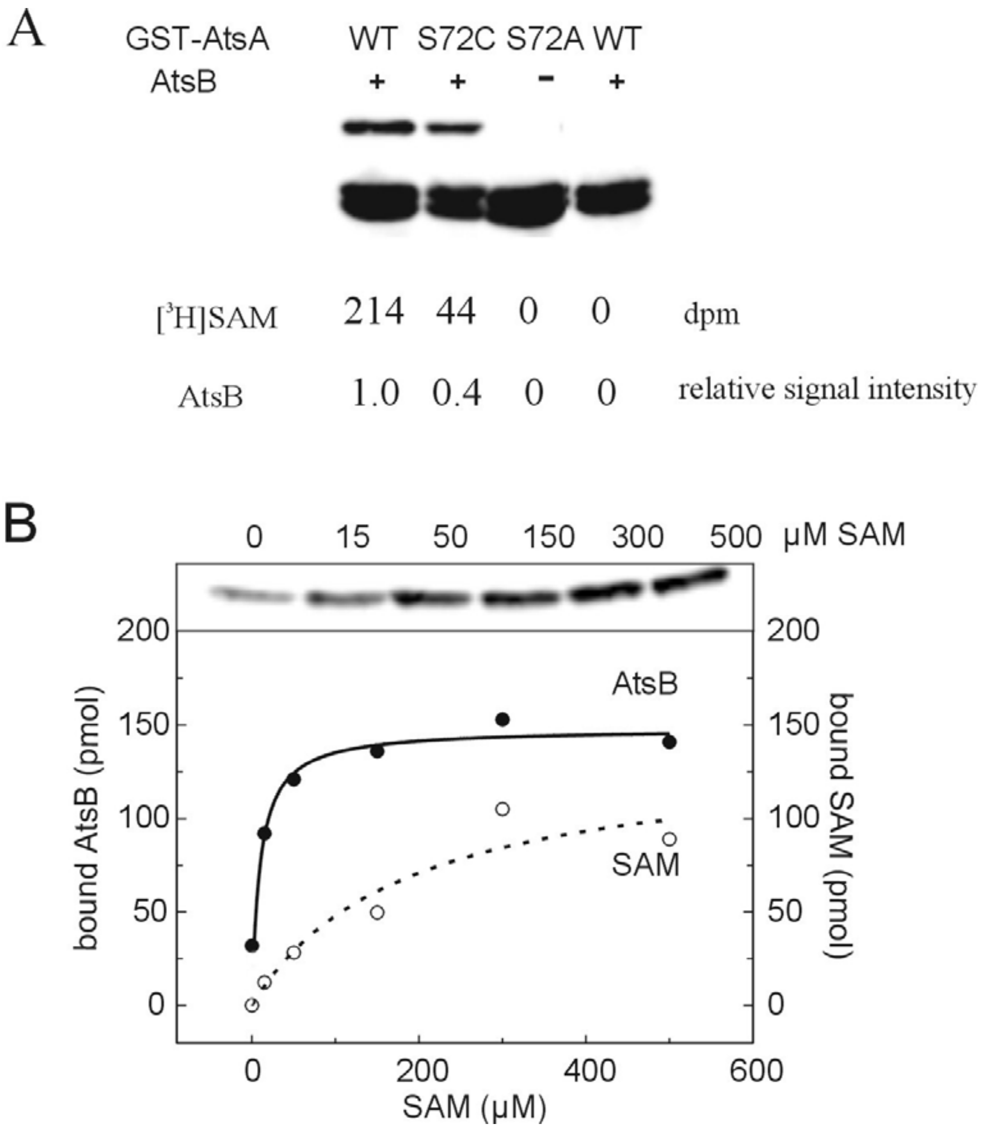


Fig. 4.14 Presence of SAM in a ternary AtsB/SAM/AtsA complex. (**A**) *In vitro* interaction reactions of AtsB and GST-AtsA, expressed in wild-type, S72C or S72A/P74A/R76A (S72A) mutant form, were incubated for 15 min at 37°C in the presence of 50 µM [^{3}H]SAM, as described in Methods. The components of these reactions binding to and eluting from GSH-agarose were analyzed by western blotting and liquid scintillation counting. The positions of the detected proteins and the amount of SAM-associated radioactivity recovered in a 33% aliquot of the GSH-eluate are given. (**B**) *In vitro* interaction reactions that had been incubated for 15 min at 37°C in the presence of the indicated [^{3}H]SAM concentrations (0 µM corresponding to 1 µCi tracer only) were analyzed by western blotting and liquid scintillation counting (see A). western blot signals for AtsB (upper panel) were quantitated by comparison with defined amounts of purified AtsB protein on the same blot and are given in pmol together with the amount of [^{3}H]SAM co-eluting with AtsB from GSH-agarose (lower panel). The pmoles SAM were calculated without correcting for endogenous SAM present in the *E. coli* lysate (see text). Only when assuming that this endogenous SAM contributed ≤10 µM to the total SAM concentration, such corrections led to meaningful results, *i. e.* increase and no decrease of SAM-binding with increasing SAM concentration (not shown).

The *in vitro* modification reaction, however, was found to be rather slow and inefficient (Fig. 4.15B). It should be noted that the peak heights of the FGly-DNP-hydrazone peptide (1767 Da) and of the unmodified peptide (1589 Da) in Fig. 4.15A can not be used to quantitate FGly-modification, as during MALDI-MS the former peptide desorbs much better from the used DNP-hydrazine matrix than the latter peptide (Peng *et al.*, 2003). The efficiency of FGly-formation could be increased up to 5-fold by repeated addition of SAM during the incubation (Fig. 4.15B). This effect is attributed to the limited thermostability of SAM. Nevertheless, from a comparison with AtsA of known FGly-content (Marquordt *et al.*, 2003) we estimate that at best about 10% of the GST-AtsA substrate was modified under conditions where GST-AtsA was in a roughly 10-fold excess over AtsB. Apart from the poor efficiency of the *in vitro* system, the data obtained show unequivocally that AtsB-mediated FGly-formation relies on SAM as an essential cofactor.

It should furthermore be noted that *in vitro* FGly-modification of GST-AtsA could not achieved using purified Strep-AtsB protein. Strep-AtsB was functional *in vivo* (see Chapter 4.2.2) and, in unpurified state, also *in vitro* (data not shown). Upon purification under aerobic conditions the modification activity was lost even after reconstituting the *in vitro* reaction with an *E. coli* lysate. In view of the oxygen sensitivity of AtsB's FeS center(s) it may be necessary to purify AtsB under anaerobic conditions. This is presently being done in cooperation with Dr. Selmer (University of Marburg).

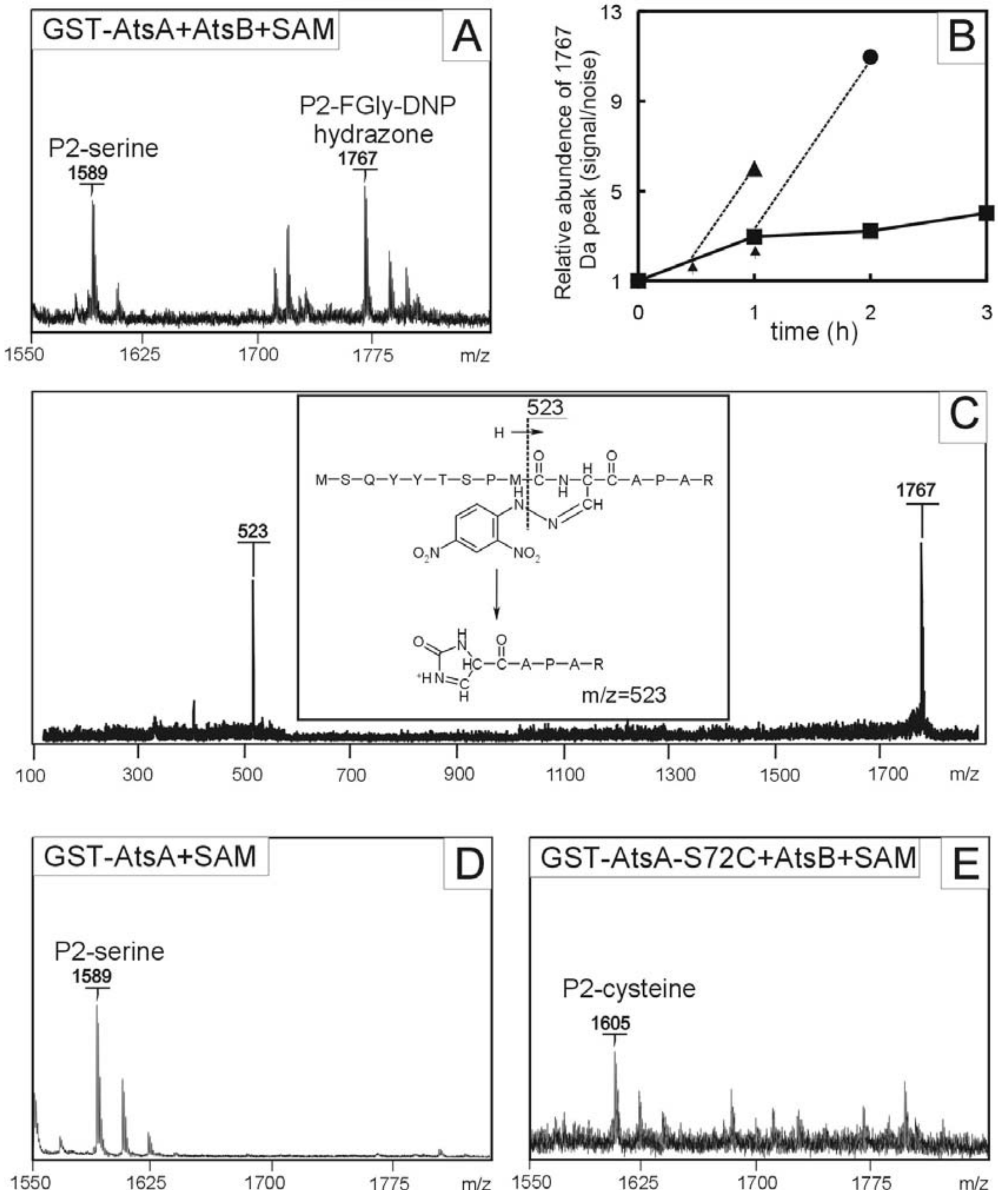


Fig. 4.15 SAM-dependent *in vitro* FGly-formation by AtsB. (**A, B**) *In vitro* reactions were set up, as described in Methods, incubating GST-AtsA and AtsB in the presence of *E. coli* protein and 500 µM SAM for up to 3 h at 37°C, as indicated in B (squares). Two samples (B, triangle and filled circle) were supplemented a second time with SAM, corresponding to 500 µM, after 30 or 60 min of incubation, as indicated by the arrows. GST-AtsA was purified on GSH-

agarose and subjected to SDS-PAGE, Coomassie staining and in-gel digestion with trypsin. The extracted tryptic peptides were analyzed by MALDI-MS using DNP-hydrazine as a matrix. (A) shows the result of the sample obtained after 2 h at 37°C with re-addition of SAM (B, circle). The FGly-containing form of tryptic peptide P2 is detected as the corresponding DNP-hydrazone (1767 Da), while the serine-containing P2 gives a signal at 1589 Da. The relative abundance of the m/z 1767 Da signal is given in B for all *in vitro* reactions as a function of their incubation time. (**C**) MALDI-PSD analysis of the 1767 Da ion led to formation of the predicted 2-oxodihydroimidazole-derivatized fragment ion (523 Da, see Peng *et al.*, 2003), thereby verifying the molecular identity of the DNP-hydrazone. The fragmentation giving rise to the 523 Da product is indicated. (**D**, **E**) Control reactions in which AtsB was omitted (D) or in which a GST-AtsA-S72C mutant protein was used (E) did not lead to FGly-modification, as only the serine- or cysteine-containing forms of P2 were detected, as indicated (D and E, respectively).

4.8 ^{83}GGE^{85} of AtsB is required for SAM-binding, AtsA-interaction and formylglycine-formation

AtsB comprises the amino acid sequence GGEPLL at positions 83-88 that distantly resembles the glycine-rich SAM binding site in the majority of non-DNA methyltransferases (Niewmierzycka and Clarke, 1999; Shields *et al.*, 2003). The sequence perfectly agrees with the semi-consensus sequence found in an alignment block of the radical SAM superfamily (Sofia *et al.*, 2001). Yet it should be noted that the individual sequence of each family member may vary considerably from the semi-consensus sequence (Sofia *et al.*, 2001).

In order to study the importance of the putative SAM binding site for AtsB-mediated FGly-formation *in vivo*, we performed an alanine scanning mutagenesis for residues ^{83}GGEPLL^{88} (Pro-86 additionally was substituted by glycine to allow for enhanced structural flexibility at this position). After co-expression of these mutants with AtsA in *E. coli*, the specific sulfatase activity of AtsA was determined as an indirect indicator of the FGly-modification activity of AtsB. It turned out that the specific sulfatase activity of AtsA, *i.e.* the ratio of activity over western blot signal, was drastically reduced when Gly-83, Gly-84 or Glu-85 of AtsB was replaced by alanine (313-, 31- and 50-fold lower specific activities, respectively, as compared to AtsA co-expressed with wild-type AtsB, Fig. 4.16A). For the other mutants, at positions 86-88, wild-type sulfatase activities were observed. The expression level of AtsB and its solubility were not affected by the mutations (Fig. 4.16A).

To directly determine the AtsB-mediated FGly-formation, the co-expressed AtsA protein was subjected to FGly analysis using the same semi-quantitative MALDI-MS method described above, which detects the FGly-modified peptide as a DNP-hydrazone derivative of 1767 Da. The intensity of this 1767 Da-signal was strongly reduced in case of AtsA co-expressed with AtsB-mutants G84A and E85A, and was not at all detectable in case of co-expression with AtsB-G83A (Fig. 4.16A). Thus, the drastically reduced AtsA activities, caused by these three AtsB-mutants, fully agreed with the impaired FGly-modification of AtsA.

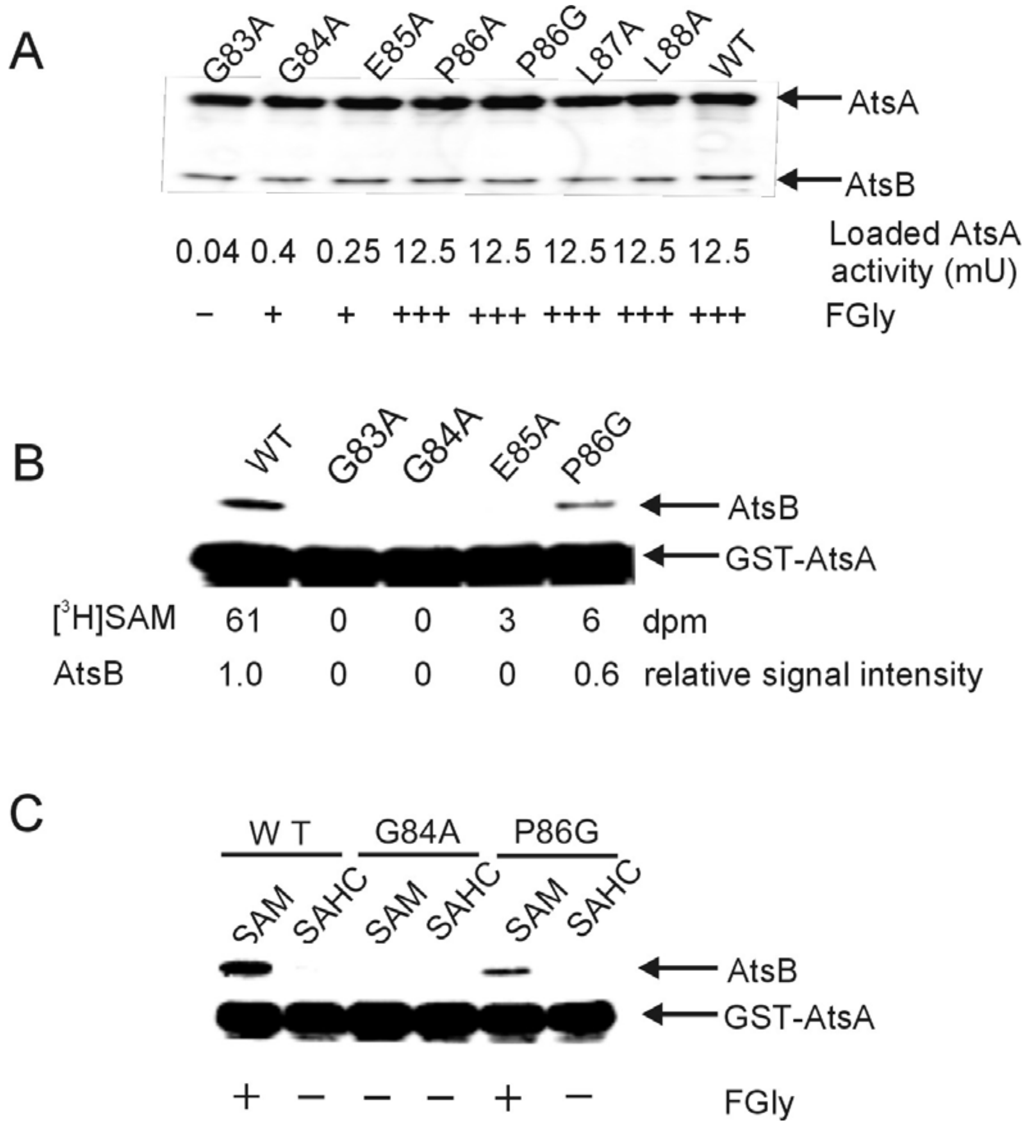

A
G83A
G84A
E85A
P86A
P86G
L87A
L88A
WT
AtsA
AtsB
0.04 0.4 0.25 12.5 12.5 12.5 12.5 12.5
Loaded AtsA activity (mU)
– + + +++ +++ +++ +++ +++
FGly
B
WT
G83A
G84A
E85A
P86G
AtsB
GST-AtsA
[³H]SAM 61 0 0 3 6 dpm
AtsB 1.0 0 0 0 0.6 relative signal intensity
C
W T
G84A
P86G
SAM
SAHC
SAM
SAHC
SAM
SAHC
AtsB
GST-AtsA
+ – – – + –
FGly

Fig. 4.16 Alanine scanning mutagenesis of the putative SAM binding site of AtsB. (**A**) AtsA was co-expressed in *E. coli* together with wild-type AtsB or with a G83A, G84A, E85A, P86A, P86G, L87A or L88A mutant form of AtsB. The western blot shows the AtsB and AtsA polypeptides, as recovered from the soluble fraction of total cell lysates. The sulfatase activities present in the samples loaded for SDS-PAGE are given below each lane. The AtsA protein, after SDS-PAGE, Coomassie staining and in-gel digestion with trypsin was analyzed for the presence of FGly by MALDI-MS using DNP-hydrazine as matrix. The relative abundance of the detected P2-FGly-DNP-hydrazone signal (cf. Fig. 4.15A and B) is given as follows: +++, high abundance (signal/noise ratio >20); +, low abundance (signal/noise ratio 2-5); -, no signal. (**B**) *In vitro* interaction reactions of GST-AtsA and AtsB, expressed in wild-type, G83A, G84A, E85A, or P86G mutant form, were incubated for 15 min at 37°C in the presence of 50 µM [^{3}H]SAM (cf. Fig. 4.14A). The components of these reactions binding to and eluting from GSH-agarose were analyzed by western blotting and liquid scintillation counting. The positions of the detected proteins, the relative amounts of AtsB, detected by quantitation of the western blot signals, and the amount of SAM-associated radioactivity recovered in a 33% aliquot of the GSH-eluate are given. (**C**) For *in vitro* FGly-modification, GST-AtsA and AtsB, expressed in wild-type, G84A or P86G mutant form, were incubated in the presence of *E. coli* protein and 500 µM SAM or SAHC, as indicated, for 2 h at 37°C, with readdition of 500 µM SAM or SAHC after 1h (cf. Fig.4.15B). The presence (+) or absence (-) of FGly in GST-AtsA is given, which was detected as DNP-hydrazone derivative by MALDI-MS analysis of its tryptic peptides.

In vitro AtsA-AtsB interaction experiments in the presence of SAM demonstrated that binding of AtsB mutants G83A, G84A and E85A to GST-AtsA was totally abolished and that binding of the AtsB-P86G mutant was reduced to 60% (Fig. 4.16B). In the corresponding eluates from the GSH-agarose columns, measurable amounts of [^{3}H]SAM only were detected for the wild-type AtsB control indicating that none of the AtsB mutants tested was able to bind SAM (Fig. 4.16B).

The AtsB-G84A and P86G mutants, being either fully (G84A) or only partially (P86G) impaired in binding of AtsA and SAM, were selected to directly test their FGly-generating activity using the *in vitro* assay described above. It turned out that the latter, but not the former, mutant was capable of FGly-generation in an SAM-dependent and SAHC-sensitive manner (Fig. 4.16C). Thus, although apparently having a clearly reduced affinity for SAM, the AtsB-P86G mutant obviously is catalytically active.

Taken together these findings strongly suggest that ^{83}GGEP^{86} indeed constitutes the SAM binding site of AtsB, or a part thereof, and that its occupation by SAM is required for the formation of a functional ternary AtsB/SAM/AtsA complex that catalyzes the FGly-modification of AtsA. Out of this tetrapeptide sequence, ^{83}GGE^{85} is essential for AtsB function. At position 86, substitution of proline by glycine was tolerated *in vivo*, while it partially impaired *in vitro* formation of the ternary complex.

4.9 Requirement of conserved cysteine residues in AtsB for formylglycine-generation *in vivo* and *in vitro*

Radical SAM proteins generate deoxyadenosyl radicals through reductive cleavage of SAM (Frey, 2001). Reduction requires an FeS center that is coordinated by a specific cluster of three conserved cysteines (C-X_3-C-X_2-C), with two aromatic residues often flanking the last cysteine. This cysteine cluster usually is located near the N-terminus and some 30-50 residues upstream of the putative SAM binding site (Sofia *et al.*, 2001). In the corresponding cluster of AtsB we mutated two cysteines, Cys-39 or Cys-42, to alanines. We also mutated three further conserved cysteines that have been predicted to coordinate two additional FeS centers, Cys-270 on the one hand, and Cys-331 and Cys-334 on the other (Schirmer and Kolter, 1998). All five mutants could be expressed in *E. coli* at levels comparable to wild-type AtsB, and the proteins were largely soluble (Fig. 4.17A). When measuring the activity of co-expressed AtsA, it was found that all five mutants were fully impaired in activating AtsA. After MALDI-MS analysis of AtsA for FGly-modification it turned out that in all five cases not even trace amounts of FGly were detectable (Fig. 4.17A).

In vitro interaction experiments showed that only AtsB mutants C39A and C42A were able to associate with GST-AtsA, albeit with reduced efficiency (55% and 16% of wild-type AtsB control, respectively, Fig. 4.17B). The other mutants showed no interaction. Binding of [^{3}H]SAM also was markedly reduced. Only AtsB-C39A still bound residual amounts of [^{3}H]SAM (about 10% of control, Fig. 5B). For all other mutants radioactivity in the GSH eluate was in the background range. Three of the five mutants also were analyzed for their capability of *in vitro*

FGly-formation. In line with the *in vivo* data, none of the cysteine mutants, including AtsB-C39A and AtsB-C42A, was able to generate FGly in the AtsA substrate (Fig. 4.17C). In conclusion, all three predicted FeS centers seem to be required for FGly-modification. Mutation of the second and third cysteine cluster abolishes *in vitro* binding of SAM and the AtsA substrate, while mutation of the first cluster allows for some residual binding (Fig. 4.17B). This residual binding is blocked in the presence of SAHC (Fig. 4.17C), suggesting that binding to GST-AtsA is SAM-dependent also for AtsB-C39A. The ternary AtsB/SAM/AtsA complex of this mutant, however, shows no turnover, indicating that AtsB-C39A is catalytically inactive.

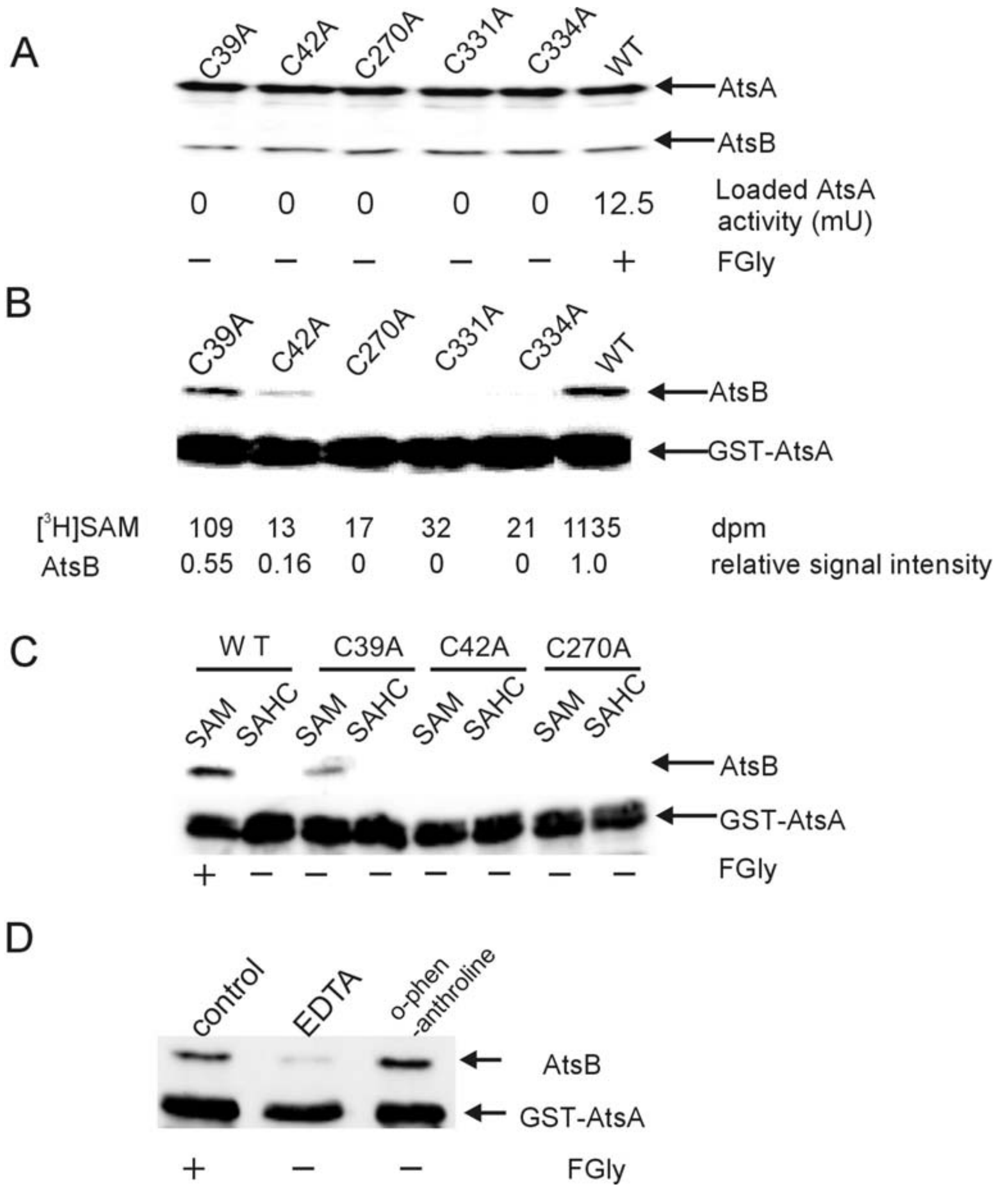

A
C39A C42A C270A C331A C334A WT
AtsA
AtsB
0 0 0 0 0 12.5 Loaded AtsA activity (mU)
− − − − − + FGly
B
C39A C42A C270A C331A C334A WT
AtsB
GST-AtsA
[3H]SAM 109 13 17 32 21 1135 dpm
AtsB 0.55 0.16 0 0 0 1.0 relative signal intensity
C
W T C39A C42A C270A
SAM SAHC SAM SAHC SAM SAHC SAM SAHC
AtsB
GST-AtsA
+ − − − − − − − FGly
D
control EDTA o-phen -anthroline
AtsB
GST-AtsA
+ − − FGly

Fig. 4.17 Mutagenesis of conserved cysteines coordinating putative FeS centers in AtsB. (**A**) AtsA was co-expressed in *E. coli* together with wild-type AtsB or with a C39A, C42A, C270A, C331A or C334A mutant form of AtsB. The western blot shows the AtsB and AtsA polypeptides, as recovered from the soluble fraction of total cell lysates. The sulfatase activities present in the samples loaded for SDS-PAGE are given below each lane. The presence (+) or absence (-) of FGly in AtsA is given, which was detected as DNP-hydrazone derivative by MALDI-MS analysis of its tryptic peptides. (**B**) *In vitro* interaction reactions of GST-AtsA and AtsB, expressed in wild-type, C39A, C42A, C270A, C331A or C334A mutant form, were incubated and analyzed as described for Fig. 4.16B. The positions of the detected proteins, the relative amounts of AtsB and the amount of SAM-associated radioactivity recovered in a 33% aliquot of the GSH-eluate are given. (**C, D**) For *in vitro* FGly-modification, GST-AtsA and AtsB, expressed in wild-type, C39A, C42A or C270A mutant form, were incubated and analyzed as described for Fig. 4.16C. The presence (+) or absence (-) of FGly in GST-AtsA is given, which was detected as DNP-hydrazone derivative by MALDI-MS analysis of its tryptic peptides. In (D) 500µM SAM and, where indicated 2mM EDTA or 2mM *o*-phenanthroline were present during the *in vitro* reaction.

Differential effects on AtsA/AtsB interaction and FGly-generating activity were also observed when adding iron chelators such as EDTA or *o*-phenonthrohine to the *in vitro* reaction. Whereas only EDTA blocked AtsA/AtsB association, both compounds fully impaired FGly-modification (Fig. 4.17D). These inhibitor experiments corroborate the conclusion drawn from the cysteine mutants of AtsB, namely that AtsB carries FeS centers that are essential for its enzymatic function as a radical SAM protein.

5 Discussion

5.1 AtsB is a prokaryotic formylglycine-generating enzyme

The present study demonstrates that the *atsB* gene product is required for FGly modification in the arylsulfatase of *K. pneumoniae*. This modification is a prerequisite for sulfatase activity, as had been shown previously for other pro- and eukaryotic sulfatases (Dierks, *et al.,* 1998; Selmer *et al.,* 1996; Schmidt *et al.,* 1995). In the absence of a functional *atsB* gene inactive sulfatase polypeptides were synthesized lacking the FGly. In the presence of *atsB*, around 50% of the recombinant sulfatase molecules, recovered from the periplasm of *E. coli*, carried the FGly leading to an overall specific activity of 73 units/mg of purified protein (Fig. 4.12A). This approximately agrees with the modification efficiency of 60% and a specific activity of 123 units/mg determined for the wild-type protein purified from *K. pneumoniae* (Miech *et al.,* 1998). These results indicate that AtsB plays an essential role in the posttranslational oxidation of the conserved serine to FGly.

Most importantly, AtsB was shown to interact with and to modify an N- and C-terminally truncated AtsA polypeptide (AtsA-(21-112)) encompassing the serine-type FGly modification motif (Figs. 4.6 and 4.15). This interaction and modification reaction depended on the presence of the serine to be modified, on the side of the substrate and on the critical cofactor SAM which is bound by the AtsB enzyme. This was shown by GST pull-down experiments and by a newly developed *in vitro* FGly generation assay. Hence, we conclude (i) that AtsB directly recognizes the critical serine 72, and (ii) that AtsA-(21-112), expressed as a C-terminal appendix of a stably folded GST domain, basically fulfils all structural requirements for proper association of its linear FGly modification motif with and modification by AtsB. This strongly indicates that AtsB itself is the oxidizing enzyme converting serine to FGly. The role of SAM as a cofactor is discussed below (see Chapter 5.4).

AslB and MM-AtsB are homologs of AtsB found in *E. coli* and in the archaebacterium *M. mazei*. When coexpressed with AslB or MM-AtsB, the AtsA protein was found to be fully or partially active (Fig. 4.1). FGly generation in AtsA was confirmed by MALDI-MS, when coexpressed with AtsB or AslB (Fig. 4.2). In conclusion, these homologs are true AtsB-orthologs. Though being under control of specific sulfatase operon promoters, these orthologs are not specific for the sulfatase of the respective operon.

In summary, AtsB is a prokaryotic FGly-generating enzyme and this modification system does not show strict substrate specificity.

5.2 AtsB-dependent and AtsB-independent formylglycine-modification

As mentioned above, the FGly modification of *Klebsiella* AtsA depends on coexpression of AtsB or its orthologs. However, FGly-modification of *Pseudomonas* sulfatase does not require coexpression of AtsB. Even after strong overexpression in *E. coli*, this cytosolic sulfatase was converted to the active enzyme, in which the cysteine 51 was quantitatively oxidized to FGly (Dierks *et al.,* 1998). Therefore, *Pseudomonas* sulfatase is modified by an AtsB - independent FGly-modification system. Since *Klebsiella* AtsA is a serine-type sulfatase and *Pseudomonas* sulfatase is a cysteine-type sulfatase, it was obvious to suppose that serine-type sulfatase modification is AtsB-dependent and cysteine-type sulfatase modification is not.

However, modification of *Methanosarcina* sulfatase, a putative sulfatase of cysteine-type, was found to be AtsB-dependent. No or only traces of FGly were generated, when *Methanosarcina* sulfatase was expressed alone in *E. coli*. When coexpressed with AtsB or MM-AtsB, around 50-60% of cysteine 55 in *Methanosarcina* sulfatase was oxidized to FGly (Fig. 4.9). AtsA-S72C, *i.e.* an artificial cysteine-type version of *Klebsiella* AtsA, can also be partially modified by AtsB *in vivo* (data not shown). Therefore, AtsB-mediated modification is not restricted to serine-type sulfatases, and some cysteine-type sulfatases can also be modified by AtsB. It should be noted that we do not know whether *Pseudomonas* sulfatase, which is an AtsB-independent sulfatase, can also be modified by AtsB and at present we have no assay system, in which AtsB-independent modification is not operative.

Both *Pseudomonas* and *Methanosarcina* sulfatase are cytosolic cysteine-type sulfatases, which share a similar FGly generation motif (*Pseudomonas* sulfatase: **C**S**P**T**R**SMLL**TG**T, *Methanosarcina* sulfatase: **C**TAG**R**AAFI**TG**R, highly conserved residues of the general FGly-motif in bold). The question arises what part(s) in these two sulfatases determine(s) the AtsB-dependence or – independence of their modification. It has to be tested whether the slight differences between the two FGly generation motifs are decisive or whether other amino acid sequences in these two sulfatases play a role. *Methanosarcina /Pseudomonas* sulfatase hybrids may help to find the answer.

5.3 Cell-biological aspects of prokaryotic formylglycine-modification

AtsB-mediated serine-type formylglycine modification requires a signal peptide in its sulfatase substrate.

Klebsiella sulfatase (AtsA) is a periplasmic protein that is synthesized with a signal peptide directing its export through the plasma membrane. In fact, signal peptides are predicted for all known serine-type sulfatases. As AtsB is a cytosolic protein (Marquodt *et al.*, 2003), an interaction of AtsB and AtsA can only occur in the cytosol (*i.e.* prior to export of AtsA). In fact, AtsB-dependent modification is operative in the cytosol, since FGly formation in a signal peptide-deleted version of AtsA still depended on AtsB. However, FGly-modification of this cytosolic AtsA was extremely inefficient and was stimulated about 100-fold by the presence of the signal peptide (Fig. 4.12). This agrees with the finding that AtsB can modify neither a *Methanosarcina* sulfatase C55S mutant (Table 5.1) nor a *Pseudomonas* sulfatase C51S mutant (Fig. 4.11), which both lack a signal peptide. To convert *Pseudomonas* sulfatase into a substrate for AtsB, an N-terminal signal peptide had to be added to the *Pseudomonas* sulfatase in addition to the C51S mutation. Hence, under *in vivo* conditions a signal peptide in the sulfatase substrate is essential for AtsB-mediated serine-modification.

This may be explained in one of three ways. The hydrophobic signal peptide may exert a direct intramolecular effect on the sulfatase's folding state and modification competence. Alternatively, AtsB binds directly, or indirectly via an adaptor protein to the signal peptide, and this may be required for the FGly-forming activity of AtsB. Furthermore, the signal peptide, after binding of an adaptor, may ferry newly synthesized AtsA to a site where AtsB is active. In the latter two cases, a chaperone or targeting factor of *E. coli*'s early secretory pathway such as trigger factor, SecB, or SecA may serve as the putative adaptor protein. Generation of the FGly-residue immediately before export of the sulfatase would minimize the exposure of the toxic formyl group to other proteins.

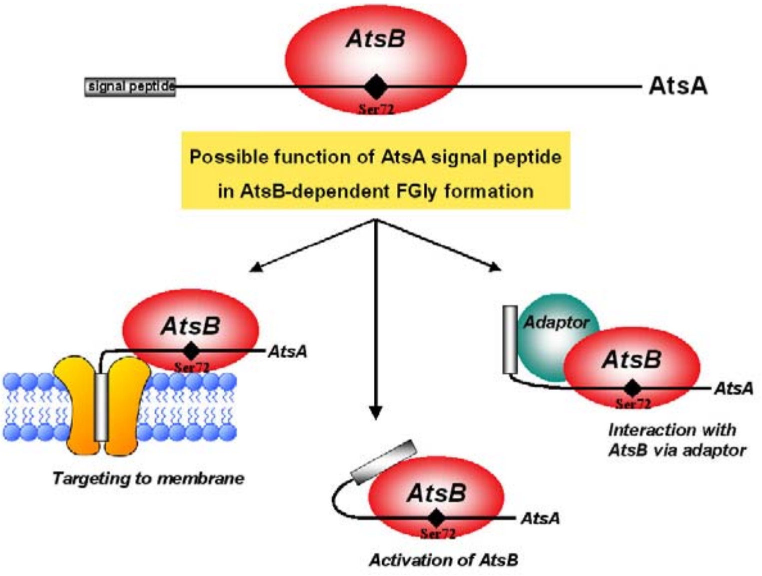

AtsB
signal peptide
Ser72
AtsA
Possible function of AtsA signal peptide in AtsB-dependent FGly formation
AtsB
AtsA
Ser72
Targeting to membrane
Adaptor
AtsB
AtsA
Ser72
Interaction with AtsB via adaptor
AtsB
AtsA
Ser72
Activation of AtsB

Figure 5.1 Possible functions of AtsA´s signal peptide in AtsB-dependent FGly formation.

AtsB-independent formylglycine modification does not require a signal peptide in the sulfatase substrate.

Pseudomonas sulfatase is a cytosolic protein, which is synthesized without a signal peptide. Therefore, AtsB-independent FGly modification does not rely on the presence of a signal peptide. To convert AtsA into a substrate of the AtsB-independent FGly modification system, the serine 72 residue has to be replaced by a cysteine and the N-terminal signal peptide has to be deleted, as is suggested by preliminary data showing reasonable activity after expression of a corresponding AtsA-S72CΔSP mutant (Table 5.1; Marquodt C, unpublished).

AtsB-mediated cysteine-type sulfatase modification is independent of a signal peptide.

As mentioned above, AtsB and its *M. mazei* ortholog can also modify the cysteine-type sulfatase of *Methanosarcina in vivo*. However, this sulfatase is a cytosolic protein. Therefore, this AtsB-dependent FGly modification does not require a signal peptide in the sulfatase substrate. Thus, obviously only AtsB-dependent serine-type sulfatase modification needs a signal peptide, whereas AtsB-dependent cysteine-type sulfatase modification does not. Obviously, AtsB is perfectly adapted on natural sulfatase substrates, as all serine-type sulfatases are periplasmic proteins containing a signal peptide, Interestingly, the presence of a signal peptide in the artificial cysteine-type sulfatase, AtsA-S72C directs it for AtsB-dependent modification (Table 5.1). Obviously, here the signal peptide and the secretory machinery target a sulfatase for AtsB-dependent modification independent of the nature of the FGly-modification motif.

Table 5.1 Prokaryotic sulfatase modification depends on AtsB or on an endogenous modification system of *E. coli* (AtsB-independent modification).

		signal peptide	AtsB-dependent modification	AtsB-independent modification
Serine-type	*Klebsiella* AtsA	+SP	50%[a]	0
		-SP	<1%	0
	Methanosarcina Sulfatase C55S mutant	+SP	n.d.	n.d
		-SP	0	0
	Pseudomonas Sulfatase C51S mutant	+SP	+++[b]	0
		-SP	0	0
Cysteine-type	*Klebsiella* AtsA S72C mutant	+SP	+++[b]	+/-[b]
		-SP	0[b]	+[b]
	Methanosarcina Sulfatase	+SP	n.d	n.d
		-SP	50%	+/-
	Pseudomonas Sulfatase	+SP	n.d[c]	+++[b]
		-SP	n.d[c]	100%

+SP: having signal peptide

-SP: no signal peptide

a: when purified from periplasm, 50% of serine 72 in AtsA is modified.
b: as concluded from the specific sulfatase activity (no FGly-determination)
c: cannot be determined in *E. coli* due to concomitant AtsB-independent modification
%: relative FGly modification degree
0: no FGly is generated
n.d: not determined
+/-: very weak modification at the detection limit
+: weak but significant modification (<10% FGly content or specific sulfatase activity <10% of the fully modified protein)
+++: high modification (≥10% FGly content or specific sulfatase activity ≥10% of the fully modified protein)

5.4 Mechanism of AtsB-mediated formylglycine-modification

AtsB acts as a radical SAM protein.

Here we show that AtsB acts as a radical SAM protein involving SAM and at least one FeS center as critical cofactors. The essential requirement for SAM was shown *in vivo*, after mutagenesis of the putative SAM binding site (see below), and also in *in vitro* experiments. Here binding of AtsB to AtsA was stimulated by SAM in a concentration-dependent manner, was impaired by its demethylated analog SAHC and was also dependent on an intact SAM binding site. In fact, ^{3}H-labelled SAM was recruited into a ternary AtsB/SAM/AtsA complex in a roughly 1:1:1 ratio (Fig. 4.14). Most significantly, we could demonstrate *in vitro* FGly-generation in the GST-AtsA substrate mediated by freshly prepared AtsB in dependence of SAM. This *in vitro* modification was furthermore dependent on temperature and incubation time, and required an intact serine-type modification motif in GST-AtsA (Fig. 4.15). It should be noted that already the formation of a stable AtsB/SAM/GST-AtsA complex was a slow and strictly temperature-dependent process (Fig. 4.13B and C). This suggests that recognition of the linear modification motif by AtsB is rate-limiting under the applied *in vitro* conditions and may require unfolding of the AtsA(residues 21-112) - appendix of the GST-domain. In fact, elongation of the AtsA-appendix at its N- or C-terminus interfered with *in vitro* interaction (not shown). *In vivo*, chaperones or components of the early secretory pathway may maintain the modification competence of newly synthesized AtsA polypeptides. Also our observation that AtsA's signal peptide, targeting the protein for export to the periplasm, stimulates FGly-modification 50- to 100-fold (Fig. 4.12), supports the idea that an export - competent, largely unfolded conformation is required for efficient FGly-modification.

AtsB's SAM binding site.

By site-specific mutagenesis we found that the sequence ^{83}GGE^{85} is essential for AtsB function. Substitution of each of these three residues by alanine drastically impaired FGly-modification activity in vivo (Fig. 4.16A). *In vitro*, none of the mutants was able to bind SAM, to interact with AtsA or to generate FGly (Fig. 4.16B and C). Mutation of the iuxtaposed Pro-86 impaired formation of the ternary AtsB/SAM/AtsA complex partially. GGEP perfectly matches a semi-consensus sequence in an alignment block of the radical SAM superfamily (Sofia *et al.*, 2001). Another member of this huge family is the HemN protein of *E. coli*, which catalyzes oxygen-independent oxidative decarboxylation of coproporphyrinogen-III. HemN carries the corresponding GGTP sequence at positions 112-115. A double mutant in which Gly-113 and, in addition, Gly-111 that directly precedes GGTP, both were substituted by valine, no longer showed enzymatic activity (Layer *et al.,* 2002). The binding of SAM by wild-type HemN and its inactivation in the mutant, however, was not assayed.

Interestingly, Shields *et al.* (2003) located a ^{98}GxG^{100} and a ^{180}EE^{181} motif in phosphatidylethanolamine N-methyltransferase that seem to cooperate in SAM binding. These separate sequence elements were found in subgroups of SAM-utilizing enzymes. From the crystal structure of several SAM-dependent methyltransferases it is known that a glycine-rich motif (motif I) and a separate motif (post I) containing a negatively charged amino acid make direct contact to SAM (Niewmierzycka and Clarke, 1999). AtsB's linear GGE-motif includes both glycines and a negative charge that, as shown here, are critical for binding of SAM.

Mechanism of AtsB function.

Like other radical SAM proteins, AtsB needs a specific FeS center to facilitate reductive cleavage of SAM. Apart from the correct position, 40-50 residues upstream of the SAM binding site, and perfect match of the ^{35}C-x3-C-x-Y-C-Y^{43} cysteine cluster of AtsB with the corresponding semiconsensus sequence of radical SAM proteins (Sofia *et al.*, 2001), we could provide experimental evidence for the essential role of this cluster. Mutagenesis of the corresponding cysteines fully inactivated AtsB's FGly-generating activity in vivo and in vitro (Fig. 4.17A-C). Interestingly, the same result was obtained after mutagenesis of two additional cysteine clusters suggesting the existence of two further important FeS centers in the C-terminal half of AtsB, which are not shared with radical SAM proteins in general. Direct evidence for the participation of metal ions came from the observation that chelating agents such as EDTA or o-phenanthroline completely blocked the FGly-modification activity of AtsB (Fig. 4.17D).

The specific FeS center of radical SAM proteins in its $[4Fe\text{-}4S]^+$ form is active for reductive cleavage of SAM (Jarrett, 2003). Since it is coordinated by only three cysteines, one iron coordination position is vacant, thereby explaining its oxygen sensitivity (Frey, 2001). On the other hand, the unique iron, that is not coordinated by a cysteine, allows SAM binding through association of the SAM sulfonium either directly with this iron or with one of its coordinating sulfides in the [4Fe-4S] center, and a possible bidentate chelation of the iron by the methionyl carboxylate and amine (Walsby *et al.,* 2002; Cosper *et al.*, 2000). These close interactions facilitate transfer of an electron from the FeS center to SAM resulting in its reductive cleavage to form methionine and the highly oxidizing deoxyadenosyl radical. The latter can directly abstract a hydrogen atom from unactivated C-H bonds

In case of AtsB, a hydrogen would be abstracted from $-C^{ß}H_2-OH$ of Ser-72 in AtsA, leading to production of deoxyadenosine and the substrate radical [-CH-OH]°. The latter has to be further oxidized, possibly by passing an electron to the oxidized $[4Fe-4S]^{2+}$ center, thereby producing FGly ($-CHO + H^+$) and regenerating the reduced $[4Fe-4S]^+$ center:

(1) $[FeS]_{red} + AdoMet^+ \rightarrow [FeS]_{ox} + Methionine + Ado^o$

(2) $Ado^o + Gly\text{-}CH_2OH \rightarrow AdoH + Gly\text{-}[CHOH]^o$

(3) $Gly\text{-}[CHOH]^o + [FeS]_{ox} \rightarrow [FeS]_{red} + Gly\text{-}CHO + H^+$

$AdoMet^+$ =S-adenosylmethionine
Ado^o = 5'-deoxyadenosyl radical
AdoH = 5'-deoxyadenosine
$Gly\text{-}CH_2OH$ = serine
Gly-CHO = FGly

SCHEME 1. Proposed mechanism for AtsB-mediated FGly formation.

To avoid formation of the highly reactive deoxyadenosyl radical in the absence of substrate, the generation of the reduced $[4Fe-4S]^+$ form and/or SAM binding have to be tightly regulated. The former can be achieved by controlled reduction through other FeS centers present in AtsB, the latter by cooperativity of substrate (AtsA) and SAM binding. For biotin synthase, a well-studied radical SAM protein, Ugulava *et al.* (2003) observed that both the rate of binding and the affinity for SAM increased significantly in the presence of substrate, and furthermore that substrate bound only in the presence of SAM. In our *in vitro* system we noticed a pronounced stimulation of AtsB binding at low, non-saturating SAM concentrations (Fig. 4.14B), which may indicate such cooperativity.

In view of the model outlined above, the critical [4Fe-4S]$^{+/2+}$ center with its unique iron contributes to SAM binding. We observed that mutating the corresponding cysteine cluster (Cys-39 and Cys-42) led to catalytically inactive AtsB protein in which SAM binding was much stronger affected than association with AtsA (Fig. 4.17B). This may reflect the contribution of the [4Fe-4S] center to SAM binding.

Further insight into molecular aspects of AtsB function most probably will require analysis under anaerobic conditions. Studies of its FeS center(s) by EPR spectroscopy in the absence and presence of SAM and a substrate peptide should clarify basic properties of those centers and their interconversions. For understanding the mechanism of the radical SAM family in general, consisting of an enormous variety of enzymes, 3D structures of some of its members will be extremely valuable.

6. Summary

Cα-formylglycine is the catalytic residue of sulfatases. FGly is generated by posttranslational modification of a cysteine (pro- and eukaryotes) or serine (prokaryotes) located in a conserved (C/S)xPxR motif. AtsB, a predicted radical SAM iron-sulfur protein, is a cytosolic protein and is strictly required for modification of serine 72 in the arylsulfatase AtsA of *K. pneumoniae*.

AtsB physically interacts with AtsA in a serine 72-dependent manner as shown in GST-pull-down experiments. The key cofactor for AtsB function is S-adenosylmethionine (SAM) which is recruited into a ternary AtsB/SAM/AtsA complex, formation of which is a prerequisite for FGly-modification, as shown by a newly developed *in vitro* assay. This strongly suggests that AtsB is the modifying enzyme for serine-type sulfatases. The sequence ^{83}GGE^{85} of AtsB and possibly a iuxtaposed iron-sulfur center, coordinated by cysteine 39 and cysteine 42, are critical for SAM-binding. Mutation of these and other conserved cysteines as well as treatment with metal chelators fully impaired FGly-formation, indicating that all three predicted FeS centers are essential for AtsB function. It is concluded that AtsB oxidizes serine to FGly by a complex radical mechanism that is initiated through reductive cleavage of SAM, thereby generating the highly oxidizing deoxyadenosyl radical which abstracts a hydrogen from the serine $-C^{\beta}H_2-OH$ side chain.

Surprisingly, AtsB and its orthologs can also modify cysteine-type sulfatases. The only natural substrate identified so far is a sulfatase from the archaebacterium *M. mazei*, which is efficiently modified by AtsB or its *M. mazei* ortholog. AtsB-mediated modification of the periplasmic serine-type, but not of the cytosolic cysteine-type sulfatases requires a signal peptide attached to the newly synthesized polypeptide. Thus, the cytosolic AtsB relies on a cytosolic

function of the signal peptide that may involve factors of the early secretory pathway. The latter may assist in maintaining an unfolded, modification competent-conformation of the sulfatase or may couple modification to export through the translocon across the plasma membrane.

For further functional and structural analysis, AtsB was purified close to homogeneity. However, the obtained protein showed no FGly modification activity, which is attributed to the oxygen sensitivity of AtsB’s iron-sulfur center(s). Future studies will require purification under anaerobic conditions.

References

Ai X, Do AT, Lozynska O, Kusche-Gullberg M, Lindahl U, Emerson CP Jr. (2003) QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling. J Cell Biol. 162:341-51.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D., Seidman, J.G., Smith, J.A. and Struhl, K.(1997). Current protocols in molecular biology, John Wiley & Sons Inc., New York.

Ballabio, A. and Shapiro, L.J. (2001). Steroid sulfatase deficiency and X-linked ichthyosis. In The Metabolic and Molecular Basis of Inherited Disease. Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. eds. (New York: McGraw-Hill).

Bassam, B. J., Caetano-Anolles, G., Gresshoff, P. M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem., 195:80-83.

Beil, S., Kehrli, H., Jamers, P., Staudenmann, W., Cook, A.M., Leisinger, T. and Kertesz, MA. (1995) Purification and characterization of the 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.

Bond CS, Clements PR, Ashby SJ, Collyer CA, Harrop SJ, Hopwood JJ, and Guss JM. (1997) Structure of a human lysosomal sulfatase. Structure. 5:277-289.

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

Cheng Q., Hwa V., and 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.

Coligan, J.E., Dunn, B.N., Ploegh, H.L., Speicher, D.W. and Wingfield, P.T. (1997). Current protocols in protein science, John Wiley & Sons Inc., New York.

Cosma MP, Pepe S, Annunziata I, Newbold RF, Grompe M, Parenti G, Ballabio A. (2003) The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases.113(4):445-456.

Cosper NJ, Booker SJ, Ruzicka F, Frey PA, Scott RA. (2000) Direct FeS cluster involvement in generation of a radical in lysine 2,3-aminomutase. 39(51):15668-15673

Daniele A, Parenti G, d'Addio M, Andria G, Ballabio A, Meroni G. (1998) Biochemical characterization of arylsulfatase E and functional analysis of mutations found in patients with X-linked chondrodysplasia punctata.Am J Hum Genet. 62:562-572.

Dhoot G.K., Gustafsson M.K., Ai X., Sun W., Standiford D.M., and Emerson C.P. Jr. (2001) Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. Science. 293:1663-1666.

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

Dierks T., Miech T., Hummerjohann J., Schmidt B., Kertesz M.A., and von Figura, K. (1998) Post formation of formylglycine in Prokaryotic Sulfatases by modification of either cysteine or serine J Biol Chem 273: 25560-25564.

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

Dierks T, Schmidt B, Borissenko LV, Peng J, Preusser A, Mariappan M, von Figura K. (2003) Multiple sulfatase deficiency is caused by mutations in the gene encoding the human C(alpha)-formylglycine generating enzyme. Cell.113:435-444.

Dodgson, K. S., White, G. F., and Fitzgerald, J. W. (1982) sulfatases of Microbial Origin, CRC Press, Boca Raton, FL.

von Figura, K., Schmidt, B., Selmer, T., and Dierks, T. (1998) A novel protein modification generating an aldehyde group in sulfatases: its role in catalysis and disease. Bioessays 20, 505-510.

von Figura, K., Gieselmann, V., and Jaeken, J. (2001). Metachromatic leukodystrophy. In The Metabolic and Molecular Basis of Inherited Disease. Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. eds. (New York: McGraw-Hill)

Franco B, Meroni G, Parenti G, Levilliers J, Bernard L, Gebbia M, Cox L, Maroteaux P, Sheffield L, Rappold GA, et al. (1995) A cluster of sulfatase genes on Xp22.3: mutations in chondrodysplasia punctata (CDPX) and implications for warfarin embryopathy. Cell 81:15-25.

Hernandez-Guzman FG, Higashiyama T, Pangborn W, Osawa Y, Ghosh D. (2003) Structure of human estrone sulfatase suggests functional roles of membrane association. J Biol Chem.278(25):22989-97.

Janknecht R, de Martynoff G, Lou J, Hipskind RA, Nordheim A, Stunnenberg HG. (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc Natl Acad Sci U S A. 88:8972-8976.

Jarrett JT. (2003) The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. Curr Opin Chem Biol.7(2):174-82.

Johnson,M.K. (1998) Iron-sulfur proteins: new roles for old clusters. *Curr. Opin. Chem. Biol.*, 2**:** 173–181.

Joseph Sambrook, David W. Russell, (2001) Molecular Cloning: A Laboratory Manual (3rd version) Cold Spring Harbor Laboratory Press

Kertesz M.A. (1999) Riding the sulfur sycle – metabolism of sulfonates and sulfate esters in Gram-negative bacteria. FEMS microbiol Review, 24: 135-175.

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

Layer G, Verfürth K, Mahlitz E, and Jahn D. (2002) Oxygen-independent Coproporphyrinogen-III Oxidase HemN from *Escherichia coli* J. Biol. Chem., 277: 34136-34142.

Lukatela G, Krauss N, Theis K, Selmer T, Gieselmann V, von Figura K, and Saenger W. (1998) Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. Biochemistry. 37:3654-64.

Marquordt C, Fang Q, Will E, Peng J, von Figura K, Dierks T. (2003) Posttranslational modification of serine to formylglycine in bacterial sulfatases.

Recognition of the modification motif by the iron-sulfur protein AtsB. 278(4):2212-2218.

Miech C, Dierks T, Selmer T, von Figura K, Schmidt B. (1998) Arylsulfatase from Klebsiella pneumoniae carries a formylglycine generated from a serine. J Biol Chem. 273(9):4835-4837.

Murooka Y., Ishibashi K., Yasumoto M., Sasaki M., Sugino H., Azakami H., and Yamashita M., (1990) A sulfur- and tyramine-regulated Klebsiella aerogenes operon containing the arylsulfatase (atsA) gene and atsB gene. J. Bacteriol. 172: 2131-2140.

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

Niewmierzycka A, Clarke S. (1999) S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase. J Biol Chem. 274(2):814-24.

Okamura H., Yamada, T., Murooka Y., and Harada T. (1976) Purification and properties of arylsulfatase of Klebsiella aerogenes: identity of the enzymes formed by non-repressed and derepressed synthesis. Agric. Biol. Chem. 40: 2071-2076.

Parenti G., Meroni G., and Ballabio A.. (1997) The sulfatase gene family. Curr Opin Genet Dev 7: 386-391.

Peng J, Schmidt B, von Figura K, Dierks T. (2003) Identification of formylglycine in sulfatases by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. 38(1):80-86.

Recksiek M, Selmer T, Dierks T, Schmidt B, and von Figura K. (1998) Sulfatases, trapping of the sulfated enzyme intermediate by substituting the active site formylglycine. J Biol Chem. 273:6096-103.

Schirmer A. and Kolter R. (1998) Computational analysis of bacterial sulfatase and their modifying enzyme. Chem. Biol. 5: 181-186.

Schmidt B, Selmer T, Ingendoh A, von Figura K. (1995) A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. 82(2):271-278.

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

Shields SJ, Oyeyemi O, Lightstone FC, Balhorn R. (2003) Mass spectrometry and non-covalent protein-ligand complexes: confirmation of binding sites and changes in tertiary structure. 14(5) 460-470.

Sofia H.J., Chen G., Hetzler B. G., Reyer-Spindola J. F., and Miller N.E. (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Research 29: 1097-1106

Schirmer A, Kolter R. (1998) Computational analysis of bacterial sulfatases and their modifying enzymes.Chem Biol. 5(8): 181-186.

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

Walsby CJ, Hong W, Broderick WE, Cheek J, Ortillo D, Broderick JB, Hoffman BM. (2002) Electron-nuclear double resonance spectroscopic evidence that S-adenosylmethionine binds in contact with the catalytically active [4Fe-4S](+) cluster of pyruvate formate-lyase activating enzyme. J Am Chem Soc. 124 (12):3143-51

Yasukawa T, Kanei-Ishii C, Maekawa T, Fujimoto J, Yamamoto T, Ishii S. (1995) Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. 270(43):25328-25331.

Acknowledgments

I would like to express my gratitude to Prof. Dr. Dr. h.c. Kurt von Figura for his excellent supervision, helpful tips, advices and comments, which led to successful experiments and important data.

I am grateful to Prof Dr. Fritz for co-reference.

My very special thanks Dr. Thomas Dierks for giving me the opportunity to do this work. This has been an exciting project for me and I thank him for guiding me. I have also benefited from the stimulating discussions and his excellent supervision. I acknowledge him for the critical reading of the thesis.

I thank Deutshe Forschungsgemeinschaft for financial support of this work.

I would like to thank Martina Balleininger, Nelli Schwabauer, Petra Schlotterhose and Nicole Leister for their excellent technical assistance and nice working atmosphere in the lab, and Klaus Neifer for DNA and protein sequence.

My very thanks to Jianhe Peng for establishing the MALDI-TOFMS FGly detection methods. I thank Ljudmila Borissenko, Michael Padva, Stephanie Schulz and Malaiyalam Mariappan for being excellent colleagues and understanding friends.

I thank all the colleagues from the Department of Biochemsitry II for creating pleasant working climate.

My very special thanks are addressed to my family and my friends, whose care, understanding, help and support made this work possible.

Publication and conferences

Marquordt C, **Fang Q,** Will E, Peng J, von Figura K, Dierks T Post-translational modification of serine to formylglycine in bacterial sulfatases: Recognition of the modification motif by the iron sulfur protein AtsB. Journal of Biological Chemistry. 2003 278(4): 2212-2218

Parts of this work was presented at the following event:

ELSO 2003 September 18-21, 2003, Berlin, Germany, poster

Curriculum vitae

Name: Qinghua Fang

Geburtsdatum: 15. August, 1975

Geburtsort: Shanghai, China

Eltern: Jiqiu Fang
Congyi Tang

Schulbildung:	1981-1987	Grundschule, Shanghai, China
	1987-1993	High School, Shanghai, China
Studium:	1993-1997	East China University of Science and Technology, Shanghai China, Food Engineering (Bachelor degree)
	1997-2000	East China University of Science and Technology, Shanghai China, Biochemical Engineering (Master degree), State Key Lab of Bioreactor, Prof. Jianjiang Zhong
Dissertation:	2000-2004	Institut für Biochemie II, Universität Göttingen, Deutschland, Prof. Dr. Dr. h.c. Kurt von Figura und Dr. Thomas Dierks