Posttranslational Formation of Formylglycine in Prokaryotic Sulfatases by Modification of Either Cysteine or Serine*

(Received for publication, June 22, 1998, and in revised form, July 8, 1998)

Thomas Dierks‡§, Claudia Miech‡, Jörg Hummerjohann¶, Bernhard Schmidt, Michael A. Kertesz‡, and Kurt von Figura

From the Institut für Biochemie und Molekulare Zellbiologie, Abt. Biochemie II, Universität Göttingen, Goserstrasse 12d, 37073 Göttingen, Germany and Institute of Microbiology, Swiss Federal Institute of Technology, ETH-Zentrum, CH-8092 Zürich, Switzerland

Eukaryotic sulfatases carry an α-formylglycine residue that is essential for activity and is located within the catalytic site. This formylglycine is generated by posttranslational modification of a conserved cysteine residue. The arylsulfatase gene of Pseudomonas aeruginosa also encodes a cysteine at the critical position. This protein could be expressed in active form in a sulfatase-deficient strain of P. aeruginosa, thereby restoring growth on aromatic sulfates as sole sulfur source, and in Escherichia coli. Analysis of the mature protein expressed in E. coli revealed the presence of formylglycine at the expected position, showing that the cysteine is also converted to formylglycine in a prokaryotic sulfatase. Substituting the relevant cysteine by a serine codon in the P. aeruginosa gene led to expression of inactive sulfatase protein, lacking the formylglycine. The machinery catalyzing the modification of the Pseudomonas sulfatase in E. coli therefore resembles the eukaryotic machinery, accepting cysteine but not serine as a modification substrate. By contrast, in the arylsulfatase of Klebsiella pneumoniae, the formylglycine is found generated by modification of a serine residue. The expression of both the Klebsiella and the Pseudomonas sulfatases as active enzymes in E. coli suggests that two modification systems are present, or that a common modification system is modulated by a cofactor.

Sulfatases are members of a highly conserved gene family (1, 2) sharing extensive sequence homology and a unique post-translational modification (3, 4). This novel protein modification generates a 2-amino-3-oxopropanoic acid, C₅-formylglycine (FGly), residue (5). In eukaryotes, FGly formation occurs in the endoplasmic reticulum by oxidation of a conserved cysteine residue and is directed by a linear sequence surrounding this cysteine (6, 7). Deficiency of FGly formation is observed in multiple sulfatase deficiency, a rare human lysosomal storage disorder that is characterized by synthesis of sulfatase polypeptides from eukaryotes and K. pneumoniae (13). The Pseudomonas sulfatase was expressed in the wild-type form, carrying a cysteine in position 51, and as a C51S mutant. The recombinant sulfatase proteins were purified and analyzed for the presence of FGly. The results obtained showed that the Pseudomonas sulfatase was modified with high specificity and that the modification system accepted only the cysteine form of this sulfatase. The modification machinery therefore shows a different specificity to the machinery converting serine to FGly in the Klebsiella sulfatase.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—P. aeruginosa PAO1 (14) was cultivated either in LB medium or a succinate minimal medium (13) with sulfate or pentanesulfonate (500 µM) as sulfur source. Escherichia coli DH5α [supE44 ΔlacU169 (860 lacZΔM15) hsdR recA1 endA1 gyrA96 thi-1 relA1] was used for subcloning, and E. coli BL21 (DE3) [hsdS gal (lacIq857 ind1 Sam7 nin5 lacUV5-T7 gene 1)] was used for protein expression, as described below. All strains were grown aerobically with constant shaking (180 rpm) at 37 °C or at 30 °C for protein overproduction.

P. aeruginosa ATS2 [ΔatsA recA7 Δtsn501] was constructed as follows. An S32-base pair deletion was created in the atsA gene of P. aeruginosa by exonuclease III/Mung bean nuclease digestion of plasmid pME4051, which carried the atsA gene on a 3.9-kilobase EcoRI-SalI fragment in

* This work was supported by the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and the Swiss National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ These authors contributed equally to this work.

§ To whom correspondence should be addressed. Tel.: 49-551-395976; Fax: 49-551-395979; E-mail: dierks@ukb2-00.uni-bc.gwdg.de.

1 FGly, C₅-formylglycine; RP-HPLC, reversed phase-high performance liquid chromatography.
Site-directed Mutagenesis of Cysteine 51—Plasmid pME4055 was constructed by cloning the atsA gene of *P. aeruginosa* as a 2.9-kilobase *SalI* fragment in pBBR1MCS (16), under the control of the lac promoter. This plasmid served as a template for polymerase chain reaction with noncoding mutagenic primers comprising a Ncol-site (5′-CATGCTGCGATCCCGCGATCGGCGTGTCGACTCGG-3′; *C51A* carried a GCC triplet instead of GCT). The polymerase chain reaction products were subcloned as Ncol fragments replacing the corresponding fragments of the template DNA, which carried an additional Ncol site in the vector. The subcloned fragments were checked by DNA sequencing to preclude any polymerase chain reaction-derived errors.

**Sulfatase Production and Purification**—For overproduction of arylsulfatase protein, an *Ndel* site was first introduced at the atsA translation start site by site-directed mutagenesis of plasmid pME4055, using mutagenic primer ATS-N (5′-GACCGCCATATGACCAAGC-3′). The polymerase chain reaction product was then cloned into the expression vector pET24a (Novagen Inc.), and the 3′-part of the gene was replaced with the entire DNA fragment from pME4055 to give plasmid pME4322. The polymerase chain reaction-derived region of this plasmid was checked by DNA sequencing. For overproduction of the C51S and C51A mutant proteins, the mutant alleles were subcloned into pME4322 as EcoRV-*SalI* fragments.

Overproduction of the sulfatase proteins was carried out in *E. coli* BL21(DE3), as described previously (17). After lysis of lysate/lysate treatment the cells were lysed using a French press, and membranes were removed by ultracentrifugation (250,000 × g, 30 min). Cell extract was then desalted into 20 mM Tris/HCl, pH 7.5, with a PD-10 gel filtration column (Amersham Pharmacia Biotech). The proteins were purified by two chromatographic steps, using a BioCAD SPRINT apparatus (Perseptive Biosystems Inc.). A first separation was obtained by anion exchange chromatography on a 1-ml Resource-Q column (Amersham Pharmacia Biotech), at a flow rate of 3 ml/min. Proteins were eluted with a gradient of 0–100 mM Na2SO4 in 20 mM Tris/HCl, pH 7.5. Fractions were then desalted using a Vivaspin 4 concentrator (Vivascience), and concentrated to 1 ml (NH4)2SO4 by addition of concentrated ammonium sulfate solution. They were then further purified by hydrophobic interaction chromatography on a 1-ml Resource-Q column (Amersham Pharmacia Biotech) at a flow rate of 3 ml/min. Proteins were eluted with a gradient of 0–100 mM Na2SO4 in 20 mM Tris/HCl, pH 7.5. The pooled fractions containing the enzyme were concentrated by ultracentrifugation with a Vivaspin 4 concentrator (Vivascience), and corrected to 1 ml (NH4)2SO4 by addition of concentrated ammonium sulfate solution. They were then further purified by hydrophobic interaction chromatography on a 1-ml Resource-Q column (Amersham Pharmacia Biotech) at a flow rate of 3 ml/min. Proteins were eluted with a gradient of 0–100 mM Na2SO4 in 20 mM Tris/HCl, pH 7.5. Fractions containing the enzyme were desalted and concentrated as before and stored at −20 °C until required.

Arylsulfatase was assayed in whole cells and in cell extracts as described previously using 4-nitrocatechol sulfate (Fluka) as substrate (12). This sulfatase can be expressed as an active enzyme, even under derepressing conditions, and is unable to utilize aromatic sulfates as a sulfur source for growth. Upon transformation of this strain with a plasmid containing the wild-type atsA gene under control of the lac promoter (pME4055), the ability of the strain to grow with aromatic sulfates as sole sulfur source was restored, and sulfatase activity could be measured during growth either in LB (Table I) or minimal medium (not shown). However, when strain ATS2 was transformed with plasmids coding for the mutated C51S or C51A versions of the arylsulfatase protein, expression of sulfatase activity was below the detection limit (Table I), and no growth was possible with aromatic sulfates as sole sulfur source. The changed residue in these constructs, cysteine 51, is equivalent to the residue that is converted to FGly in the sulfatases of eukaryotes and of *Klebsiella pneumoniae*, suggesting that FGly plays a crucial role in the *Pseudomonas* enzyme, as well as in the other sulfatases.

In the *Klebsiella* sulfatase, the FGly is generated from a serine (12). This sulfatase can be expressed as an active enzyme in *E. coli* (22). To test whether the loss of arylsulfatase activity caused by the C51S and C51A mutations in the *Pseudomonas* sulfatase was only observed after expression in *P. aeruginosa*, the same plasmids were used to transform *E. coli* DH5α. After induction with isopropyl thiogalactopyranoside, arylsulfatase activity was measured in the cell extracts (Table I). Although *E. coli* carries at least one sulfatase-related gene, the *asA* gene (29), this species has not yet been found to express active endogenous sulfatases. Expression of the native *Pseudomonas* arylsulfatase in *E. coli* led to significant levels of enzyme activity in the cells, whereas the mutant forms again showed no enzyme activity (Table I). This demonstrates that the *P. aeruginosa* arylsulfatase can be produced in a stable, active form by *E. coli*.

For further protein chemical studies, the *atsA* gene and its mutant derivatives were placed under the control of the T7 promoter, and the proteins were overproduced in *E. coli* BL21(DE3). After overexpression of the *Pseudomonas* sulfatase or its C51S derivative in this strain, the target proteins constituted 20–30% of total cell protein (not shown). The wild-type protein and the C51S mutant were purified to homogeneity (4–10-fold purification); the purified wild-type enzyme showed a catalytic activity of 47 μmol/min per mg, whereas the activity of the C51S mutant was extremely low (13 nmol/min per mg).

### RESULTS

Arylsulfatase expression in *P. aeruginosa* PA01 is repressed during growth in LB medium because of the presence of excess inorganic sulfate (Table I; Ref. 13). Derepression is observed during growth in minimal medium with alternative sulfur sources such as sulfate esters, sulfonates, or methionine, and under these conditions arylsulfatase activities of up to 56 nmol/min per mg of total cell protein are observed (21). *P. aeruginosa*

| Strain          | Plasmid          | Encoded protein | Arylsulfatase activity (nmol/min/mg protein) |
|-----------------|------------------|-----------------|---------------------------------------------|
| *P. aeruginosa* |                  |                 |                                             |
| PA01            | pBBR1MCS         | AtsA            | 0                                           |
| ATS2            | pME4055          | AtsA            | 22                                          |
|                 | pPAS-S1          | C51S-AtsA       | 0                                           |
|                 | pPAS-A3          | C51A-AtsA       | 0                                           |
| *E. coli* DH5α  | pBBR1MCS         | AtsA            | 0                                           |
|                 | pME4055          | AtsA            | 0                                           |
|                 | pPAS-S1          | C51S-AtsA       | 0                                           |
|                 | pPAS-A3          | C51A-AtsA       | 0                                           |

Strains harboring the indicated plasmids were grown in LB medium. Gene expression in *E. coli* was induced in the mid-exponential growth phase with isopropyl-thio-galactopyranoside (0.5 mM) over 3 h, and arylsulfatase activity was then measured. Arylsulfatase activity in *P. aeruginosa* was measured in the late exponential phase.
The C51A mutant could not be purified, because it was degraded by the cells on overexpression (not shown).

To examine whether residue 51 was converted to FGly, the purified proteins were denatured in guanidine hydrochloride and incubated with NaB\(^{3\text{H}}\)H\(_4\), which reduces the aldehyde group of FGly and generates a \(^{3\text{H}}\)serine residue (4, 5, 12). The samples were then subjected to reductive carboxymethylation of cysteines. After removal of all low molecular weight compounds by gel filtration, aliquots were analyzed by SDS-polyacrylamide gel electrophoresis, followed by Coomassie staining and fluorography (Fig. 1). The wild-type protein was found to carry a \(^{3\text{H}}\)label but the C51S mutant was not labeled by this treatment (Fig. 1B). The two carboxymethylated sulfatases were digested with trypsin and their tryptic peptides were separated by RP-HPLC. The fractions were analyzed for radioactivity and by mass spectrometry. Radioactivity was recovered only from the wild-type sample (Fig. 2) and found to be associated with a single trypptic peptide (Fig. 2A) of 1521 Da (Fig. 3A). This mass is predicted for the \(^{3\text{H}}\)serine 51-containing form of the tryptic peptide 3 comprising residues 42–55 of the Pseudomonas sulfatase, i.e. after reduction of FGly-51. A peptide of the C51S mutant eluting with the same retention time (Fig. 2B) also showed a mass of 1521 Da (Fig. 3B) but carried no \(^{3\text{H}}\)-label (Fig. 2B). Sequencing of the two peptides led to the amino acid sequence predicted for the reduced wild-type and the C51S form of peptide 3 (Fig. 4, A and B), both carrying a serine in position 51. In the tenth sequencing cycle, corresponding to residue 51, the \(^{3\text{H}}\)-radioactivity was released from the wild-type peptide (Fig. 4C). The association of the radioactivity with residue 51 strongly suggests that cysteine 51 in the newly translated wild-type protein had been oxidized to FGly, which was then reduced to \(^{3\text{H}}\)serine by treatment with NaB\(^{3\text{H}}\)H\(_4\).

In control samples, in which the treatment with NaB\(^{3\text{H}}\)H\(_4\) had been omitted, peptide 3 showed a mass of 1519 Da for the wild-type peptide (Fig. 3C) and of 1521 Da for the C51S peptide (Fig. 3D), as predicted for the FGly-51 and serine 51-containing forms of peptide 3. When p-nitroaniline was used as a matrix for matrix-assisted laser desorption ionization mass spectrometry of the same peptides, masses of 1639 Da and 1521 Da were determined (Fig. 3, E and F). The increase of the wild-type peptide 3 by 120 Da is because of formation of a Schiff base between the peptide and p-nitroaniline (4, 5, 12), thereby confirming the presence of the aldehyde group in the wild-type and its absence in the mutant peptide. Further support for this conclusion was obtained when sequencing the peptides. Whereas the entire sequence of the C51S containing peptide 3 could be determined (not shown, cf. Fig. 4B), almost no signal was obtained for the C-terminal residues 51–55 of the wild-type peptide (Fig. 4D). This became most obvious in the case of proline 53, the only amino acid within this sequence (FGlySPTR) that can be recovered with good efficiency during sequencing (see Fig. 4A and B). The presence of a FGly residue in a peptide is known to block Edman degradation at the position of the FGly (4, 5, 12).

**DISCUSSION**

The presence of FGly in the aroylsulfatase of P. aeruginosa is a prerequisite for sulfatase activity, as has been observed earlier for the eukaryotic sulfatases (5, 11). FGly formation in the Pseudomonas aroylsulfatase is catalyzed in E. coli by specific oxidation of cysteine residue 51. Generation of FGly appeared to be quantitative, because by mass spectrometry we could not detect the unmodified peptide 3 comprising carboxymethylcysteine 51 (1595 Da) among its trypptic peptides. Modification of the Pseudomonas enzyme was only observed on a cysteine residue, and serine could not substitute for cysteine as a substrate for the reaction. This specificity, observed here in E. coli, is most likely also true in P. aeruginosa, because also in this species sulfatase activity could only be detected after expression of the wild-type enzyme, and not with the C51S sulfatase. Interestingly, the arylsulfatase protein was degraded by the E.
coli host after replacement of the cysteine 51 by an alanine residue, suggesting that changes in the modification status of the protein may also have an effect on its correct folding in vivo.

In *Klebsiella pneumoniae*, by contrast, it was found that the FGly in the *Klebsiella* sulfatase is generated by modification of a serine residue (12). This modification reaction is also catalyzed in *E. coli*, because the *Klebsiella* enzyme can be expressed in *E. coli* as an active sulfatase (22). Generation of FGly, i.e. serine semialdehyde, from serine is most likely to be a one-step oxidation process, whereas FGly generation from cysteine has been proposed to occur in two steps involving an oxidation and a hydrolysis reaction (3, 5). It is highly unlikely that the cysteine is converted to serine before being oxidized to FGly, because substitution of the critical cysteine residue of the *Pseudomonas* sulfatase by serine abolished FGly formation. Thus, the modification of both cysteine and serine is achieved by direct oxidation of the respective residue found in the primary translation product.

The modifying machinery catalyzing this oxidation is highly specific for the respective residue, as was demonstrated for the cysteine-converting system of prokaryotes (this study) and of eukaryotes (6, 7, 11). The serine-converting system, which modifies the serine of the *Klebsiella* sulfatase but not that of the *Pseudomonas* C51S sulfatase, may involve a modification machinery independent of the cysteine-converting system. Alternatively a cofactor may confer specificity to a modification machinery catalyzing FGly formation from either cysteine or serine. Because the *Pseudomonas* sulfatase is a cytosolic enzyme, the cysteine-specific modification system must exist in the cytosol. On the other hand, the *Klebsiella* sulfatase contains a leader peptide and is exported into the periplasm. The serine-converting system may therefore not be localized in the cytosol, but in the plasma membrane or in the periplasm.

In both *P. aeruginosa* and *K. pneumoniae*, expression of arylsulfatase is coupled to the sulfur status of the cell, and is repressed when preferred sulfur sources such as sulfate or cysteine are present (21, 24). When the *atsA* gene was expressed under lac control, however, arylsulfatase activity was observed even in the presence of excess sulfate. Expression of the bacterial genes encoding the modification system(s) therefore appears to be independent of the sulfur supply to the cells. To date, the FGly modification has only been found in sulfatase enzymes of both eukaryotic and prokaryotic origin. However, 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 suggest that other as yet unidentified bacterial proteins also undergo a similar FGly modification.

The cysteine-modifying system of prokaryotes shows exactly the same specificity as the eukaryotic system. We anticipate, therefore, that a genetic approach to identify the modifying enzyme(s) in bacteria may be of help in elucidating the genetic defect in human multiple sulfatase deficiency, which leads to synthesis of catalytically inactive sulfatases lacking the FGly.

Acknowledgment—We thank Klaus Neifer for peptide sequencing.

REFERENCES

1. Parenti, G., Meroni, G., and Ballabio, A. (1997) Curr. Opin. Genet. Dev. 7, 386–391
2. Franco, B., Meroni, G., Parenti, G., Levilliers, J., Bernard, L., Gebbia, M., Cox, L., Morateaux, P., Sheffield, L., Rappold, G. A., Andria, G., Petit, C., and Ballabio, A. (1995) Cell 81, 15–25
3. von Figura, K., Schmidt, B., Selmer, T., and Dierks, T. (1996) Bioessays 18, 505–510
4. Selmer, T., Hallmann, A., Schmidt, B., Sumper, M., and von Figura, K. (1996) Eur. J. Biochem. 238, 341–345
5. Schmidt, B., Selmer, T., Ingendoh, A., and von Figura, K. (1995) Cell 82, 271–278
6. Dierks, T., Schmidt, B., and von Figura, K. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 11963–11968
7. Dierks, T., Leece, M. R., Schmidt, B., and von Figura, K. (1998) FEBS Lett.
Formylglycine Formation in Prokaryotic Sulfatases

17. Eichhorn, E., van der Ploeg, J. R., Kertesz, M. A., and Leisinger, T. (1997) J. Biol. Chem. 272, 23931–23936
18. Bradford, M. M. (1976) Anal. Biochem. 72, 248–254
19. Dierks, T., Volkmer, J., Schlenstedt, G., Jung, C., Sandholzer, U., Zachmann, K., Schlotterhose, P., Neifer, K., Schmidt, B., and Zimmermann, R. (1996) EMBO J. 15, 6931–6942
20. Chamberlain, J. P. (1979) Anal. Biochem. 98, 132–135
21. Hummerjohann, J., Kuttel, E., Quadrini, M., Leisinger, T., and Kertesz, M. A. (1998) Microbiology 144, 1375–1386
22. Muruoka, Y., Ishibashi, K., Yasumoto, M., Sasaki, M., Sugino, H., Azakami, H., and Yamashita, M. (1990) J. Bacteriol. 172, 2131–2140
23. Daniels, K. L., Plunkett, G., III, Burland, V., and Blattner, F. R. (1992) Science 257, 771–778
24. Dodgson, K. S., White, G. F., and Fitzgerald, J. W. (1982) Sulfatases of Microbial Origin, CRC Press, Inc., Boca Raton, FL