DOI: 10.1002/cbic.200800801

Formylglycine Aldehyde Tag—Protein Engineering through a Novel Post-translational Modification

Marc-André Frese and Thomas Dierks*[a]

Site-specific labeling and conjugation of proteins are central aims in protein engineering. To avoid perturbation of protein function, localization or interactions, the reactive attachment site needs to be defined, physiologically compatible and small in size. Recently, the application of a genetically encoded aldehyde tag was reported.[1,2] It makes use of a short, six-amino-acid recognition motif, which in vivo directs the post-translational oxidation of a specific cysteine to Cα-formylglycine (FGly). FGly is the essential active-site residue found in sulfatase enzymes.

The discovery of FGly as a novel amino acid, generated post-synthetically,[3] was prompted by the finding that, in patients suffering from multiple sulfatase deficiency, all sulfatases are produced as inactive polypeptides. FGly turned out to be the key catalytic residue of sulfatases that acts as an aldehyde hydrate in a novel hydrolytic reaction mechanism.[4] Patient sulfatases lack FGly and carry cysteine instead, because primary translation products remain unmodified. The defective FGly modification step occurs in the endoplasmic reticulum (ER) at a stage when newly synthesized sulfatases have not yet folded into their native structure.[5] The modification is directed by the short linear motif CxPxR.[6] An auxiliary motif (xxxLTGR), immediately downstream, was shown to improve FGly formation significantly. The arylsulfatase A-derived 16-mer sequence PVSLCTPSRAALLTGR was able to direct FGly generation when transplanted into heterologous proteins.[5,6]

The 16-mer peptide and modifications thereof served as affinity ligands that allowed the FGly-generation enzyme (FGE) to be purified from the soluble content of ER membranes.[7] The FGE-encoding gene was discovered simultaneously through a sophisticated genetic approach.[8] FGE turned out to be the first functionally characterized member of a large enzyme family found in both pro- and eukaryotes. In fact, FGly in bacterial sulfatases had been discovered earlier and, surprisingly, could be generated from both cysteine and serine residues.[9,10] Thus, bacteria are much more versatile and have evolved three different systems to generate FGly, namely 1) the cysteine-specific FGE, 2) the serine- and cysteine-specific AtsB, also termed anaerobic sulfatase maturating enzyme (anSME),[11,12] and 3) a system X, unknown as yet, that efficiently modifies cysteine in certain sulfatases when expressed in *E. coli*.[9,13]

The FGE- and AtsB/anSME-mediated mechanisms of FGly generation differ profoundly. On the one hand FGE catalyzes a metal-independent oxygenase reaction involving cysteine sulfenic acid as the reactive species.[14,15] On the other hand, the AtsB/anSME family consists of radical S-adenosylmethionine proteins that initiate catalysis through Fe-dependent generation of an adenosyl radical.[16–18] These entirely different FGE and AtsB mechanisms and the third unidentified *E. coli* system provide three enzymatic tools with divergent substrate specificities and reactive conditions for generating FGly.

When attempting to generate the FGly aldehyde tag in engineered proteins, bacterial expression systems are advantageous. Due to the small size of the tag, the DNA sequence encoding the recognition motif is easily incorporated into the target protein cDNA by standard molecular biology techniques. Based on this strategy, Carrico et al.[2] expressed different model proteins, equipped with a LCTPSR recognition motif at either the N or C terminus, in *E. coli* together with FGE from *Mycobacterium tuberculosis*. Regardless of the position of the tag, the modification of the target cysteine residue to FGly was highly efficient (90–99%). A significant conversion of cysteine to FGly was observed even without coexpression of FGE due to the endogenous FGly-generating activity (system X) of *E. coli*. Surprisingly, mutations in which the conserved proline or arginine residues of the LCTPSR motif were replaced by alanine were well tolerated by the *E. coli* system;[1] this raises questions as to which minimal motif is required and how the specificity is restricted to target cysteine residues.

Using an alanine-scanning peptide substrate library, Rush et al.[1] investigated in detail the specificities of the FGEs from *M. tuberculosis* and *Streptomyces coelicolor*[19] to identify minimal sequence requirements. Interestingly, the two enzymes showed clearly divergent tolerance towards substitutions within the recognition motifs. The much broader spectrum of substrates accepted by *M. tuberculosis* FGE was explained by a reduced surface complementarity towards the sulfatase CxPxR peptide and, as a consequence, increased accessibility of the active site, as supported by homology modeling.[1] It could thus be possible, when applying such less-specific FGEs, to design motifs that are tailored for each recombinant protein in order to show minimum interference with protein structure and function.

Coexpression of the target protein together with a prokaryotic FGE is proba-

[a] *M.-A. Frese, Prof. Dr. T. Dierks*
Faculty of Chemistry, Biochemistry I
Bielefeld University
Universitätsstrasse 25
33615 Bielefeld (Germany)
Fax: (+49) 521-106-6014
E-mail: thomas.dierks@uni-bielefeld.de

bly the most convenient way to obtain aldehyde-tagged proteins. In vivo conditions enable post-translational FGly formation within the partially folded protein, while making use of the endogenous folding machinery in order to reach the native structure. However, in vitro generation of FGly by incubation of substrate peptides with purified FGE has also been reported.[1,7,20] In the case of completely folded proteins, the accessibility of the recognition motif to FGE is critical, thus suggesting that this approach is probably restricted to peptides and, possibly, denatured proteins. In vitro FGly generation under defined conditions might have the advantage of minimizing side reactions of the aldehyde that possibly lead to adducts with endogenous cellular compounds.

Because the recognition motif can be placed either at the N or C terminus or at internal positions, the chemically reactive aldehyde group can be used for site-specific modifications, depending on the desired applications. Therefore, the FGly aldehyde tag opens up a broad range of new possibilities for bioorthogonal protein conjugation, labeling and immobilization. Based on standard aldehyde chemistry, other molecules can be covalently linked to the aldehyde-tagged protein. For example, aminooxy or hydrazide reagents specifically react with the aldehyde group of the protein, yielding the corresponding oxime or hydrazone products.[2] An overview of possible applications of the aldehyde tag is given in Figure 1.

Beside the conjugation of the tagged protein with fluorescent dyes or biotin, which allows the specific detection of the protein in localization and interaction studies, conjugation with aminooxy-PEG (poly(ethylene glycol)) has been described.[2,21] PEGylation or conjugation to other water-soluble polymers can improve the therapeutic index of many pharmacologically relevant proteins and increase their serum half-life and proteolytic stability.[22] In contrast to the unspecific PEGylation of lysine or cysteine residues, the use of the FGly aldehyde tag enables the generation of homogenous, defined products, as desired for therapeutic approaches. Furthermore, the aldehyde function also allows the attachment of carbohydrates. By replacing endogenous glycosylation sites with the recognition motif, mammalian proteins could be produced in *E. coli* and subsequently linked to chemically synthesized glycans, thereby obtaining glycoproteins similar to those produced by eukaryotic cells.[21] A further promising therapeutic application is the directed conjugation of drugs to protein ligands such as antibodies, which could be used in tumor targeting.

Technical applications of the FGly aldehyde tag are also conceivable. The specific and oriented immobilization of tagged proteins to surfaces or beads will be useful in the context of protein microarrays and biosensors. Interestingly, it has also been shown that the aldehyde modifications can be reversed or even replaced. Due to the differences in thermodynamic stability, a biotin residue linked across an *N*-acylhydrazone to a FGly-containing protein was successfully replaced with either methoxyamine to remove the label or with an aminooxy-modified FLAG peptide.[2] In both cases, the more stable oximes were formed, thus enabling sequential reactions at the reactive position. The same strategy might also allow application of the aldehyde tag in protein purification, for example, by using an immobilized hydrazide as an affinity matrix and methoxyamine as an eluent.

In summary, the FGly aldehyde tag represents an easy way to specifically equip proteins with a genetically encoded aldehyde function. This method expands the molecular toolbox for protein engineering with promising possibilities in basic research, biotechnology and pharmacology.

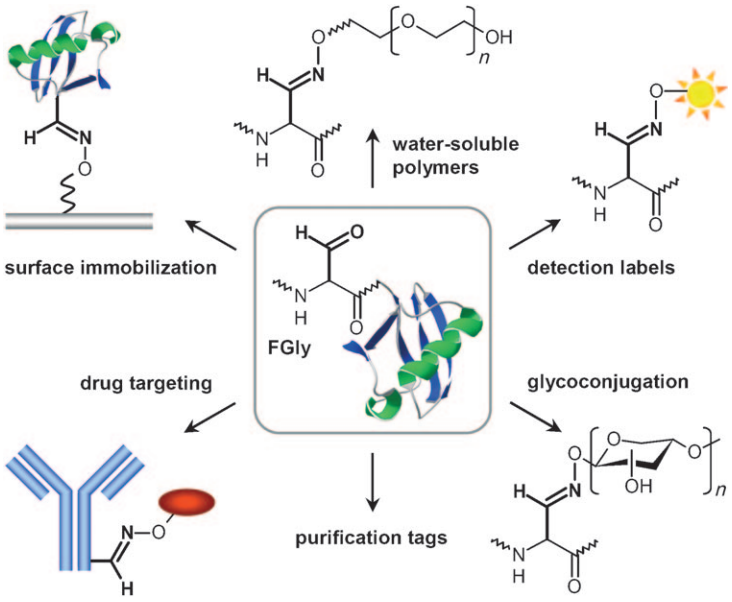

Figure 1. Possible applications of genetically encoded formylglycine aldehyde tags in protein engineering.

Acknowledgements

We would like to thank Norbert Sewald (Bielefeld University) for critically reading the manuscript.

Keywords: aldehyde tag · enzymes · formylglycine · protein engineering · protein modifications

[1] J. S. Rush, C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130*, 12240–12241.
[2] I. S. Carrico, B. L. Carlson, C. R. Bertozzi, *Nat. Chem. Biol.* **2007**, *3*, 321–322.
[3] B. Schmidt, T. Selmer, A. Ingendoh, K. von Figura, *Cell* **1995**, *82*, 271–278.
[4] I. Boltes, H. Czapinska, A. Kahnert, R. von Bülow, T. Dierks, B. Schmidt, K. von Figura, M. A. Kertesz, I. Usón, *Structure* **2001**, *9*, 483–491.
[5] T. Dierks, B. Schmidt, K. von Figura, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11963–11968.
[6] T. Dierks, M. R. Lecca, P. Schlotterhose, B. Schmidt, K. von Figura, *EMBO J.* **1999**, *18*, 2084–2091.

[7] T. Dierks, B. Schmidt, L. Borissenko, J. Peng, A. Preusser, M. Mariappan, K. von Figura, *Cell* **2003**, *113*, 435–444.
[8] P. Cosma, S. Pepe, I. Annunziata, R. Newbold, M. Grompe, G. Parenti, A. Ballabio, *Cell* **2003**, *113*, 445–456.
[9] T. Dierks, C. Miech, J. Hummerjohann, B. Schmidt, M. A. Kertesz, K. von Figura, *J. Biol. Chem.* **1998**, *273*, 25560–25564.
[10] C. Miech, T. Dierks, T. Selmer, K. von Figura, B. Schmidt, *J. Biol. Chem.* **1998**, *273*, 4835–4837.
[11] C. Szameit, C. Miech, M. Balleininger, B. Schmidt, K. von Figura, T. Dierks, *J. Biol. Chem.* **1999**, *274*, 15375–15381.
[12] O. Berteau, A. Guillot, A. Benjdia, S. Rabot, *J. Biol. Chem.* **2006**, *281*, 22464–22470.
[13] A. Benjdia, G. Dehò, S. Rabot, O. Berteau, *FEBS Lett.* **2007**, *581*, 1009–1014.
[14] T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner, M. Rudolph, *Cell* **2005**, *121*, 541–552.
[15] D. Roeser, A. Preusser-Kunze, B. Schmidt, K. Gasow, J. Wittmann, T. Dierks, K. von Figura, M. Rudolph, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 81–86.
[16] C. Marquordt, Q. Fang, E. Will, J. Peng, K. von Figura, T. Dierks, *J. Biol. Chem.* **2003**, *278*, 2212–2218.
[17] A. Benjdia, S. Subramanian, J. Leprince, H. Vaudry, M. K. Johnson, O. Berteau, *J. Biol. Chem.* **2008**, *283*, 17815–17826.
[18] T. L. Grove, K. H. Lee, J. St. Clair, C. Krebs, S. J. Booker, *Biochemistry* **2008**, *47*, 7523–7538.
[19] B. L. Carlson, E. R. Ballister, E. Skordalakes, D. S. King, M. A. Breidenbach, S. A. Gilmore, J. M. Berger, C. R. Bertozzi, *J. Biol. Chem.* **2008**, *283*, 20117–20125.
[20] A. Preusser-Kunze, M. Mariappan, B. Schmidt, S. L. Gande, K. Mutenda, D. Wenzel, K. von Figura, T. Dierks, *J. Biol. Chem.* **2005**, *280*, 14900–14910.
[21] I. S. Carrico, B. L. Carlson, C. R. Bertozzi, P. Wu, *International patent application* **2008**, WO 2008/0363350A2.
[22] J. M. Harris, R. B. Chess, *Nat. Rev. Drug Discovery* **2003**, *2*, 214–221.

Received: December 2, 2008
Published online on January 7, 2009

