New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo

Jason S. Rush and Carolyn R. Bertozzi*
Departments of Chemistry and Molecular and Cell Biology and Howard Hughes Medical Institute,
University of California, Berkeley, California 94720

Received June 14, 2008; E-mail: crb@berkeley.edu

Published on Web 08/23/2008

Formylglycine generating enzyme (FGE) was identified in 2003 as the postranslational machinery that activates type I sulfatases in eukaryotes.1 The enzyme oxidizes a cysteine residue within a ~13 amino acid consensus sequence, also termed the “sulfatase motif”, forming an aldehyde-bearing formylglycine (FGly) residue (Figure 1) that is critical for the sulfatases’ catalytic function.2 In eukaryotes, FGE requires a minimal submotif, CxPxR,3,4 that is highly conserved among all type I sulfatases. However, in prokaryotes either CxP/AxR5 motifs or serine-based SxPxR6 motifs are found within sulfatases. Prokaryotic FGEs, first characterized from Mycobacterium tuberculosis and Streptomyces coelicolor,7 recognize CxPxR, while anaerobic sulfatase-maturating enzymes (anSMEs) act on both CxAxR and SxPxR.5 FGEs and anSMEs have distinct sulfatase substrates and catalytic mechanisms.7,8

In addition to its intriguing biological function, FGE has also attracted attention as a tool for protein engineering. Conversion of cysteine to FGly introduces a uniquely reactive aldehyde group at a specific site dictated by the sulfatase motif. Recently, we reported that a six-residue sulfatase submotif (LCTPSR) can be introduced at a specific site dictated by the sulfatase motif.9 Recently, we reported that a six-residue sulfatase submotif (LCTPSR) can be introduced at a specific site dictated by the sulfatase motif.9

In eukaryotes, FGE requires a minimal submotif, CxPxR,3,4 that is highly conserved among all type I sulfatases. However, in prokaryotes either CxP/AxR5 motifs or serine-based SxPxR6 motifs are found within sulfatases. Prokaryotic FGEs, first characterized from Mycobacterium tuberculosis and Streptomyces coelicolor,7 recognize CxPxR, while anaerobic sulfatase-maturating enzymes (anSMEs) act on both CxAxR and SxPxR.5 FGEs and anSMEs have distinct sulfatase substrates and catalytic mechanisms.7,8

In addition to its intriguing biological function, FGE has also attracted attention as a tool for protein engineering. Conversion of cysteine to FGly introduces a uniquely reactive aldehyde group at a specific site dictated by the sulfatase motif. Recently, we reported that a six-residue sulfatase submotif (LCTPSR) can be introduced at a specific site dictated by the sulfatase motif.9

Figure 1. Reaction catalyzed by FGE.

PARASLTTGQ and LCTPSRGSLFTGR, from S. coelicolor and M. tuberculosis, respectively). Each residue within the sequences was probed by alanine substitution to generate a total of 28 peptides including the two wild-type sequences (native alanine residues within the S. coelicolor sequence were substituted with glycine). The percent conversion of cysteine to FGly was quantified for each alanine- (or glycine)-substituted peptide relative to that of the corresponding wild-type sequence.

As shown in Figure 3, the two FGEs displayed different tolerances for alanine mutations within the sulfatase motifs. Substitution at any position in the native sequence recognized by S. coelicolor FGE resulted in significant reduction in FGly formation (Figure 3a, blue bars). Replacement of Thr3, Pro4, Arg6, or Leu9 with alanine was particularly detrimental. A similar specificity profile was observed with the library derived from the M. tuberculosis sulfatase motif (Figure 3b). Human FGE, which has a 51% amino acid sequence identity to S. coelicolor FGE, also has a strict requirement for Pro and Arg within the CxPxR sequence.3,4 However, the human enzyme is known to tolerate substitutions corresponding to Thr3 or Leu9,4 indicating species-specific variation in substrate preference. Surprisingly, M. tuberculosis FGE displayed a much greater tolerance for alanine substitution in both sulfatase motif libraries (Figure 3a and b, red bars). Notably, replacement of Pro4 or Arg6 with alanine was well tolerated, as were substitutions in the C-terminal region.

Despite the 46% amino acid sequence identity shared by M. tuberculosis and S. coelicolor FGEs, their response to alanine substitutions in peptide substrates is very different. To gain insight into the molecular basis of substrate discrimination, we generated structural models of FGE-peptide complexes using the S. coelicolor enzyme’s crystal structure7 and a homology model of M. tuberculosis FGE (Figure 4).14 These models indicated that the substrate’s conserved Pro residue binds within a pocket that varies considerably

Figure 2. A high-throughput assay for FGE activity.
in size between the species homologues. The pocket in the *M. tuberculosis* FGE model (Figure 4b) appears more open, potentially to be more confined around the bound Pro residue.

The data in Figure 3 suggest that FGEs from certain prokaryotes are capable of modifying alternative aldehyde tag sequences that diverge from the canonical motif. In previous work, we showed that *E. coli* possesses an FGE-like activity that converts Cys to FGly in heterologous proteins possessing the canonical sequence LCTPSR. Although its molecular identity is not known, the FGE-like activity’s presence in this popular protein expression host enables the production of aldehyde-tagged proteins without need for exogenous FGE. To determine whether *E. coli*’s FGE-like activity exhibits substrate promiscuity, we expressed the maltose-binding protein (MBP) possessing various aldehyde tag sequences at the C-terminus downstream of a His6 tag (Figure 5). Control proteins bearing the corresponding C-to-A mutation or the wild-type sulfatase motif (LCTPSR) were expressed similarly. The isolated proteins were reacted with Alexa Fluor C5-aminoxyacetamide and analyzed by SDS-PAGE and fluorescence imaging (Figure 5).

The *E. coli* machinery converted all three sequences tested—LCTPSR (wild-type), LCTASR, and LCTASA—at comparable levels, while no signal was observed for any of the C-to-A mutants.Alanine substitution of the conserved Pro and Arg residues within the canonical sequence did not significantly reduce conversion efficiency. This striking observation suggests that a wide range of aldehyde tag sequences are recognized in *E. coli*, offering a practical system for expression of modified proteins.

In summary, peptide library screening revealed noncanonical sequences that are recognized by *M. tuberculosis* FGE in vitro and the *E. coli* FGE-like activity in vivo. This finding expands the range of aldehyde tag sequences for protein engineering. An important future goal is to identify the molecular nature of *E. coli*’s machinery.

Acknowledgment. We thank M. Breidenbach and B. Carlson for technical assistance. This work was supported by grants to C.R.B. from the National Institutes of Health (GM059907 and Nanomedicine Development Center).

Supporting Information Available: Experimental procedures, spectral data, and assay data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

1. Dierks, T.; Schmidt, B.; Borissenko, L. V.; Peng, J. H.; Preusser, A.; Mariappan, M.; von Figura, K. *Cell* 2003, 113, 435–444.
2. Schmidt, B.; Selmer, T.; Ingendoh, A.; von Figura, K. *Cell* 1995, 82, 271–278.
3. Dierks, T.; Lecca, M. R.; Schlotterhose, P.; Schmidt, B.; von Figura, K. *EMBO J.* 1999, 18, 2084–2091.
4. Knaut, A.; Schmidt, B.; Dierks, T.; von Bulow, R.; von Figura, K. *Biochemistry* 1998, 37, 13941–13946.
5. Berteau, O.; Guillot, A.; Benjdia, A.; Rabot, S. *J. Biol. Chem.* 2006, 281, 22464–22470.
6. Szameit, C.; Miech, C.; Balleininger, M.; Schmidt, B.; von Figura, K.; Dierks, T. *J. Biol. Chem.* 1999, 274, 15375–15381.
7. Carlson, B. L.; Ballister, E. R.; Skordalakes, E.; King, D. S.; Breidenbach, M. A.; Gilmore, S. A.; Berger, J. M.; Bertozzi, C. R. *J. Biol. Chem.* 2008, 283, 20117–20125.
8. Benjdia, A.; Leprince, J.; Guillot, A.; Vaudry, H.; Rabot, S.; Berteau, O. *J. Am. Chem. Soc.* 2007, 129, 3462–3463.
9. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* 2007, 3, 321–322.
10. For example, methylobacterium species and *Synechococcus sp.* WH 5701 possess putative sulfatases with the motifs CTAGR and CTSGR, respectively.
11. (11) See Supporting Information for synthetic details.
12. The assay was optimized using authentic synthetic standards. The presence of FGly was confirmed via oxime formation followed by MALDI mass spectrometry. See Supporting Information for details.
13. Roesser, D.; Preusser-Kunze, A.; Schmidt, B.; Gasow, K.; Wittmann, J. G.; Dierks, T.; von Figura, K.; Rudolph, M. G. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 81–86.
14. The homology model was constructed using Modeller and protein database code 1Y4J.

JA804530W