Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific *In Vivo* Bioorthogonal Ligation with *trans*-Cyclooctenes

Jason L. Seitchik[1], **Jennifer C. Peeler**[1], **Michael T. Taylor**[2], **Melissa L. Blackman**[2], **Timothy W. Rhoads**[3], **Richard B. Cooley**[3], **Christian Refakis**[1], **Joseph M. Fox**[2], and **Ryan A. Mehl**[3,*]

[1]Franklin & Marshall College, Department of Chemistry, Lancaster, PA 17604-3003, USA

[2]Brown Laboratory, Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, USA

[3]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA

Abstract

Bioorthogonal ligation methods with improved reaction rates and less obtrusive components are needed for site-specifically labeling proteins without catalysts. Currently no general method exists for *in vivo* site-specific labeling of proteins that combines fast reaction rate with stable, nontoxic and chemoselective reagents. To overcome these limitations we have developed a tetrazine-containing amino acid, **1**, that is stable inside living cells. We have site-specifically genetically encoded this unique amino acid in response to an amber codon allowing a single **1** to be placed at any location in a protein. We have demonstrated that protein containing **1** can be ligated to a conformationally strained *trans*-cyclooctene, **2,** *in vitro* and *in vivo* with reaction rates significantly faster than most commonly used labeling methods.

Understanding protein networks and structural changes is key to revealing the secrets of biology. Currently available *in vitro* bioorthogonal ligation methods for site-specifically labeling proteins are limited by the reaction rates or bulk of reactive components. Also, no general method exists for *in vivo* site-specific labeling of proteins with a catalyst-free reaction or reactants that are stable, nontoxic and chemoselective.

A wide variety of functionalities have been genetically incorporated for the site-specific labeling of proteins, but these methods are not effective *in vivo* because they have modest coupling rates[1]. A robust but slow labeling reaction involves azides with unstrained alkynes and Cu^{+1} catalyst ($k_2 = 10^{-2} – 10^{-4}$ $M^{-1}s^{-1}$).[2] Strained, fluorinated alkynes amino acids can be used without Cu^{+1} but still have modest coupling rates ($k_2 = 1$ $M^{-1}s^{-1}$)[1a,1c,3]. Genetically encoded cyanobenzathiazole condensations and photo-click reactions have been developed with rate constants reaching 10 $M^{-1}s^{-1}$ but have side reactions *in vivo* or need short wavelengths of light.[4] A wide variety of reactions have also been developed but show varying levels of side reactions *in vivo* (using cyclooctynes, aldehydes, phosphine labeling reagents)[5]. Alternative genetic incorporation methods such as self-labeling tag sequences[6] and fluorescent protein fusions[7] are non-toxic and can exhibit fast coupling rates, but the components are often bulky and can rarely be moved to multiple locations within the labeled protein without compromising the function of the system of interest.[8] A bioorthogonal ligation reaction on proteins that is going to advance *in vivo* labeling will need i) components that are unreactive to cell culturing conditions, ii) to be rapid and quantitative, iii) to be encoded genetically for any site on the protein, iv) to result in a stable linkage

Corresponding author: ryan.mehl@oregonstate.edu.

facilitating labeling with diverse probes. In order to satisfy these labeling criteria and overcome current limitations we set out to develop a genetically encodable functionality based on the tetrazine-alkene coupling reaction.

Inverse electron-demand Diels-Alder reactions of *s*-tetrazines are venerable reactions in organic synthesis.[9] Recently, we introduced the bioorthogonal reaction between tetrazines and *trans*-cyclooctene derivatives,[10] a method of rapid bioconjugation which has been applied by a number of groups.[11] The foundation of this bioorthogonal reaction is a photochemical, flow-chemistry method that provided the first general and scalable route to functionalized *trans*-cyclooctenes.[12] These strained alkenes have allowed for tetrazines with reduced reactivities to be used that are more stable toward nucleophiles (ie water and thiols) but still yield fast coupling reactions (coupling rates with $k_2 > 1000$ $M^{-1}s^{-1}$).[10,11,13]

To develop an unnatural amino acid that could partner in the tetrazine ligation, one of the reactive partners needs to be attached to an amino acid, while the other needs to be attached to a useful label. The *trans*-cyclooctene functionality is stable under biological conditions and is relatively easy to functionalize.[11a,b,11d,11f] Accordingly, we set out to develop a tetrazine containing amino acid that is stable to cell culturing conditions but still sufficiently reactive with strained alkenes. The tetrazines that have been studied in conjugations with *trans*-cyclooctene, such as amido-substituted 3,6-di(2-pyridyl)-*s*-tetrazines and mono-aryl-*s*-tetrazines, are stable in biological milieu, but display slow background reactivity that would be incompatible with cellular growth conditions.[11c,13–14] Thus, we aimed to design a tetrazine derivative that would be stable in cells, 4-(6-methyl-*s*-tetrazin-3-yl)aminophenylalanine, **1**. Adding electron donating methyl and secondary amine substituents to the 3 and 6 position of the tetrazine should improve its stability against nucleophilic attack. The tetrazine-UAA **1** was synthesized in five steps from commercially available materials (see supporting information for details).

While the electron donating 3-amino substituent of **1** makes this tetrazine stable and soluble in biological conditions, it also decreases the rate of Diels-Alder reactions. Accordingly, we chose to combine this amino acid with derivatives of compound **2**, a strained *trans*-cyclooctene (s-TCO) that we recently developed.[13] With tetrazines, compound **2** displays relative rates that are significantly faster than our original system based on 5-hydroxy-*trans*-cyclooctene.[10]

We set out to evolve the *Methanococcus jannaschii* (*Mj*) tyrosyltRNA synthetase (RS)/tRNA$_{CUA}$ pair since it has previously been used to incorporated UAAs of similar structure in response to an amber codon.[15] To evolve the orthogonal *Mj*TyrRS/tRNA$_{CUA}$ pair capable of incorporating **1** in *E. coli*, we used a library of the synthetase (RS) gene that was randomized for the codons corresponding to six active-site residues within 7 Å of the bound tyrosine.[15] We performed two rounds of alternating positive and negative selection on this library (see supporting information for details). Plasmids from surviving clones were transformed into cells with a plasmid containing a GFP gene interrupted with an amber codon.[6g,16] 96 colonies were assessed for UAA-dependent expression of GFP. The top 20 performing clones showed greater than 300 mg/L of GFP-1 expression in the presence of **1** and no detectable GFP fluorescence over background in the absence of 1 (Sup. Fig. 2). Sequencing these 20 clones revealed one unique RS sequence contain the mutations (Y32E, L65A, A107E, F108P, Q109S, D158G, L162G).

To further characterize the incorporation of **1** into proteins in response to the amber codon, the selected RS was cloned into a *pDule* vector that contains one copy of *Mj* tRNA$_{CUA}$ to create *pDule-mtaF*.[6g,16–17] Expression of GFP gene interrupted by an amber codon at site 150 in the presence of *pDule-mtaF* was efficient and dependent on the presence of **1** (Fig.

1b). Using 1 mM **1**, 124 mg of GFP-1 was purified per liter of media, while GFP-wt yielded 220 mg/L under similar conditions (no GFP is produced in the absence of **1**). To further demonstrate that **1** can be incorporated into recombinant proteins using *pDule-mtaF*, we compared the masses of GFP-1 to GFP-wt using ESI-Q-Tof mass analysis. The native GFP-wt has the expected mass of 27827 ±1 Da and GFP-1 exhibits the expected mass increase to 27971±1 Da verifying that **1** is incorporated at single site (Sup. Fig. 2). The site of **1** incorporation was confirmed by analysis of the tandem mass spectrometry (MS/MS) fragmentation series of the relevant tryptic peptide (Fig. 1C).[18] Overall, the results of protein expression with affinity purification, SDS-PAGE, MS, and MS/MS analysis demonstrate the high fidelity and efficient incorporation of **1** at a genetically programmed site in GFP using *pDule-mtaF*. Since expressions contained **1** for 40 hrs in media at 37°C, this also confirms that stability of **1** to cell growth and protein purification conditions.

The conjugated pi systems of tetrazines cause them to absorb visible light between 510 and 550 nm, enabling them to quench many attached fluorescent probes.[5g] This quenching has proven useful because once the tetrazine undergoes cycloaddition with alkenes the conjugated system is removed and the coupled product ceases to quench nearby fluorophores. When **1** is encoded at site 150 in GFP, approximately 12 Å from the protein chromophore, the protein fluorescence is quenched 11 fold (Figure 2A,B). Incubating GFP-1 (3μM) with 39 μM **2** shows a complete return of fluorescence in less than 30 seconds indicating GFP-3 was formed. ESI-Q-Tof of the desalted reaction mixture confirmed the quantitative conversion of GFP-1 (expected 27971 Da; observed 27971±1 Da) into GFP-3 (expected 28094 Da; observed 28093 ±1 Da (Sup Figure 2).

It has been shown that tetrazine-alkene reaction rates are solvent dependent and are accelerated in polar solvents.[10] In order to determine this rate constant for this reaction at a specific site in a protein the increase in GFP-3 fluorescence was followed in a PBS buffer at pH 7, 22°C. The kinetics of the reaction were performed under pseudo-first-order conditions as verified by a single exponential fit for product formation ([GFP] was 0.10μM and [**2**] ranged from 27.0 to 6.8 μM). The *in vitro* second order rate constant for GFP-1 with **2** is 880±10 $M^{-1}s^{-1}$ (Figure 3A). This verifies that this site-specific tetrazine-strained cyclooctene reaction is orders of magnitude faster than established site-specific labeling methods.

Thus far, tetrazine-alkene coupling reactions on engineered proteins have only been characterized *in vitro* and on cell surfaces. While it is possible to genetically encode ketone, azide and alkyne functionality site-specifically on amino acids, little labeling has been done on these engineered proteins *in vivo* with these amino acids. To determine if this bioorthogonal reaction is rapid and quantitative *in vivo*, *E. coli* cells containing expressed GFP-1 were incubated with 0.1 mM **2** in PBS buffer at room temperature. As expected, the fluorescence returned in less than a minute, indicating that GFP-3 had been formed. The cells were cooled and washed twice with PBS at 0°C to remove unreacted **2**. ESI-Q-Tof MS of affinity purified his-tag protein resulted in a molecular mass matching only GFP-3 (Fig 2C). This verifies that **2** can be taken up by *E. coli* cells, the reaction is facile and the conjugated product is stable *in vivo*. This demonstrates the first site-specific *in vivo* reaction for this type of bioorthogonal ligation.

In vivo site-specific reaction rate constants have yet to be determined for bioorthogonal reactions. In order to determine the rate constant for this reaction inside living cells, *E. coli* expressing GFP-1 was pelleted and resus-pended in PBS buffer. *In vivo* kinetic studies were performed similarly to the *in vitro* studies and the five pseu-do-first-order rate constants were used to determine the bimolecular rate constant. The *in vivo* second order rate constant for GFP-1 with **2** is 330±20 $M^{-1}s^{-1}$ (Figure 3B). Since identical conditions were used for *in*

vitro and *in vivo* kinetic measurements, the 60% rate constant reduction is most likely a result of the *in vivo* cellular environment or hampered uptake of **2**. Limiting cellular uptake of **2** would result in non-exponential kinetics for the formation of GFP-3, which was not observed (Figure 3B). Also, the rate constant for the overall process measured in cells is comparable to the *in vitro* rate constants indicating the uptake step cannot be significantly slower than the reaction step. Molecular crowding or electrostatic changes induced by the internal cellular environment could account for the rate change. The *in vivo* coupling reaction was also verified in rich media and on live cells. No toxicity was evident as judge by comparing cell viability of cells harboring the sTCO reagent to those that did not (Sup. Fig. 5). While the *in viv*o reaction is slower than the *in vitro*, the coupling rate is still sufficient to label proteins in their native environment.

To verify that this labeling strategy is general, **2** was conjugated to a diacetyl-fluorescein dye forming **4** (Sup Scheme 2). The fluorescently labeled trans-cyclooctene, **4,** was then reacted with GFP-1, similarly to reactions with **2**. Quantitative labeling of GFP-1 would result in GFP-5 mass of 28653 Da (Figure 2D), but in aqueous conditions deacetylated products would be expected. While the fluorescein-labeled protein ionized poorly via ESI-MS, MALDI MS was used to confirm labeling (expected 28653 Da; observed 28635 ± 100 Da). This confirms that protein-containing **1** is reactive with labeled strained trans-cyclooctenes.

In conclusion, we report a general method for the site-specific incorporation of a novel, stable tetrazine containing amino acid, **1,** into recombinant proteins in *E. coli*. We demonstrated that protein containing **1** can be quantitatively labeled with strained *trans*-cyclooctenes and that the resulting ligation product is stable to *in vivo* cellular conditions. We determined the *in vitro* and *in vivo* rate constants for this bioorthogonal coupling to be significantly faster than currently used site-specific labeling reactions.[1a,2–4] The speed of this catalyst-free reaction and the stability of reacting components and ligation product will make it an advantageous tool for studying complex protein networks and conformational changes *in vivo* and *in vitro*. Our future efforts will focus on developing this reaction for use in labeling proteins in live yeast and mammalian cells, and engineering protein interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. (a) Agard NJ, Baskin JM, Prescher JA, Lo A, Bertozzi CR. ACS Chem Biol. 2006; 1:644. [PubMed: 17175580] (b) Sletten EM, Bertozzi CR. Acc Chem Res. 2011; 44:666. [PubMed: 21838330] (c) Debets MF, van Berkel SS, Dommerholt J, Dirks AT, Rutjes FP, van Delft FL. Acc Chem Res. 2011; 44:805. [PubMed: 21766804]

2. (a) Hao Z, Song Y, Lin S, Yang M, Liang Y, Wang J, Chen PR. Chem Commun (Camb). 2011; 47:4502. [PubMed: 21387054] (b) Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A,

Chin JW. J Am Chem Soc. 2009; 131:8720. [PubMed: 19514718] (c) Fekner T, Li X, Lee MM, Chan MK. Angew Chem Int Ed Engl. 2009; 48:1633. [PubMed: 19156778] (d) Deiters A, Schultz PG. Bioorg Med Chem Lett. 2005; 15:1521. [PubMed: 15713420] (e) Deiters A, Cropp TA, Mukherji M, Chin JW, Anderson JC, Schultz PG. J Am Chem Soc. 2003; 125:11782. [PubMed: 14505376]

3. Plass T, Milles S, Koehler C, Schultz C, Lemke EA. Angew Chem Int Ed Engl. 2011; 50:3878. [PubMed: 21433234]

4. (a) Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW. J Am Chem Soc. 2011; 133:11418. [PubMed: 21736333] (b) Wang J, Zhang W, Song W, Wang Y, Yu Z, Li J, Wu M, Wang L, Zang J, Lin Q. J Am Chem Soc. 2010; 132:14812. [PubMed: 20919707] (c) Wang Y, Song W, Hu WJ, Lin Q. Angew Chem Int Ed Engl. 2009; 48:5330. [PubMed: 19544336]

5. (a) Carrico ZM, Romanini DW, Mehl RA, Francis MB. Chem Commun (Camb). 2008:1205. [PubMed: 18309418] (b) Mehl RA, Anderson JC, Santoro SW, Wang L, Martin AB, King DS, Horn DM, Schultz PG. J Am Chem Soc. 2003; 125:935. [PubMed: 12537491] (c) Wang L, Zhang Z, Brock A, Schultz PG. Proc Natl Acad Sci U S A. 2003; 100:56. [PubMed: 12518054] (d) Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, Lo A, Codelli JA, Bertozzi CR. Proc Natl Acad Sci U S A. 2007; 104:16793. [PubMed: 17942682] (e) Sletten EM, Bertozzi CR. Angew Chem Int Ed Engl. 2009; 48:6974. [PubMed: 19714693] (f) Hao Z, Hong S, Chen X, Chen PR. Acc Chem Res. 2011; 44:742. [PubMed: 21634380] (g) Devaraj NK, Weissleder R. Acc Chem Res. 2011; 44:816. [PubMed: 21627112]

6. (a) Fernandez-Suarez M, Baruah H, Martinez-Hernandez L, Xie KT, Baskin JM, Bertozzi CR, Ting AY. Nat Biotechnol. 2007; 25:1483. [PubMed: 18059260] (b) Griffin BA, Adams SR, Tsien RY. Science. 1998; 281:269. [PubMed: 9657724] (c) Halo TL, Appelbaum J, Hobert EM, Balkin DM, Schepartz A. J Am Chem Soc. 2009; 131:438. [PubMed: 19105691] (d) Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K. Nat Biotechnol. 2003; 21:86. [PubMed: 12469133] (e) Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV. ACS Chem Biol. 2008; 3:373. [PubMed: 18533659] (f) Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL. Nat Chem Biol. 2007; 3:707. [PubMed: 17891153] (g) Stokes AL, Miyake-Stoner SJ, Peeler JC, Nguyen DP, Hammer RP, Mehl RA. Molecular Biosystems. 2009; 5:1032. [PubMed: 19668869] (h) Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT. Proc Natl Acad Sci U S A. 2005; 102:15815. [PubMed: 16236721]

7. (a) Giepmans BN, Adams SR, Ellisman MH, Tsien RY. Science. 2006; 312:217. [PubMed: 16614209] (b) Shaner NC, Steinbach PA, Tsien RY. Nat Methods. 2005; 2:905. [PubMed: 16299475]

8. Hinner MJ, Johnsson K. Curr Opin Biotechnol. 2010; 21:766. [PubMed: 21030243]

9. (a) Boger DL. Chem Rev. 1986; 86:781.(b) Thalhammer F, Wallfahrer U, Sauer J. Tetrahedron Lett. 1990; 31:6851.

10. Blackman ML, Royzen M, Fox JM. J Am Chem Soc. 2008:13518. [PubMed: 18798613]

11. (a) Li ZB, Cai HC, Hassink M, Blackman ML, Brown RCD, Conti PS, Fox JM. Chem Commun. 2010; 46:8043.(b) Selvaraj R, Liu S, Hassink M, Huang C-w, Yap L-p, Park R, Fox JM, Li Z, Conti PS. Bioorganic and Medicinal Chemistry Letters. 2011; 21:5011. [PubMed: 21601452] (c) Rossin R, Verkerk PR, van den Bosch SM, Vulders RCM, Verel I, Lub J, Robillard MS. Angew Chem Int Edit. 2010; 49:3375.(d) Devaraj NK, Hilderbrand S, Upadhyay R, Mazitschek R, Weissleder R. Angew Chem Int Edit. 2010; 49:2869.(e) Devaraj NK, Upadhyay R, Hatin JB, Hilderbrand SA, Weissleder R. Angew Chem Int Edit. 2009; 48:7013.(f) Reiner T, Keliher EJ, Earley S, Marinelli B, Weissleder R. Angew Chem Int Ed. 2011; 50:1922.

12. Royzen M, Yap GPA, Fox JM. J Am Chem Soc. 2008; 130:3760. [PubMed: 18321114]

13. Taylor MT, Blackman ML, Dmitrenko O, Fox JM. J Am Chem Soc. 2011; 133:9646. [PubMed: 21599005]

14. Karver MR, Weissleder R, Hilderbrand SA. Bioconj Chem. 2011; 2:2263.

15. Xie J, Schultz PG. Methods. 2005; 36:227. [PubMed: 16076448]

16. Miyake-Stoner SJ, Refakis CA, Hammill JT, Lusic H, Hazen JL, Deiters A, Mehl RA. Biochemistry. 2010; 49:1667. [PubMed: 20082521]

17. Miyake-Stoner SJ, Miller AM, Hammill JT, Peeler JC, Hess KR, Mehl RA, Brewer SH. Biochemistry. 2009; 48:5953. [PubMed: 19492814]

18. (a) Xu H, Freitas MA. BMC Bioinformatics. 2007; 8:133. [PubMed: 17448237] (b) Xu H, Freitas MA. J Proteome Res. 2008; 7:2605. [PubMed: 18543962] (c) Xu H, Yang L, Freitas MA. BMC Bioinformatics. 2008; 9:347. [PubMed: 18713471]
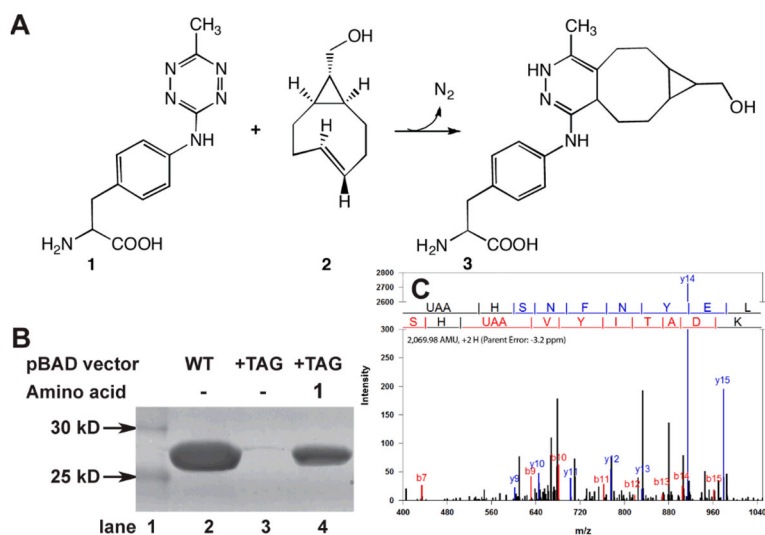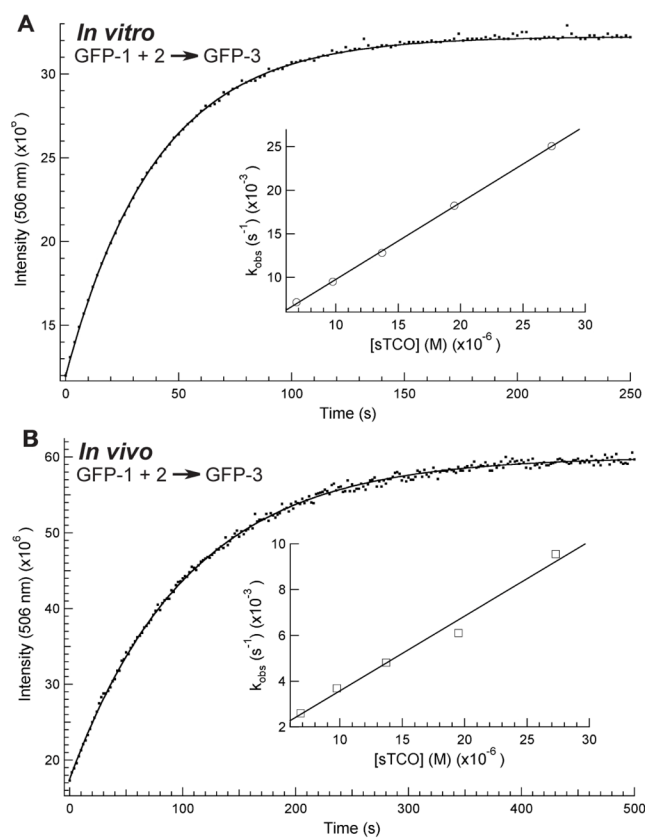
Figure 1.
Genetic incorporation of 4-(6-methyl-*s*-tetrazin-3-yl)aminophenylalanine, **1,** into proteins. (A) Amino acid **1** reacts with sTCO, **2**, then loses nitrogen gas to form in the stable conjugate **3**. (B) The evolved *Mj*RS/tRNA$_{CUA}$ pair in *pDule-mtaF* allows for site-specific incorporation of **1** in response to an amber codon. Lane 2 shows expression levels of GFP-wt from pBad-GFP-His$_6$. Production of GFP-1 from pBad-GFP-150TAG-His$_6$ is dependent on **1** in the growth media, lane 3 without **1** present, lane 4 with 1 mM **1** present. Protein was purified by Co^{+2} affinity chromatography, separated by SDS-PAGE and stained with Coomassie. (C) MS/MS fragmentation of tryptic peptides derived from GFP-1 samples demonstrates the efficient high-fidelity incorporation of a single **1** UAA in response to an amber stop codon using the mtaF synthetase. The spectra confirm **1** incorporation at codon 150. The fragmentation sites are illustrated above the spectrum.
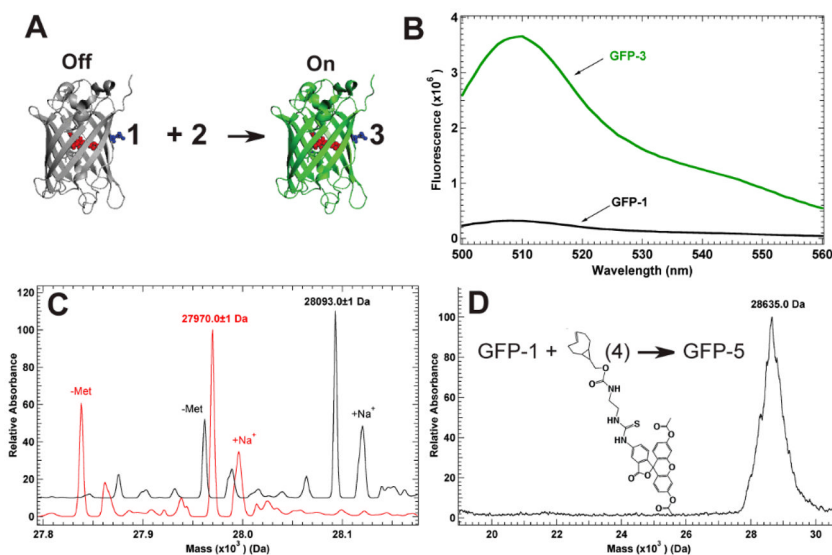
Figure 2.
Characterization of tetrazine transcyclooctene reaction on GFP. (A) Quenched GFP-1 is "turned on" by coupling with **2**, forming fluorescence GFP-3. (B) Excitation at 488 nm produces low flourescene for GFP-1, while the reaction forming GFP-3 produces full fluorescence for GFP. (C) ESI-Q-Tof MS analysis of product from GFP-1 with **2** demonstrates specific and quantitative labeling of GFP-1. (D) MALDI MS analysis of GFP-1 reaction with diacetyl-fluorescein labeled sTCO.

Figure 3.
Rate constant determination for *in vitro* and *in vivo* reaction of GFP-1 and **2**. (A) *In vitro* kinetics resulted in rate k= 880±10 $M^{-1}s^{-1}$ (B) *In vivo* kinetics resulted in k= 330±20 $M^{-1}s^{-1}$. For both experiments unimolecular rate constants were calculated by fitting the rate of GFP-3 formation to a single exponential at five different concentrations of **2**. Inset is bimolecular rate constant determination using unimolecular $k_{obs}=k[\mathbf{2}]$.