J|A|C|S
ARTICLES

Published on Web 10/04/2010

A Biosynthetic Route to Photoclick Chemistry on Proteins

Jiangyun Wang,*,† Wei Zhang,† Wenjiao Song,‡ Yizhong Wang,‡ Zhipeng Yu,‡ Jiasong Li,§ Minhao Wu,§ Lin Wang,§ Jianye Zang,§ and Qing Lin*,‡

National Key Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260, United States, and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, 230026, China

Received May 21, 2010; E-mail: qinglin@buffalo.edu; jwang@ibp.ac.cn

Abstract: Light-induced chemical reactions exist in nature, regulating many important cellular and organismal functions, e.g., photosensing in prokaryotes and vision formation in mammals. Here, we report the genetic incorporation of a photoreactive unnatural amino acid, *p*-(2-tetrazole)phenylalanine (*p*-Tpa), into myoglobin site-specifically in *E. coli* by evolving an orthogonal tRNA/aminoacyl-tRNA synthetase pair and the use of *p*-Tpa as a bioorthogonal chemical "handle" for fluorescent labeling of *p*-Tpa-encoded myoglobin via the photoclick reaction. Moreover, we elucidated the structural basis for the biosynthetic incorporation of *p*-Tpa into proteins by solving the X-ray structure of *p*-Tpa-specific aminoacyl-tRNA synthetase in complex with *p*-Tpa. The genetic encoding of this photoreactive amino acid should make it possible in the future to photoregulate protein function in living systems.

Introduction

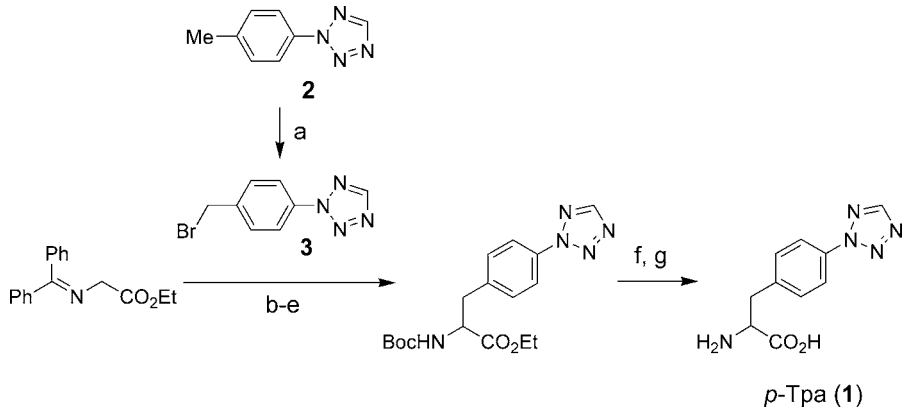
Light regulates many important cellular and organismal functions such as vision and phototropic response through specific light-induced chemical reactions. For example, prokaryotes such as *Brucella melitensis*, *Brucella abortus*, *Erythrobacter litoralis*, and *Pseudomonas syringae* contain light-activated histidine kinases that function as light-sensory modules in plant and algal phototropins and in fungal blue-light receptors. These kinases undergo a specific, photoinduced flavin addition to a cysteine group on the kinase in response to light, which in turn regulates bacterial gene expression, chemotaxis, phototaxis, and virulence.[1] Similarly, in mammals, vision formation involves a photoinduced isomerization of the 11-*cis*-retinyl Schiff base in rhodopsin.[2] In efforts to harness the power of light to image or regulate cell signaling, natural light-sensitive proteins such as rhodopsin and phytochrome have been genetically fused with the target protein in order to achieve light-dependent regulation with a high spatiotemporal precision in live cells and in mice.[3] However, fusion of large photoreceptors to proteins may significantly perturb the target protein function. To overcome this limitation, an alternative approach is to incorporate a photoreactive amino acid into the protein site-selectively using the amber codon suppression technique,[4] allowing subsequent photochemical regulation. To this end, nonnatural photoreactive amino acids containing aryl ketone,[5] aryl azide,[6] and aryl diazirine[7] groups have been successfully incorporated site-specifically into proteins. While these photoreactive groups have provided invaluable tools for protein photoregulation,[8] the photogenerated reactive intermediates tend to insert into the closest C−H (or X−H) bonds nonselectively, preventing wide application in biological studies.

† Chinese Academy of Sciences.
‡ State University of New York at Buffalo.
§ University of Science and Technology of China.

(1) (a) Swartz, T. E.; Tseng, T.-S.; Frederickson, M. A.; Paris, G.; Comerci, D. J.; Rajashekara, G.; Kim, J.-G.; Mudgett, M. B.; Splitter, G. A.; Ugalde, R. A.; Goldbaum, F. A.; Briggs, W. R.; Bogomolni, R. A. *Science* **2007**, *317*, 1090–1093.
(2) (a) Gartner, W. *Angew. Chem., Int. Ed.* **2001**, *40*, 2977–2981. (b) Borhan, B.; Souto, M. L.; Imai, H.; Shichida, Y.; Nakanishi, K. *Science* **2000**, *288*, 2209–2212. (c) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739–745.
(3) (a) Airan, R. D.; Thompson, K. R.; Fenno, L. E.; Bernstein, H.; Deisseroth, K. *Nature* **2009**, *458*, 1025–1029. (b) Levskaya, A.; Weiner, O. D.; Lim, W. A.; Voigt, C. A. *Nature* **2009**, *461*, 997–1001. (c) Shu, X.; Royant, A.; Lin, M. Z.; Aguilera, T. A.; Lev-Ram, V.; Steinbach, P. A.; Tsien, R. Y. *Science* **2009**, *324*, 804–807. (d) Gradinaru, V.; Zhang, F.; Ramakrishnan, C.; Mattis, J.; Prakash, R.; Diester, I.; Goshen, I.; Thompson, K. R.; Deisseroth, K. *Cell* **2010**, *141*, 154–165.
(4) (a) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249. (b) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775–782. (c) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
(5) (a) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020–11024. (b) Mori, H.; Ito, K. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16159–16164. (c) Ye, S.; Köhrer, C.; Huber, T.; Kazmi, M.; Sachdev, P.; Yan, E. C.; Bhagat, A.; RajBhandary, U. L.; Sakmar, T. P. *J. Biol. Chem.* **2008**, *283*, 1525–1533.
(6) (a) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027. (b) Carrico, I. S.; Maskarinec, S. A.; Heilshorn, S. C.; Mock, M. L.; Liu, J. C.; Nowatzki, P. J.; Franck, C.; Ravichandran, G.; Tirrell, D. A. *J. Am. Chem. Soc.* **2007**, *129*, 4874–4875.
(7) (a) Suchanek, M.; Radzikowska, A.; Thiele, C. *Nat. Methods* **2005**, *2*, 261–268. (b) MacKinnon, A. L.; Garrison, J. L.; Hegde, R. S.; Taunton, J. *J. Am. Chem. Soc.* **2007**, *129*, 14560–14561. (c) Tippmann, E. M.; Liu, W.; Summerer, D.; Mack, A. V.; Schultz, P. G. *Chembiochem.* **2007**, *8*, 2210–2214.
(8) For a recent example, see: Baruah, H.; Puthenveetil, S.; Choi, Y.-A.; Shah, S.; Ting, A. Y. *Angew. Chem., Int. Ed.* **2008**, *47*, 7018–7021.

 10.1021/ja104350y

Scheme 1. Synthesis of *p*-(2-Tetrazole)phenylalanine (**1**)[a]



[a] Conditions: (a) NBS, BPO, CCl_4, reflux, 61%; (b) LiHMDS/THF, −78 °C; (c) **3**, THF, −78 °C to r.t.; (d) 2 N HCl, MeOH; (e) $(Boc)_2O$, TEA, THF; 69% for steps b−e; (f) NaOH, EtOH/H_2O; (g) TFA/DCM; 95% for steps f and g.

We have recently reported aryl tetrazoles as a new class of photoreactive groups that undergo selective photoinduced cycloaddition reactions with electron-deficient alkenes as well as terminal unactivated alkenes (termed "photoclick chemistry") in biological systems.[9] Since our earlier efforts relied on native chemical ligation to introduce aryl tetrazoles into the target protein site-selectively, the resulting tetrazole-containing proteins needed to be microinjected into cells in order to study the photochemical regulation of protein function,[10] severely limiting the utility of these photoreactive aryl tetrazole groups. One way to circumvent this limitation is to genetically encode the cognate reaction partner alkene into proteins site-selectively[11] and use aryl tetrazoles as reagents in the photoclick chemistry for photochemical modulation or visualization of the alkene-encoded proteins. Indeed, we recently demonstrated that an alkene amino acid, homoallylglycine, can serve as a metabolic reporter for spatiotemporally controlled imaging of newly synthesized proteins in live mammalian cells.[12] However, because of fast diffusion of the reactive small-molecule aryl tetrazoles in cells, it would be more desirable to encode these aryl tetrazole moieties directly into the macromolecular protein target in order to facilitate photochemical regulation of the targeted protein with a higher spatiotemporal precision. To this end, herein we report the first genetic incorporation of a tetrazole amino acid, *p*-(2-tetrazole)phenylalanine (*p*-Tpa), site-specifically into a protein in *E. coli*, the structural basis for selective recognition of this photoreactive amino acid by its cognate aminoacyl-tRNA synthetase, and the performance of a site-specific photoclick reaction on proteins.

Results and Discussion

Synthesis of Photoreactive *p*-(2-Tetrazole)phenylalanine. We began our investigation with *p*-(2-tetrazole)phenylalanine (*p*-Tpa) because of its small size and close resemblance to tyrosine. The synthesis of *p*-Tpa was carried out according to Scheme 1.[13] Briefly, *p*-(2-tetrazole)benzyl bromide (**3**) was obtained in 61% yield by refluxing 2-tolyltetrazole (**2**) with *N*-bromosuccinimide (NBS) and benzoyl peroxide (BPO) in carbon tetrachloride overnight. The Boc-protected ethyl ester of tetrazole amino acid was obtained in four steps with the key step involving the alkylation of glycine Schiff base with an excess amount of **3** for an overall yield of 69% (Scheme 1). Subsequent removal of the protecting groups afforded *p*-(2-tetrazole)phenylalanine (**1**) in 95% yield over two steps.

Selection of *p*-Tpa-Specific Aminoacyl-tRNA Synthetase. To selectively incorporate *p*-Tpa at a defined site in proteins in *E. coli*, an orthogonal tRNA/aminoacyl-tRNA synthetase pair that uniquely specifies *p*-Tpa in response to TAG amber codon needs to be developed. Since a large number of nonnatural amino acids carrying large hydrophobic groups at the para position of tyrosine have been successfully incorporated site-specifically into proteins in vivo using the orthogonal tRNA/aminoacyl-tRNA synthetase pairs derived from the *Methanococcus jannaschii* amber suppressor tyrosyl-tRNA (*Mj*$tRNA^{Tyr}_{CUA}$)/tyrosyl-tRNA synthetase (*Mj*TyrRS) pair,[4c] we decided to evolve the *p*-Tpa-specific aminoacyl-tRNA synthetase using the *Mj*$tRNA^{Tyr}_{CUA}$/*Mj*TyrRS system. An *Mj*TyrRS library, pBK-lib-jw1, was constructed based on a previously reported *Mj*TyrRS library[14] in which six residues (Tyr32, Leu65, Phe108, Gln109, Asp158, and Leu162) were randomized through an overlapping extension PCR using synthetic oligonucleotide primers in which the intended randomization sites was encoded by NNK (N = A+T+C+G; K = T+G). The resulting library was subjected to a second round of randomization during which any one of the six residues (Ile63, Ala67, His70, Tyr114, Ile159, Val164) is either mutated to Gly (see Table S1 in the Supporting Information for the mutagenesis primers) or kept unchanged to

(9) (a) Lim, R. K.; Lin, Q. *Chem. Commun.* **2010**, *46*, 1589–1600. (b) Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin, Q. *Angew. Chem., Int. Ed.* **2008**, *47*, 2832–2835. (c) Song, W.; Wang, Y.; Qu, J.; Lin, Q. *J. Am. Chem. Soc.* **2008**, *130*, 9654–9655. (d) Wang, Y.; Song, W.; Hu, W. J.; Lin, Q. *Angew. Chem., Int. Ed.* **2009**, *48*, 5330–5333.
(10) Song, W.; Yu, Z.; Madden, M. M.; Lin, Q. *Mol. Biosyst.* **2010**, *6*, 1576−1578.
(11) (a) van Hest, J. C.; Kiick, K. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 1282–1288. (b) Zhang, Z.; Wang, L.; Brock, A.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2840–2842. (c) Wang, J.; Schiller, S. M.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2007**, *46*, 6849–6851. (d) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187–1197. (e) Ai, H.-w.; Shen, W.; Brustad, E.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2010**, *49*, 935–937.
(12) Song, W.; Wang, Y.; Yu, Z.; Rivera Cera, C. I.; Qu, J.; Lin, Q. *ACS Chem. Biol.* **2010**, *5*, 875–885.
(13) Wang, Y.; Lin, Q. *Org. Lett.* **2009**, *11*, 3570–3573.
(14) Xie, J.; Liu, W. S.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2007**, *46*, 9239–9242.

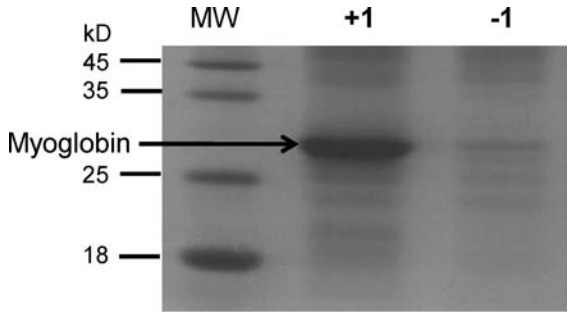


Figure 1. Coomassie blue-stained SDS-PAGE gel showing expression of the TAG4 mutant myoglobin in the presence and absence of 1 mM *p*-Tpa (**1**).

generate the pBK-lib-jw1 library. This library was then subjected to both positive and negative selections. In the positive selection, cell survival was dependent on the suppression of an amber codon introduced at a permissive site in the chloramphenicol acetyl transferase (CAT) gene when cells cotransformed with the pBK-lib-jw1 library and a mutant *Mj*tRNA$^{Tyr}_{CUA}$were grown in the presence of 1 mM *p*-Tpa and 80 mg/L chloramphenicol. Positively selected clones were then transformed into cells containing *Mj*tRNA$^{Tyr}_{CUA}$and a gene encoding the toxic barnase protein with three amber mutations introduced at the permissive sites. These cells were grown in the absence of *p*-Tpa to remove any clones that utilize endogenous amino acids (negative selection). Three rounds of positive and two rounds of negative selections of the pBK-lib-jw1 library produced one aminoacyl-tRNA synthetase clone, *p*-TpaRS-1, that grew at 120 μg/mL of chloramphenicol in the presence of 1 mM *p*-Tpa, but only at 40 μg/mL chloramphenicol in the absence of *p*-Tpa. Sequencing of this clone revealed the following six mutations: Tyr32Leu, Leu65Ile, Gln109Met, Asp158Gly, Leu162Val, and Val164Gly. Invariably, these mutations represent substitution of larger amino acids by smaller ones, thus creating additional space to accommodate the large tetrazole group. The majority of mutations (five out of six) were apparently derived from the first round of randomization, as these six residues make primary contact with a bound tyrosine in the active site of *Mj*TyrRS (Figure S1, Supporting Information).

In Vivo Incorporation of *p*-Tpa into Myoglobin in *E. coli.* To examine whether the evolved amber suppressor *Mj*tRNA$^{Tyr}_{CUA}$/*p*-TpaRS-1 pair charges *p*-Tpa efficiently in *E. coli*, we substituted the Ser-4 codon with the amber codon TAG in sperm whale myoglobin carrying a C-terminal His_6 tag. The expression of myoglobin was then carried out in *E. coli* transformed with *Mj*tRNA$^{Tyr}_{CUA}$and *p*-TpaRS-1, and the transformants were allowed to grow in either minimal medium or 2YT medium supplemented with 1 mM *p*-Tpa. As a control, expression of myoglobin was also carried out in the absence of *p*-Tpa. The His_6-tagged myoglobin proteins were purified with a Ni-NTA column and then analyzed by SDS-polyacrylamide gel electrophoresis (Figure 1). We found that the full-length myoglobin was obtained only in the presence of **1** (compare lane 2 to lane 1 in Figure 1), indicating that *p*-Tpa incorporation is highly specific. The yield of the *p*-Tpa-encoded mutant myoglobin in either medium was approximately 2 mg/L. To demonstrate that only *p*-Tpa was inserted into the myoglobin protein in response to the TAG codon, we analyzed the *p*-Tpa-encoded mutant myoglobin by ESI-TOF mass spectrometry and found that mutant myoglobin gave an observed mass of 18481.0 Da (Figure 2), matching very well with the calculated mass of 18482 Da. As a control, wild-type myoglobin showed an intact mass of 18354.0 Da in the same analysis, matching perfectly with the calculated mass of 18354 Da (Figure S2, Supporting Information). Importantly, no peaks corresponding to incorporation of any of the 20 natural amino acids into the myoglobin Ser-4 site were observed (the small peak with a mass of 18349 Da corresponds to *p*-Tpa-substituted myoglobin with the first Met deleted), indicating a high fidelity of biosynthetic incorporation of *p*-Tpa by the selected *Mj*tRNA$^{Tyr}_{CUA}$/*p*-TpaRS-1 pair.

Functionalization of *p*-Tpa-Encoded Myoglobin in Vitro. To assess whether *p*-Tpa can serve as a bioorthogonal reporter for photoclick chemistry, we first investigated the reactivity of 2-tolyltetrazole (**2**), a simple analogue of *p*-Tpa, in a cycloaddition reaction with an electron-deficient dipolarophile, dimethyl fumarate. We found that 2-tolyltetrazole reacts efficiently with dimethyl fumarate, with a cycloaddition rate constant (k_2) of $0.082 \pm 0.011\ M^{-1}\ s^{-1}$ in ACN/PBS buffer (1:1) (Figure 3 and Table 1). Compared to the diaryltetrazoles,[9b] the cycloaddition rate constant for monoaryl tetrazole **2** appears to decrease by about 100-fold, presumably due to lower HOMO energy of the photogenerated nitrile imine intermediate.[9d] We then prepared an FITC-modified fumarate (**5**) and examined the ability of compound **5** to selectively label the *p*-Tpa-encoded myoglobin via photoclick chemistry (Figure 4a). Briefly, we incubated FITC-fumarate **5** with either mutant myoglobin or wild-type myoglobin in PBS buffer, and the mixtures were subjected to 5-min photoirradiation using a handheld 254-nm UV lamp followed by incubation at room temperature for additional 1.5 h. The reaction mixtures were then analyzed by SDS-PAGE and in-gel fluorescence analysis. We found that only the *p*-Tpa-encoded myoglobin showed fluorescent labeling (Figure 4b), indicating that the reaction is selective toward the tetrazole moiety. To establish that the fluorescent labeling requires photoactivation, we incubated FITC-fumarate **5** with the *p*-Tpa-encoded myoglobin for 1.5 h without prior photoirradiation. As expected, no fluorescent band was detected (Figure 4c), confirming that the photoclick reaction mediates the fluorescent labeling.

Since FITC-fumarate can be used to fluorescently label the *p*-Tpa-containing proteins, we sought to use this technique to probe the potential mis-incorporation of *p*-Tpa into wild-type myoglobin as well as other cellular proteins in *E. coli*. To this end, BL21 cells transformed with the wt-myoglobin expression plasmid were allowed to grow in either GMML or LB medium supplemented with 1 mM *p*-Tpa. Then, the purified wild-type myoglobin and the crude cell lysates were subjected to the photoclick reactions with FITC-fumarate, and the reaction mixtures were analyzed by SDS-PAGE and in-gel fluorescence. No fluorescent bands were detected (Figure S3, Supporting Information), suggesting there was no misincorporation of *p*-Tpa proteins into cellular proteins in the absence of the *Mj*tRNA$^{Tyr}_{CUA}$/*p*-TpaRS-1 pair.

X-ray Structure of the *p*-TpaRS-1:*p*-Tpa Complex. To gain structural insights into the recognition of *p*-Tpa by the evolved *p*-TpaRS-1, we solved the crystal structure of *p*-TpaRS-1 bound to *p*-Tpa to a resolution of 2.50 Å (Figure 5a; see Table S2 in the Supporting Information for details of data collection and refinement statistics). As expected, the evolved p-TpaRS1 forms a homodimer[15] in the crystal, with the two subunits of the homodimer related by a crystallographic 2-fold symmetry axis (Figure S4a, Supporting Information). The backbone conformation of *p*-TpaRS-1 superimposes well with that of wild-type

(15) Steer, B. A.; Schimmel, P. *J. Biol. Chem.* **1999**, *274*, 35601–35606.

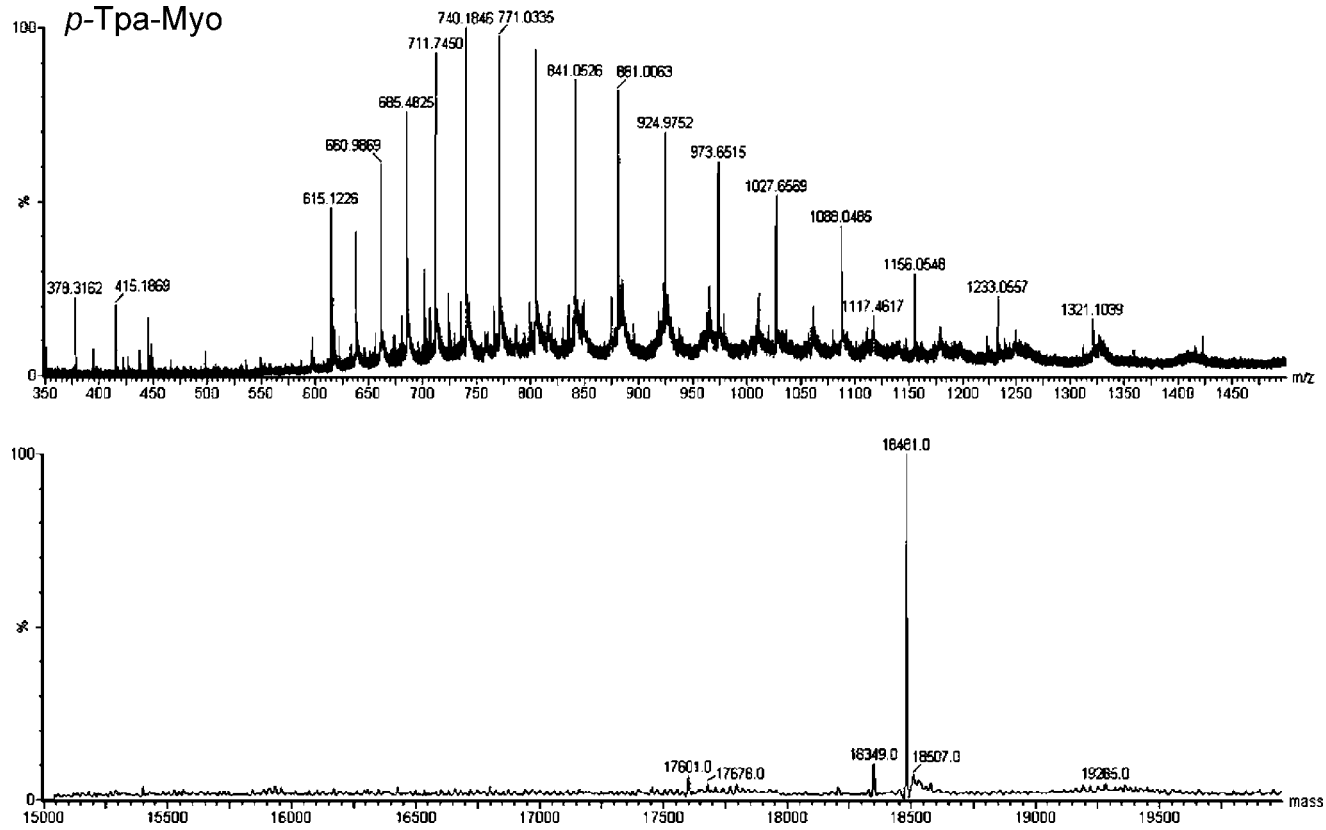


Figure 2. ESI-TOF mass spectrometry analysis of *p*-Tpa-encoded mutant myoglobin. The charge ladder (top) and the deconvoluted intact mass (bottom) were shown. Calculated intact mass = 18482 Da, found 18481.0 Da. The small peak with the intact mass of 18349.0 corresponds to the *p*-Tpa-encoded mutant myoglobin with the first Met deleted.

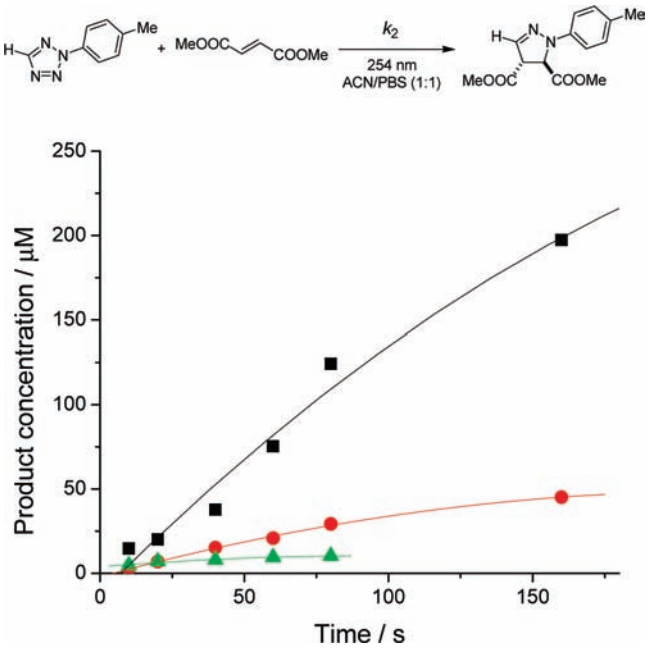


Figure 3. Time courses of the pyrazoline cycloadduct formation for the photoinduced cycloaddition reactions between 2-tolyltetrazole and dimethyl fumarate at three different concentrations: ■ = 1 mM 2-tolyltetrazole and 20 mM dimethyl fumarate; ● = 200 μM 2-tolyltetrazole and 20 mM dimethyl fumarate; ▲ = 200 μM 2-tolyltetrazole and 4 mM dimethyl fumarate.

Table 1. Determination of the Kinetic Constant of the Photoinduced Cycloaddition Reaction between 2-Tolyltetrazole and Dimethyl Fumarate

2-tolyltetrazole:dimethyl fumarate	k_{obs}(s^{-1})	k_2 (M^{-1} s^{-1})	average $k_2 \pm$ SD (M^{-1} s^{-1})
1000 μM:20 mM	1.44×10^{-3}	7.20×10^{-2}	0.082 ± 0.011
200 μM:20 mM	1.61×10^{-3}	8.05×10^{-2}	
200 μM:4 mM	3.77×10^{-4}	9.43×10^{-2}	

*Mj*TyrRS[16] (see Figure S4b in the Supporting Information for structural overlay). However, the *p*-TpaRS-1 active site has been significantly enlarged to favor the binding of *p*-Tpa relative to tyrosine (Figure 5b). Notably, the two aryl rings in *p*-Tpa adopt a coplanar geometry with a dihedral angle of 1.1°, in excellent agreement with our previous small-molecule crystallographic study.[17] Among the six identified mutations, the Tyr32Leu and Asp158Gly mutations appear to be critical in recognizing *p*-Tpa as a substrate, as these two residues in the wild-type *Mj*TyrRS are engaged in the hydrogen bonding with the tyrosine hydroxyl group and occupy the tetrazole space (compare Figure S1 to Figure S4c). On the other hand, while not making direct contact with *p*-Tpa, the Val164Gly mutation acquired during the second round of mutation converts the GVD loop, which is located close to the critical *p*-Tpa contact helix α8 where two important mutations Asp158Gly and Leu162Val reside,[16] into a more

(16) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat. Struct. Biol.* **2003**, *10*, 425–432.
(17) Wang, Y.; Hu, W. J.; Song, W.; Lim, R. K.; Lin, Q. *Org. Lett.* **2008**, *10*, 3725–3728.

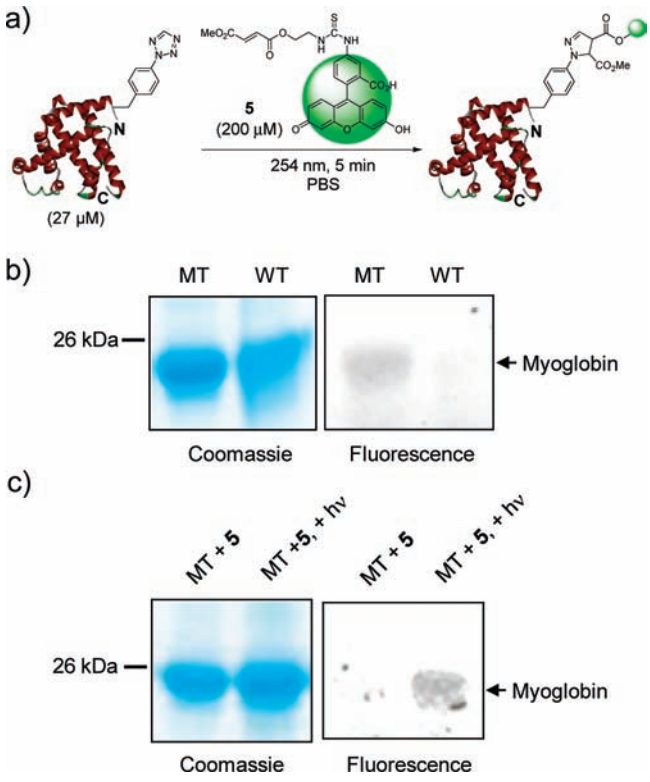


Figure 4. Selective fluorescent labeling of the *p*-Tpa-encoded myoglobin via photoclick chemistry: (a) Reaction of a mixture of *p*-Tpa-encoded mutant myoglobin (or wild-type, 27 μM) and FITC-fumarate **5** (200 μM) in PBS buffer supplemented with 3 mM TCEP and exposed to 5-min photoirradiation of a handheld 254-nm UV lamp followed by additional 1.5-h incubation at room temperature. (b) Coomassie blue stain (left) and inverted fluorescence image (right) after resolving the reaction mixtures on SDS-PAGE gel. (c) Fluorescent labeling of *p*-Tpa-encoded mutant myoglobin (MT) with FITC-fumarate **5** requires photoactivation: A solution of *p*-Tpa-encoded mutant myoglobin (27 μM) and FITC-fumarate **5** (200 μM) in PBS buffer supplemented with 3 mM TCEP was either simply incubated or subjected to photoirradiation by a hand-held 254-nm UV lamp for 5 min followed by 1.5-h incubation at room temperature. The inverted fluorescence image (right) were shown along with the Coomassie blue stain (left).

flexible GGD loop, presumably to allow the synthetase to tolerate all these mutations without loss of the aminoacylation activity. Overall, the substrate specificity of *M. jannaschii* tyrosyl-tRNA synthetase was altered by active site mutations to favor *p*-Tpa over tyrosine without major backbone structural reorganization, which was also observed in other *Mj*TyrRS engineering work.[14,18]

Conclusion

In summary, we have demonstrated the site-specific incorporation of *p*-Tpa into a protein via a biosynthetic route and the use of *p*-Tpa as a photoreactive "handle" for fluorescent labeling of the target protein. The structural basis for biosynthetic incorporation of *p*-Tpa was also elucidated through X-ray structural analysis of the *p*-Tpa/*p*-TpaRS-1 complex. With the appropriately selected dipolarophiles, this photoreactive amino acid may allow photoregulation of protein function in living cells with a spatiotemporal precision. Moreover, by using two different blank codons[19] to encode the tetrazole and alkene amino acids[11b,e] separately into either two interacting proteins or two adjoining sites in a single protein, it might now be possible to photoregulate specific protein−protein interactions or protein conformations in a living organism.

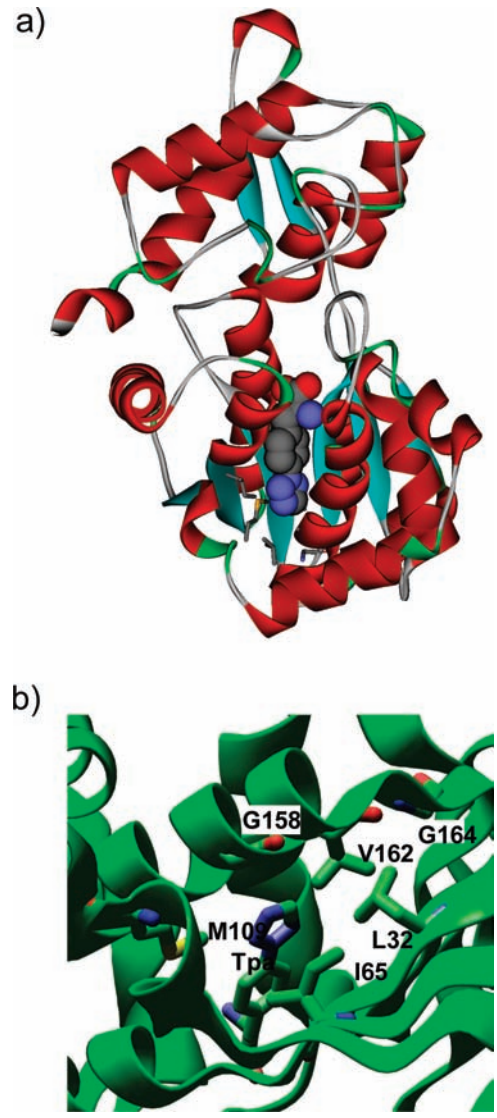


Figure 5. Crystal structure of *p*-TpaRS-1 bound to *p*-Tpa: (a) A ribbon model showing the overall fold with the bound *p*-Tpa ligand in the CPK model. (b) A close-up view of the *p*-Tpa binding pocket (rotated 180° along the horizontal axis relative to part a with the six mutated residues rendered in tube models. The complex structure has been deposited into the Protein Data Bank with an access code of 3N2Y.

Experimental Section

General Methods. Chemicals were purchased from Alfa Aesar and Sigma-Aldrich and used without further purification. The phage-resistant DH10B strain (GeneHogs, Invitrogen) was used for all in vivo manipulation unless noted otherwise. Cells were incubated at 37 °C. PCR was performed with the Expand High Fidelity PCR System (Roche). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Oligonucleotides were purchased from Shenggong Inc.

***p*-(2-Tetrazole)phenylalanine (1).** The tetrazole amino acid **1** was synthesized as described previously:[13] ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.26 (s, 1H), 8.71 (s, 3H), 8.04 (d, J = 8.5 Hz, 2H),

(18) Liu, W.; Alfonta, L.; Mack, A. V.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2007**, *46*, 6073–6075.

(19) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464*, 441–444.

7.61 (d, $J = 8.5$ Hz, 2H), 4.21 (s, 1H), 3.31−3.30 (m, 2H); ^{13}C NMR (75.4 MHz, DMSO-d_6) δ 170.1, 154.0, 37.7, 135.2, 131.4, 120.2, 53.0, 35.0; HRMS (ESI) calcd for $C_{10}H_{12}O_2N_5$ 234.0986 [M + H^+], found 234.0985.

2-Tolyltetrazole (2). The compound was synthesized as described previously:[13] ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.23 (s, 1H), 7.99 (d, $J = 8.5$ Hz, 2H), 7.48 (d, $J = 8.5$ Hz, 2H), 2.41 (s, 3H); ^{13}C NMR (75.4 MHz, DMSO-d_6) δ 153.7, 140.1, 133.9, 130.5, 119.9, 20.7; HRMS (ESI) calcd for $C_8H_9N_4$ 161.0822 [M + H^+], found 161.0821.

FITC-Fumarate (5). To a solution of Boc-aminoethyl methyl fumarate (250 mg, 0.916 mmol) in 10 mL of CH_2Cl_2 was added 1 mL of trifluoroacetic acid. The mixture was stirred until the starting material was consumed based on TLC. The solvent was removed under vacuum together with chloroform for three times. The residue was redissolved in 30 mL of THF, and the free primary amine was obtained by adding triethylamine (254 μL, 1.83 mmol). To the mixture was added dropwise a solution of fluorescein isothiocynate (isomer I, 90%; 179 mg, 0.460 mmol) in 10 mL of THF. After 12 h the solution was concentrated and the residue was purified by reverse-phase HPLC to give the desired product as an orange solid (30 mg, 12% yield): ^{1}H NMR (500 MHz, CD_3OD-d_4) δ 8.20 (br, 1H), 8.14−8.12 (m, 1H), 7.77 (d, $J = 8.5$ Hz, 1H), 7.56 (d, $J = 8.5$ Hz, 1H), 7.19 (d, $J = 8.0$ Hz, 1H), 6.91 (d, $J = 2.0$ Hz, 2H), 6.71−6.69 (m, 4H), 6.56 (dd, $J = 8.5, 2.0$ Hz, 2H), 4.47 (t, $J = 5.0$ Hz, 2H), 4.03−3.90 (m, 2H), 3.79 (s, 3H); ESI-MS calcd for $C_{28}H_{23}N_2O_9S$ [M + H^+] 563.1, found 563.0.

Selection of the *p*-Tpa-Specific Aminoacyl-tRNA Synthetase. *E. coli* DH10B cells harboring the pREP(2)/YC plasmid was used as the host strain for the positive selections. Cells were transformed with pBK-lib-jw1 library, recovered in SOC for 1 h, and washed twice with the glycerol-supplemented minimal medium with leucine (GMML) before plating on GMML-agar plates supplemented with kanamycin, chloramphenicol, tetracycline, and *p*-Tpa at 50 μg/mL, 60 μg/mL, 15 μg/mL, and 1 mM, respectively. The plates were incubated at 37 °C for 60 h, the surviving cells were scraped, and the plasmid DNA was extracted and purified by gel electrophoresis. The pBK-lib-jw1 DNA that passed the positive selection was transformed into electrocompetent cells harboring the negative selection plasmid pLWJ17B3, recovered for 1 h in SOC, and then plated on LB-agar plates containing 0.2% arabinose, 50 μg/mL ampicillin, and 50 μg/mL kanamycin. The plates were then incubated at 37 °C for 8−12 h, and pBK-lib-jw1 DNA from the surviving clones was extracted as described above. The library was then carried through a subsequent round of positive selection, followed by one round of negative selection and the final round of positive selection (with chloramphenicol at 70 μg/mL). At this stage, 96 individual clones were selected, suspended in 100 μL of GMML in a 96-well plate, and replica-spotted on two sets of GMML plates: one set of GMML-agar plates was supplemented with tetracycline (15 μg/mL), kanamycin (50 μg/mL), and chloramphenicol at concentrations of 60, 80, 100, and 120 μg/mL with 1 mM *p*-Tpa; the other set of plates was identical but did not contain *p*-Tpa, and the chloramphenicol concentrations were 0, 20, 40, and 60 μg/mL. After incubation at 37 °C for 60 h, one clone was found to survive at 120 μg/mL chloramphenicol in the presence of 1 mM *p*-Tpa but only at 20 μg/mL chloramphenicol in the absence *p*-Tpa. This clone was picked and sequenced to afford *p*-TpaRS-1.

Synthetase Expression and Crystallization Setup. DNA fragments encoding the *p*-TpaRS-1 were amplified by PCR and cloned into the NdeI and XhoI sites of the expression vector pET22b, which was then transformed in BL21(DE3) cells and grown to an OD of 0.8. After induction overnight with 1 mM IPTG, cells were pelleted by centrifugation and resuspended in lysis buffer (50 mM HEPES, pH 7.9, 500 mM NaCl, 10 mM β-mercaptoethanol, 10% glycerol, 0.1% Triton X-100, 5 mM imidazole). Cells were sonicated, and the cell lysate was pelleted by centrifugation. The supernatant was collected and incubated with Ni-NTA agarose beads for 1 h at 4 °C, filtered, and washed with wash buffer (50 mM HEPES, pH 7.9, 500 mM NaCl, 10 mM β-mercaptoethanol, 5 mM imidazole). The synthetase was eluted with a wash buffer containing 250 mM imidazole in buffer A (25 mM Tris, pH 8.5, 25 mM NaCl, 10 mM β-mercaptoethanol, 1 mM EDTA), purified by anion exchange chromatography (MonoQ; Amersham Biosciences) using a salt gradient from 25 mM to 1 M NaCl, dialyzed in a buffer containing 20 mM Tris, pH 8.5, 50 mM NaCl, 10 mM 2-mercaptoethanol, and concentrated to 20−30 mg/mL. Crystals of the mutant TyrRS-amino acid complex were grown at 16 °C by using the sitting-drop vapor-diffusion technique against a mother liquor composed of 16−20% polyethylene glycol (PEG) 300, 3−5% PEG 8000, 100 mM Tris (pH 8.8), and 10% glycerol and a 1:1 mixture of concentrated synthetase (15 mg/mL) and 1 mM *p*-Tpa.

X-ray Crystallography. X-ray diffraction data of *p*-TpaRS-1 were collected at beamline BL17U of Shanghai Synchrotron Radiation Facility (SSRF) at a single wavelength of 0.9789 Å. The data were reduced and scaled using the HKL2000 package. The structure of *p*Tpa-*p*TpaRs-1 complex was solved by molecular replacement using wild-type *M. jannaschii* tyrosyl-tRNA synthetase (PDB code: 1J1U)[16] as a search model in which water molecules and other heteroatoms were deleted and processed by Phaser of the CCP4 package. Structural refinement was carried out by Refmac5 in the CCP4 program suite. In the refinement process, the program Coot in the CCP4 program suite was used for the model building (main chain tracing), ligand and water finding, and real space refinement of side chains and zones. The coordinate of *p*-Tpa was generated by SMILES Translator and Structure File Generator (http://cactus.nci.nih.gov/translate) and program Coot.

Expression and Characterization of Wild-Type and *p*-Tpa-Encoded Myoglobin. For expression of the *p*-Tpa-encoded mutant myoglobin, plasmid pBAD/JYAMB-4TAG-Myo was cotransformed with pBK vector expressing *p*-TpaRS-1 into GeneHog *E. coli* cells. Cells were amplified in 2YT media (5 mL) supplemented with kanamycin (50 μg/mL) and tetracycline (15 μg/mL). Starter culture (1 mL) was used to inoculate 100 mL of liquid GMML or 2YT supplemented with appropriate antibiotics and 1 mM of *p*-Tpa. Cells were then grown at 37 °C to OD_{600} of 0.5, and protein expression was induced by addition of 0.2% arabinose. After another 4−12 h, cells were harvested by centrifugation. The *p*-Tpa-encoded myoglobin proteins were purified by Ni-NTA affinity chromatography under native conditions. For expression of wild-type myoglobin, BL21(DE3) cells transformed with pBad-wtMyo were allowed to grow either in minimal medium containing 1% glycerol and 0.3 mM leucine (GMML medium) or LB medium supplemented with 15 μg/mL of tetracycline in the presence or absence of 1 mM *p*-Tpa at 37 °C. When cells reached an optical density (OD_{600}) of ∼0.5, 0.2% arabinose was added to induce protein expression. After overnight culture, cells were harvested and wild-type myoglobin proteins in the supernatant were purified by Ni-NTA affinity chromatography. Protein mass was determined on Micromass Q-TOF micro mass spectrometer equipped with electrospray ionization. Ten microliters of 0.1 mg/mL myoglobin proteins in 10 mM of ammonium bicarbonate was mixed with 1% formic acid at a v/v ratio of 1:1. Calibration was performed in positive ion mode using wild-type sperm whale myoglobin. Molecular masses were obtained by deconvoluting multiply charged protein mass spectra using MassLynx software (Micromass). Theoretical mass of wild-type protein was calculated using Protein Mass Calculator (http://www.peptideguide.com/peptide-calculator.html). Theoretical masses for myoglobin proteins incorporating *p*-Tpa at the fourth position were calculated by adding the molecular weight of *p*-Tpa and subtracting the molecular weight of serine from that of wild-type myoglobin.

Reaction of *p*-Tpa-Encoded Myoglobin with FITC-Fumarate. A 2-μL solution of mutant myoglobin (5 μg/μL) (or wild-type myoglobin, 5 μg/μL; final concentration = 27 μM) was incubated with 2 μL of fumarate-FITC (2 mM in DMSO) and 2 μL of TCEP (30 mM in H_2O) in 14 μL of PBS buffer. After irradiation of the mixtures in a 96-well microtiter plate with a hand-

held 254-nm UV lamp for 5 min followed by incubation at room temperature for an additional 1.5 h, the reactions were quenched by adding 1 μL of 1 N HCl. After adjustment of the pH to 7.0 with 1 N NaOH, 5 μL of 6× SDS sample buffer was added and the mixtures were boiled at 95 °C for 5 min. The samples were loaded onto a NuPAGE 12% Bis-Tris gel for protein electrophoresis. The resolved gel was subjected to both Coomassie blue staining and in-gel fluorescence analysis.

Kinetic Analysis of Cycloaddition Reaction. Appropriate amounts of 2-tolyltetrazole and dimethyl fumarate were dissolved in acetonitrile to obtain the indicated concentrations. Separate reactions were set up by incubating 5000 μL of reaction mixture in a quartz test tube and irradiated with a hand-held 254 nm UV lamp. Aliquots of 200-μL reaction mixtures from each sample were withdrawn and injected into the HPLC at various time points. The amounts of cycloaddition product P in the reaction mixtures were monitored by UV absorbance at 254 and 370 nm. The concentrations of product P were determined by comparing the integrated areas with a standard curve that correlates the absorbance areas at 254 nm with the known concentrations of standard pyrazoline sample P (0−1000 μM) run under identical HPLC conditions. The photolysis of tetrazole was very rapid; almost all of 2-tolyltetrazole (A) disappeared after 3 min, suggesting that the ring-opening step is much faster than the cycloaddition step ($k_1 \gg k_2$).[9b] Thus, the overall reaction follows pseudo-first-order reaction kinetics in the presence of an excess amount of dimethyl fumarate. Under this condition, the rate of the cycloaddition obeys the first-order rate law: $\ln([A]_0/([A]_0 - [P]_t)) = k_{obs}t$. By plotting product formation $[P]_t$ versus time (t) and fitting the data to the above equation, the pseudo-first-order rate constant, k_{obs}, was derived. The second-order rate constant for the cycloaddition, k_2, was calculated according to the following equation: $k_{obs} = k_2$[fumarate].

Acknowledgment. We gratefully acknowledge the Major State Basic Research Program of China (2010CB912301, 2009CB825505, 2006CB910903), National Science Foundation of China (30870592, 90913022) to J.W., and the National Institutes of Health (GM 085092) to Q.L. for financial support.

Supporting Information Available: Protein structure data collection and refinement statistics, supplemental figures, experimental procedures, and compound characterization data. This material is available free of charge via the Internet at http://pubs.acs.org.

JA104350Y