Synthetic Biology

DOI: 10.1002/ange.201308794

Proximity-Enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids**

*Zheng Xiang, Vanessa K. Lacey, Haiyan Ren, Jing Xu, David J. Burban, Patricia A. Jennings, and Lei Wang**

Abstract: *The selective generation of covalent bonds between and within proteins would provide new avenues for studying protein function and engineering proteins with new properties. New covalent bonds were genetically introduced into proteins by enabling an unnatural amino acid (Uaa) to selectively react with a proximal natural residue. This proximity-enabled bioreactivity was expanded to a series of haloalkane Uaas. Orthogonal tRNA/synthetase pairs were evolved to incorporate these Uaas, which only form a covalent thioether bond with cysteine when positioned in close proximity. By using the Uaa and cysteine, spontaneous covalent bond formation was demonstrated between an affibody and its substrate Z protein, thereby leading to irreversible binding, and within the affibody to increase its thermostability. This strategy of proximity-enabled protein crosslinking (PEPC) may be generally expanded to target different natural amino acids, thus providing diversity and flexibility in covalent bond formation for protein research and protein engineering.*

The ability to covalently crosslink proteins spontaneously would provide a tremendous opportunity for enhancing existing protein properties or engineering new ones because covalent linkages are more stable and selective than the noncovalent interactions between protein side chains. In natural proteins, the disulfide bond formed between two Cys residues plays a crucial role in protein folding, stability, and activity for a vast array of proteins.[1] The inherent redox sensitivity and reversibility of the disulfide linkage, however, also set limitations for protein expression, engineering, and application.[2] Adding new covalent bonds into proteins should overcome such limitations and broaden the scope of protein manipulation. To this end, we have developed a general strategy in which an unnatural amino acid (Uaa) is designed to react with a natural amino acid in a protein through proximity-enhanced reactivity.[3] The Uaa does not react with free natural amino acids under physiological conditions, thereby permitting in vivo genetic incorporation. When the Uaa is placed with appropriate orientation in proximity to its target natural amino acid in a protein, the increased local effective concentration facilitates the reaction of the Uaa with the target residue to build a covalent bond. By using this strategy, we have recently shown that new inter- or intra-protein covalent bonds could be formed between a Cys residue and *p*-2′-fluoroacetylphenylalanine.[3] Herein, we report the genetic incorporation of a series of haloalkane Uaas and demonstrate their ability to form covalent bonds with Cys in proximity-enabled protein crosslinking (PEPC), thus suggesting that this overall strategy may be applicable to a host of Uaas to generate various new covalent bonds in proteins. To demonstrate the potential of PEPC, we show that interprotein PEPC can enable irreversible protein binding and that intraprotein PEPC can be used to increase a protein's thermostability.

We designed and synthesized a series of tyrosine analogues containing different halogen atoms linked with aliphatic chains of varying length (Figure 1a, Scheme S1 in the Supporting Information). By exploiting specific protein activities, small-molecule alkyl halides have been used for activity-based protein profiling and protein tagging.[4] The reactivity between alkyl halides and Cys has also been employed to synthesize cyclic peptides.[5] We therefore expect the alkyl halides, after being incorporated into proteins, to react with the sulfhydryl group of a Cys residue when they are in close proximity. To date, no haloalkane Uaa has been genetically incorporated into proteins in living cells.[6]

To produce an aminoacyl-tRNA synthetase specific for haloalkane Uaas, we generated a mutant synthetase library based on MmOmeRS,[7] which was created by directed evolution from the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (MmPylRS)for incorporating the Uaa *O*-methyltyrosine. Based on the crystal structure of MmOmeRS in complex with *O*-methyltyrosine,[7] we randomized five residues (Val346, Trp348, Ser399, Val401, Trp417) in MmOmeRS to create a library of 3×10^7 mutants (Figure 1b).[8] Three rounds of selection of this library with CprY yielded one clone that showed CprY-dependent survival in chloramphenicol (see the Supporting Information). Sequencing revealed that the selected synthetase, named MmXYRS, contains the following substitutions: V346A, W348A, L401V, and W417T

[*] Dr. Z. Xiang, Dr. V. K. Lacey, Dr. H. Ren, Prof. Dr. L. Wang
The Jack H. Skirball Center for Chemical Biology and Proteomics
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: lwang@salk.edu
Homepage: http://wang.salk.edu/
Dr. J. Xu, D. J. Burban, Prof. Dr. P. A. Jennings
Department of Chemistry and Biochemistry
University of California, San Diego
La Jolla, CA 92093 (USA)

[**] We thank C. Hoppman for helpful discussions, Y.-X. Guo and M. Bowman for help with FPLC purification, J.-K. Weng for pHIS8.3-MBP plasmid, and W. Fischer and W. Low for help with mass spectrometry. L.W. acknowledges support from the California Institute for Regenerative Medicine (RN1-00577-1) and US National Institutes of Health (1DP2OD004744-01, P30CA014195).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201308794.

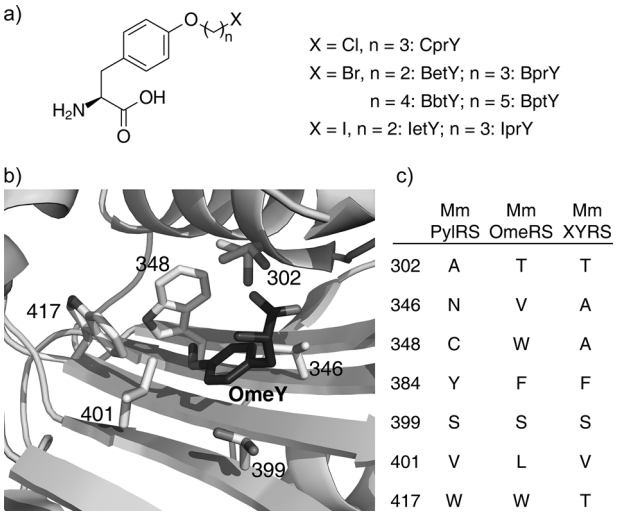

Figure 1. a) Structures of the haloalkane Uaas. Uaas are abbreviated in the following format: CprY, *O*-(3-Chloro*pr*opyl)-L-tyrosine. b) Active site of MmOmeRS in complex with *O*-methyltyrosine (OmeY, dark) with the residues randomized in the library shown as stick representations. c) Amino acid sequences of the evolved MmXYRS in comparison to WT MmPylRS and the parental MmOmeRS.

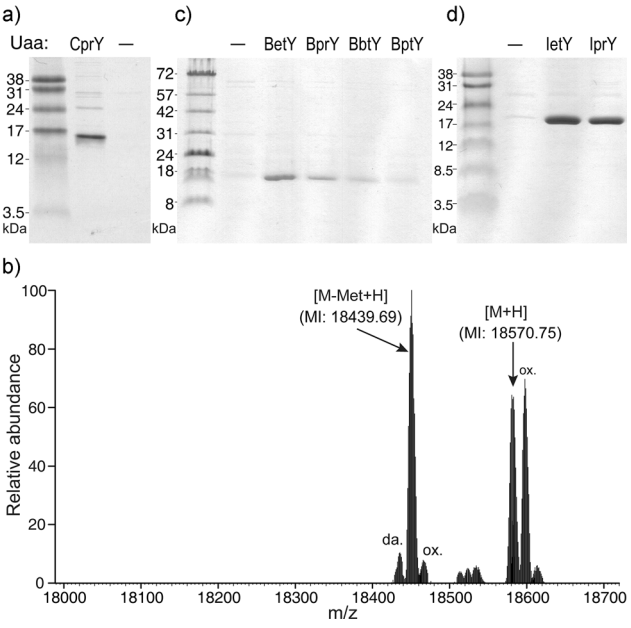

Figure 2. Incorporation of haloalkane Uaas into myoglobin in *E. coli*. SDS-PAGE analysis of myoglobin(4TAG) expressed using the $tRNA^{Pyl}_{CUA}$ and MmXYRS in the presence of 1 mM CprY (a; tris-glycine gel); BetY, BprY, BbtY, and BptY (c; tris-glycine gel); and IetY and IprY (d; tricine gel). Samples were normalized for equal cell numbers for each lane. b) High resolution ESI-FTMS of the intact myoglobin produced with 1 mM CprY (a). Monoisotopic masses of major peaks indicate that only CprY was incorporated into myoglobin at the TAG-encoded position. Peaks corresponding to oxidation (ox.) and deamidation (da.) of CprY-myoglobin were also detected and labeled.

(Figure 1c). Ser399 was retained, but the codon was changed (see the Supporting Information).

To measure the in vivo translational efficiency and fidelity of the evolved MmXYRS, we expressed the genes for $tRNA^{Pyl}_{CUA}$ and MmXYRS in *E. coli* together with a sperm whale myoglobin gene with a C-terminal His tag and a TAG codon at position 4. After Ni^{2+} affinity chromatography, a strong band running at ~17 kD was observed by SDS-PAGE only from cells supplemented with 1 mM CprY, thus indicating that full-length myoglobin was produced by the *E. coli* (Figure 2a). We then analyzed the purified myoglobin with Fourier transform ion trap mass spectrometry (ESI-FTMS). A peak with a monoisotopic mass of 18570.75 Da was observed (Figure 2b), which corresponds to intact myoglobin with a single CprY residue at position 4 (expected $[M+H]^+ =$ 18570.76 Da). A second measured peak corresponds to CprY-myoglobin lacking the initiating Met residue (expected $[M\text{-Met}+H]^+ = 18439.72$ Da, measured 18439.69 Da). There were no peaks that corresponded to myoglobin variants resulting from the incorporation of any endogenous amino acid instead of CprY. Moreover, there were no peaks corresponding to myoglobin with the halogen substituted by cysteine, glutathione, 2-mercaptoethanol, or imidazole. These results indicate that CprY was specifically incorporated into myoglobin and that the CprY residue was stable and unmodified during protein synthesis in *E. coli* and throughout the purification process.

We next examined whether MmXYRS could incorporate other haloalkane Uaas with structures similar to CprY. Indeed, SDS-PAGE analyses of myoglobin expressed in the presence of different Uaas showed that both bromo- and iodo-containing Uaas were incorporated by the $tRNA^{Pyl}_{CUA}$/MmXYRS pair (Figure 2c,d). ESI-FTMS analyses of the purified myoglobin samples clearly demonstrated a high fidelity of incorporation for these Uaas (Figure S1 and Table S1 in the Supporting Information).

With MmXYRS and a set of haloalkane Uaas in hand, we embarked on testing intermolecular PEPC between the Z protein and the Z_{SPA} affibody (Afb), which binds to the Z protein with a K_d of 6 μM.[9] We chose to introduce Cys into the Z protein at Asn6 and incorporate the Uaa into the Afb at Asp36; these two sites are brought into close proximity when the two proteins bind (Figure 3a).[3a,9] To clearly distinguish the Z protein from the Afb by molecular weight by SDS-PAGE, we fused the maltose binding protein (MBP) to the Z protein and its mutant, Z(N6C). The purified Afb(D36Uaa) proteins were incubated with the purified MBP–Z(N6C) in a 2:1 ratio in PBS buffer (pH 7.4) at 37°C for 1 h, and the reaction mixture was subjected to SDS-PAGE under denaturing conditions. We observed a band with a molecular weight corresponding to MBP–Z in complex with the Afb (Figure 3b), thus indicating that the two proteins were covalently crosslinked. By contrast, when the Afb(D36Uaa) mutants were incubated with MBP–Z(WT), no band corresponding to the complex was detected (Figure 3b), thus showing that the crosslinking reaction was specific for Cys6. For the Uaas with different halogen atoms linked by the same alkyl chain, the order of crosslinking efficiency is I > Br > Cl, a result consistent with the order of the halide leaving ability in S_N2 reactions.

To explore intramolecular PEPC, we introduced both the haloalkane Uaa and the Cys residue into the same Afb

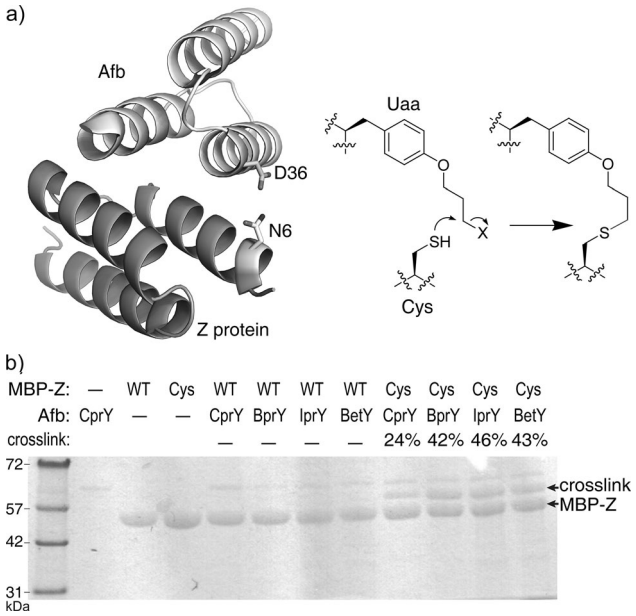

Figure 3. Intermolecular PEPC of the Afb to its substrate Z protein fused to MBP (MBP–Z). a) Crystal structure of the Afb–Z-protein complex (PDB ID 1LP1), with Asp36 in the Afb and Asn6 in the Z protein (for substitution by the Uaa and Cys, respectively) shown as stick representations. b) SDS-PAGE analysis of Afb–Z crosslinking. The identities of residue 6 of the Z protein in MBP–Z and residue 36 of the Afb are indicated. Crosslinking efficiency was calculated based on the relative band intensities of the crosslinked complex and the MBP–Z fusion protein.

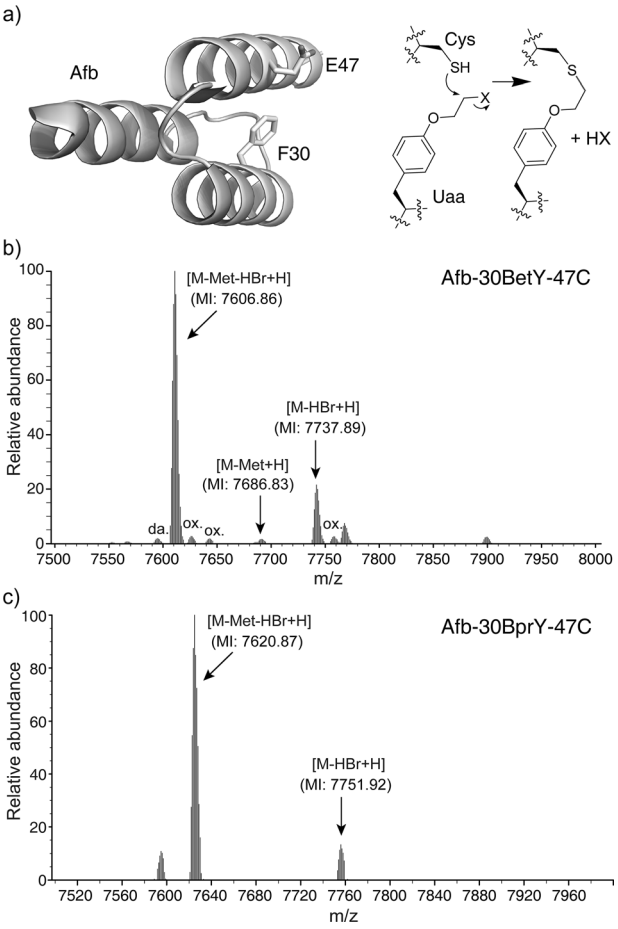

Figure 4. Intramolecular PEPC within the Afb. a) Structure of the Afb (PDB ID 1LP1) showing Phe30 and Glu47, which were used for introducing the Uaa and Cys, respectively. b) High resolution ESI-FTMS analysis the Afb with BetY incorporated at residue 30 and Cys at residue 47. The crosslinking reaction of the Uaa with Cys results in the loss of HBr. Crosslinked products: [*M*-HBr + H], expected 7737.90 Da, measured 7737.89 Da; [*M*-Met-HBr + H], expected 7606.86 Da, measured 7606.86 Da. Non-crosslinked products: [*M* + H], expected 7817.82 Da, not detected; [*M*-Met + H], expected 7686.78 Da, measured 7686.83. c) High resolution ESI-FTMS analysis of the Afb with BprY incorporated at residue 30 and Cys at residue 47. Crosslinked products: [*M*-HBr + H], expected 7751.91 Da, measured 7751.92 Da; [*M*-Met-HBr + H], expected 7620.87 Da, measured 7620.87 Da. Non-crosslinked products: [*M* + H], expected 7831.84 Da, not detected; [M-Met + H], expected 7700.80 Da, not detected.

protein and determined whether covalent crosslinking could spontaneously occur. Based on the structure of the Afb,[9] two proximal sites, F30 and E47, were chosen for incorporating the Uaa and Cys, respectively (Figure 4a). The Uaa was incorporated at the Phe site to minimize potential perturbation of the Afb structure. *E. coli* cells harboring the expression plasmids were grown in rich medium 2xYT supplemented with 1 mM of BetY or BprY. After induction with 0.2% L-arabinose for 5 h at 30 °C, the proteins were purified by Ni^{2+} affinity chromatography. For each mutant protein, only one band, running at the same position as the WT affibody in SDS-PAGE (Figure S2), was detected. As another control, we incorporated BetY at position 30 while maintaining the WT E47 (Afb-30BetY). ESI-FTMS analyses of Afb-30BetY indicate that BetY was selectively incorporated at the TAG position with high fidelity and the Br group was not substituted by any other molecule or amino acid side chain of the Afb (Figure S3). By contrast, when Cys was introduced at position 47, the mutant proteins Afb-30BetY-47C and Afb-30BprY-47C clearly showed strong peaks with monoisotopic masses corresponding to the Afb with a covalent Cys–Uaa bond at the introduced positions (Figure 4b,c), thus indicating successful intramolecular crosslinking. For the Afb-30BetY-47C mutant, a peak with a monoisotopic mass corresponding to non-crosslinked Afb ([*M*-Met + H]) was also detected, albeit with very weak intensity (less than 2% of the crosslinked Afb). Afb-30BprY-47C showed no peak for non-crosslinked Afb, thus indicating complete intramolecular crosslinking.

To assess the effects of intramolecular PEPC on protein properties, we measured the thermal stability of the FPLC-purified WT Afb, Afb-30BetY-47C, and Afb-30BprY-47C by using circular dichroism (CD) spectroscopy. All CD spectra showed negative minima at 222 nm and a signal sign change from negative to positive at 200 nm (Figure S4a), results consistent with the helical structure of the Afb. The melting curves were analyzed based on a two-state unfolding model of a monomeric protein (Figure S4b–d).[10] WT Afb and Afb-30BprY-47C showed similar T_m values of (46.7 ± 0.2) °C and (46.5 ± 0.5) °C, respectively. By contrast, Afb-30BetY-47C showed a markedly higher T_m value of (60.4 ± 0.7) °C. These results suggest that an intramolecular covalent linkage of an appropriate length improves the thermal stability of Afb.

In summary, we genetically encoded a series of haloalkane Uaas that possess proximity-enabled bioreactivity toward Cys. In contrast to the Uaas incorporated to date, which have been either chemically inert or bioorthogonal, these uniquely bioreactive Uaas enable both inter- and intra-molecular PEPC in a controlled and specific manner. In contrast to disulfide bonds, the covalent bond formed between a haloalkane Uaa and Cys is irreversible and stable under reducing conditions. The series of haloalkane Uaas used also increases the diversity and flexibility in side chain length, type, and reactivity compared to our previously reported *p*-2′-fluoroacetylphenylalanine approach.[3a] We showed that interprotein PEPC enables irreversible protein binding and that cotranslational PEPC within an Afb can increase its thermostability. We are developing new Uaas to expand proximity-enabled bioreactivity to other natural amino acid residues and are applying PEPC to study protein interactions and develop protein therapeutics.[3b] We expect that this spontaneous PEPC will provide new avenues to novel protein properties and functions for a wide range of applications in basic and synthetic biology.

Received: October 9, 2013
Published online: January 21, 2014

Keywords: protein crosslinking · protein engineering · proximity-enabled reactivity · tRNA · unnatural amino acids

[1] a) H. Kadokura, F. Katzen, J. Beckwith, *Annu. Rev. Biochem.* **2003**, *72*, 111 – 135; b) C. S. Sevier, C. A. Kaiser, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 836 – 847.
[2] M. Berkmen, *Protein Expression Purif.* **2012**, *82*, 240 – 251.
[3] a) Z. Xiang, H. Ren, Y. S. Hu, I. Coin, J. Wei, H. Cang, L. Wang, *Nat. Methods* **2013**, *10*, 885 – 888; b) I. Coin, V. Katritch, T. Sun, Z. Xiang, F. Y. Siu, M. Beyermann, R. C. Stevens, L. Wang, *Cell* **2013**, *155*, 1258 – 1269.
[4] a) B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383 – 414; b) G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, *ACS Chem. Biol.* **2008**, *3*, 373 – 382.
[5] a) L. Yu, Y. Lai, J. V. Wade, S. M. Coutts, *Tetrahedron Lett.* **1998**, *39*, 6633 – 6636; b) Z. Dekan, I. Vetter, N. L. Daly, D. J. Craik, R. J. Lewis, P. F. Alewood, *J. Am. Chem. Soc.* **2011**, *133*, 15866 – 15869.
[6] a) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498 – 500; b) J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon, P. G. Schultz, *Nat. Biotechnol.* **2004**, *22*, 1297 – 1301; c) Q. Wang, A. R. Parrish, L. Wang, *Chem. Biol.* **2009**, *16*, 323 – 336; d) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413 – 444.
[7] J. K. Takimoto, N. Dellas, J. P. Noel, L. Wang, *ACS Chem. Biol.* **2011**, *6*, 733 – 743.
[8] V. K. Lacey, G. V. Louie, J. P. Noel, L. Wang, *ChemBioChem* **2013**, *14*, 2100 – 2105.
[9] M. Hogbom, M. Eklund, P. A. Nygren, P. Nordlund, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3191 – 3196.
[10] N. J. Greenfield, *Nat. Protoc.* **2007**, *1*, 2527 – 2535.

