

Genetic Incorporation of Olefin Cross-Metathesis Reaction Tags for Protein Modification

Bhaskar Bhushan,[†] Yuya A. Lin,^{†,§} Martin Bak,[†] Anuchit Phanumartwath,[†] Nan Yang,[†] Matthew K. Bilyard,[†] Tomonari Tanaka,^{†,||} Kieran L. Hudson,[†] Lukas Lercher,[†] Monika Stegmann,[‡] Shabaz Mohammed,^{†,‡,||} and Benjamin G. Davis^{*,†,||}

[†]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, U.K.

[‡]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K.

Supporting Information

ABSTRACT: Olefin cross-metathesis (CM) is a viable reaction for the modification of alkene-containing proteins. Although allyl sulfide or selenide side-chain motifs in proteins can critically enhance the rate of CM reactions, no efficient method for their site-selective genetic incorporation into proteins has been reported to date. Here, through the systematic evaluation of olefin-bearing unnatural amino acids for their metabolic incorporation, we have discovered *S*-allylhomocysteine (Ahc) as a genetically encodable Met analogue that is not only processed by translational cellular machinery but also a privileged CM substrate residue in proteins. In this way, Ahc was used for efficient Met codon reassignment in a Met-auxotrophic strain of *E. coli* (B834 (DE3)) as well as metabolic labeling of protein in human cells and was reactive toward CM in several representative proteins. This expands the use of CM in the toolkit for “tag-and-modify” functionalization of proteins.

Alkenes can be installed into proteins by incorporation of some unnatural amino acids (uAAs).^{1–4} However, although aliphatic alkene-containing amino acids such as homoallylglycine (Hag)⁴ are reactive in both self-metathesis and cross-metathesis (CM) reactions as monomeric, protected amino acids in organic solvents⁵ and can be metabolically incorporated into proteins, they are unreactive in CM reactions in aqueous media.⁶ On the other hand, *chemically* installed⁷ uAAs such as *S*-allylcysteine (Sac)⁶ and *Se*-allylselenocysteine (Seac)⁸ have been demonstrated to be CM-reactive under aqueous conditions, with the allylic chalcogen heteroatom S (and even more so Se) providing crucial coordination to the catalyst metal center.^{6,8,9} Schultz and co-workers have reported the incorporation of *O*-crotylserine in yeast via amber stop codon suppression and have demonstrated its reactivity in on-protein intramolecular ring-closing metathesis.¹⁰ However, CM with genetically incorporated residues has yet to be demonstrated.

An alternative approach to “nonsense” codon reassignment involves the direct commandeering of “sense” codons for amino acids such as methionine (Met) and subsequent reassignment to incorporate unnatural amino acids (e.g., Hag, norleucine, trifluoromethionine, homopropargylglycine,

and azidohomoalanine) as Met analogues.^{4,11–13} This effective reassignment of the Met codon exploits the flexibility of native methionyl-tRNA synthetase (MetRS) in accepting these analogues as substrates for tRNA loading.¹⁴ Here, by probing the flexibility of MetRS, we show metabolic labeling of proteins with an unnatural allylchalcogen-containing amino acid (the Met analogue *S*-allylhomocysteine (Ahc)) that is also metathesis-reactive.

The incorporation efficiency of amino acids and their analogues *in vivo* is controlled, to a significant extent, by their activation by the corresponding aminoacyl-tRNA synthetases (aaRSs).^{4,15,16} Fersht and Dingwall¹⁷ first demonstrated that a Met analogue, *L*-ethionine, may be mischarged by the MetRS without being metabolically edited. By contrast, *L*-homocysteine, a natural competitor of Met, is transformed to the corresponding thiolactone to avoid misreading in translation.¹⁸ Side-chain length and heteroatom position can therefore be critical determinants in “metabolic recognition” of such analogues. Therefore, in our design of putative analogues (Figure 1), we explored these aspects: tolerance of MetRS toward positioning of the side-chain heteroatom required for CM *plus* concomitant variation of the side-chain length.

Two types of substrates were designed and synthesized: those with a γ -heteroatom (1 and 2) and those with a δ -

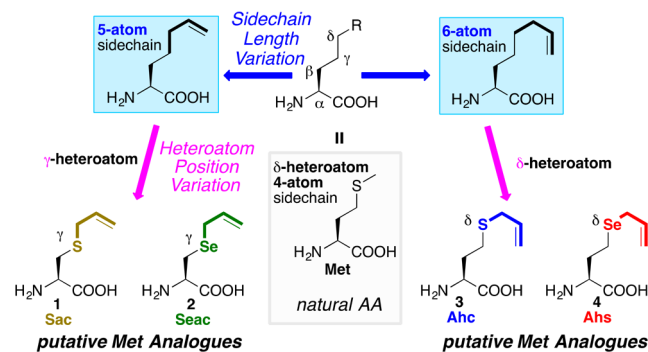


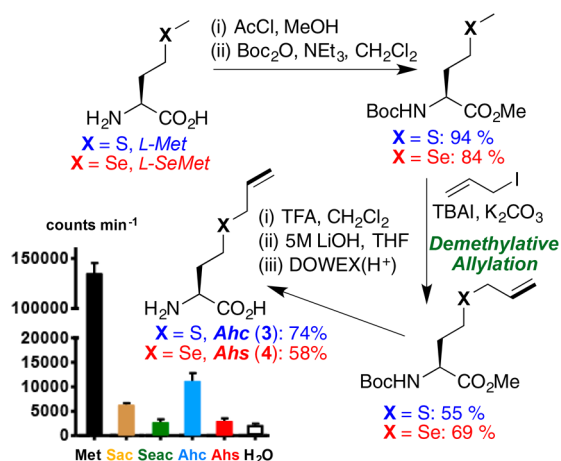
Figure 1. Design of the γ -heteroatom Met analogues *S*-allylcysteine (Sac, 1) and *Se*-allylselenocysteine (Seac, 2) and the δ -heteroatom analogues *S*-allylhomocysteine (Ahc, 3) and *Se*-allylhomoselenocysteine (Ahs, 4).

Received: September 2, 2018

Published: October 16, 2018

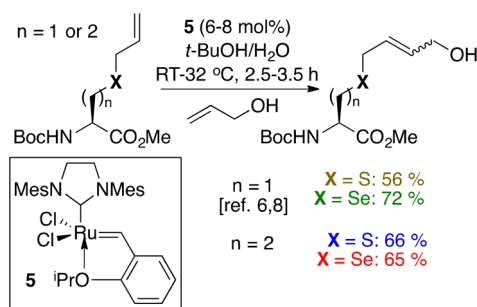
heteroatom (3 and 4). Previous work in our group has demonstrated that the *chemical* installation of Seac in proteins enhances the rate of on-protein CM and allows for an increased breadth of methylation partners compared with Sac.^{8,9} Thus, the metabolic incorporation efficiencies of selenium analogues of 1 and 3 (2 and 4, respectively) were also considered. We reasoned too that since none of these uAAs possess a free side-chain SH/SeH (unlike, e.g., L-homocysteine), they would not be “edited” through conversion to the corresponding chalcogen lactones. 1 and 2 were prepared according to our prior methods.¹⁹ 3 and 4 were synthesized in homochiral form (L-3, L-4) from L-Met and L-SeMet, respectively, using a demethylative allylation strategy that proved to be direct and efficient (Scheme 1). 3 could also be readily synthesized from the commercially available DL-homocysteine thiolactone (DL-3) as a racemate (see the Supporting Information (SI)).

Scheme 1. Synthesis of S-Allylhomocysteine (Ahc, 3) and Se-Allylhomoselenocysteine (Ahs, 4) and Relative Processing of Met and Analogues 1–4 by MetRS Determined through ATP-PPi Exchange; End Point 20 min (See the SI)



First, to evaluate the chalcogen-assisted CM reactivity of these motifs, protected forms of amino acids 1–4 were tested as small-molecule models (Scheme 2) as reported previously.^{6,8} Pleasingly, all proved to be CM-reactive toward allyl alcohol, under aqueous conditions typical^{6–9} of prior successful protein CM reactions in reasonably short reaction times; little or no homodimerization was observed. Encour-

Scheme 2. Cross-Metathesis on the Model Amino Acids Sac (1), Seac (2), Ahc (3), and Ahs (4)



aged by these results, we next explored the incorporation of 1–4 into proteins.

Molecular mechanics analysis was used to explore structural constraints in interactions with and processing by MetRS. Docking (see the SI) of 1–4 and their corresponding adenylates into the active site of *Escherichia coli* MetRS (derived from the Met-bound MetRS structure in Protein Data Bank (PDB) entry 1PG0) used gradient-descent minimization of the energy of conformationally randomized ligands and an AMBER-derived force field (Figure 2 and the SI).²⁰ Pleasingly,

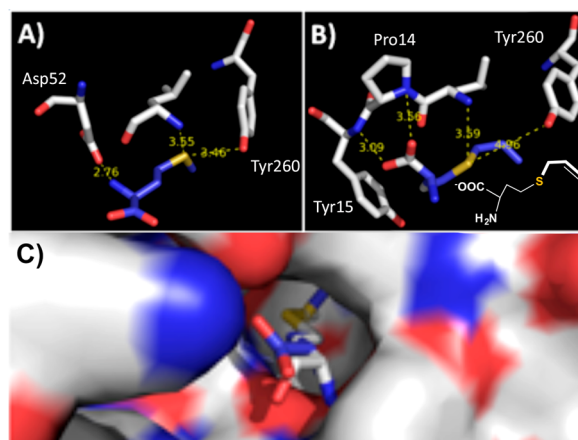


Figure 2. (A, B) Interactions of docked amino acids with MetRS for (A) Met and (B) Ahc (see Figure S1 for further poses of analogues 1–4). (C) Docked Met (blue backbone) adopts a similar pose in the same Met pocket of MetRS identified by crystallography (PDB entry 1PG0, white backbone).

in their minimized poses all of the uAA analogues were found to occupy the Met-binding site of MetRS (Figure 2C). Consistent with previous studies,¹⁶ these revealed major contributors for binding of native Met (e.g., electrostatic interaction of Met-N α with the side chain of Asp52) similar to those known. Notably, in all cases (Met and uAAs 1–4) the heteroatom (S/Se) of the side chain was held by hydrogen bonding to the backbone amide NH of Leu13 at the heart of the Met binding site. However, flexibility in the hydrophobic “end wall” of the site (determined by Tyr260) led to accommodation of a range of side-chain termini (Me or allyl; Figure 2): heteroatom-to-Tyr260-O ω distances were displaced by up to ~0.5 Å. Moreover, primary binding contact with Asp52 was lost or distorted. Notably, for Ahc (3) compensatory interactions were predicted between Ahc-C α (=O)O and both of the backbone amide NHs of Tyr15 and Pro14 (Figure 2B). Together these data highlighted the critical role of side-chain heteroatom positioning and encouragingly suggested sufficient flexibility of MetRS (in accommodating extended termini Me/allyl) as well as compensatory binding modes in certain cases (i.e., Ahc).

To experimentally test these predictions, *E. coli* MetRS was expressed,¹⁹ and the relative activities for loading of Met and analogues 1–4 onto tRNA were determined using ATP-PPi exchange (see Scheme 1 and the SI).²¹ Met, as the natural substrate, proved to be most effective, as expected. However, from the panel of olefinic uAAs 1–4, only the S-allyl uAAs 1 and 3 showed significant activity above background; Ahc (3) showed the greatest activity (~6-fold greater than 1; Scheme 1). Determination and comparison of the Michaelis–Menten

parameters for Met and **3** (Table 1, SI and Figure S2) suggested that they display similar binding to MetRS (as

Table 1. Michaelis–Menten Parameters for MetRS^a

	Met	Ahc (3)
k_{cat} (s ⁻¹)	23.2 ± 2.7	3.1 ± 2.7
K_M (μM)	870 ± 188	1030 ± 235
k_{cat}/K_M (s ⁻¹ μM ⁻¹)	0.026 ± 0.006	0.0030 ± 0.0007

^aAlso see Figure S2 and the SI for further details.

judged by K_M), although **3** has a lower k_{cat} . Thus, despite the clear differences in structure, the turnover (as judged by k_{cat}/K_M) of **3** by MetRS was less than an order-of-magnitude lower than for the native substrate Met (Table 1). Following these in silico and in vitro predictions and validations of putative uAA processing, in vivo incorporation of each Met analogue was investigated. The nucleosome constituent protein histone H3^{22,23} was tested as the first target model protein using the Met-auxotrophic B834(DE3) *E. coli* strain. Efficient incorporation (>95%) was observed only for **3** by both LC–MS and tryptic digest-MS/MS (Figure 3 and the SI).

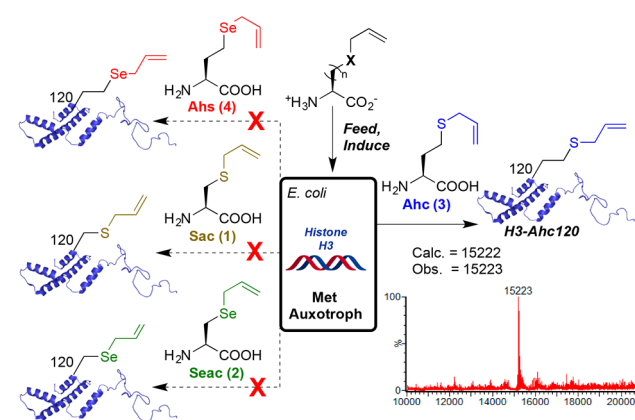


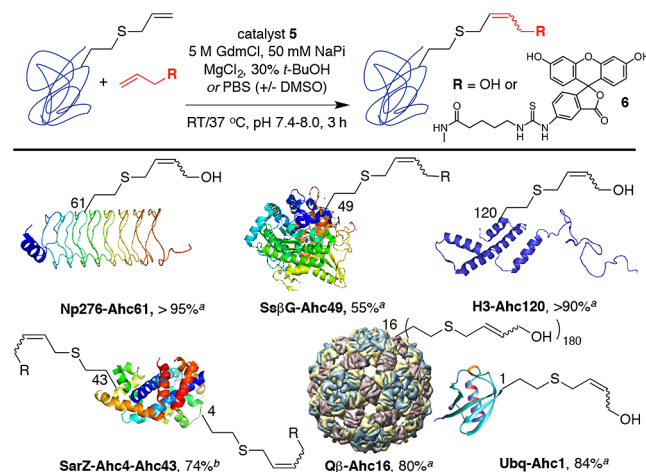
Figure 3. Successful incorporation of Ahc (**3**) into histone H3; Sac (**1**), Seac (**2**), and Ahs (**4**) were not incorporated (see the SI).

These experiments highlighted **3** as the most suitable for genetic incorporation. Proteins of varying structure and function were chosen to test the scope. Thus, in addition to H3, single-site incorporations of Ahc were successfully carried out in right-handed β -helix pentapeptide repeat protein Np276²⁴ and TIM-barrel β -glycosidase Ss β G.²⁵ Excellent levels (>95%) of multisite incorporation of Ahc were also demonstrated in monomeric α -helix bundle DNA-binding protein SarZ²⁶ as well as multimeric bacteriophage coat protein Q β .²⁷ For all proteins, the site and level of incorporation was confirmed by LC–MS of intact protein and tryptic-MS/MS analysis. Ahc-incorporated proteins retained both secondary and tertiary structure (as determined by circular dichroism; see the SI) and function (retention of glycosidase activity by Ss β G and DNA binding by SarZ; see the SI and Figures S6–S8). Typical yields (in mg/L) for Ahc-incorporated proteins were as follows: SarZ, ~0.5 (in Selenomet media), ~2.9 (Selenomet); Ss β G, ~4.5 (Selenomet); IgG-Fc, ~1 (MCQ-free using HEK293T); H3, ~2.4 (Selenomet); Np276, ~1.1 (Selenomet).

With Ahc-containing proteins in hand, their reactivity toward CM was tested; allyl alcohol and fluorescein–olefin **6**

were used as model metathesis partners. Evaluations of reaction conditions and optimizations were conducted through systematic variation of various parameters (Tables S1 and S2) and highlighted the utility of *t*BuOH or DMSO as cosolvents lacking α -protons (where needed), the potential utility of PEG-500 as a cosolvent, the strong benefit of MgCl₂ as an additive,⁶ the utility of phosphate-based buffers and mild denaturant (e.g., through guanidinium) to increase solubility and stability, and the use of Ru scavenging during workup for LC–MS monitoring. Following this process and consistent with the observed reactivity for Ahc in model amino acids, CM was observed on all of the tested Ahc-containing proteins, with conversions ranging from 55% (for bulkier **6**) to >95% for allyl alcohol, including for the intact Q β virus-like particle bearing 180 reaction sites (Scheme 3) and for SarZ, which has two reaction sites. Ss β G was found to retain its functional activity following CM reactions (see Table S4).

Scheme 3. CM on Ahc-Incorporated Proteins: Single-Site Np276, Histone H3, Ss β G, and Ubq and Multisite SarZ and Q β



^aConversion based on MS with allyl alcohol. ^bConversion based on fluorescence assay (Figs S3–S5) with **6**.

Finally, direct incorporation of Ahc into mammalian cells could have useful application in the generation of probes in cellulo and in chemical proteomic strategies (e.g., non-canonical amino acid tagging^{28,29}). We next tested and demonstrated such incorporation using Ahc in human (HEK293T) cells (Figure 4 and the SI). As a test protein from human cells, not only was the installation of Ahc into the Fc region of immunoglobulin G (IgG) determined by both MS and MS/MS (see the SI), but the incorporated Ahc was also found to be reactive, allowing direct CM labeling of IgG-Fc through reaction with olefin–biotin **7**.³⁰ **7** also proved to be effective in reaction with other proteins (e.g., SarZ and Ss β G; see the SI). Given the potential use of biotin tags in affinity proteomic methods,^{28,29} this both highlights the capability of Ahc to serve as a general Met analogue across different translational systems/cell types and also suggests its utility in the future interrogation of human proteomes.

In conclusion, our results demonstrate that the previously unexplored amino acid S-allylhomocysteine (Ahc) is an effective Met surrogate that is incorporated into proteins not only by the translational apparatus of Met-auxotrophic *E. coli* (with good efficiency >95%), thereby allowing genetic control

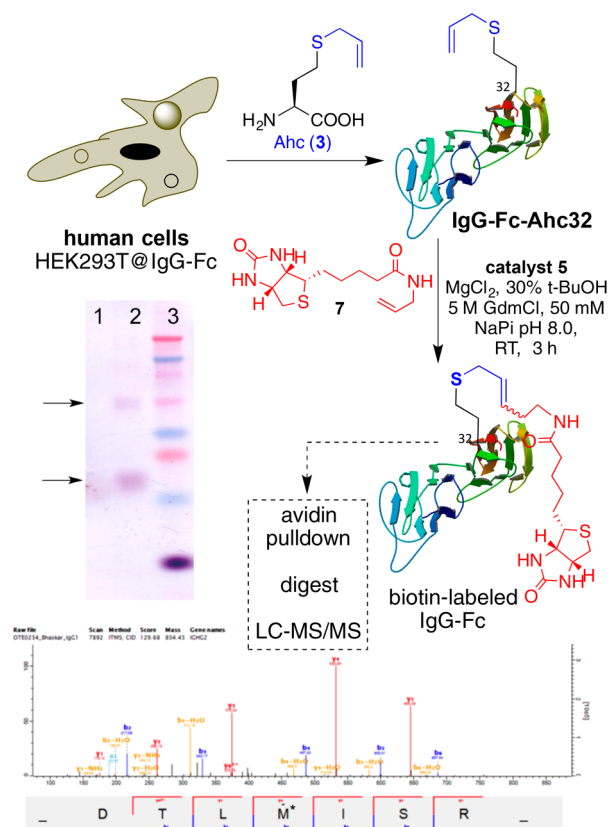


Figure 4. Use of Ahc in human cells and CM labeling of Ahc-incorporated IgG-Fc with olefin–biotin 7. Insets show MS/MS of isolated IgG-Fc_{29–36} peptide DTL-M*-ISR containing analogue M* = 3 and anti-biotin Western (nonreducing, size markers at 20, 30, 40, 50, 60, 80, and 110 kDa; arrows show IgG-Fc monomer and dimer) demonstrating incorporation and CM reaction with 7, respectively. The dotted box shows a potential affinity proteomic workflow.

of olefin cross-metathesis in proteins using sense (Met) codon reassignment, but also even into human cells. Such incorporation could potentially be improved further by the use of MetRS variants.³ It should be noted that there remain some limitations of the CM method; we observed here, for example, that in one case the use of tBuOH to improve the solubility of catalyst 5 led to loss of activity of the highly solvent-sensitive protein SarZ (see SI). In addition, while reactions using 5 typically proceed with strong *E* selectivity,⁹ some *E/Z* heterogeneity may also exist in conjugates. Thus, with the continuing development of water-soluble^{31,32} or *Z*-selective³³ catalysts for cross-metathesis and the genetically controlled capabilities disclosed here, there is now good motivation for the development of effective in vivo Ahc-enabled CM in cellular systems, thereby providing access to a broad range of metal-mediated biological applications.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b09433.

Full procedures and protein ESI-MS (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*ben.davis@chem.ox.ac.uk

ORCID

Tomonari Tanaka: 0000-0002-7050-9043

Shabaz Mohammed: 0000-0003-2640-9560

Benjamin G. Davis: 0000-0002-5056-407X

Present Addresses

§Y.A.L.: Department of Chemistry, National Sun Yat-sen University, No. 70 Lienhai Road, Kaohsiung, Taiwan.

||T.T.: Department of Biobased Materials Science, Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto, Japan.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Rhodes Trust (B.B.), the EPA Cephalosporin Scholarship (B.B.), the EPSRC (Y.A.L., L.L.), CRUK (L.L.), and the Royal Society Wolfson Research Merit Award (B.G.D.) for funding; Florian Seebeck and Ritu Raj for discussions; and James Wickens and Sonia de Munari for technical assistance.

■ REFERENCES

- (1) van Hest, J. C. M.; Tirrell, D. A. Efficient introduction of alkene functionality into proteins in vivo. *FEBS Lett.* **1998**, *428*, 68.
- (2) Song, W.; Wang, Y.; Yu, Z.; Vera, C. I. R.; Qu, J.; Lin, Q. A Metabolic Alkene Reporter for Spatiotemporally Controlled Imaging of Newly Synthesized Proteins in Mammalian Cells. *ACS Chem. Biol.* **2010**, *5*, 875.
- (3) Ngo, J. T.; Tirrell, D. A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.* **2011**, *44*, 677.
- (4) van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *J. Am. Chem. Soc.* **2000**, *122*, 1282.
- (5) Gibson, S. E.; Gibson, V. C.; Keen, S. P. Cross metathesis of the amino acid homoallylglycine. *Chem. Commun.* **1997**, 1107.
- (6) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J.; Davis, B. G. Allyl sulfides are privileged substrates in aqueous cross-metathesis: application to site-selective protein modification. *J. Am. Chem. Soc.* **2008**, *130*, 9642.
- (7) Chalker, J. M.; Lin, Y. A.; Boutureira, O.; Davis, B. G. Enabling olefin metathesis on proteins: chemical methods for installation of S-allyl cysteine. *Chem. Commun.* **2009**, 3714.
- (8) Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G. Rapid cross-metathesis for reversible protein modifications via chemical access to Se-allyl-selenocysteine in proteins. *J. Am. Chem. Soc.* **2013**, *135*, 12156.
- (9) Lin, Y. A.; Chalker, J. M.; Davis, B. G. Olefin cross-metathesis on proteins: investigation of allylic chalcogen effects and guiding principles in metathesis partner selection. *J. Am. Chem. Soc.* **2010**, *132*, 16805.
- (10) Ai, H. W.; Shen, W.; Brustad, E.; Schultz, P. G. Genetically encoded alkenes in yeast. *Angew. Chem., Int. Ed.* **2010**, *49*, 935.
- (11) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19.
- (12) Duewel, H.; Daub, E.; Robinson, V.; Honek, J. F. Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein 19F NMR. *Biochemistry* **1997**, *36*, 3404.
- (13) Bogosian, G.; Violand, B. N.; Dorwarding, E. J.; Workman, W. E.; Jung, P. E.; Kane, J. F. Biosynthesis and Incorporation into Protein of Norleucine by *Escherichia coli*. *J. Biol. Chem.* **1989**, *264*, 531.
- (14) Budisa, N. Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew. Chem., Int. Ed.* **2004**, *43*, 6426.

(15) Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One* **2007**, *2*, e972.

(16) Datta, D.; Vaidehi, N.; Zhang, D.; Goddard, W. A. Selectivity and specificity of substrate binding in methionyl-tRNA synthetase. *Protein Sci.* **2004**, *13*, 2693.

(17) Fersht, A. R.; Dingwall, C. An editing mechanism for the methionyl-tRNA synthetase in the selection of amino acids in protein synthesis. *Biochemistry* **1979**, *18*, 1250.

(18) Jakubowski, H. Proofreading in vivo: editing of homocysteine by methionyl-tRNA synthetase in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 4504.

(19) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **2001**, *19*, 751.

(20) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785.

(21) Ghosh, G.; Pelka, H.; Schulman, L. H. Identification of the tRNA anticodon recognition site of *Escherichia coli* methionyl-tRNA synthetase. *Biochemistry* **1990**, *29*, 2220.

(22) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z.-W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000**, *406*, 593.

(23) Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **2007**, *128*, 693.

(24) Vetting, M. W.; Hegde, S. S.; Hazleton, K. Z.; Blanchard, J. S. Structural characterization of the fusion of two pentapeptide repeat proteins, Np275 and Np276, from *Nostoc punctiforme*: Resurrection of an ancestral protein. *Protein Sci.* **2007**, *16*, 755.

(25) Aguilar, C. F.; Sanderson, I.; Moracci, M.; Ciaramella, M.; Nucci, R.; Rossi, M.; Pearl, L. H. Crystal structure of the β -glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* **1997**, *271*, 789.

(26) Kaito, C.; Morishita, D.; Matsumoto, Y.; Kurokawa, K.; Sekimizu, K. Novel DNA binding protein SarZ contributes to virulence in *Staphylococcus aureus*. *Mol. Microbiol.* **2006**, *62*, 1601.

(27) Kozlovskaya, T. M.; Cielēns, I.; Dreilīņa, D.; Dišlers, A.; Baumanis, V.; Ose, V.; Pumpēns, P. Recombinant rna phage $q\beta$ capsid particles synthesized and self-assembled in *Escherichia coli*. *Gene* **1993**, *137*, 133.

(28) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9482.

(29) Howden, A. J. M.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **2013**, *10*, 343.

(30) Wittrock, S.; Becker, T.; Kunz, H. Synthetic Vaccines of Tumor-Associated Glycopeptide Antigens by Immune-Compatible Thioether Linkage to Bovine Serum Albumin. *Angew. Chem., Int. Ed.* **2007**, *46*, 5226.

(31) Hong, S. H.; Grubbs, R. H. Highly Active Water-Soluble Olefin Metathesis Catalyst. *J. Am. Chem. Soc.* **2006**, *128*, 3508.

(32) Skowerski, K.; Szczepaniak, G.; Wierzbicka, C.; Gulajski, L.; Bieniek, M.; Grela, K. Highly active catalysts for olefin metathesis in water. *Catal. Sci. Technol.* **2012**, *2*, 2424.

(33) Herbert, M. B.; Grubbs, R. H. Z-Selective Cross Metathesis with Ruthenium Catalysts: Synthetic Applications and Mechanistic Implications. *Angew. Chem., Int. Ed.* **2015**, *54*, 5018.