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

Bhaskar Bhushan,† Yuwa A. Lin,†§ Martin Bak,† Anuchit Phanumartwiwath,† Nan Yang,† Matthew K. Bilyard,† Tomonari Tanaka,†‡,§ Kieran L. Hudson,† Lukas Lercher, Monika Stegmann,† Shabaz Mohammed,† Yuya A. Lin,† Tomonari Tanaka,† Monika Stegmann,† Shabaz Mohammed,† and Benjamin G. Davis*†‡

1Department of Chemistry, University of Oxford, Oxford OX1 3QU, U.K.
2Department 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). However, although aliphatic alkene-containing amino acids such as homosallylglycine (Hag) are reactive in both self-metathesis and cross-metathesis (CM) reactions as monomeric, protected amino acids in organic solvents and can be metabolically incorporated into proteins, they are unreactive in CM reactions in aqueous media. On the other hand, chemically installed uAAs such as S-allylcysteine (Sac) and Se-allylselenocysteine (Seac) 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. Schultz and co-workers have reported the incorporation of O-crotylseline in yeast via amber stop codon suppression and have demonstrated its reactivity in on-protein intramolecular ring-closing metathesis. 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, homoproparglyglycine, and azidohomoalanine) as Met analogues. 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. 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). Fersht and Dingwall first demonstrated that a Met analogue, l-ethionine, may be mischarged by theMetRS without being metabolically edited. By contrast, l-homocysteine, a natural competitor of Met, is transformed to the corresponding thiolactone to avoid misreading in translation. 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-...
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 metathesis partners compared with Sac.\textsuperscript{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.\textsuperscript{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 readily synthesized from the commercially available DL-homocysteine thiolactone (DL-3) as a racemate (see the Supporting Information (SI)).

Processing of Met and Analogues 1

Supporting Information (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.\textsuperscript{6,8} Pleasingly, all proved to be CM-reactive toward allyl alcohol, under aqueous conditions typical of prior successful protein CM reactions in reasonably short reaction times; little or no homodimerization was observed. Encouraged 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 AMBER-derived force field (Figure 2 and the SI).\textsuperscript{20} Pleasingly, 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,\textsuperscript{16} these revealed major contributors for binding of native Met (e.g., electrostatic interaction of Met-Na 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\textsubscript{w} 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-Cα 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,\textsuperscript{17} 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).\textsuperscript{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 judged by $K_D$), 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 candidates, we next tested and demonstrated such incorporation using Ahc in human cells, not only was the installation of Ahc into the protein SarZ, which has two reaction sites. Ssβ/G was found to retain its functional activity following CM reactions (see Table S4).

Table 1. Michaelis–Menten Parameters for MetRS$^a$

<table>
<thead>
<tr>
<th></th>
<th>Met (3)</th>
<th>Ahc (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ ($s^{-1}$)</td>
<td>23.2 ± 2.7</td>
<td>3.1 ± 2.7</td>
</tr>
<tr>
<td>$K_M$ ($\mu M$)</td>
<td>870 ± 188</td>
<td>1030 ± 235</td>
</tr>
<tr>
<td>$k_{cat}/K_M$ ($s^{-1}$ $\mu M^{-1}$)</td>
<td>0.026 ± 0.006</td>
<td>0.0030 ± 0.0007</td>
</tr>
</tbody>
</table>

$^a$Also see Figure S2 and the SI for further details.

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 $\beta$-helix pentapeptide repeat protein Np276$^{24}$ and TIM-barrel $\beta$-glycosidase Ssβ/G.$^{25}$ Excellent levels (>95%) of multisite incorporation of Ahc were also demonstrated in monomeric $\alpha$-helix bundle DNA-binding protein SarZ$^{26}$ as well as multimeric bacteriophage coat protein QB.$^{27}$ For all proteins, the site and level of incorporation was confirmed by LC−MS of intact protein and tryptic-MS/MS (see the SI). Typical yields (in mg/L) for Ahc-incorporated proteins were as follows: SarZ, ∼0.5 (in Selenomut media), ∼2.9 (Selenomet); Ssβ/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 $tBuOH$ or DMSO as cosolvents lacking $\alpha$-protons (where needed), the potential utility of PEG-500 as a cosolvent, the strong benefit of MgCl$_2$ as an additive,$^b$ 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 Qβ virus-like particle bearing 180 reaction sites (Scheme 3) and for SarZ, which has two reaction sites. Ssβ/G was found to retain its functional activity following CM reactions (see Table S4).


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$^{30,31}$ also proved to be effective in reaction with other proteins (e.g., SarZ and Ssβ/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
of olefin cross-metathesis in proteins using sense (Met) codon reassignment, but also even in human cells. Such incorporation could potentially be improved further by the use of MetRS variants.\textsuperscript{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,\textsuperscript{9} some $E/Z$ heterogeneity may also exist in conjugates. Thus, with the continuing development of water-soluble\textsuperscript{31,32} or $Z$-selective\textsuperscript{33} catalysts for cross-metathesis and the genetically controlled capabilities disclosed here, there is now good motivation for the development of effective in vivo Ace-enabled CM in cellular systems, thereby providing access to a broad range of metal-mediated biological applications.

\section*{ASSOCIATED CONTENT}

\begin{itemize}
  \item \textbf{Supporting Information}
  \begin{itemize}
    \item The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b09433.
  \end{itemize}
  \item \textbf{AUTHOR INFORMATION}
  \begin{itemize}
    \item Corresponding Author
      *ben.davis@chem.ox.ac.uk
  \end{itemize}
\end{itemize}


(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.


