Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Site-selective covalent reactions on proteinogenic amino acids

Wing Ho So[1,3], Yu Zhang[1,3], Wei Kang[1,3], Clarence TT Wong[1], Hongyan Sun[2] and Jiang Xia[1]

To achieve precise control of the signaling events or to achieve unmistakable synthesis of biomolecules, nature has evolved organic reactions involving proteinogenic amino acids with unparalleled site selectivity. For example, dedicated enzymes accurately dictate the site of post-translational modifications in signaling proteins, and ribosomes precisely link the C-terminal carboxylic acid of one unprotected amino acid with the N-terminal amino group of the other amino acid through spatially confined proximity. For many years, chemists have been striving to achieve site selectivity on biomolecules by mimicking nature. Driven by the development of chemoselective protein conjugation reactions, enzymology and protein–protein interactions, the past decade has witnessed a boom in site-selective protein conjugation reactions. (In this review, a site-selective protein conjugation reaction is defined as an organic reaction that targets *a single amino acid* instead of a kind of amino acids in a protein or a proteome under physiological conditions, for example, a single cysteine residue among all of the cysteines.) In this review, we summarize the recent advancements of bioconjugation reactions that demonstrate this feature of precise site selectivity, focusing on the reactions of the proteinogenic amino acids (excluding those at non-coded or non-proteinogenic amino acids that are introduced to proteins through genetic manipulations).

Addresses
[1] Department of Chemistry, The Chinese University of Hong Kong, Shatin, Hong Kong Special Administrative Region
[2] Department of Biology and Chemistry, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region

Corresponding authors: Wong, Clarence T.T. (ClarenceTTW@cuhk.edu.hk), Sun, Hongyan (hongysun@cityu.edu.hk) and Xia, Jiang (jiangxia@cuhk.edu.hk)
[3] These authors contributed equally to this work.

Current Opinion in Biotechnology 2017, **48**:220–227

This review comes from a themed issue on **Pharmaceutical biotechnology**

Edited by **Tiangang Liu** and **Chu-Young Kim**

http://dx.doi.org/10.1016/j.copbio.2017.06.003

Introduction

The emergence of chemoselective protein modifications can be traced back to the first half of the twentieth century. In a pioneering example, Bender and coworkers synthesized an active enzyme 'thiol-subtilisin' by converting the serine residue in the active site of subtilisin (a serine protease) to a cysteine. The hydroxyl group on the serine first undergoes a nucleophilic attack with phenylmethanesulfonyl fluoride. Displacement reaction with thioacetate followed by hydrolysis gives a cysteine residue [1]. Since then, elegant organic reactions have been developed for reactions with non-enzymatic proteins, or to convert proteinogenic amino acids to chemical handles for biorthogonal reactions under physiological conditions. On the other hand, biochemists have learned to modify enzymes to achieve site-selective protein conjugation reactions. Using another route, chemists incorporated non-proteinogenic/unnatural amino acids with orthogonal reactivity to proteins through genetic manipulations and markedly expanded the chemical space of protein conjugation reactions: biorthogonal reactions absent in nature ensure unsurpassed precision of the reactions. For example, in a three-step approach, O-phosphoserine (Sep)-containing recombinant protein was generated through an orthogonal translation system. The Sep residue can then be selectively converted to dehydroalanine. An alkylation reaction then installs different posttranslational modifications at the dehydroalanine site [2]. Technological developments of genetic incorporation and biorthogonal reactions of non-proteinogenic chemical groups have been thoroughly reviewed elsewhere [3–10]. This review will focus on the site-selective reactions on *natural proteinogenic* amino acids. Here, we distinguish residue-specific chemoselective reactions versus site-selective reactions by defining the former as the reactions with one particular type of chemical groups (for example all cysteines), and the latter as those that target a single residue among its kind (for example, reactions with only one selected cysteine among all the cysteines in a protein).

In the first part of this review, we summarize residue-specific chemoselective protein conjugation reactions, because the invention of chemoselective reactions fuels the development of site-selective reactions. In the second part, we review the strategies to achieve site selectivity by harnessing the catalytic power of enzymes. Finally, we discuss affinity-guided (a.k.a. ligand-guided) strategies to achieve site-selective protein conjugation reactions based on the principle of proximity-induced reactivity.

Part I. Chemoselective protein conjugation reactions

From single cysteines to di-and tetra-cysteines

Classical protein conjugation reactions rely on the chemical properties of the side chains of individual residues. Twenty proteinogenic amino acids offer access to functional groups ranging from thiol to indole. In contrast to the harsh conditions of most organic reactions, protein conjugation reactions are often limited to aqueous buffer and ambient temperature. Most of the protein conjugation reactions are polar reactions involving nucleophilic side chains and electrophilic modifiers. Take cysteine, the most nucleophilic residue among all proteinogenic amino acids as an example. Chemoselective cysteine-specific reactions are the classical method of protein labeling [2–9,11] (Figure 1a). Nevertheless, new reagents are being reported: a recent example shows the use of hypervalent iodine reagents to react with cysteines through the concept of 'umpolung' [12]. Another cysteine-reactive reagent is dibromomaleimide, a dual Michael acceptor capable of reacting sequentially with two different thiol molecules [13]. Similar dibromomaleimide molecules can staple the di-cysteine moieties from the reduced disulfide bond [14•] (Figure 1b). On the other hand, the FlAsH and ReAsH labeling reagents react specifically with a tetracysteine motif in the peptide sequence CCPGCC in proteins, but not with the individual thiol groups or disulfide bonds [15••] (Figure 1c). From single cysteines to di-cysteines and to tetracysteines, the increasing sequence requirement demonstrates one strategy for enhancing site selectivity.

N-terminal cysteine reactions

An N-terminal cysteine harbors a unique 1-amino, 2-thiol moiety with special chemical reactivity distinct from internal thiols. Taking advantage of this uniqueness, chemists have developed native chemical ligation (NCL) to build a native amide bond between two peptide chains [16•] (Figure 1d). A golden standard in protein semi-synthesis, the NCL method has inspired the development of terminal selective conjugation chemistry [17] (Figure 1e). N-terminal cysteine also reacts with aldehydes to give a thiazolidine ring, and with a cyanobenzothiazole group to give a stable linkage for protein labeling and surface immobilization [18,19] (Figure 1d). Recently, a minimalist probe 4F-2CN was reported to react with N-terminal cysteine to give a cyclized product with green fluorescence that can find useful applications in protein labeling and bioimaging [20]. Collectively, these reactions represent another category of site-selective cysteine reactions that confine the reactive site to N-terminal residues.

Chemoselective reactions for lysine, arginine, methionine and tyrosine in proteins have also been developed based on the chemical properties of the side chain groups [21–30] (Figure 1f–I). Although one could harness the stoichiometry of protein labeling reactions by controlling the number of solvent-accessible reactive residues on protein surface, reactions driven by chemical reactivity lack an intrinsic mechanism to achieve site selectivity on a protein. Nevertheless, the advancement of chemoselective conjugation reactions provides the chemical foundation for the site-selective protein conjugation reactions described below.

Part II. Enzymatic reactions to control site selectivity

Enzymes are naturally selected catalysts with exceptional specificity on their substrates. The combination of biological recognition and proximity-induced chemistry yields reactions that are site-specific both at the active site of the enzymes and the substrates or cofactors. In fact, competitive covalent enzyme inhibitors react only with the active site residue of the enzyme. Protein chemists have therefore borrowed the knowledge from enzyme biochemistry to devise site-specific reactions for protein covalent labeling and other uses.

Enzyme as a protein tag

The first set of enzymatic labeling reactions utilize the enzyme itself as the tag. For example, an enzyme is fused to the protein of interest. An enzyme-specific probe can then be engineered from the covalent inhibitors by attaching an indicator moiety (such as a fluorescent molecule or a biotin) at one position of the inhibitor that does not interfere with the reaction between the probe and enzyme tag. Most reactions involve a nucleophilic attack of the active site residue toward the warhead of the inhibitor, resulting in a stable covalent linkage between the inhibitor and the enzyme (the covalent linkages are often buried so that they are not easily hydrolyzed). Another characteristic of these enzymes is that they normally have an open substrate-binding channel, enabling the inhibitors to carry one label tag (a fluorescent molecule or a biotin) at the distal end of the reactive warhead. Examples in this category include the SNAP/CLIP tags, Halo tag, and β lactamase tag [31•,32,33] (Figure 2a).

Enzyme as a catalyst

The second strategy to achieve site-selective protein reactions is to engineer the peptide substrate of the enzymatic reaction as a protein tag, and utilize the enzyme to assist the transfer of a small molecule to the peptide substrate tag at a selected site (Figure 2b). For example, a biotin ligase BirA was used to covalently transfer a biotin ketone analogue to a specific lysine residue in a ligase substrate fused to other proteins of interest [34]. The amide-linked ketone moiety can then be site-selectively labeled with ketone-reactive probes. Chemists also engineered a lipoic acid ligase to render it acceptable to a reactant derivative through an amide bond [35]. The same strategy has found success in a