Site-specifically modified fusion proteins for molecular imaging

Hui Wang1, Xiaoyuan Chen1

1The Molecular Imaging Program at Stanford (MIPS), Department of Radiology and Bio-X Program, Stanford University School of Medicine, Stanford, California

TABLE OF CONTENTS

1. Abstract
2. Introduction to molecular imaging probes
3. Chemical modification of protein probes for molecular imaging
4. Site-specific mutagenesis modification of protein probes
5. Fusion proteins for site-specific modification
   5.1. Cys-tag
   5.2. Chelator-tag
   5.3. BirA-based site-specific biotinylation
   5.4. Sfp phosphopantetheinyl transferase-based site-specific modification
   5.5. Formylglycine-generating enzyme-based site-specific modification
   5.6. The hAGT based site-specific modification
   5.7. Transglutaminase-based site-specific modification
   5.8. Intein-based site-specific modification
6. Conclusion and perspectives
7. References

1. ABSTRACT

To visualize and quantify specific molecular pathways in vivo, it is essential to develop molecular imaging probes with high target affinity and specificity. The introduction of radioisotope, fluorophores and other detectable labels onto a protein ligand in a site-specific manner without compromising its function and binding affinity has been a valuable tool for such molecular imaging. Site-specific labeling of proteins poses an enormous challenge, because the correct functional groups of one protein amidst all of the other expressed proteins containing the same range of amino acids must be accurately targeted. Chemists have developed many diverse strategies for accomplishing site-specific modification; unnatural amino acids have been incorporated into targeted protein for site-specific modification. Enzymatic post-translational modification of proteins is an alternative approach and beneficial complement for site-specific labeling. A small peptide tag is a substrate for the enzyme that can be appended to N-, C-terminus or internal of target protein for site-specific labeling. In this review, we will discuss several valuable tags for addressing this problem, and comment on the potential usage of tagged fusion protein for molecular imaging.

2. INTRODUCTION TO MOLECULAR IMAGING PROBES

In the past decade, numerous novel treatments have aimed to prolong survival, induce remission, and provide better quality of life for cancer patients. Among the successful examples are the molecularly targeted cancer treatment drugs such as the anti-HER-2/neu antibody trastuzumab for Erb-2 expressing breast cancers (1), and the kinase inhibitor imatinib (Glivec®) for chronic myelogenous leukemia and gastrointestinal stromal tumors (2). Despite their efficacy in certain settings, molecularly targeted drugs are resource-intensive to develop, being expensive and requiring protracted time, and numerous patients during development, which result in the inevitably high cost of the approved therapy (3). Moreover, molecular targets have been identified and characterized in relatively few oncology patients. Improving the phenotypic characterization of the underlying molecular lesions would therefore expand the overall effect of targeted agents in oncology.

The development of targeted therapy strategy highlights the need for a faster, more efficient and cost-effective development to better define patients likely to
Site-specifically modified fusion proteins for molecular imaging

benefit from treatment. As addressed by the recent Food and Drug Administration (FDA) Critical Path Initiative, collaborative interactions among such scientific knowledge areas as bioinformatics, genomics, materials science, and imaging technologies are needed to design and implement better drug development tools (3). Important among these tools are molecular imaging probes that image specific molecular pathways in vivo, enabling visualization of phenotypic expression of key targets in cancer disease processes.

Probes to image molecular targets have included small molecules, peptides, proteins, and antibodies labeled with radionuclides for positron emission tomography (PET) and single photon emission computed tomography (SPECT) imaging, near-infrared (NIR) fluorescence imaging, molecular magnetic resonance imaging (mMRI) that employs paramagnetic and superparamagnetic-based moieties, and microbubbles for contrast enhanced ultrasound (CEU).

Despite their promise for characterizing targets and the functional consequences of drug-target interactions, the development of molecular imaging probes has faced several significant hurdles. For example, the targets are expected to have low (nanomolar to micromolar) concentrations; thus, adequate intracellular delivery and acceptable signal-to-noise amplification are necessary. Furthermore, the molecular imaging probe should have only a trace level of mass in tissues of the target(s) so as not to exert mass or pharmacologic effects. In addition, the molecular imaging probe should have very high target affinity and specificity so that the image contrast will truly reflect the receptor density or enzyme activity found in target organ or tissue. This review article focuses on protein based molecular imaging probes and the importance of site-specific labeling to maintain the receptor binding affinity and specificity after appropriate modification, bioconjugation, or labeling.

3. SITE-SPECIFIC CHEMICAL MODIFICATION OF PROTEIN PROBES

Fluorophores, isotopes, and other detectable labels are most commonly introduced into a protein through in vitro modifications with suitable amine or thiol reactive reagents. Proteins typically contain multiple amines. Even with the reduced pKₐ of the amino-terminus of the protein compared to the lysine side chain, protein labeling is generally nonspecific. Cysteine labeling is typically more specific than amine labeling as many natural cytosolic proteins lack cysteine residues and a single cysteine can be added by site-directed mutagenesis without affecting the function of the protein. However, cysteine-based methods alone cannot easily perform multiple modifications to a protein (such as introducing two fluorophores for fluorescence resonance energy transfer (FRET) analysis) except when the reactivities of the two cysteines are quite different, and not all protein targets allow for the introduction of cysteine without impairing function. Recent work by Francis et al. has identified reactions that can specifically target tryptophan(4) and tyrosine(5) residues. Using rhodium carbenoids, the selective modification of the indole functionality in tryptophan (at N-1 or C-2) can be achieved. When applied to myoglobin or subtilisin Carlsberg (with 2 and 1 surface tryptophan, respectively), modification occurs exclusively at the tryptophan residues in about 50-60% yield. The results do suggest that transition metal complexes can be used for specific bioconjugation of aromatic residues in proteins. However, this reaction requires low pH (1.5-3.5) and an organic cosolvent to solubilize the organic reagents, both of which may denature the target protein. The same group also developed conditions to perform a Mannich-type of reaction to selectively target tyrosine residues using aldehyde and aniline at pH 6.5(5). By attaching a fluorophore to the aniline, conjugation of rhodamine to chymotrypsinogen A is achieved without altering the protein’s activity. Another strategy for selective tyrosine modification using diazonium salts has been applied to modify the exterior of the rodlike tobacco mosaic virus. The addition of a variety of groups, including biotin, fluorophores, and crown ethers modulates the properties of this protein scaffold.

Site-specific chemical modification is strictly defined as a process which yields a stoichiometrically altered protein with the quantitative covalent derivatization of a single, unique amino acid residue without either modifying the other amino acid residue or changing the conformation of the protein. In practice, this objective is rarely obtained because of several major problems. First, only very few reagents are specific for the modification of a single functional group. Second, even with a reagent that appears to be functional group specific, the modification of only one of the several residues within a functional group class is unlikely unless the specific residue is uniquely reactive. Finally, it is also unlikely that the site-specific modification of a protein can be achieved without any conformational change.

4. SITE-SPECIFIC MODIFICATION VIA INCORPORATED UNNATURAL AMINO ACID IN TARGET PROTEIN

Biological methods such as site-directed mutagenesis can be used to systematically alter proteins at the genetic level. Through unnatural amino acid mutagenesis, chosen modifications can be added to proteins during their cellular synthesis. As far as we know, all known organisms specify the same 20 amino acids (with the rare exceptions of selenocysteine (7) and pyrrolysine (8)). Although a 20-amino-acid code might be sufficient for life, it might not be optimal. The methodology to incorporate amino acid with defined steric and electronic properties at unique sites in proteins will provide powerful new tools for exploring protein structures and functions in vitro and in vivo. To-date, over 30 unnatural amino acids (Figure 1)-including those containing spectroscopic probes, post-translational modifications, metal chelators, photoaffinity labels and other chemical moieties- have been selectively incorporated into proteins with high fidelity and efficiency in response to unique three and four base codons (9). For example, Ohno et al. (10) utilized the yeast amber suppressor tRNA^177/mutated tyrosyl synthetase pair as a
5. Fusion proteins for site-specific modification

Recombinant DNA construction of protein fusions in which the protein segment has a dedicated site for labeling or could be recognized by a coenzyme ligase fused to a protein of interest, may offer distinct advantages over conventional protein fusions and chemical or mutagenesis-based site-specific modification of target protein. Some obvious benefits of the fusion protein approach over chemical modification and site-specific mutagenesis include the following: (1) the fusion proteins can be specifically labeled by growth of cells cultures in the presence of labeled coenzymes, or be site-specifically labeled under certain mild conditions; (2) if the coenzyme possesses specific binding toward an immobilized ligand, the fusion protein can be readily purified, perhaps in a native form; (3) the protein segment that is fused to the protein of interest would be much smaller than those commonly used, with less alteration of the chemical and biological properties of the protein of interest; and (4) the tag is usually fused to the N- or C-terminus of the targeted protein, with minimum effect on the protein structure and function.
Site-specifically modified fusion proteins for molecular imaging

The addition of unnatural moieties, such as fluorophores, affinity labels, spin-label probes, isotope, and bioorthogonal functional groups to proteins has also proven useful for a variety of processes and applications both in vitro and in vivo, including molecular imaging. In this review, we will focus on the strategies for constructing fusion proteins containing small peptide tag for site-specific modification as a beneficial complementary of the abovementioned methods and the potential application of such fusion protein in molecular imaging.

5.1. Cys-tag

As previously mentioned, thiol-reactive reagents such as the maleimide-based reagents and the haloacetic-based reagents are somewhat more selective because of the lower amounts of cysteine residues in peptides and proteins. In cases where all thiol groups are involved in disulfide bond formation, one can introduce a single surface-accessible thiol into the target protein by site-directed mutagenesis.(11)

Introducing N-terminus or C-terminus cysteine residue for site-specific labeling using thiol-specific reagents provides an alternative approach. Olfson et al. (12) have previously produced an anti-carcinoembryonic antigen (anti-CEA) diabody assembling V L-eight amino acid linker-VH. Following radiolabeling at random sites on the protein with radiodiode or radiometals, this fragment exhibits rapid tumor targeting in a nude mouse/LS174T human colon carcinoma xenograft model in biodistribution and imaging studies (13). In order to allow site-specific radiolabeling using thiol-specific reagents, they (12) described four mutant anti-CEA diabodies engineered by substitution or addition of unique cysteine residues. Four variants of anti-CEA were constructed, expressed and purified, two of which (with the C-terminal sequence – LGGC or –SGGC) were found to exist as a stable disulfide-purified, two of which (with the C-terminal sequence – SGGC) were found to exist as a stable disulfide.

Other tags such as optical or fluorescent probes can provide a facile and reliable method for permanent deposition of functionally active growth factors on synthetic or protein scaffolds that has applications for advanced tissue engineering. Furthermore, to image the VEGFRs expression in angiogenic vasculature with the single-chain VEGF, this cysteine-tagged single-chain VEGF was site-specifically labeled with contrast agents for near-infrared fluorescence imaging, single-photon emission computed tomography or photon emission tomography (18). These probes retained VEGF activity and receptor binding ability in vitro. Imaging of orthotopic mouse mammary 4T1 carcinomas revealed that 99mTc-HYNIC-C-VEGF fully retained VEGF bioactivity and receptor binding ability in vivo. Imaging of orthotopic mouse mammary 4T1 carcinomas revealed that 99mTc-HYNIC-C-VEGF was preferentially accumulated at the tumor rim, where the most extensive angiogenesis takes place. They also reported a similarly enhanced circumferential uptake with near-infrared fluorescence and positron emission tomography (PET) imaging of subcutaneous 4T1 tumors (16). Using the same 15-amino acid Cys-tag, Backer et al. (17) expressed a fusion protein containing N-terminal Cys-tag and single chain VEGF (Cys-tag scVEGF); this fusion protein was conjugated to fibroenectin (FN) through a common thiol-directed bi-functional cross-linking agent. FN-scVEGF conjugates, when coated on cell culture plate, retained soluble VEGF activity on stimulating VEGFR2 expressing 293/KDR cell growth. These results demonstrated that site-specific immobilization via Cys-tag can provide a facile and reliable method for permanent deposition of functionally active growth factors on synthetic or protein scaffolds that has applications for advanced tissue engineering. Furthermore, to image the VEGFRs expression in angiogenic vasculature with the single-chain VEGF, this cysteine-tagged single-chain VEGF was site-specifically labeled with contrast agents for near-infrared fluorescence imaging, single-photon emission computed tomography or photon emission tomography (18). These probes retained VEGF activities in vitro and underwent selective and highly specific focal uptake into the vasculature of tumors and surrounding host tissue in vivo. The fluorescence contrast agent showed long-term persistence and co-localized with endothelial cell markers, indicating that internalization was mediated by the receptors. The multimodal imaging of VEGF receptors with these probes will be useful for clinical diagnosis and therapeutic monitoring, and will help accelerate the development of new angiogenesis-directed drugs and treatments.

Besides its usefulness in site-specific radiolabeling, fluorophore coupling, and PEGylation via Cys-tag incorporated into target protein, another interesting application of Cys-tag is to introduce multiple, covalent modifications involving several different functionalities in a site-specific manner (Figure 2). Cysteines that are oxidized in a disulfide bridge, or coordinated to a metal can be protected from covalent modification by thiol-reactive
Site-specifically modified fusion proteins for molecular imaging

Another strategy is to introduce two functional labels into the target protein by combining chemical and enzymatic modifications. Transglutaminase (TGase) catalyzes the acyl transfer reaction between the γ-carboxyamide group of an acceptor glutamine (Gln) residue and a primary amine donor (22). This TGase-base site-specific modification of protein will be discussed in the following part of this review. Jager et al. (23) combined traditional chemical thiol labeling with enzymatic modification of an engineered, reactive Gln by TGase to obtain highly pure, stoichiometrically D/A-labeled protein. In this study, a 64-residue, single domain serine protease inhibitor, chymotrypsin inhibitor 2 (C12), served as a model system. A unique Cys residue (Cys400 was introduced at a solvent-exposed position into C12 for site-specific labeling with Alexa488-maleimide (A488 hereafter). A substrate sequence-recognition tag (Pro-Lys-Pro-Gln-Gln-Phe) for tissue TGase was appended at the N-terminus of C12. TGase then catalyzed the acyl transfer reaction between the primary amine of Alexa746 cadaverine (A647-(CH 2)6-NH 2, A647 hereafter) and the γ-carboxyamide moiety of the reactive Gln (Gln4) in the TGase tag. The three internal Gln were imbedded into a rigid tertiary structure context not recognized as substrate for tissue TGase. These results showed that chemical labeling of cysteine, followed by enzymatic modification of reactive Gln in this fusion protein made stoichiometrically D/A labeled substrate suitable for single-molecule FRET experiments. Thermodynamic data indicated that neither the presence of TGase tag nor D/A labeling perturbed protein stability. This two-step chemical/enzymatic labeling may thus present a simple, low-cost, and widely available strategy for the D/A labeling of protein in FRET-based single-molecule protein folding studies.

The strategies to introduce two or more functional group into target protein in a site-specific manner may also provide a platform to site-specifically double label a target protein with amine-reactive reagents and sulfhydryl-reactive agents simultaneously for dual-modality molecular imaging. For example, amine reactive chelator DOTA can be conjugated to one site and subsequently labeled with
Site-specifically modified fusion proteins for molecular imaging

Figure 3. Examples of short amino acid sequences that form stable \([^{99m}\text{Tc}]\text{O}^{3+}\) complexes. Reproduced with permission from (27).

Cu-64 for PET imaging, whereas Cy5-maleimide can be conjugated to a thiol reactive site for optical imaging. This will be a complement the current multimodality imaging that uses a triple/dual fusion reporter gene strategy. However, introducing unpaired engineered cysteines is problematic for routine production and storage, because it can interfere with the disulfides in the target protein and cause misfolding, leading to poor production yields and covalent aggregations. Lastly, trans-chelation to other sulfur-containing ligands has been shown to be an important route of radiopharmaceutical instability (24).

5.2. Chelator tag

\(^{99m}\text{Tc}\) is a widely used isotope for SPECT and scintigraphic imaging. \(^{99m}\text{Tc}\) is characterized by optimal physical properties such as a half-life of 6 h, a nearly monoenergetic \(\gamma\)-emission of 140 keV, and its instant availability through a generator system. The creation of \(^{99m}\text{Tc}\) binding site by the conjugation of a chelator to the targeting molecule is associated with instability of the labeled proteins. Conjugates of a cleavable, ester-linked bifunctional chelator RP-1 with scFv has been used to prepare \(^{99m}\text{Tc}\)-labeled scFv species (25, 26) for \textit{in vivo} biodistribution and scintigraphic imaging studies. However, such methods require specialized expertise in coupling chemistry and typically result in heterogeneous products with possible damage to the labeled protein (27). Intact IgG can be labeled directly with \(^{99m}\text{Tc}\) through the hinge-region cysteine residue, but unmodified scFv cannot be directly labeled, as the conserved cysteine pairs in each variable domain form essential disulfide bonds.

Several short amino acid sequences (Figure 3) (27) that provide coordinating nitrogen and sulfur atoms have been incorporated into the N- or C-term of bioactive molecules before radiolabeling with \(^{99m}\text{Tc}\). This approach has several attractive features. First, the \(^{99m}\text{Tc}\)-binding amino acid sequence can be added to or incorporated within proteins by DNA recombinant techniques. Second, these amino acid sequences form stable \([\text{TcO}]^{3+}\) complex when \(^{99m}\text{Tc}\) pertechnetate is reduced by SnCl\(_2\) in the presence of a peptide containing one of the chelating amino acid sequences. Third, the charge and polarity of the complex can be modified by changing the chelating peptide sequence. For example, George et al. (28) introduced a Gly\(_n\)Cys as a peptide chelate into the C-terminus of 741F8 scFv, which targeted the tumor-associated antigen c-erbB-2. This Gly\(_n\)Cys peptide is typically coordinated with \(^{99m}\text{Tc}\) as oxotechnetium through its amide nitrogens and the terminal cysteine, forming a square planar complex. The glycines allow maximal flexibility of the peptide to facilitate chelation of \(^{99m}\text{Tc}\) without disrupting the structure of adjacent framework regions. These results showed that the complex formed between the peptide and \(^{99m}\text{Tc}\) was stable both \textit{in vitro} and \textit{in vivo}, and could be used to image tumor very efficiently. In addition, a (Gly)\(_3\)-Cys-Ala sequence was genetically inserted at the C-terminus of L19, which is a single-chain antibody fragment that specifically targets the ED-B sequence of fibronectin (29). The resulting fusion protein AP39 could be directly labeled with \(^{99m}\text{Tc}\). In this molecule, technetium forms a Tc (V) oxo metal complex with three amide nitrogen atoms and one thiol sulfur atom in the Gly-Gly-Cys sequence. Bogdanov et al. (30) showed specific Tc chelation to a construct of green fluorescent protein with the added C-terminal sequence LGGGGCGGGCGI. Tait et al. (31) expressed ACGGGHM-Annexin V, AGCGGHM-Annexin V and ACGGHMH-Annexin V fusion proteins. Two derivatives form N\(_2\)S and one forms N\(_3\)S, chelation site for subsequent \(^{99m}\text{Tc}\) radiolabeling for imaging purpose. All three proteins could be labeled with \(^{99m}\text{Tc}\) to specific activities of at least 50-100 \(\mu\text{Ci}/\mu\text{g}\). The proteins retained membrane binding activity after the radiolabeling procedure, and quantitative analysis indicated a dissociation constant \((K_d)\) of 7 \(\text{nmol/L}\). The radiolabeled proteins were stable when incubated with phosphate-buffered saline or serum \textit{in vitro}. Annexin V site specifically labeled with \(^{99m}\text{Tc}\) at the N-terminus showed twice as much apoptosis-specific liver uptake as did Annexin V derivatized randomly via amino groups. Another widely used specific peptide sequences, mercaptoacetyltiglycine (MAG3) (32-35), have also been introduced into recombinant proteins in order to create endogenous chelation sites for technetium labeling (34-36).

All the Tc complexes mentioned above that have been used for Tc labeling of proteins are “classical Werner type” coordination compounds without direct metal-carbon bonds. Technetium is usually in the oxidation state +5 as [M=O]\(^{7+}\) core, wrapped by tetradeutate N- or S-ligands, which provide high thermodynamic stability because of their chelate effect. Waibel et al. (37) reported a new approach using Tc(I)-carbonyl compound to form an extremely stable complex to the His-tag, which is one of the most convenient attachment sites and has been used widely as a purification tag for immobilized metal ion affinity chromatography.

Besides technetium, a tetracysteine (TC) tag: Cys-Cys-Xaa-Xaa-Cys-Cys, where Xaa is an noncysteine amino
acid (38–40) is reported able to specifically chelate biaisenical compounds such as FlAsH, a fluorescein derivative, and ReAsH, a resoefm derivative, for site-specific protein modification. The main advantage of TC tag over autofluorescent proteins is its relatively small size, which can be a mere six amino acids (CCPGGC in single-letter amino acid code).

5.3. BirA-based site specific biotinylation

Protein biotinylation is a powerful technique for many molecular biology and biomedical applications due to the high affinity and the specificity of the biotin-avidin interaction. The biotin-avidin interaction has been extensively exploited for labeling biochemical reagents such as antibodies due to the availability of methods to covalently attach biotins to reagents in vitro. Likewise, in vitro biotin labeling has been applied to a variety of drug-targeting and viral gene therapy vector-targeting strategies. While the existing biotinylation technology is certainly robust, it is limited by the fact that the in vivo labeling procedure is not site-specific, requires a prior purification of the labeled substrate, and can lead to inactivation of the biotinylated protein.

Given the difficulties with chemical biotinylation, more recent efforts have been devoted to developing systems in which protein biotinylation can be achieved directly in living cells without chemical reagents. Heterologous expression systems emulating such a “metabolic” approach offer distinct advantages over chemical biotinylation of target proteins. For example, recombinant biotinyl domain fusion proteins can be biotinylated in living cells without the use of specialized reagents and without non-specific and heterogeneous biotinylation at sites other than the acceptor lysine of the biotinyl domain.

Biotin-dependent carboxylases (41) are a class of enzymes that undergo post-translational modification in which the biotin moiety is covalently linked to a single lysine residue via an amide bond. This two-step reaction, as summarized in the following equations, is catalyzed by a class of enzymes termed the biotin holoenzyme synthetases (BHS).

\[
\text{biotin + ATP} \rightarrow \text{biotin-5’-AMP + PPI} \quad (1)
\]

\[
\text{biotin-5’-AMP + apocarboxylase} \rightarrow \text{holocarboxylase + AMP} \quad (2)
\]

In the first step, the enzyme catalyzes formation of an activated intermediate, biotinyl-5’-adenylate (bio-5’-AMP) from biotin and ATP. Subsequent nucleophilic attack at the activated carboxylate of biotin results in formation of the amide between the biotin moiety and the target lysine residue. The biotinylation reaction is highly specific, with only the biotin-dependent carboxylases serving as substrates in vitro. Here, we reviewed the application of the Escherichia coli enzyme biotin ligase (BirA) (42) and different tags recognized by BirA for site-specific biotinylation. The tags range from 123-amino acid (PSTCD) to a 15-amino acid peptide called the acceptor peptide (AP) (43–46). These tags have been fused to target proteins and site-specifically biotinylated in vitro and in vivo successfully.

The PSTCD is a domain of the 1.3S Propionibacterium shermanii transcarboxylase that is naturally biotinylated at lysine 89 of the domain. This PSTCD, when expressed in both E.coli and Saccharomyces cerevisiae as a fusion protein, is metabolically biotinylated by each organism’s cellular biotin ligase enzyme. Chapman-Smith et al. (47) have fused the DNA sequence encoding the PSTCD to the carboxyl terminal of β-galactosidase, neomycin phosphotransferase, Tn5 and Tn903, and chloramphenicol acetyltransferase encoded by Tn9. All the fusion proteins encoded by this chimeric gene became biotinylated in E.coli as long as the biotin ligase recognition sequence remained intact. The use of this approach was expanded to include higher eukaryotic hosts such as cultured mammalian cells and mice (48).

The size of PSTCD was 123 amino acids. To maximize the usefulness of this biotin tag in mammalian cells, deletion constructs were created to reveal the minimum domain of the PSTCD that could be biotinylated. A series of deletion constructs of the PSTCD was made by inverse PCR of the original 123-amino acid biotinylated domain fused to the GFP and knob proteins (49, 50). They found that only the 70- and 63-amino-acid N-terminally truncated forms of the PSTCD domain were still biotinylated when fused to GFP (51).

Although the truncated PSTCD could be used as a tag for the site-specific modification of protein, some features of this approach including the introduction of a bulky (>75 residues) biotinyl domain, and incomplete biotinylation of the target protein, are less than ideal. The first of these issues was resolved by the successful screening of peptide libraries for small peptides capable of accepting biotin from E.coli BirA. Beckett et al. (43) reported that a numbers of peptides that serving as substrates for BirA-catalyzed biotinylation were identified from screens of four biased peptide libraries fused to the lac repressor. Among these a 23-mer peptide, MAGGLNDIFEAQKIEWHEDTGGS, was as effective as the natural substrate in BirA-catalyzed biotin transfer.

To establish the minimal target for BirA-catalyzed biotinylation, a series of truncations of this 23-mer were synthesized and a 14-mer, GLNDIFEAQKIEWH, efficiently mimicked the biotin acceptor function of the much larger protein domain normally recognized by BirA (43, 44). Although this sequence worked well, Cull et al (44) recommended using a slightly extended 15-mer, termed AviTag (GLNDIFEAQKIEWHE) because it is consistently biotinylated at a rate slightly better than that of the natural substrate.

Ting et al. (52) fused AviTag (acceptor peptide) to CFP targeted to the surface with a signal peptide and the transmembrane (TM) domain of the platelet-derived growth factor receptor (AP-CFP-TM). After transfection of this construct into HeLa cells, addition of BirA to the cell medium together with ATP and biotin, biotinylation was detectable using fluorescent dyes or quantum dots (QD) conjugated streptavidin. The main advantage of AP is its small size, which means it is unlikely to perturb folding,
Site-specifically modified fusion proteins for molecular imaging

secretion, and protein–protein interactions during trafficking to the surface, compared with fluorescent proteins (GFP is 238 residues) (48, 53) and other proteins directing site-specific labeling such as O6-alkylguanine-DNA alkyltransferase (54) or acyl carrier protein (55).

To expand the usefulness of enzymatic protein biotinylation, Ting et al. (56) screened the phage library and identified a 15-mer peptide sequence (termed AP2) which can be specifically recognized and biotinylated by Saccharomyces cerevisiae (yeast) biotin ligases (yBL), in the process discovering a new biotin ligase-AP pair, (BL2-AP2). This new pair has a similar specificity and kinetics to the original BirA-AP pair, but represents an orthogonal system, in that BirA does not recognize the new AP2 while conversely, the new BL2 does recognize the original AP. Such BirA-AP/BL2-AP2 orthogonal systems have been applied to perform two-color QD labeling on HeLa cells.

5.4. Sfp phosphopantetheinyl transferase-based site-specific modification

The Sfp phosphopantetheinyl transferase covalently transfers 4′-phosphopantetheinyl (Ppant) groups from CoA to conserved serine residues on peptide carrier protein (PCP) and acyl carrier protein (ACP) domains in nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) in Bacillus subtilis (57). The post-translational modification of PCP and ACP domains by Ppant is crucial for the activation of NRPS and PKS, because the flexible 20 Å Ppant group functions as a swinging arm providing the anchoring point for the attachment of the peptide or polyketide chain as it elongates along the biosynthetic assembly line (58).

Although the substrate for Sfp must contain CoA, the enzyme is otherwise quite tolerant of a range of small molecules attached to CoA. In fact, Sfp transfers small molecules of diverse structures from CoA to the conserved serine on the PCP or ACP domain (58, 59). Sfp-catalyzed post-translational modification of PCP or ACP takes advantage of the substrate promiscuity of Sfp and the small size of PCP and ACP domains (~80 residues) to provide an attractive method for one-step site-specific protein labeling with small molecules. The E.coli phosphopantetheinyl transferase AcpS is functionally similar to Sfp, although AcpS preferentially modifies ACP domains (60).

Sfp and AcpS have been used for site-specific protein labeling in cell lysates or on live cell surfaces. In these applications the target protein is expressed as a fusion to PCP or ACP at either the N or the C terminus, and Sfp is used to enzymatically attach a small-molecule probe - Ppant conjugate to the expressed PCP or ACP tag. This method has allowed the site-specific labeling of target proteins by small molecules of diverse structures and functionalities, such as biotin, fluorophores, porphyrins, sugars and peptides (Figure 4) (59, 61, 62). The labeling reaction proceeds with high efficiency and can be carried out in cell lysates or in culture media. A sufficient number of cell surface protein molecules can be labeled for imaging purposes within 15 min at submicromolar concentrations of the enzyme and CoA-conjugated small-molecule probes.

Thus far, protein labeling catalyzed by Sfp and AcpS has been used for high-throughput printing of protein microarrays (63), site-specific attachment of small molecules to phage particles (64) and live cell imaging of cell surface proteins (55, 65, 66).

Most recently, the ybbR tag, a short (11-residue) peptide (DSLEFIASKLA), was found to be an efficient substrate for Sfp-catalyzed protein labeling, thereby replacing the full-length PCP or ACP domain for the construction of smaller fusions of the target protein. It was so named because part of its sequence was derived from the cell wall.
ybbR ORF in the *B. subtilis* genome. The site of Sfp-catalyzed ybbR tag labeling was mapped to the underlined Ser residue, and the ybbR tag was found to have a strong tendency for adopting an α-helical conformation in solution. ybbR tag can be fused to the N or C terminus of target proteins or inserted in a flexible loop in the middle of a target protein for site-specific protein labeling by Sfp. The short size of the ybbR tag and its compatibility with various target proteins, the broad substrate specificity of Sfp for labeling the ybbR tag with small-molecule probes of diverse structures, and the high specificity and efficiency of the labeling reaction make Sfp-catalyzed ybbR tag labeling an attractive tool for expanding protein structural and functional diversities by posttranslational modification (67).

5.5. Formylglycine-generating enzyme-based site-specific modification

Over a decade ago, Schmidt *et al.* (68) discovered that most sulfatases contain an aldehyde-bearing formylglycine residue within their active site that is required for catalytic activity. The formylglycine residue is produced by co- or post-translational modification of a conserved cysteine residue found within the "sulfatase motif" (Figure 5) (69). This motif contains a highly conserved CXPXR submotif, where X is usually serine, threonine, alanine or glycine. Formylglycine-generating enzyme (FGE) oxidizes cysteine to formylglycine and has recently been identified in eukaryotes (70) and prokaryotes. Interestingly, FGE recognizes its substrate mainly by its primary sequence. This means that a small peptide tag based on the sulfatase motif, called the "aldehyde tag", might direct formation of formylglycine independent of its context. Furthermore, many organisms have endogenous FGE activities (71), suggesting that this post-translational modification system could serve as a general means for engineering proteins for site-specific labeling (72-74).

Carrico *et al.* (69) expressed recombinant proteins with either N- or C-terminal aldehyde tags in *E. coli*. The tags they used included a full-length sulfatase tag (LCTPSRGSLFTGR) and a 6-amino-acid tag (LCTPSR) containing only core conservative residues. These constructs were co-expressed with an additional FGE from *Mycobacterial tuberculosis* to maximize the formylglycine formation. After purification, a panel of proteins was robustly site-specific labeled with aminooxy- or hydrazide-functionalized probes, including fluorescent dyes, biotin, and poly (ethylene glycol) (PEG).

5.6. The hAGT based site-specific modification

Human O6-alkylguanine-DNA alkyltransferases (hAGT) is a 207-amino acid enzyme. The physiological function of hAGT is the removal of alkyl groups from guanine bases of DNA to avoid mutations during cell division. O6-Alkylguanine adducts are formed in DNA by many alkylating agents, including important environmental carcinogens such as nitrosamines. These adducts are repaired by a unique single step mechanism mediated by AGT (75-77). AGT transfers the alkyl group from O6-alkylguanine to a cysteine residue located within the protein, and each molecule can act only once because the S-alkylcysteine formed is not regenerated. This transfer protects against the mutagenic and cytotoxic effects of alkylating agents AGT undergoes a classic suicide reaction as the cysteine residue in its active site forms a covalent and stable thioether bond with the alkyl group, allowing the release of the de-alkylated, original guanine base, and the content of AGT is a critical factor in determining cellular sensitivity to these agents (78, 79). The AGT family members share a highly conserved -PCHRV- amino acid sequence surrounding the cysteine acceptor site and contain a number of other conserved amino acids in the active site and DNA binding domain of the protein (Figure 6).

For AGT-based site-specific modification of protein (54), covalent labeling with a small molecule is achieved through hAGT, which irreversibly transfers the alkyl group from its substrate, O6-alkylguanine-DNA, to one of its cysteine residues. This covalent and site-specific modification of AGT was studied in great detail. (80, 81). They observed that AGT can accept any chemical compound when it is linked to the guanine base through a benzyl group in different hosts such as E coli, yeast, and mammalian cell cultures.
Figure 6. (A) Mechanism of DNA repair by hAGT; (B) Covalent labeling of an X-hAGT fusion protein using O^6^-benzylguanine (BG) derivatives of type 1; (C) Structure of the BG derivatives BGBT, BGAF, and BGFL used in this work; (D) Scheme of synthesis of BGBT, BGAF, and BGFL. Reproduced with permission from (54).

The fact that the gene encoding hAGT can be expressed as a soluble protein in various host cells led to the development of the SNAP-tag™ technology (82, 83). Directed evolution experiments and protein engineering...
with hAGT resulted in a smaller enzyme with much faster reaction kinetics and devoid of interactions with other biomolecules such as double stranded DNA. The reaction mechanism of the SNAP-tag™ system allows the covalent and specific attachment of target proteins to any desired compound or solid surface. For example, Tirat et al. (84) fused ubiquitin conjugating enzyme Rad6B to either N- or C- terminus of SNAP-tag™ using DNA recombinant technology. After labeling, the AGT-tagged Rad6B proteins were characterized by mass spectrometry, peptide map and a SDS-PAGE band shift assay, with all results demonstrating the high specificity of labeling by using the SNAP-tag™ technology. Furthermore, modified AGT-Rad6B fully retained its functional activity. They further tested an AGT substrate consisting of an Alexa555 fluorophore bound to benzylguanine without any spacer (85) and similar results were obtained. The major problem of this approach, however, is the rather long sequence of hAGT, which might have some effect on target protein conformation and function when used as a fusion tag.

5.7. Transglutaminase-based site-specific modification

The basic premise for this strategy lies in the specificity of the transglutaminase. Transglutaminase catalyzes the reaction:

\[
R-CONH_2 + R'-NH_2 \rightleftharpoons R-CONH-R' + NH_3^2\text{Ca}^2+
\]

Here, R-CONH₂ represents the acceptor, a Gln residue in proteins, and R'-NH₂, the donor, an alkylamine; in vivo, the common amine donor is probably a protein-bound lysine side chain (86), and an amide bond cross-linkage is formed between the proteins. It appears that transglutaminase, especially guinea pig liver transglutaminase (TGase), has very stringent sequence specificity and structural requirements for the amine acceptor site around the Gln residues. According to the specific (87) determinant for proteins, substrate Gln residues for TGase must first satisfy an accessibility criterion, either by being in a highly flexible region of the polypeptide chain or by being clearly exposed to the solvent in a more structured region of the polypeptide chain. Due to the conformational requirement of transglutaminase-catalytic labeling, the direct Gln-specific modification or design of the mutation for the modification may be difficult to apply to most proteins.

Sato et al. (88) were able to site-specifically label IL-2 protein with dansylcadaverine by introducing a very short substrate sequence for TGase at the N-terminus of IL-2. The chimeric protein was expressed and then enzymatically labeled with the fluorophore only at a Gln site in the appended N-terminal sequence. A model system was constructed. The sequence selected for the preparation of the chimeric protein (named rTG1-IL-2) is PKPQQFM, derived from Substance P, which is known to be a very good substrate in vitro for TGase. The CD spectra and IL-2 bioactivity studies showed that the overall structure of rTG1-IL-2 was similar to that of rhIL-2. In fact, the addition of the full sequence of Substance P (RPKQQPQGLM) to the N-terminus of hIL-2 revealed that almost 50% of the IL-2 bioactivity was lost on a molar basis. Thus, the study selected the essential part of the Substance P molecule consisting of the seven-residue sequence to avoid a conformational change of hIL-2.

5.8. Intein-mediated site-specific protein modification

Inteins are naturally occurring protein fragments that are capable of excising themselves out from a precursor protein and ligating the two flanking protein fragments (N-extein and C-extein), via a peptide bond, to form a mature protein. This process, termed intein-mediated protein slicing, occurs during the posttranslational stage of protein synthesis. Intein-mediated protein splicing is fascinating in that it is an autocatalytic event and requires neither cofactors nor auxiliary enzymes in the process (89).

Intein-mediated labeling reactions require two protein domains (90): the N-terminal domain which contains the protein sequence to be labeled (the target domain), and a C-terminal domain containing the intein (the intein domain). These domains are usually joined in a single fusion protein containing both the intein and target domains, a target-intein fusion protein. For those intein domains beginning with an N-terminal cysteine residue, the molecular basis of intein-mediated activation is a spontaneous reversible N→S transacylation step forming thioester that occurs at this cysteine residue. The ligation of probes to target domains has been achieved using a range of cysteine derivatives under appropriate conditions (91). Although the thioester is more reactive toward nucleophiles than a backbone amide, a direct reaction with peptides containing an N-terminal cysteine residue was unsuccessful, a result that has been rationalized in terms of steric hindrance from the bulky intein domain (92).

The usage of several thiols including thiophenol (92), ethanethiol, and MESA (93) has been reported to enhance intein-mediated ligation reactions, presumably by trans-thioesterification to the less hindered intermediate. Under these conditions, cysteine derivatives are thought to form the kinetic product, then rearrange to the thermodynamically more stable amide (Figure 7) (89). The intein-mediated activation of the C-terminus of a protein domain has been shown to provide a general method for attaching a wide range of molecules, including proteins (49, 94), peptides (92), fluorescent labels (95, 96), carbohydrates (97), oligonucleotides (98), and affinity tags and metal chelators (99). Intein-mediated ligation has also been used for the assembly of selectively labeled proteins for NMR studies (99) and to prepare cyclic peptides and proteins (100, 101). For in-depth coverage of intein-mediated protein ligation and its application, readers are directed to a detailed review by Muralidharan and Mui (89). In our review, we will focus solely on the recent expanded applications of the IPL technology in the fields of bioimaging.

Tan et al. (102) developed strategies for site-specific biotinylation of proteins. In their strategy, the protein of interest was fused through its C-terminus to an intein, which had its other terminus fused to a chitin-binding domain (CBD). Upon expression in the host cell, the cell lysate containing the fusion protein was first loaded onto a column packed with chitin beads followed by the addition of a thiol cleaving reagent. The fusion protein underwent self-cleavage, catalyzed by the fused intein, and resulted in the protein having a reactive α-thioester group at the C-terminus. At this stage, the thiol side chain from the addition of cysteien-biotin attacked the α-thioester functionally and the resulting thioester-linked intermediate
underwent spontaneous rearrangement, allowing a native peptide bond to generate proteins that were site-specifically biotinylated at their C-terminus. This strategy is simple yet efficient, and capable of biotinylating various proteins from different biological resources. This strategy has also been extended to in vivo biotinylation of proteins in bacterial and mammalian systems. The simple addition of the cell-permeable cysteine–biotin probe to the growth media followed by further incubation of the cells resulted in substantial biotinylation of the protein inside the cells. The efficiency of intein-mediated protein biotinylation depends greatly on the intein fused to target protein: as much as 2-10 times improvement in protein biotinylation may be achieved by the simple switch in the intein fused (102). However, the limitation of this strategy is the biotinylation of the endogenous protein in both mammalian cells and bacterial. From western blots, three other biotinylated proteins besides the targeted protein were detected from the lysate of mammalian cells and biotinylated acetyl-CoA carboxylase was found from bacterial lysate.

In fact, this intein-mediated protein biotinylation could be extended to cell-free protein system (103). A cell-free system has many potential advantages over traditional in vivo recombinant protein expression in cells. Cellular toxicity of overexpression of certain proteins, possible degradation by endogenous protease and formation of inclusion bodies of protein all might be avoided in a cell-free system. Although the cell-free biotinylation approach seems the most simple out of the three strategies mentioned here, the efficiency of protein expression and subsequent biotinylation depend on a few other factors including the DNA quality and concentration, as well as the nature of the target protein (104).

One potential problem with intein-mediated ligation chemistry is the possible oxidation of both the activating thiol reagent and the modified cysteine derivative to disulfides. The use of thiol reducing agents such as DTT is not possible, as these compounds effectively cleave and derivatize the target domain. This problem has been circumvented by the use of the water soluble phosphine...
Site-specifically modified fusion proteins for molecular imaging

TCEP, which efficiently reduces disulfides to thiols. Another potential problem with intein-mediated protein modification is the relatively large size of intein. To date, most intein used were over 100 amino acid, for example Mxe GyrA intein, 198 aa (105); Mib RecA intein, 229 aa (107, 108); Mth RIRI intein, 134 aa (94); and Ssp DnaB intein (109), 154aa (109).

6. CONCLUSIONS AND PERSPECTIVES

Modern drug discovery strongly depends on the availability of target proteins in sufficient amounts and with desired properties. For some applications, proteins have to be equipped with additional functionalities for protein purification, fluorescent or radioactive labels for detection, glycosylation and phosphorylation for biological activity, and many more properties. This functionality can result from the expression of the protein as a fusion protein with an additional polypeptide, the labeling the protein with synthetic molecules, the incorporation of unnatural amino acids or a combination of these different approaches. For a protein modification technique to find widespread use among biologists, it must be practical, general and have many properties. This functionality can result from the expression of the protein as a fusion protein with an additional polypeptide, the labeling the protein with synthetic molecules, the incorporation of unnatural amino acids or a combination of these different approaches. For a protein modification technique to find widespread use among biologists, it must be practical, general and available. Appending a peptide tag to target protein fulfills all these requirements. Recombinant proteins have been expressed for decades as fusion proteins with an additional polypeptide for purification or characterization of target proteins. To extend the traditional use, the tag can also be used to label the fusion protein with synthetic probes in a site-specific and covalent manner.

This review summarizes current methods for site-specific modification of target protein. These methods have been widely used in proteomics, protein-protein interaction, protein purification and immobilization, and protein microarray studies, among others. However, only the cysteine-containing tag has been used for molecular imaging purpose. The AviTag can be foreseeably applied to produce site-specifically biotinylated proteins for labeling avidin-derivatized fluorophores, quantum dots, microbubbles, and chelators for optical, ultrasound, and radionuclide imaging. This strategy may also be extended by adding an ybbR tag onto the target protein to provide Sfp-catalyzed protein labeling of biotin, fluorophores, and metal chelators for molecular imaging applications. Furthermore, two or more orthogonal tags can be introduced into one protein and used to couple different detectable probes for multimodality imaging. In combining the convenience of expressing genetically tagged fusion proteins and the diversity of synthetic chemistry, site-specifically labeled protein probes will become ever more useful and popular in the molecular imaging area.

7. REFERENCES

Site-specifically modified fusion proteins for molecular imaging


Site-specifically modified fusion proteins for molecular imaging


**Key Words:** Site-Specific Modification, Fusion Protein, Molecular Imaging, Cys-Tag, Chelator-Tag, Biotin Ligase (Bira), Stp Phosphopantetheinyl Transferase, Formylglycine-generating enzyme, Human O6-alkylguanine-DNA alkyltransferases (hAGT), Transglutaminase, Intein, Review

**Send correspondence to:** Xiaoyuan Chen, PhD, the Molecular Imaging Program at Stanford (MIPS), Department of Radiology and Bio-X Program, Stanford University School of Medicine, 1201 Welch Rd, P095, Stanford, CA 94305, Tel: 650-725-0950, Fax: 650-736-7925, E-mail: shawchen@stanford.edu

http://www.bioscience.org/current/vol13.htm