Selective Labeling of Proteins with Chemical Probes in Living Cells

Selective labeling of proteins with small molecules introduces novel chemical and physical properties into proteins, enabling the target protein to be investigated or manipulated with various techniques. Different methods for labeling proteins in living cells have been developed by using protein domains, small peptides, or single amino acids. Their application in cells and in vivo has yielded novel insights into diverse biological processes.

Protein Domain Tags

Many protein domains, as part of their natural functions, interact with small molecules in a highly specific manner (FIGURE 1). For example, enzymes catalyze reactions by mediating the precise positioning of substrates in their active sites. Similarly, small molecule enzyme inhibitors such as natural toxins and therapeutic drugs exert their effects by binding tightly to active sites. Specialized proteins are also chemically modified in detoxification processes. Each of these types of interactions—enzyme-substrate, target-inhibitor, and scavenger-toxin—has been exploited to attach chemical probes to proteins of interest with high specificity.

Covalent attachment of chemical probes has the advantage of greater irreversibility compared with noncovalent binding. A modified 33-kDa bacterial dehalogenase domain has recently been developed to perform efficient covalent self-labeling. Dehalogenases break down haloalkanes in a two-step process involving alkyl transfer to an active site Asp side chain (and the loss of the halide ion) followed by hydrolysis of the alkyl-enzyme bond. The hydrolysis step requires an active site His residue (59), mutation of which prevents the alkyl-enzyme conjugate. An optimized hydrolysis-deficient haloalkalase (marketed as Halotag, Promega) can be specifically labeled using cell-permeable haloalkanes bearing fluorophores or biotin (39). Since euukaryotes lack dehalogenases, background staining in cells is very low.

Before the use of dehalogenase, Johnsson and colleagues (32) were the first to react small molecules with a protein domain for labeling purposes, using the 21-kDa O6-alkylguanine-DNA alkyltransferase (AGT). AGT recognizes guanine bases alkylated on the O6 position and transfers the alkyl group to one of its reactive cysteine side chains, regenerating guanine. Cell-permeable O6-benzylguanine derivatives bearing fluorophores, biotin, hapten, and even other protein ligands have been used to label AGT fusions (32). A mutant AGT with increased reactivity for free benzylguanine derivatives and decreased binding to DNA has been designed (25) and is available along with a variety of probes from Covalys. The 50-fold faster reaction of the mutant vs. wild-type AGT allows specific visualization of fusion proteins even in the nuclei of cells expressing endogenous AGT (31). Also, although wild-type AGT changes conformation upon alkylation to allow its degradation (13), mutant AGT fusions have been followed for days without unexpected signal decay, implying that degradation may not occur in fusions or with mutant AGT (31). AGT has proven to be a reliable labeling method in time-lapse studies of centromere and sarcoplasmic reticulum protein movements in living cells (31, 69).

Enzymes and irreversible active-site inhibitors, i.e., suicide substrates, can serve as protein-ligand pairs for covalent labeling without further protein engineering as long as the inhibitors do not react with endogenous enzymes. In the only example of this concept to date, fluorophore-conjugated analogs of the suicide substrate p-nitrophenyl phosphate (pNPP) were...
used to label proteins fused to the fungal cell wall enzyme cutinase (5). However, because pNPP is cell-impermeant, this system is only suitable for labeling surface proteins.

Although less desirable for long-term studies, protein domains can also be labeled noncovalently with chemical ligands. This was first demonstrated using fluorescent derivatives of methotrexate to visualize fusions to mammalian dihydrofolate reductase (DHFR) in DHFR-deficient cells (29). Recently, Cornish and colleagues created a variety of cell-permeant derivatives of the antibiotic trimethoprim for visualizing fusions with the 18-kDa E. coli DHFR (7). Trimethoprim shows high affinity to *E. coli* DHFR (*K_d* = 1 nM) and minimal (*K_d* = 3 μM) binding to mammalian DHFR (66) and, unlike methotrexate, is not cytotoxic. A similar method arose from research on the immunosuppressant FK506 and its 12-kDa protein target FKBP12. The FK506 analog SLF binds with 94 pM affinity to the FKBP12(F36V) mutant but not endogenous FKBP12, allowing specific labeling of FKBP12(F36V) fusions in cells by SLF derivatives (43).

A protein domain that allows labeling specifically in vesicles or the cell surface is a single-chain antibody, either selected for binding to a small molecule that can be conjugated to a fluorophore (18) or selected for the ability to increase the quantum yield of fluorescent probes upon binding (71). The latter tag, like fluorescent proteins, is restricted to optical readouts but allows for fluorophores with superior optical characteristics or for membrane-impermeant fluorophores for selective labeling of surface proteins.

The ever-growing list of enzyme-inhibitor pairs characterized by the pharmaceutical industry may serve as a source for future labeling systems. Suitable pairs are those that exhibit orthogonality and versatility, i.e., neither enzyme nor drug has effects in mammalian cells, and the inhibitors tolerate functional derivatization while retaining reactivity. Although most interactions are noncovalent, some well-established drugs such as certain β-lactam antibiotics are suicide substrates and may serve as the basis for future covalent labeling systems. Another route to finding new protein-chemical pairs may be the engineering of chemical ligands. This was first demonstrated using fluorescent derivatives of methotrexate to visualize fusions to mammalian dihydrofolate reductase (DHFR) in DHFR-deficient cells (29). Recently, Cornish and colleagues created a variety of cell-permeant derivatives of the antibiotic trimethoprim for visualizing fusions with the 18-kDa E. coli DHFR (7). Trimethoprim shows high affinity to *E. coli* DHFR (*K_d* = 1 nM) and minimal (*K_d* = 3 μM) binding to mammalian DHFR (66) and, unlike methotrexate, is not cytotoxic. A similar method arose from research on the immunosuppressant FK506 and its 12-kDa protein target FKBP12. The FK506 analog SLF binds with 94 pM affinity to the FKBP12(F36V) mutant but not endogenous FKBP12, allowing specific labeling of FKBP12(F36V) fusions in cells by SLF derivatives (43).

**Small Peptide Labeling**

Protein domains specifically designed for labeling can be especially useful when combined with small chemical ligands. One approach is to use biarsenical systems, which are covalent but may be reversible and have a wide array of small molecules available to choose from. The first anionic biarsenical labeling molecule is the basic component of the biarsenical reagent, labeled diaminobenzidine (21). The tetraarsenical, ReAsH (1), with each arsenic adjacent cysteine in a specific binding pocket, is readily covalently bound to proteins. It is well established as a suicide substrate.

Another route to finding new protein-chemical pairs may be the engineering of chemical ligands. This was first demonstrated using fluorescent derivatives of methotrexate to visualize fusions to mammalian dihydrofolate reductase (DHFR) in DHFR-deficient cells (29). Recently, Cornish and colleagues created a variety of cell-permeant derivatives of the antibiotic trimethoprim for visualizing fusions with the 18-kDa E. coli DHFR (7). Trimethoprim shows high affinity to *E. coli* DHFR (*K_d* = 1 nM) and minimal (*K_d* = 3 μM) binding to mammalian DHFR (66) and, unlike methotrexate, is not cytotoxic. A similar method arose from research on the immunosuppressant FK506 and its 12-kDa protein target FKBP12. The FK506 analog SLF binds with 94 pM affinity to the FKBP12(F36V) mutant but not endogenous FKBP12, allowing specific labeling of FKBP12(F36V) fusions in cells by SLF derivatives (43).

**FIGURE 1.** Classes of protein domain or peptide tags used for chemical labeling of fusion proteins

A: domain tags self-label either covalently or noncovalently. B: peptide tags either self-label with varying degrees of reversibility or utilize enzyme-mediated covalent attachment. The standard methods of fluorescent protein or epitope tags are shown for comparison. Examples of tags of each class are listed to the right.
antibodies that recognize a desired chemical structure. This provides for covalent as well as noncovalent interactions, since antibodies can be engineered to react with ligands (10, 78). The recent availability of methods for selecting camel antibody single-chain variable domains makes possible the isolation of relatively small (11 kDa) domains, which may retain function in the cytosol in some cases (63).

Small Peptide Tags

Protein domains mediate chemical labeling with high specificity, but their size and folding requirements generally restrict their use to the NH$_2$ and COOH termini of proteins of interest. Smaller tags have less potential to disrupt protein folding or function and in certain cases can be placed in loops or domains within proteins. The inherent challenge with peptide chemical labeling pairs is molecular recognition in a small space, i.e., the designing of an interface that is tight and specific and yet on the protein side is encodable within a few amino acids.

The first and most extensively used system for specific labeling of a peptide by functionalized small molecules is the tetracysteine-biarsenical system. The basic components, rationally designed by Griffin and colleagues (24), are peptide tags with the consensus sequence CCXXCC and chemical probes, such as the green fluorophore FlAsH or the red fluorophore ReAsH (1), with two arsenic atoms positioned so that each arsenic atom bonds with two thiol groups on adjacent cysteines. These arsenic-sulfur bonds are covalent but mediated by outer valence orbitals and are reversible in the presence of excess thiols. The bistetradentate nature of the complex results in extremely high (10$^{-10}$ pM) affinity vs. interactions that involve only one arsenic atom (1). Recently, sequences surrounding the CCXXCC core have been optimized for improved ReAsH and FlAsH affinity, with their effects presumably being mediated by increased hydrophobic contacts between peptide and ligand outside of the arsenic-sulfur bonds (43). In being isolated from peptide library screen for a specific binding activity, these sequences may be considered to have characteristics of aptamers (i.e., a short peptide sequence evolved for a specific binding activity; see below).

The tetracysteine-biarsenical system has been used with a wide array of biosensor-conjugated functional groups. Aside from their usual use as fluorophores for visualization, FlAsH and in particular ReAsH can generate reactive oxygen species for light-assisted protein inactivation in functional studies (42, 75) or diaminobenzidine precipitation in electron microscopy (21). The tetracysteine-biarsenical system has also been extended to attach chemical cross-linkers (38) and calcium sensors (74) to particular sites on proteins. Recently, a variation of the tetracysteine-biarsenical system has been developed that uses a fluorophore with more widely spaced arsenic atoms and a peptide with sequence CCKAEACC (8). These components do not interact with those of the traditional tetracysteine-biarsenical systems, and therefore the two systems can be used in the same cell in an orthogonal manner.

“Cysteine is the most frequently used residue for selective chemical modification due to its relatively low abundance in proteins and the increased nucleophilicity of the thiol group relative to other amino acid side chains.”

The primary drawbacks of the tetracysteine-biarsenical system are background labeling and toxicity. CCXXCC motifs are not found in the genome, but multiple proteins contain motifs that differ from CCXXCC by only one cysteine, and substantial labeling of cytoplasmic proteins can occur to various degrees depending on cell type (23, 41, 70). Typically, high micromolar concentrations of dithiols are used during labeling to reduce background accumulation and afterward as a destaining step (23, 41), although these reagents reduce the effective affinity of the desired interaction to the nanomolar range (1). FlAsH and ReAsH may also bind to cellular components through nonspecific hydrophobic interactions (23). The system is therefore most useful for overexpressed or highly concentrated proteins of interest, such as actin concentrated on polyribosomes (62) or viral proteins during assembly (64). Furthermore, potential toxicities from residual off-target binding of arsenic or the dithiol antidotes themselves remain a concern (2, 35). These drawbacks, as well as the need for long labeling and destaining times, are likely to make use of the tetracysteine-biarsenical system in thick tissues or animals difficult.

Following the introduction of the tetracysteine-biarsenical system, other strategies for specifically attaching chemical probes to peptides have been developed. These fall into three categories: rationally designed peptide-probe pairs, peptide aptamers evolved for binding to chemical probes, and peptides from natural biosynthetic pathways that serve as specific sites for covalent attachment of chemical groups by enzymes. In analogy to the tetracysteine system, Tsien and colleagues developed a method for labeling polyhistidine peptides with a complex of two zinc ions and a zinc-chelating dye called HidZFP (26). This interaction exhibits moderately tight affinity (~40 nM) but is currently limited by suboptimal dye characteristics. In a similar concept, the zinc chelator DpaTyr binds to a string of four Asp residues with 1.4$\mu$M affinity and can be attached to various fluorophores (54). This low-affinity interaction can be converted into a...
### Table 1. Comparison of protein and peptide tags

<table>
<thead>
<tr>
<th>Tag</th>
<th>Size (aa)</th>
<th>Mechanism</th>
<th>Kd</th>
<th>Probe Type</th>
<th>Demonstrated Functionalities</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehalogenase</td>
<td>297</td>
<td>Ester</td>
<td>0</td>
<td>Alkylhalides</td>
<td>Fluorescence, biotinylation</td>
<td>Sold as Halotag, Promega</td>
</tr>
<tr>
<td>Alkylguanosine transferase</td>
<td>182</td>
<td>Thioether</td>
<td>0</td>
<td>Alkylguanosines</td>
<td>Fluorescence, biotinyl, and methotrexate handles</td>
<td>Sold as SNAP tag, Covalsy</td>
</tr>
<tr>
<td>Cutinase</td>
<td>200</td>
<td>Phospho</td>
<td>0</td>
<td>Alkyl-p-nitrophenyl-phosphonates</td>
<td>Fluorescence, quantum dots</td>
<td>EC only</td>
</tr>
<tr>
<td>DHFR</td>
<td>157</td>
<td>Noncovalent</td>
<td>1 nM</td>
<td>Trimethoprim conjugates</td>
<td>Fluorescence</td>
<td></td>
</tr>
<tr>
<td>FKB12(F3A6)</td>
<td>108</td>
<td>Noncovalent</td>
<td>0.094 nM</td>
<td>SLF' conjugates</td>
<td>Fluorescence</td>
<td></td>
</tr>
<tr>
<td>Single-chain antibody</td>
<td>230-300</td>
<td>Noncovalent</td>
<td>5 nM</td>
<td>pHDe conjugates, fluorophores</td>
<td>Fluorescence</td>
<td></td>
</tr>
<tr>
<td>Tetracysteine</td>
<td>12</td>
<td>Thiol-aromatic bond</td>
<td>0.01-1 nM</td>
<td>Biarsenical conjugates</td>
<td>Fluorescence, singlet oxygen generation, chemical crosslinking, calcium sensing</td>
<td></td>
</tr>
<tr>
<td>Polyhistidine</td>
<td>6</td>
<td>Noncovalent</td>
<td>40-1400 nM</td>
<td>Zinc chelator conjugates</td>
<td>Fluorescence, chemical reactivity with nucleophiles</td>
<td>EC only</td>
</tr>
<tr>
<td>Lanthamide-binding tag</td>
<td>14</td>
<td>Noncovalent</td>
<td>220 nM</td>
<td>Lanthanides</td>
<td>Phosphorescence, spectroscopy</td>
<td>EC only</td>
</tr>
<tr>
<td>TRS14</td>
<td>38</td>
<td>Noncovalent</td>
<td>0.025 nM</td>
<td>Texas Red derivatives</td>
<td>Fluorescence, calcium sensing</td>
<td></td>
</tr>
<tr>
<td>Biotin acceptor peptide</td>
<td>15-75</td>
<td>Amidic</td>
<td>0</td>
<td>Biotin and analogs</td>
<td>Biotin handle, click chemistry for multiple functionalities</td>
<td>Requires ATP</td>
</tr>
<tr>
<td>Lipic acid acceptor peptide</td>
<td>22</td>
<td>Amidic</td>
<td>0</td>
<td>Lipic acid conjugates</td>
<td>Click chemistry for multiple functionalities</td>
<td>Requires ATP</td>
</tr>
<tr>
<td>Phosphopanteth-esteryl</td>
<td>15</td>
<td>Amidic</td>
<td>0</td>
<td>CoA conjugates</td>
<td>Organic fluorescence, biotin handle</td>
<td>EC only</td>
</tr>
<tr>
<td>Sortag</td>
<td>5</td>
<td>Amidic</td>
<td>0</td>
<td>Pentaglycine conjugates</td>
<td>Organic fluorescence, biotin handle</td>
<td>EC only</td>
</tr>
<tr>
<td>G-tag</td>
<td>4-7</td>
<td>Amidic</td>
<td>0</td>
<td>Primary amines</td>
<td>Organic fluorescence, biotin handle</td>
<td>EC only</td>
</tr>
<tr>
<td>Farnesyltag</td>
<td>4</td>
<td>Amidic</td>
<td>0</td>
<td>Phenylphosphate conjugates</td>
<td>Click chemistry for multiple functionalities</td>
<td>EC only</td>
</tr>
<tr>
<td>Aldehyde tag</td>
<td>6</td>
<td>Enzymatic</td>
<td>0</td>
<td>None</td>
<td>Chemical reactivity with nucleophiles</td>
<td>EC, luminal only</td>
</tr>
</tbody>
</table>

**Mechanism**
- Ester linkage
- Thioether linkage
- Phosphoester linkage
- Noncovalent
- Thiol-aromatic bond
- Amidic
- Amidic
- Amidic
- Amidic
- Amidic
- Amidic
- Amidic
- Amidic

**Probe Type**
- Alkylhalides
- Alkylguanosines
- Alkyl-p-nitrophenyl-phosphonates
- Trimethoprim conjugates
- SLF' conjugates
- pHDe conjugates, fluorophores
- Biarsenical conjugates
- Zinc chelator conjugates
- Lanthanides
- Texas Red derivatives
- Biotin and analogs
- Lipic acid conjugates
- CoA conjugates
- Pentaglycine conjugates
- Primary amines
- Phenylphosphate conjugates
- None

**Demonstrated Functionalities**
- Fluorescence, biotinylation
- Fluorescence, biotinyl, and methotrexate handles
- Fluorescence, quantum dots
- Fluorescence
- Fluorescence, singlet oxygen generation, chemical crosslinking, calcium sensing
- Fluorescence, chemical reactivity with nucleophiles
- Phosphorescence, spectroscopy
- Fluorescence, calcium sensing
- Fluorescence, biotin handle
- Click chemistry for multiple functionalities
- Organic fluorescence, biotin handle
- Organic fluorescence, biotin handle
- Chemical reactivity with nucleophiles

**Notes**
- Sold as Halotag, Promega
- Sold as SNAP tag, Covalsy
- EC only
- EC only
- EC only
- Requires ATP
- Requires ATP
- EC only
- EC only
- EC only
- EC only
- EC, luminal only

This table provides a comparison of protein and peptide tags, including their size, mechanism, and demonstrated functionalities.

Covalent one by placing appropriate reactive chemical groups in the peptide tag and in the chelator probe (31). However, applicability inside eukaryotic cells has yet to be demonstrated for these zinc-chelating probes and may be difficult given their large sizes and therefore poor membrane permeability, especially for DpsY derivatives. Furthermore, zinc is required at micromolar levels but is maintained in the cytosol at only picomolar levels (6).

As an alternative to the rational design, other groups have turned to in vitro selection of aptamers from randomized libraries. A 14-aa aptamer has been evolved to bind lanthanide ions (20), which can be visualized by luminescent or spectroscopic methods. One study took advantage of the small size of the lanthanide-binding tag to estimate distances between insertion sites on extracellular regions of proteins by measuring resonance energy transfer (65). A potentially more generalizable aptamer is TR512, recently isolated by Nolan and colleagues with impressively strong (20 pM) affinity to the central three-ring core of the dye Texas Red (44). This structure can be derivatized while retaining binding, as demonstrated for the Texas Red-Labeling with more than one aptamer may be required for visualization of hydrophobic targets, but the technology has proven to be versatile.

In a different group of enzymes that have been adapted to specifically bind zinc, a series of zinc-chelating probes were designed (51). However, applicability inside eukaryotic cells has yet to be demonstrated for these zinc-chelating probes, especially for small molecules to be used inside eukaryotic cells. Furthermore, the central three-ring core of the dye Texas Red (44) can be derivatized while retaining binding, as shown by colleagues with impressively strong (25 pM) affinity.

The central three-ring core of the dye Texas Red (44) can be derivatized while retaining binding, as shown by colleagues with impressively strong (25 pM) affinity to the central three-ring core of the dye Texas Red (44). This structure can be derivatized while retaining binding, as demonstrated for the Texas Red-Labeling with more than one aptamer may be required for visualization of hydrophobic targets, but the technology has proven to be versatile.
demonstrated inside cells with a cell-permeable form of the Texas Red-based calcium indicator X:Rhod-5F (44). Labeling with other Texas Red derivatives has yet to be performed, and there is a possibility that the rather hydrophobic Texas Red core will exhibit nonspecific binding, but the TR512-Texas Red system does theoretically feature desirable characteristics of high affinity, flexibility, self-labeling, and intracellular compatibility. In a different approach from self-labeling, a variety of enzymes from microbial biosynthetic pathways have been adapted as tools to conjugate small molecules to protein surfaces. Conceptually, this solves the issue of specificity encoding in a peptide by requiring that specificity determinants reside on both the peptide substrate and in the active site of the conjugating enzyme. The best characterized among these enzymes is biotin ligase, which attaches biotin onto the biotin carboxyl carrier subunit of acetyl-CoA carboxylase. A 67-aa mammalian biotin acceptor sequence supports biotinylation by endogenous cytoplasmic biotin ligases (55) and may also allow biotinylation of secreted proteins by endogenous processes despite the lack of known biotin ligases in the secretory pathway (73). The E. coli biotin ligase BirA can also specifically conjugate biotin or biotin analogs onto a lysine residue in the natural 75-aa substrate or in a 15-aa biotin acceptor peptide (BAP) identified by peptide library screening (4, 12). BAP is not an acceptor for eukaryotic biotin ligases, so its biotinylation is completely dependent on BirA expression. As expected for its small size, BAP can function as an acceptor when inserted into protein domains, but efficiency is affected by variability in access or recognition by BirA (12). Long-term lack of toxicity has been demonstrated by the generation of transgenic mice expressing BirA and BAP-tagged protein (15). Transplanted cells expressing biotinylated proteins on their surfaces can be detected in living mice after intravenous injection of streptavidin-conjugated imaging reagents (73).

Assuming efficient access of biotin ligase to its small molecule co-substrates ATP and biotin, the kinetics of BirA biotinylation of BAP are acceptably fast for labeling a large proportion of tagged proteins in an acceptable amount of time. Conjugation rates are governed by the equation [enzyme][substrate][Kcat]/[Km] when [substrate] << [enzyme], a safe assumption given Km values are in the hundreds of micromolar for the BirA interaction with BAP or intact protein substrates (4). When measured on BirA (preloaded with activated ATP/biotin) and the BAP substrate, kcat/Km is ~1 per microsecond per minute (4). Assuming concentrations of several micromolar for both enzyme and peptide, initial conjugation rates will also be several micromolar per minute. Therefore, this reaction is efficient enough to label a substantial fraction of unreacted peptides over minutes under ideal conditions.

Biotinylation on cell surfaces is made feasible by the lack of endogenous extracellular biotinylated proteins and by access to streptavidin-conjugated reagents for downstream assays (28, 56). Intracellular use, on the other hand, requires probes that can be conjugated to target peptides preferentially over native biotin. Biotin analogs containing chemically reactive groups such as ketones, azides, and alkynes, which are not present in mammalian cells, may allow specific chemical reaction with a variety of appropriately activated compounds in the cytoplasm. However, several parameters must be improved to make this labeling pathway robust enough for routine use. For one, adequate orthogonality with endogenous biotinylation mechanisms needs to be achieved. Ting and colleagues have characterized an alkynyl-containing biotin analog that is not utilized as a substrate by mammalian biotin ligases but is accepted by yBt with only a modest loss in catalytic efficiency (68) and therefore can be attached to a yBt-specific acceptor peptide (9). Likewise, an azide analog is only utilized by Pyrococcus horikoshii biotin ligase (PbBlt), although this proceeds at a 1,000-fold slower rate than biotin, and no small acceptor peptide is yet available (68).

Still, neither the yBt-alkyne nor PbBlt-azide system is currently orthogonal as endogenous biotin in the cell would still serve as a superior substrate, so further mutation may be desired to reduce cross-utilization of native biotin. It is also not clear to what extent biotin analogs are able to enter cells through the same transport mechanisms utilized by biotin. A final limitation is that the subsequent steps of performing chemoselective chemistry on azides and alkynes in the cytoplasm of living cells using currently available reagents are rather inefficient (3). Improving and adapting biotinylation for intracellular protein tagging will therefore continue to be an interesting challenge for biochemists and chemists.

A similar system using the E. coli IbpC lipoic acid ligase LplA, which is structurally related to biotin ligase and functions analogously to attach a lipoic acid prosthetic group to pyruvate dehydrogenase, has also been developed by Ting and colleagues and may provide an easier path to orthogonality (19). Impressively, both a 22-aa LplA acceptor peptide (LAP) and an azide-containing lipoic acid probe were successfully engineered through rational design, although LAP is not as efficient a substrate as the native LplA substrate protein Ep2, and the azide probe shows a relatively high Km for LplA (127 μM vs. 1.7–4.5 μM for lipoic acid). In the presence of high probe concentrations, attachment of the azide probe to extracellularly presented LAP proceeds as quickly as biotinylation. Because mammalian cells contain pyruvate dehydrogenase to the mitochondrial matrix (57), LplA-mediated labeling in the cytoplasm should be possible with minimal background labeling or toxicity. Indeed, some lipoic acid derivatives, including the azide probe, are able to enter cells (A. Y. Ting, personal communication). But intracellular probe concentrations are still lower than...
can be achieved extracellularly, so efficient labeling inside cells may require further evolution of small peptide substrates and/or the LpIA enzyme. Other transferases capable of attaching chemical groups to peptides are limited to extracellular uses for the foreseeable future. The Bacillus subtilis phosphopantetheinyl transferase (PPTase) Sfp catalyzes the transfer of a phosphopantetheinyl group from the metabolic intermediate CoA to bacterial acyl carrier proteins (ACPs) or the related B. subtilis carrier protein (PCP). Another PPTase, E. coli AcpP, functions similarly on a subset of ACPs. Both PPTases tolerate the presence of various functional substrants on the terminal thiol of phosphopantetheinyl and have been successfully used to mediate the covalent functionalization of phosphopantetheinyl ligands from CoA derivatives to ACPs and PCPs, e.g., for multicolor sequential fluorophore labeling or for attaching fluorophores to extracellular regions of proteins (46, 76). Yin and Walsh have identified 12-amino acid sequences that serve as specific substrates for Sfp and AcpP with \( k_m/K_M \) values of 0.19 and 0.015 per micromolar per minute, respectively, in saturating CoA concentrations (86). The former is about 12 times slower than the native substrate protein PCP (85) and several times slower than biotinylation by BirA but is simpler in not requiring ATP as a co-substrate. It is currently limited to extracellular tagging because CoA is cell impermeant, and intracellular use would require proper conjugation of functional groups to CoA by biosynthetic pathways. Other systems for extracellular tagging include the bacterial enzyme sortase, which cleaves the Thr-Gly amide bond in LPXTG sequences and ligates the new bacterial enzyme sortase, which cleaves the Thr-Gly amide bond in LPXTG sequences and ligates the new

**FIGURE 2. Principle of native chemical ligation**

Peptide A with a COOH-terminus thioester group reacts with peptide B with an NH2-terminal Cys residue through thioester exchange, resulting in a thioester-linked initial product. This product then undergoes an intramolecular rearrangement and gives the final polypeptide that is linked by a native peptide bond. Multiple modifications can be introduced into either peptide through chemical synthesis.

**TABLE 1. Comparisons**

<table>
<thead>
<tr>
<th>Protein and Peptide Tag</th>
<th>Comparisons</th>
</tr>
</thead>
</table>
| Dehalogenases and AGT | Mediate specific and irreversible self-labeling with sensitive chemical probes, allowing detection of protein expressed within endogenous levels, but are large. Tetracysteines are capable of stable self-labeling with a large variety of chemical probes at various locations within proteins, but toxicity and background are common issues. Aptamers allow self-labeling and flexible insertion sites in proteins, but they have yet to be evaluated in multiple contexts. Enzymatically mediated conjugation is highly specific and irreversible but requires that the tag be in a location on the protein that is accessible to enzyme, at least transiently. However, requisite co-expression of an enzyme may be a feature rather than a complication depending on the experimental design, e.g., it could allow confinement of labeling to particular times or tissues via restricted enzyme expression. Each system also has a different set of requirements in terms of extracellular or luminal or intracellular environments and a different set of compatible ligands. Currently, none of the characterized enzyme-peptide conjugation systems functions efficiently in the cytosol, leaving AGT, dehalogenase, and tetracysteine as the only covalent labeling methods suitable for intracellular use. Finally, the fusion of tags requires testing that structure or function is not perturbed. Larger tags are more likely to be a problem, especially when NH2 and COOH termini are used as sites for targeted添加新的化学反应外，额外的标签的位置对结构或功能的干扰也可能是一个问题，尤其是在NH2和COOH末端被用作添加额外化学反应的位点的情况下。
methylated lysines and thus provide convenient means to study how lysine methylation affects chromatin structure and function. Nonetheless, chemical modification relies on the unique reactivity of an amino acid side chain. The intrinsic selectivity is low unless no Cys is present or unwanted Cys can all be mutated in the target protein. To label buried residues, the protein may need to be denatured to increase the accessibility and then renatured for functional assay. For intracellular and in vivo labeling, a reduced environment is necessary for Cys labeling, and careful controls need to be performed to exclude complications resulting from labeling other proteins. Thus the overall selectivity and efficiency are limited.

**Introduction of Chemical Modifications by Native Chemical Ligation**

Native chemical ligation and its extension, expressed protein ligation, chemically couple two or more peptide fragments into a full-length protein (14, 47). The COOH-terminal α-disulfide of one peptide reacts with the NH2-terminal Cys of another peptide to form a disulfide derivative.

**FIGURE 3. Genetically encoding unnatural amino acids (UAAs) in live cells**

A new set of components consisting of an orthogonal tRNA, an orthogonal aminoacyl-tRNA synthetase, and a unique codon is required. The synthetase specifically aminoacylates the UAA onto the orthogonal tRNA only. During protein translation, the acylated tRNA delivers the UAA in response to a unique codon that does not encode any of the common 20 amino acids. The UAA cannot be a substrate for the endogenous synthetases and needs to be efficiently transported into the cytoplasm when added to the growth medium or be biosynthesized by the host cell. Various modifications can be site-specifically introduced into proteins in the format of an UAA or by labeling the UAA side chain via an orthogonal reaction.
native peptide link, leaving no chemical artifacts behind (FIGURE 2). Both the thioester peptide and the N-Cys peptide can be prepared either recombiningly or synthetically. Desired modifications can be introduced in the synthetic peptide fragment by using chemical synthesis and incorporated into the target protein after ligation. This method has been successfully used to introduce various probes and modifications for studying protein structure and function in vitro (48). In one elegant intracellular study, a palmitoylated (PalFar) and a hexadecylated (HDFar) Nras protein were prepared using native chemical ligation and microinjected into MDCK cells (61). The hexadecyl group mimics the properties of the palmitoyl group but is not cleavable by enzymes catalyzing de- and repalmitoylation. Although the PalFar-Nras rapidly traffics from the plasma membrane to the Golgi, the HDFar-Nras accumulates nonspecifically in cellular membranes, suggesting that depalmitoylation is directly involved in Nras trafficking.

The native chemical ligation method has the power to introduce diverse structures that are synthetically accessible, and multiple modifications can be incorporated in one synthetic peptide. Nonetheless, because peptides longer than 50 residues are difficult to synthesize using solid-phase peptide synthesis, appropriate sites for cleavage and ligation must be chosen carefully, and modifying internal sites in large proteins is cumbersome. It is also possible to ligate the peptides in living cells. A synthetic fragment can be injected into cells to react with an endogenously produced protein fragment (22). For cell biology and in vivo studies, microinjection of either the in vitro ligation product or the synthetic fragment for intracellular ligation can be a drawback. New methods for protein delivery in cells are awaited to circumvent this issue (49).

Site-Specific Incorporation of Unnatural Amino Acids

Site-specific incorporation of unnatural amino acids (UAAs) into proteins is another powerful technique to selectively modify and label proteins. Schultz et al. developed a general in vitro biosynthetic method in which an amber suppressor tRNA is chemically acylated with an UAA (52). The tRNA and the target gene containing an amber stop codon at the desired mutation site are then added to cell extracts that support transcription and translation. The tRNA recognizes the amber codon and incorporates the attached UAA. A variety of UAAs have been incorporated into proteins using this approach, regardless of the incorporation site or protein size, and have been applied to a large number of problems in protein chemistry (11). An extension of this method involves the microinjection of the chemically acylated tRNA and amber codon-containing mutant mRNA into Xenopus oocytes (53). The endogenous oocyte protein synthesis machinery supports translation and incorporation of the UAA. This change enables the structure-function studies of integral membrane proteins, which are generally not amenable to in vitro expression systems (16, 46). In practice, chemical acylation of the suppressor tRNA is technically demanding. In addition, acylated tRNA is consumed stoichiometrically and cannot be regenerated in cells or cell extracts, which leads to low expression of the target protein.
Genetically encoding an UAA in a manner similar to that of common amino acids would enable site-directed mutagenesis in living cells with UAAs. A general method to expand the genetic code to include UAAs involves the generation of a new tRNA-codon-synthetase set that is specific for the UAA and does not cross talk with other sets for common amino acids (79, 80) (FIGURE 3). The new synthetase is evolved to charge specifically an UAA onto the new tRNA. This tRNA recognizes a codon that does not encode any common amino acids (e.g., a stop codon or an extended codon). In E. coli, the new tRNA-synthetase pair enables the UAA to be site-specifically incorporated into proteins at the unique codon with high fidelity and efficiency.

Genetically encoded UAAs can directly introduce modifications in proteins to facilitate the investigation of biological processes in native settings (FIGURE 4). In addition, p-carboxymethyl-L-phenylalanine is an analog for phosphothreonine but is resistant to hydrolysis by protein tyrosine phosphatases. When it is used to replace Tyr701 in human signal transducer and activator of transcription-1 (STAT1), the resulting mutant protein is constitutively active, which demonstrates and binds to the same DNA sequences as phosphothreonine701STAT1 (84). In a second example, protein phosphorylation in yeast is controlled by a photocaged serine, 4,5-dimethoxy-2-nitrobenzylserine (DMNB-Ser). This UAA blocks phosphorylation and is incorporated at different phosphoserine sites in the transcription factor Pho4. Upon photodecaging (λ = 365 nm), the protein is covalently cross-linked and is incorporated into human GAB2 at positions proximal to the ligand-binding pocket. Upon light exposure (λ = 365 nm), the protein is covalently linked to the epidermal growth factor receptor. Lastly, to probe the inactivation mechanism of the K⁺ channel Kv1.4, UAAs with extended side chains are incorporated into Kv1.4 expressed in HEK239 cells. By comparing the K⁺ currents of the mutant channels, it was found that the bulkiness of residues in the inactivation peptide is essential for fast channel inactivation, a finding that had not been possible using conventional mutagenesis (82).

In addition to introducing modifications directly, UAAs with unique reactive groups can be incorporated as a chemical handle for further protein derivatization, providing new labeling reactions that are orthogonal to those of endogenous amino acids. The ketone, azide, acrylate, and thioester groups have all been genetically encoded and used to modify proteins in vitro (83). The ketone group has also been incorporated into the membrane protein LamB in E. coli and directly labeled with membrane-impermeable dyes. However, intracellular labeling is still challenging, as chemical reactions have not been well developed that occur efficiently under mild cellular conditions and are fully orthogonal to cellular components.

Over 40 UAAs have been genetically incorporated into different proteins in bacteria, yeast, and mammalian cells, and the list is growing (81, 83). These amino acids contain chemical reactive groups, photoreactive active groups, biophysical probes, or moieties mimicking posttranslational modifications. A major goal is to generate stable cell lines or transgenic animals capable of inheriting UAA alterations for long-term studies. However, this method cannot be used to incorporate UAAs that are toxic or incompatible with the protein biosynthesis machinery (e.g., n-amine acids). Encoding the UAA with a stop codon may lead to UAA incorporation into other proteins whose genes end with the same stop codon, which may bring about complications and needs to be investigated. To fully exploit the power of these UAAs, efforts are required to improve their incorporation efficiency in eukaryotic cells, to develop extended codons or unnatural base codons for the incorporation of multiple UAAs simultaneously, and to expand the method to multicellular organisms.

Conclusion

The last few years have seen an explosion in new techniques for site-specific chemical labeling of proteins. These techniques span a wide range in terms of submolecular precision, temporal control, labeling specificity, and ease of implementation. No one system excels in all characteristics, and the choice of system involves finding the best fit for the experimental problem. The techniques with the longest history, AGT, tetra-cysteines, bioinertification, cysteine modification, native protein ligation, and UAA incorporation, have already demonstrated their utility in cultured cells in a variety of questions. For methods that are completely genetically encodable, we expect it will not be too long before some are applied to study proteins in their physiological contexts through the generation of transgenic animals. Meanwhile, ongoing work promises to increase the quality and number of protein labeling techniques.


41. Machleidt T, Robers M, Hanson G. Protein labeling.


