Fluorescent labelling in living cells
Anselm Fabian Lowell Schneider¹ and Christian Peter Richard Hackenberger¹,²

The labelling of proteins with green fluorescent protein enabled the visualization of proteins in living cells for the first time. Since then, much progress has been made in the field. Modern strategies allow the labelling of proteins in live cells through a range of specialized methods with sophisticated chemical probes that show enhanced photophysical properties compared to fluorescent proteins. This review briefly summarizes recent advances in the field of fluorescent chemical protein labelling inside living cells and illustrates key aspects on the requirements and advantages of each given method.

Addresses
¹ Leibniz-Institut für Molekulare Pharmakologie im Forschungsverbund Berlin e.V. (FMP), Campus Berlin-Buch, Robert-Roessle-Str. 10, D-13125 Berlin, Germany
² Humboldt-Universität zu Berlin, Institut für Chemie, Brook-Taylor-Str. 2, D-12489 Berlin, Germany

Corresponding author: Hackenberger, Christian Peter Richard (hackenbe@fmp-berlin.de)

Introduction
Labelling of proteins inside living cells is of tremendous importance in the characterization of protein dynamics, mobility and function and thus in the understanding of biological processes. Labelling of proteins in vivo is not a new area, with proteins being expressed as fusion constructs with fluorescent proteins such as GFP to monitor their cellular localization as far back as 1995 [1]. The creation of fusion proteins with GFP has been tremendously useful to study protein localization, movement, interactions and function inside living cells. However, labelling with GFP has significant drawbacks as proteins fused to the fluorescent protein can have altered localization and, due to its considerable size, altered function as well. More sophisticated strategies for the fluorescent labelling of proteins in living systems are being developed constantly towards obtaining methods that are fast, selective and minimally disruptive to the living system.

Chemical labelling of proteins is more versatile than labelling via fusion proteins and opens the possibility of labelling with a wide array of functionalities, most notably fluorescence, with bright organic dyes of a broad range of colours. Countless techniques for chemical protein labelling exist and are discussed broadly in several great review articles [2,3]. In this article, only techniques that can be applied to the labelling of proteins inside living cells are compared and contrasted in order to guide the reader to identify the optimal labelling strategy for a given scenario.

The specific chemical labelling of proteins inside living cells poses a special challenge, as labelling reactions have to run under mild conditions in the cell interior and be very selective for the target protein. As such, very specialized techniques exist, ranging from the labelling of target proteins via entire protein domains, over the usage of specialized peptide tags to the labelling of endogenous proteins in living systems (see Table 1 for an overview).

Protein domain recognition strategies
Proteins have the ability to show high affinity binding to natural as well as synthetic ligands. This can be of use in protein labelling, wherein the protein of interest is fused to a protein showing high affinity for a ligand that is itself labelled. The small ligands can penetrate cells and bind the corresponding protein inside the cell, in turn leading to the indirect labelling of the protein of interest as part of the fusion protein. Labelling of a protein with ligand-binding or self-labelling enzymatic protein domain brings along a substantial increase in size that can potentially be disruptive to the protein’s localization and function but in turn promises high specificity and reaction rate with a great variety of labels. Selected labelling methods using protein domains are shown in Figure 1.

Protein–ligand–interactions
The FK506 binding protein (FKBP) binds a number of natural and synthetic compounds with high affinity thus rendering them good candidates for such a labelling method—however, FKBP ligands can also interact with endogenous FKBP's. FKBP is a mutant of the FK506 binding protein that binds the synthetic ligand SLF with a more than 1000-fold selectivity in comparison to the wildtype protein [4]. This interaction can be used to label...
proteins of interest in live mammalian cells, by expressing the protein of interest as a fusion to FKBP\(^*\) and then applying a fluorescently labelled cell-permeable SLF\(^*\), which leads to labelling of the FKBP\(^*\) fusion protein inside living cells [5].

Trimethoprim (TMP) is an inhibitor of the 18 kDa *Escherichia coli* dihydrofolate reductase (eDHFR). TMP shows high affinity towards eDHFR that is not substantially disrupted by derivatisation of TMP while showing much lower affinity for mammalian analogues of the enzyme. Because of this, it is possible to visualize fusion proteins expressed with eDHFR in mammalian cells with the help of fluorescently labelled TMP that can penetrate cells and bind eDHFR [6].

Since its inception, this system has seen substantial improvements. Recent approaches utilize the introduction of a reactive cysteine residue in the eDHFR protein near the TMP binding site while the fluorescent TMP label is outfitted with an acrylamide electrophile leading to the formation of a covalent bond between the two thus making the labelling essentially irreversible (Figure 1a) [7]. A more modular approach has been developed as well, in which the TMP carries an azide label that is then covalently attached to the protein of interest via the TMP-eDHFR interaction inside living cells before it is addressed in a second step with dibenzocyclooctyne (DBCO) or bicyclononyne (BCN) carrying probes [8].

**Self-labelling protein domains**

The 20 kDa large DNA repair protein O\(^6\)-alkylguanine-DNA alkyltransferase (AGT) catalyzes the attachment of O\(^6\)-alkylguanine or O\(^6\)-benzylguanine or a derivative to a cysteine residue on the enzyme. As a fusion to a protein of interest, AGT (marketed as SNAP tag) can catalyze the labelling of the fusion construct with a fluorescent, cell permeable derivative of O\(^6\)-benzylguanine (Figure 1b) [9]. An orthogonal system, the CLIP tag, has also been developed that utilizes derivatives of O\(^2\)-benzylcytosine instead [10]. Because of the orthogonality of SNAP and CLIP, they can be used together in double labelling approaches such as pulse-chase experiments [10] or as split-tags that activate and become fluorescently labelled upon protein–protein interaction, allowing the individual monitoring of proteins after they dissociate [11]. The Halo tag (33 kDa) is another example of a self-labelling enzyme that was engineered from the bacterial enzyme haloalkane dehalogenase to attach a haloalkane derivative to a reactive aspartate residue in the enzyme’s active center [12].

One issue with self-labelling enzymes is that washing steps are often required to remove the unbound fluorophore from the cells in order to avoid a high background signal, which complicates labelling experiments with a quick readout. Therefore, labels have been developed for SNAP, CLIP and Halo tags that turn on fluorescence upon labelling, making it possible to image without a

### Table 1

<table>
<thead>
<tr>
<th>Labeling technique</th>
<th>One-step or two-step labeling</th>
<th>Covalent or non-covalent</th>
<th>Size of genetic tag</th>
<th>Modification site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein domain-based labelling strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBP(^*) – SLP</td>
<td>One-step</td>
<td>Non-covalent</td>
<td>12 kDa</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>eDHFR – TMP</td>
<td>One-step or two-step</td>
<td>Non-covalent</td>
<td>18 kDa</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>SNAP tag, CLIP tag</td>
<td>One-step or two-step</td>
<td>Covalent</td>
<td>20 kDa</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>Halo tag</td>
<td>One-step or two-step</td>
<td>Covalent</td>
<td>33 kDa</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>PYP</td>
<td>One-step or two-step</td>
<td>Covalent</td>
<td>14 kDa</td>
<td>N- or C-terminus</td>
</tr>
</tbody>
</table>

### Peptide-based labelling strategies

<table>
<thead>
<tr>
<th>Labeling technique</th>
<th>One-step or two-step labeling</th>
<th>Covalent or non-covalent</th>
<th>Size of genetic tag</th>
<th>Modification site</th>
</tr>
</thead>
<tbody>
<tr>
<td>His tag = N(^{2+}) – NTA</td>
<td>One-step</td>
<td>Non-covalent</td>
<td>10 amino acids</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>D4 tag – DpaTyr</td>
<td>One-step</td>
<td>Non-covalent</td>
<td>4–12 amino acids</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>Tetracycline (CCXCC)</td>
<td>One-step</td>
<td>Covalent</td>
<td>6 amino acids</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>dC10x</td>
<td>One-step</td>
<td>Covalent</td>
<td>22 amino acids</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>Sortase A</td>
<td>Two-step</td>
<td>Covalent</td>
<td>6 amino acids</td>
<td>C-terminus</td>
</tr>
<tr>
<td>LpIA</td>
<td>One-step or two-step</td>
<td>Covalent</td>
<td>13 amino acids</td>
<td>N- or C-terminus</td>
</tr>
<tr>
<td>Cell-penetrating DnaE inteins</td>
<td>One-step or two-step</td>
<td>Covalent</td>
<td>~11 kDa depending on choice of intein pair</td>
<td>C-terminus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Labeling technique</th>
<th>One-step</th>
<th>Covalent</th>
<th>1 amino acid</th>
<th>N- or C-terminus, or internal</th>
</tr>
</thead>
<tbody>
<tr>
<td>UAA</td>
<td></td>
<td></td>
<td></td>
<td>Internal (proximal to ligand binding site)</td>
</tr>
<tr>
<td>LDT</td>
<td>One-step</td>
<td>Covalent</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>
washing step [13] even if the labelling itself is quite slow and the background is high.

The more recently developed PYP (photoactive yellow protein)-tag, based on a protein derived from purple bacteria, is smaller in size than the SNAP-, CLIP- and Halo-tags (14 kDa) and can form a stable thioester bond with derivatives of 4-hydroxycinnamic acid [14]. In order to develop a rapid staining method with low background even without washing, fluorescent probes were designed for the PYP-tag that show only low fluorescence in the polar environment of the cell interior but much higher fluorescence in the low-polar environment of the PYP-tag binding pocket (Figure 1c). Thereby it was possible to complete the labelling of intracellular proteins without washing after only 6 min [15].

**Peptide recognition strategies**

Peptides, due to their smaller size, promise to be less invasive than entire protein domains when attached to a protein of interest. However, techniques utilizing an affinity peptide as a recognition motif to label a target protein are often less selective and lead to lower specificity of the labelling. Selected labelling strategies are shown in Figure 2.

**Metal-ion-affinity**

Metal chelation capabilities of peptide tags such as the Ni\(^{2+}\)-chelation of the poly-histidine-tag in combination with N-nitrotriacetic acid (NTA) are commonly used for affinity chromatography of proteins. It can also be used as a tool for protein labelling with the advantage that only a small peptide tag is used that is only minimally disruptive.
to protein function. The strong and specific interaction between NTA and the His tag has been applied to the non-covalent fluorescent labelling of histagged membrane proteins on the surface of live cells using an NTA-chromophore conjugate [16].

More recently, this technique has also been applied to the labelling of poly-histidine-tagged proteins inside living cells through the usage of a cell-penetrating multivalent NTA complex. For this, the cell-penetrating TAT (trans-activator of transcription)-peptide was synthesized with an N-terminal His6-tag used in combination with a trivalent and fluorescently labelled NTA (trisNTA). Upon binding of the cell-penetrating peptide to the trisNTA, the complex can enter the cells, where the trisNTA will preferentially bind the His10 tagged protein of interest (due to the 10-fold higher binding affinity of NTA to the longer tag) resulting in non-covalently labelling (Figure 2a) [17*].

The tetra-aspartate (D4 tag, DDDD) is an alternative to the His tag and can be labelled through the use of zinc complexes, such as the DpaTyr complex based on 2,2′-dipicolylamine on a tyrosine scaffold. This has been used...
in the non-covalent labelling of membrane proteins on living cells using a triply repeated D4 tag [18]. It has also been applied to the covalent labelling of proteins inside live E. coli cells, for which the D4 tag is outfitted with a cysteine while the DpaTyr label carries an N-α-chloroaacetyl moiety, leading to the formation of a covalent bond upon binding [19].

**Self-labelling-tags**

FlAsH (fluorescein arsenical helix binder) is a fluorescent chemical label that allows the specific modification of tetracycline (CCXXCC) motifs in proteins. First reported in 1998 by Roger Tsien, it was the first notable alternative to tagging proteins with GFP. The peptide tag is short and due to its small size and labelling occurring on the amino acid side chains, it can be introduced at any position in a protein instead of just the termini.

The FlAsH label itself is also small (<1 kDa), therefore cell-permeable and can be used to specifically label proteins inside living cells (Figure 2b) [20]. A range of different biarsenic dyes have been created that vastly increase the range of applications for the system, for example allowing pulse-chase experiments [21], or monitoring protein dynamics or folding inside living cells [22]. Nevertheless, FlAsH and similar biarsenic compounds show high background labelling as well as cell toxicity, requiring the addition of 1,2-ethanediol and thus making them less-than-ideal labels for experiments inside living cells.

Another prominent tag is the fluorette tag. It binds a TexasRed fluorophore and through a reactive cysteine residue can form a covalent bond with an N-α-chloroaacetylactamide group on the fluorophore, enabling one-step labelling of proteins in live bacterial cells that are tagged with the genetically encodable tag [23]. The fluorette tag is small and the labelling is specific and rapid, although it is limited to the TexasRed fluorophore.

The more recently developed dC10α-tag represents a self-labelling peptide tag with an alpha-helical structure presenting two cysteine residues, one on each end of the helix. The corresponding label contains a fluorophore and two distinct maleimide functionalities quenching the fluorescence through photoinduced electron transfer (PeT) until both maleimides have undergone thiol addition to the dC10α-tag (Figure 2c) [24]. In order to prevent unspecific additions of the maleimide reagent to glutathione or other free thiols in the cell interior, the reactivity of the maleimide was attenuated using a methoxy substituent on the maleimide which decreases unspecific reactions while maintaining high reaction rate with the two reactive cysteines in the dC10α-tag. The authors selectively label actin in the cytosol and histone H2B in the nucleus of live cells with only minor unspecific labelling [24].

**Enzyme catalyzed label transfer**

Sortase A is a transpeptidase enzyme that requires an LPXTG recognition sequence on the N-terminal fragment and an oligoglycine sequence on the C-terminal fragment and catalyzes the formation of a peptide bond between the two fragments [25]. Bioconjugation using sortase A is versatile and can be used to attach a variety of functionalities to proteins. However, because of the peptidic nature of the recognition sequences, sortase A is mostly limited to protein labelling in vitro, but has been used for protein labelling on live cell-surfaces [26] and even in the ligation of GFP to proteins of interest in various cellular compartments inside live mammalian cells [27].

An engineered variant of sortase A that increases the activity while eliminating the need for calcium in the reaction was now used in the labelling of proteins inside living E. coli cells with substrate mimics containing a bioorthogonal handle [28]. For this, the sortase A variant is simply coexpressed with the protein of interest and the substrate mimic containing an azide functionality is added to the cells. After cell lysis, the clarified lysate was treated with an azide-reactive dye and gel electrophoresis could show that only the protein of interest was labelled to a high degree. This study could also enable two-step labelling of proteins inside mammalian cells in the future.

The E. coli enzyme lipoic acid ligase (LplA) recognizes a short peptide sequence and catalyzes the ligation of lipoic acid onto proteins important for the E. coli metabolism [29]. The enzyme is somewhat promiscuous and can be used to ligate an azide derivative to an engineered acceptor peptide sequence, enabling two-step labelling with cyclo-octyne carrying probes on the cell surface of living cells [30]. More recently, the LplA has been engineered to directly accept a coumarin derivative as substrate and thereby label the acceptor peptide fluoroscently in one step [31]. A coumarin derivative carrying two acetoxy-methyl (AM) groups to mask the negative charges on the molecule was able to enter mammalian cells where the AM protecting groups are removed by endogenous esterases. The molecule then serves as substrate for the LplA and a 13 amino acid long peptide sequence fused to the protein of interest can be labelled in the cytosol and nucleus of living cells. The method reportedly shows higher specificity and lower toxicity than arsenic-based probes while keeping the tag length short [31]. The usage of coumarin is not ideal in cellular contexts as the wavelength required to excite the fluorophore is harsh on cells. Through computational design it was now also possible to create a LplA mutant that is able to instead ligate a red fluorophore (Figure 2d) [32*].

Split inteins are a useful tools for protein engineering and labelling, because during protein trans splicing (PTS) the
intein fragment is excised leaving behind a protein with only a small peptide sequence. The naturally split DNA polymerase III (DnaE) intein pair from *Nostoc punctiforme* (Npu) has an exceptionally high reaction rate and is extraordinarily robust, showing reactivity at temperatures between 6 and 37°C, with different extein sequences and in the presence of up to 6 M urea [33]. Because of its natural robustness, the DnaE intein pair has previously been used to label proteins on the surface of live cells with synthetic moieties containing bioorthogonal handles in a two-step labelling strategy [34].

The C-terminal fragment of the DnaE intein in combination with the N-terminal fragment of a split intein from the cyanobacterium *Anabaena variabilis* (Ava) has now also been applied to the labelling of histones inside live cells. For this, the C-terminal intein sequence including the desired cargo were fused to a cell-penetrating peptide (CPP) via a disulfide bond [35**]. The construct enters the cell and the CPP is removed by reductive cleavage of the disulfide. The cargo diffuses into the nucleus and can then undergo trans-splicing with the target protein tagged with the N-terminal intein sequence, leading to the excision of the intein and leaving behind a labelled histone protein (Figure 2c). This was used to semi-synthetically generate a native histone modification inside live cells and evaluate its epigenetic effect [35**].

**Endogenous protein labelling strategies**

Strategies that allow labelling of proteins without the use of an affinity tag are the most modern approaches and perhaps the most elegant.

**Unnatural amino acids**

Unnatural amino acids (UAAs) allow the introduction of a bioorthogonal handle onto a protein with only a single amino acid change as opposed to a peptide or a protein domain. A variety of bioorthogonal handles for labelling can be introduced using a range of unnatural amino acids discussed in detail in recent review articles [36,37]. However, despite some recent advances, the usage of unnatural amino acids requires extensive engineering of the living system in question and can still often lead to low yields of labelled protein.

**Ligand-directed ligation**

Ligand-directed tosyl chemistry (LDT-chemistry) can be used to modify an endogenous protein at a specific site. This is done through the usage of a ligand bearing a tosyl ester. The ligand binds the target protein at a specific site and through a nucleophilic attack from a nearby amino acid side-chain will label the target protein at a residue proximal to the binding site. The label is removed from the ligand, which can then liberate the binding site and leave a functional, labelled protein behind (Figure 3) [38]. The technique has already been used more extensively *in vitro* and on cell surface proteins [39], where it has now also been used in a bimolecular fluorescence quenching and recovery (BFQR) assay to identify modulators of a cell-surface receptor [40]. In living cells, the technique has recently been applied to the labelling of native FKBP12 in order to characterize the protein’s interaction with the FKBP-rapamycin-binding domain from mTOR (FRB) [41].

Because the technique does not require a protein or peptide tag it is possibly less invasive than other labelling methods, and it allows the labelling of proteins on a site near the active center instead of the termini of the protein. However, LDT-chemistry requires extensive knowledge of the system in question as the ligand molecules have to be carefully designed. The reaction also does not go to completion and LDT-chemistry is slow: in
the case of FKBP12 incubation with the ligand in vitro at 37°C shows a labelling yield of 70% after 48 hours [41].

Faster ligand-directed labelling approaches are being developed [42,43] but they still cannot match other labelling methods in speed.

Conclusion
The labelling of proteins inside living cells has made unprecedented advances in recent years. Flexible chemical labelling of proteins offers great advantages over the labelling with fluorescent proteins, with sophisticated probes being developed for fluorescence imaging, NMR and mass spectrometry.

Yet, the rule of thumb that the smaller the tag, the lower the specificity of the labelling reaction still often holds true, and even a label as minimal as a His tag can perturb a proteins structure or function [44]. LDT-chemistry can circumvent this problem, but suffers from its own set of problems, most prominently the amount of labour required and the low labelling rate. Therefore, work remains to be done when it comes to creating quick, stable, versatile and specific labelling strategies that do not lead to the perturbation of the living systems in question.

Conflict of interest
There is no conflict of interest relating to this article.

Acknowledgements
The authors acknowledge support from the DFG (SFB 765 and SPP 1623), the Fonds der Chemischen Industrie (Doctoral fellowship to A.S.), the Einstein Foundation, and the Boehringer-Ingelheim Foundation (Plus 3 award).

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


4. A comprehensive overview of the chemistries and corresponding labelling strategies that are available for bioorthogonal, chemical labelling of proteins in different contexts.


The authors develop a near-infrared fluorophore that has ideal characteristics for in vivo high resolution imaging with its high permeability, brightness and stability. They go on to apply this fluorophore using different labelling techniques in various cells and tissues.


The authors demonstrate a potent labeling strategy that should be immediately applicable to a great variety of proteins as the only requirement for its usage is a His-tag.


The article offers great insights on the work done to generate an enzyme mutant that can ligate a substrate of choice to a peptide tag. The authors then apply the ligase in different cell types and compare their labeling method with other, commonly used methods.


The authors propose a labeling strategy that is minimally disruptive as it leaves behind a protein with only a minimal label through intein transsplicing. The authors demonstrate their strategy with different labels and characterize the outcome in detail.


A great example of what is possible using ligand-directed ligation of a fluorescent reporter to a target protein of interest and subsequent evaluation.


