Labeling Proteins for Single-Molecule Imaging

Introduction

Fluorescent microscopy techniques rely on the fact that molecules of interest fluoresce or can be made to fluoresce. Most proteins of interest (POI) require a fluorescent tag, as intrinsically fluorescent POIs are the exception.

A fluorescent signal allows a proteins precise location to be determined as well as its behavior to be observed. This can then be studied in relation to other fluorescently tagged proteins. In this way, researchers are able to use fluorescence microscopy to visualize underlying biological processes.

Proteins can be labeled in a variety of ways, such as by other – fluorescent – proteins or by using artificial dye molecules. However, depending on spatial restrictions, some of these methods can be more beneficial than others. Sometimes, protein epitopes are not accessible by larger structures, or environmental conditions, such as pH, can restrict the use of a technique.

This document explains the most common protein labeling techniques, gives advice for selecting the best fluorophores for a range of samples and examines how camera choice can help overcome some fluorophore weaknesses.

Considerations for protein labeling

In general, two common approaches are used for protein labeling; in-/direct labeling using dye molecules connected to the protein by a linker and the use of genetically encoded fluorescent proteins.

When using these labeling techniques, there are three important criteria to consider that determine the final resolution and accuracy with which a structure can be described:

1. **Size**

   The size, but more importantly the location, of the fluorophore, is critically important. The size of a dye molecule is largely irrelevant as it is typically a fraction of a nanometer (nm), the real problem arises from the way the dye molecule is targeted to the POI, it must be targeted in a way that does not affect the activity of the POI.

2. **Brightness, blinking & photostability**

   Fluorescent proteins are usually dimmer (lower quantum yield), bleach quicker and have slower blinking kinetics than organic dye molecules. The blinking frequency and number of blinking cycles a molecule can undergo, as well as the ability to be pushed easily into the dark triplet state and be reactivated, is important for techniques such as

STORM and PALM. Fluorescent proteins may manage to undergo 10-100s of cycles whereas organic dyes can blink 1000s of times before breaking down and being permanently bleached (Figure 1). STORM and PALM super-resolution microscopy crucially rely on the number of emitted photons and the ability of the dye molecule to blink, therefore organic dyes are preferred.
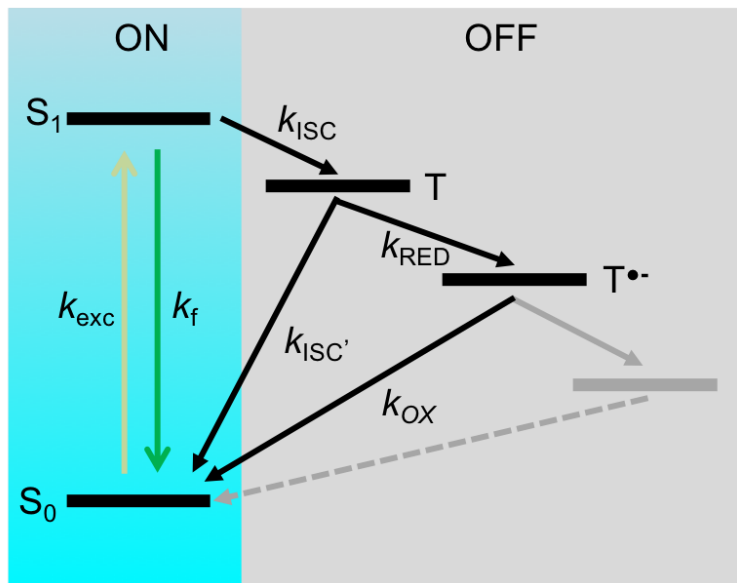

Figure 1: Underlying photophysical processes of reversible photoswitching of Alexa Fluor and ATTO dyes.
After excitation of the fluorophores (k_{exc}) into their first excited singlet state S_1, the excited-state energy is either released through fluorescence emission with rate k_f, or the triplet state is occupied (k_{ISC}). The triplet state T is either depopulated ($k_{ISC'}$) or reduced by thiolate RS (k_{RED}) to form radical anion $T^{\bullet-}$, which is reoxidized by molecular oxygen with rate k_{OX} or reacts to other nonfluorescent species. (modified from Heilemann *et al.*, Angew. Chem. Int. Ed., 2009)

3. Location of the POI

Surface proteins can be easily labeled with antibodies. But one of the biggest disadvantages of the antibody labeling approach is that antibodies can't easily cross the plasma membrane.

To enable the researcher to label intracellular structures and compartments, cells need to be permeabilized, i.e. the plasma membrane is made to no longer function as a barrier between extra- and intracellular space. As a consequence, cells will die quickly when permeabilization is performed live, which in turn requires the cell to be fixed. Fixation, usually performed by fixatives (organic solvents, cross-linkers), kills cells and arrest them at a certain time point which prevents the researcher from performing any dynamic measurement.

Methods of Labeling Proteins

Antibodies

Antibodies are proteins and part of the immune system. They help to identify intruders, such as bacteria and viruses. Each antibody uniquely recognizes a specific antigen structure (i.e. an epitope such as proteins/polysaccharides) on its matching target. The enormous specificity arises from the antibody detecting the matching epitope.

The most commonly applied method of antibody labeling is by using an unlabeled primary antibody (AB) to target a protein of choice and then targeting a fluorescently labeled secondary AB to the primary antibody (Figure 2). Considering the structure of a conventional AB – consisting of 2 heavy chains and 2 light chains with a total size of ca. 150 kDa – is roughly 10 nm in length, this will add a fluorescent structure 20 nm away from the actual location of the protein.
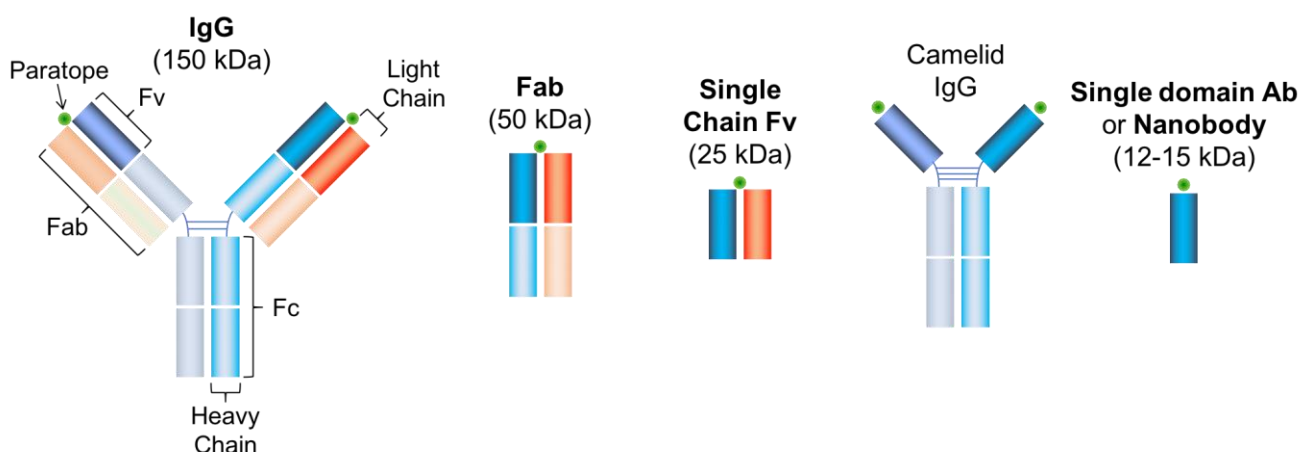

Figure 2: Antibodies. Different types of ABs with decreasing size. Conventional IgG consisting of 2 heavy and two light chains is roughly 150 kDa in size. A Fab fragment is 50 kDa and specific to only one epitope. The Single Chain Fv consists only of the variable domain and one paratope, altogether 25 kDa in size. The camelid IgG can be modified and turned into a single domain antibody with only 12-15 kDa.

Moreover, a conventional polyclonal AB – a mixture of antibodies recognizing various antigens on the same target structure - will recognize multiple sites on a POI. This will add further uncertainty to the determined position. A monoclonal AB – one identical antibody species only detecting one target structure - will decrease the likelihood of this artifact.

Nevertheless, secondary ABs are usually polyclonal to enhance the signal and increase the number of localization events per structure. The fact that this adds uncertainty is usually ignored as the number of photons is more important – although this depends on the detection system and its sensitivity.

To get the fluorophore closer to the protein of interest, a labeled primary antibody is a good option. This is called direct immunostaining. Nevertheless, it takes away the flexibility to change the fluorophore of choice and to adapt to various imaging modalities. It also is more expensive.

It is possible to get the fluorophore closer to the POI by using only the labeled Fab-fragment – the part of the antibody which contains one light chain, half a heavy chain and only one paratope - an AB which is only half of the size of a single AB (5 nm). The benefit of this approach is the improved specificity of the fragment as it only consists of a single domain and potential crosslinking as thereby prevented. Moreover, smaller ABs can bind much easier to otherwise inaccessible protein domains.

If the Fab-fragment is still too large, a labeled Fv-domain can be used. The Fv-domain consists only of the antigen-recognition site of a Fab-fragment and roughly halves the size again. It retains the benefits of the Fab-fragment in terms of specificity and crosslinking.

As a valid alternative, ABs can be produced in species which consist only of 2 heavy chains (Heavy chain ABs, e.g. camelids, sharks, crocodiles). If this is still too big, the heavy-chain AB can be modified to function as a nanobody. The nanobody consists of a labeled single Fv domain, is less than a tenth of the size of a conventional AB (12-15 kDa) and is therefore only a couple of nm long.

One problem when using antibodies is that they cannot move across cell membranes. Therefore, cells must be permeabilized before labeling can begin, which usually means they are dead, and hence live cell imaging is not possible.

Biotinylation

Biotinylation involves covalently binding biotin to a POI, nucleic acid or other molecule. Biotin has very high-affinity interaction partners in avidin and streptavidin, as well as highly specific antibodies. Avidin and streptavidin can then be modified to express a fluorescent tag.

In contrast to some other techniques, biotinylation seems not to interfere with the underlying physiological processes enabling live cell imaging. To avoid aggregation/clustering of biotinylated targets, monovalent streptavidin can be used.

Epitope tags

Epitope tags (Figure 3) can be an option if no good antibody is available for immunofluorescence labeling of the protein of interest. Those epitope tags can be expressed as a fusion construct – genetically encoded and directly attached to the protein of interest. The short epitope tags then can be labeled by high-affinity antibodies which are commercially available.
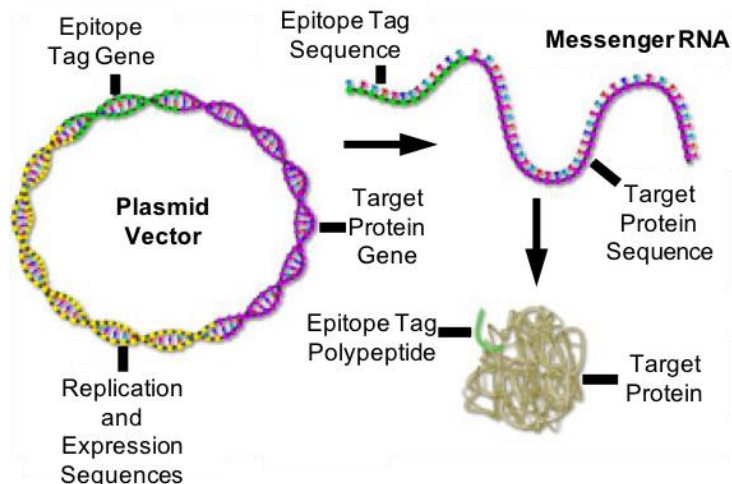
Figure 3: Epitope tags. The epitope tag gene gets incorporated into a plasmid vector and upon expression gets transcribed and translated into an easily recognizable polypeptide (https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/confocal/applications/epitopetagging)

Examples of these tags are:

- **Flag-tag** (sequence - DYKDDDDK): First epitope tag to be published and is patented by Sigma-Aldrich. A designed peptide.

- **HA-tag** (sequence - YPYDVPDYA): Fragment 108-106 of human influenza hemagglutinin, which is the antigenic glycoprotein responsible for the binding of flu virus to the surface of cells.

- **V5-tag** (sequence – GKPIPNPLLGLDST): Derived from a small epitope (Pk) present on the P and V proteins of simian virus 5 (SV5).

- **Myc-tag** (sequence - EQKLISEEDL): Fragment of the myc transcription factor and proto-oncogene.

The epitope tag needs to be at an exposed position at the POI, which sometimes can be tricky to achieve and requires careful design. It may also have an unlikely, but possible negative side-effect on the functionality of the POI.

Small Molecule Probes

Rather than using fluorescent proteins, which have their drawbacks, it is possible to introduce a non-fluorescent protein which binds a ligand tagged with a fluorescent dye molecule. Two classes can be distinguished; self-labeling (Figure 4) and enzyme-mediated (Figure 7).

1. **Self-labeling genetically encoded tags**

Based on very high affinity and specificity of the involved reagents, self-labeling genetically encoded tags can be used when an antibody against a protein of interest isn't available. A protein of interest gets modified by attaching an additional recognition side which is accessible for a fluorescent probe.
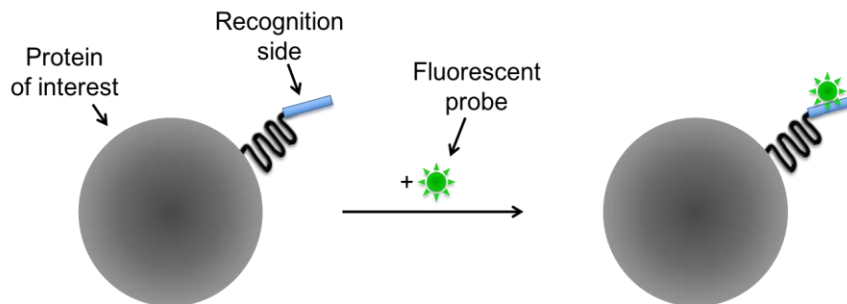
Figure 4: Self-labeling genetically encoded tags. The protein of interest is modified to have a recognition site which specifically binds a matching fluorescent probe (Modified from Fernandez-Suarez and Ting, Nature Reviews Molecular Cell Biology, 2008)

As one of the first examples, FlAsH was used as a chemical surrogate to fluorescent proteins for labeling proteins in living cells with organic fluorophores.

FlAsH (Figure 5) and **ReAsH** are membrane permeable biarsenical derivatives of fluorescein (green) and resorufin (red) which are initially nonfluorescent. Upon binding to recombinant proteins containing a tetracysteine tag (CCXXCC) they become fluorescent.
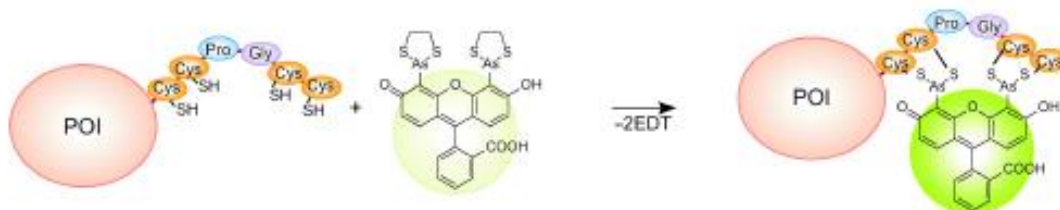

Figure 5: FlAsH. Initially not fluorescent, the two ethane dithiols attached to the Fluorescein will be released during the binding process, and the tagged protein of interest becomes visible.

A big issue is non-specific labeling of thiol-rich biomolecules. Cysteine-rich proteins can, by this, cause a large background fluorescence, burying the intended signal.

The **HaloTag** (Figure 6) is a newer – also membrane-permeable - alternative, consisting of 33 kDa hydrolase that is expressed as a fusion tag on the protein of interest. Upon addition of a ligand comprising of a reactive linker (chloroalkane) and a functional group (e.g. fluorescent dye), this will bind rapidly and irreversibly to the HaloTag.

Various options are available which allows the user to select the fluorophore characteristics they are interested in. Versions having Oregon Green and diAcFAM, for example, are fluorogenic, i.e. are non-fluorescent until cleavage by intracellular esterases.
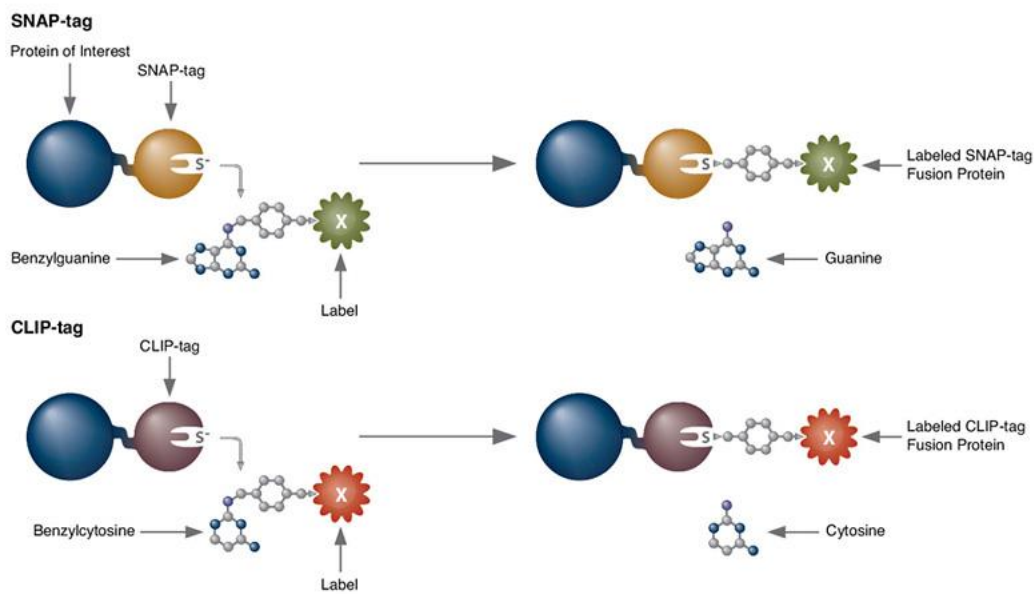
Figure 6: SNAP- and CLIP-tags. Mechanism: SNAP-tag or CLIP-tag fused to the protein of interest labels itself with "X" releasing guanine or cytosine. (copied from https://www.neb.com/tools-and-resources/feature-articles/snap-tag-technologies-novel-tools-to-study-protein-function)

The **SNAP-tag** is a 20 kDa mutant of a DNA repair enzyme that is expressed as a fusion tag on a protein of interest and works similarly to the HaloTag i.e. it binds irreversibly to a synthetic probe consisting of a reactive linker (benzylguanine) and a label (e.g. fluorescent dye).

The **CLIP-tag** is a modified version of the SNAP-tag, which has been altered to recognize a benzylcytosine linker.

Both, SNAP- and CLIP-tags can be expressed in the same cell and fused to different proteins for dual-color labeling.

2. Enzyme-mediated labeling genetically encoded tags

Enzyme-mediated labeling requires the presence of an enzyme to covalently bind the reactive group to the linker (Figure 7). As examples, ACP- and MCP-tags will be used which are based on the acyl carrier protein (ACP).

A benefit of the ACP- and MCP-tags is that, at 9 kDa, they are smaller than SNAP/CLIP-tags at but require an additional enzyme to catalyze the labeling step.

The use of ACP/MCP-tags is problematic in that their substrates are non-cell-permeable and can't be used for intracellular proteins. Also, the same fluorescently-conjugated substrates will label both ACP- and MCP-tags.

The specificity is therefore determined by the synthase used for labeling. However, while the specific ACP-Synthase (4′-phosphopantetheinyl transferase) will modify predominantly the ACP-tag, the non-specific SFP-Synthase (4′-phosphopantetheinyl transferase) will label both ACP- and MCP-tags. Hence, sequential incubation is required for

dual-color labeling and it must be hoped that all reaction sides were labeled during the ACP-step, as otherwise false-positives could be observed after the second step.
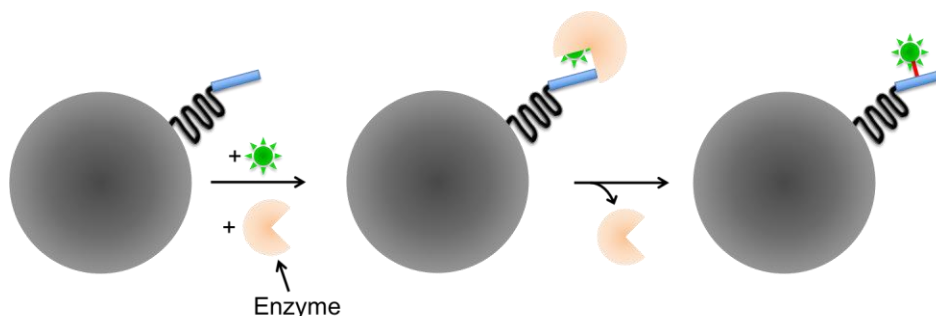

Figure 7: Enzyme-mediated labeling genetically encoded tags. The required enzyme catalyzes the binding reaction between reaction site and fluorescent probe (modified from Fernandez-Suarez and Ting, Nature Reviews Molecular Cell Biology, 2008).

Bioorthogonal Labeling

To avoid interfering with the cell's metabolism, researchers have developed a method which circumvents direct cross-reaction with physiological processes by stepping into the word of bioorthology (Figure 8).

Bioorthogonal reactions occur between two functional groups that do not react with biological molecules, but selectively react with each other e.g. tetrazines and strained alkenes/alkynes.

To enable a protein of choice to be useful for bioorthogonal labeling techniques, genetic code expansion is required. Genetic reprogramming facilitates the site-specific incorporation of unnatural amino acids bearing bioorthogonal functional groups into proteins in bacteria, mammalian cells, and animals.
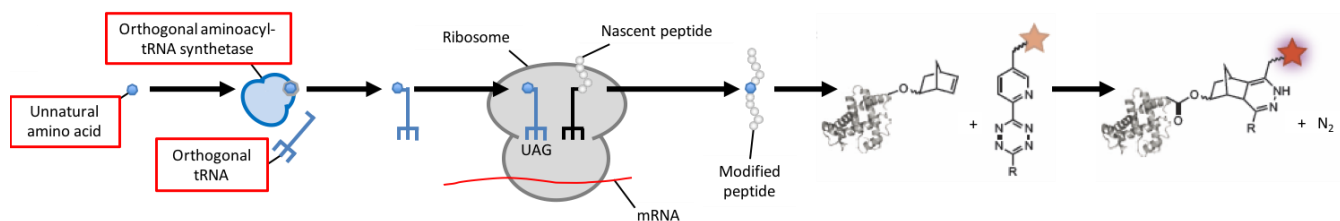

Figure 8: Bioorthogonal labeling scheme. The stop-codon UAG is used to encode for the unnatural amino-acid. To achieve this a suitable tRNA synthetase and orthogonal tRNA need to be available. Once the amino-acid is incorporated in the modified peptide, it can be made visible by using the according label. Usually, this process is fluorogenic. (Modified from Davis and Chin, Nature Reviews Molecular Cell Biology, 2012 & Romanini and Cornish, Nature Chemistry, 2012).

Incorporating an unnatural amino acid requires an orthogonal aminoacyl-tRNA synthetase/tRNA pair. Attaching a tetrazine linker to fluorophores quenches their fluorescence prior to labeling with the unnatural amino acid, thereby reducing background fluorescence.

As with fluorescent proteins, protein expression can be tightly controlled by regulating amino acid availability. If no unnatural amino acid is fed to the cell, no tagged protein will be produced and hence won't interfere with the cell's metabolism.

Fluorophores

The methods mentioned above detail how a protein of interest can be detected but labeling always requires the presence of a fluorescent probe. Throughout the last century, various methods have been developed to highlight structures for microscopy techniques. Of particular interest are labels which can easily be detected by the eye, but, more recently, by sensitive detection devices such as scientific cameras.

Most of the fluorescent dyes used in microscopy have certain wanted and unwanted characteristics. Usually, they get excited by a lower wavelength (higher energy light), get elevated from a ground state to an excited state and - while falling back to the ground state - emit photons of a longer wavelength (lower energy light). This, in turn, can be detected by suitable devices with the help of optics suited to the label of choice.

Various labels will be discussed below, and their advantages and disadvantages briefly described.

Organic dyes

1. **FITC/TRITC**

The fluorescence activity of organic dyes can usually be attributed to their large conjugated aromatic electron system, which is excited by short wavelength light and emits light at longer wavelengths. The first ones to be reported were fluorescein isothiocyanate (FITC) and tetramethylrhodamine isothiocyanate (TRITC).

FITC and TRITC are now considered to have very poor characteristics and are out-performed by more modern dyes.

2. **Cyanines**

Based on cyanine, popular dyes include Cy2, Cy3, Cy5, and Cy7 (Figure 9). All of which can be linked to nucleic acids or proteins via their reactive groups. For proteinous labeling, maleimide groups are used for example. Interestingly – concerning fluorescence – Cy5 is sensitive to its electronic surrounding. This can be utilized for enzyme measurement. Conformational changes of the attached protein lead to positive or negative alterations in fluorescence emission.
Like FITC/TRITC, cyanine dyes are mostly outperformed by more modern fluorescent dyes in terms of brightness, photostability, and quantum yield.

3. Alexa Fluor® dyes

Alexa Fluor® dyes provide researchers with the largest spectral variety available (Figure 9). Alexa molecules are negatively charged and hydrophilic fluorescent. Their name is usually an indication for the appropriate laser excitation wavelength, Alexa Fluor®488, for example, is excited by 488 nm light.
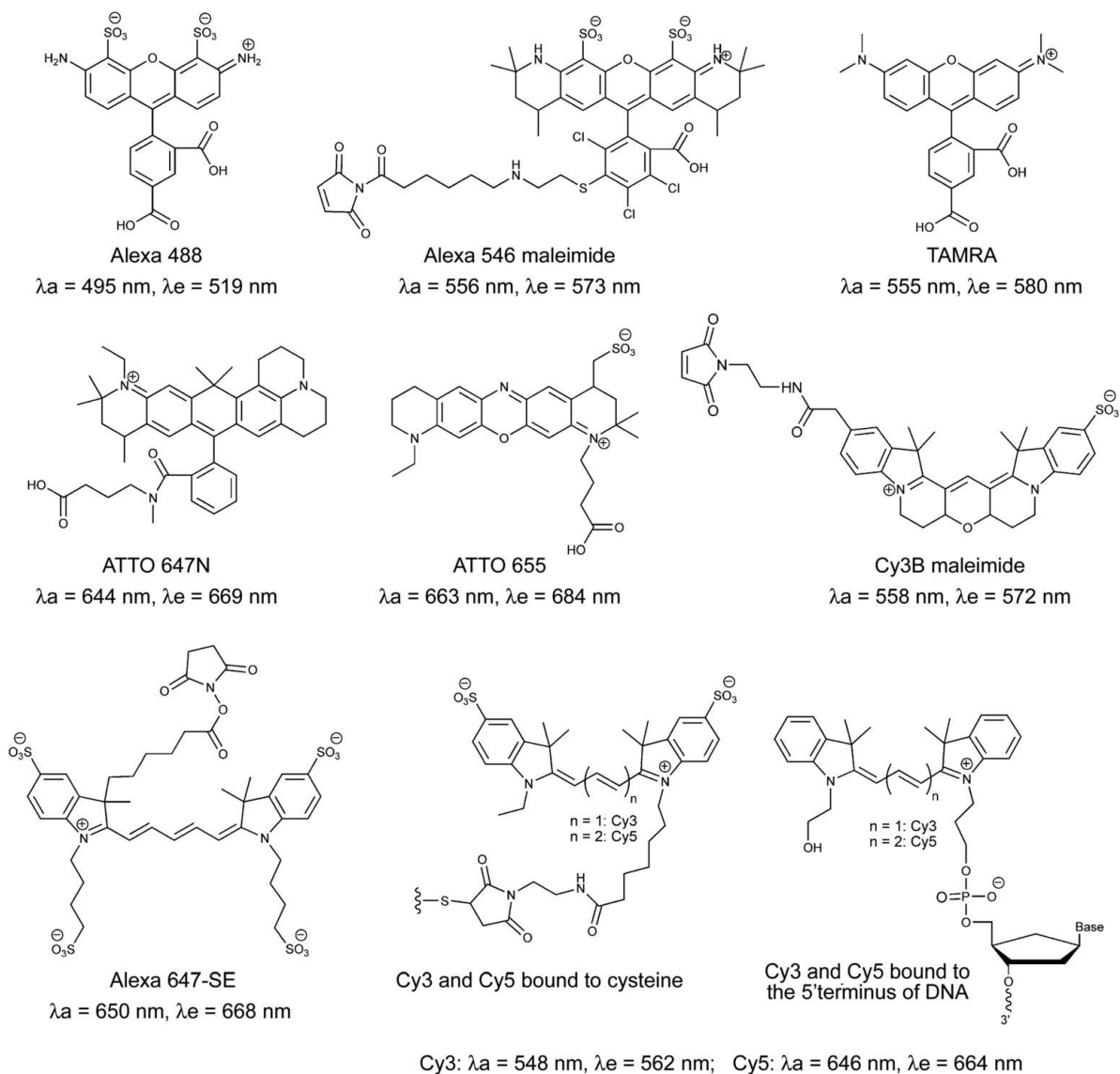

Figure 9: Chemical structures of representative dyes used in single molecule fluorescence. Alexa 488, Alexa 546, and TAMRA are rhodamines, Cy3B, Alexa 647, Cy3 and Cy5 are carbocyanines, ATTO 655 is an oxazine, and ATTO647N is a carbopyronin. Maleimide derivatives react with the thiol group of cysteine residues as exemplified with Cy3/5 (bottom-center). Succinimidyl esters (SE) are commonly used to label primary amino groups. Note that the phosphoramidites of Cy3 and Cy5 used in the solid-state synthesis of DNA oligonucleotides do not contain sulfonate substituents (bottom-right). la and le refer to the absorption and emission maxima of the free dyes as reported by the vendors (Cy3 and Cy5: GE Life Sciences, Alexa dyes, and TAMRA: Life Technologies, ATTO dyes: Atto-Tec GmbH; from Stennett *et al* (2014)).

They are superior to older dyes, Alexa Fluor®488 is a fluorescein derivate but displays better stability, brightness, and lower pH sensitivity.

The entire range of Alexa Fluor® dyes originates from sulfonated forms of different basic fluorescent substances like fluorescein, coumarin, cyanine or rhodamine.

Quantum Dots

Quantum dots are an alternative to fluorescent dyes and proteins, they display a much brighter signal and are very photostable (Figure 10).

Quantum dots consist of a heavy metal core coated with various layers of shielding to protect the cell from the detrimental effects of the heavy metal. The surface can be functionalized easily to recognize, for example, antibodies or streptavidin.
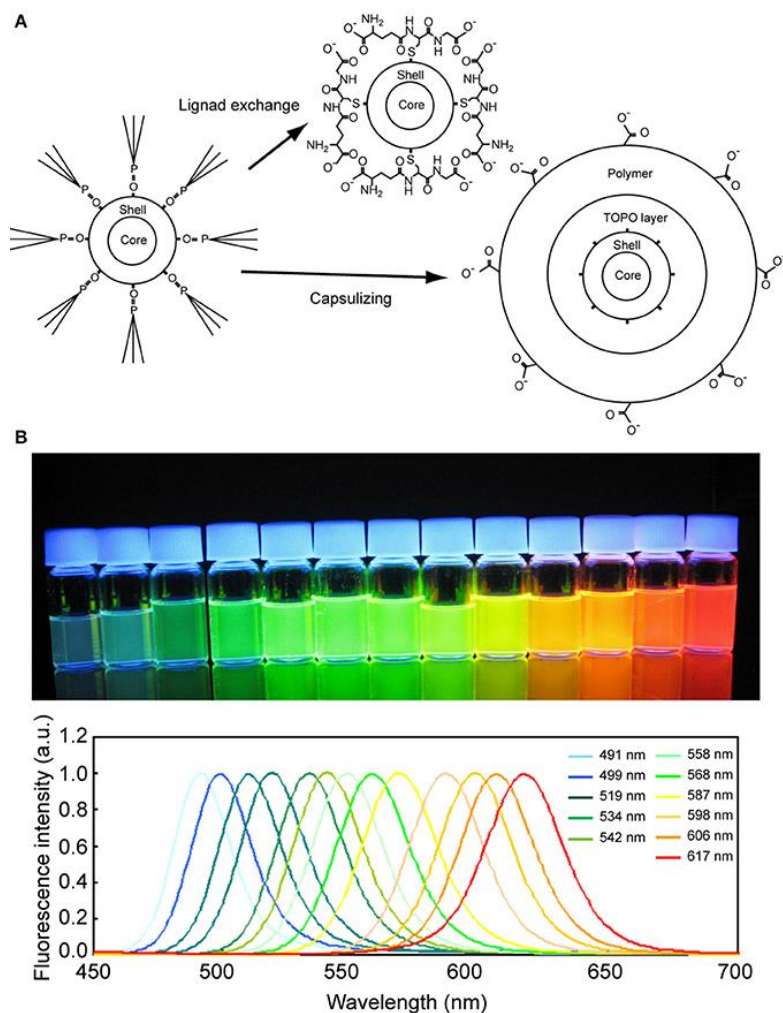

Figure 10: Quantum dots. **(A)** Schematic drawing of the surface modification of a Qdot. **(B)** Fluorescence photograph (upper) and spectra (lower) of Qdots of various diameters. The Qdots were excited by a UV light of 365 nm wavelength. (from Ichimura *et al.*, 2014).

The initial size varies between 5 and 10 nm, which is of crucial importance as the size determines the emission characteristics of the quantum dots. They feature a large absorption range but very narrow emission spectra, which makes them ideal for multiplexing. A positive side effect of the heavy metal core is that quantum dots are also ideal for correlative fluorescence and electron microscopy.

The functionalization potentially increases the size of quantum dots significantly. Hence, this can restrict access to smaller compartments and structures and gives the labeled molecules a tendency to aggregate. Furthermore, quantum dots can blink which might be an undesired effect but might actually be useful for localization based super-resolution techniques such as STORM.

Fluorescent proteins

Aside from the development of antibodies for use in research, the utilization of fluorescent proteins in the late 1970s was one of the biggest breakthroughs in science. Fluorescent proteins are very small and highly specific, as they can be genetically incorporated into a protein of interest. This means that one protein of interest is labeled by one fluorescent protein – a perfect ratio (although in some cases tandem/tandem-dimer variants can be used to increase the brightness). This, in turn, is a good predictor of individual fluorophores' behavior.

The targeted protein of choice can either be labeled by overexpression of a fusion-construct or production of a transgenic cell-line/animal. The result does, under normal circumstances and careful planning, not interfere with the physiological function of the used organism.

The very first fluorescent proteins being reported was the "Green Fluorescent Protein" (GFP, Figure 11) in 1962 by Osamu Shimomura. GFP is derived from *Aequorea Victoria*, a jellyfish found in the Pacific Ocean. GFP emits green light (ca. 525 nm) when excited with blue light (ca. 490nm). Additionally, fluorescent proteins have a size benefit as green fluorescent protein is less than 3 nm.
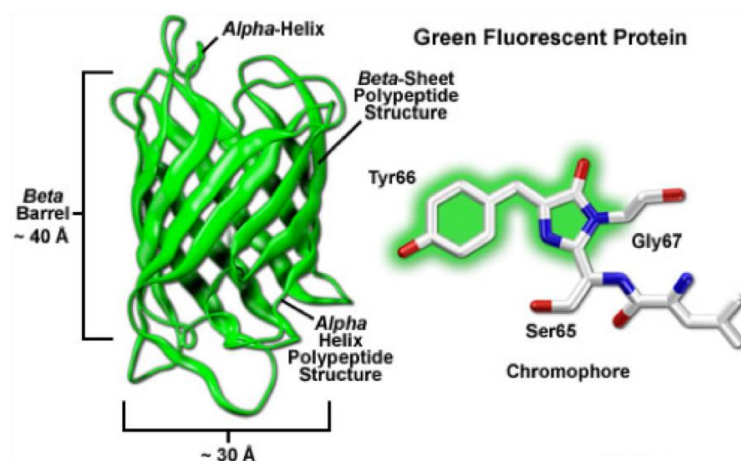

Figure 11: Green Fluorescent Protein. Basic structure of beta-barrel and chromophore. (http://zeiss-campus.magnet.fsu.edu/)

The basic structure of a fluorescent protein consists of a chromophore which is embedded in a so-called beta-barrel, protecting the chromophore reasonably well from certain negative effectors such as too low/high pH which will alter the characteristics of the specific fluorescent protein. Various discoveries, modifications, and optimizations have led to variants which display different spectral properties (throughout the entire visual spectrum), show higher photostability, are brighter and monomeric.

Selecting the Right Camera

Observing fluorescently labeled structures in fixed and living samples is the essence of modern biological science.

With fixed samples, there is usually no limit to the number of photons which can be harvested from individual structures of interest as those are labeled with a very high abundance of antibodies or with bright dye molecules. Researchers can benefit from large field of views available from state-of-the-art Scientific CMOS cameras allowing them to increase their sample throughput massively. The use of lower magnification lenses with high numerical apertures are also beneficial and in consequence, would require cameras with smaller pixels to allow for Nyquist sampling.

Imaging live samples with high light levels risks altering the behavior of the structure of interest, resulting in less physiological data. Therefore, it is crucial to use imaging devices that detect the maximum number of photons emitted from the structure with as little excitation light as possible. This requires a camera with high sensitivity, particularly at the wavelength of interest. Higher sensitivity also allows for lower exposure times to be used to increase imaging speed and, therefore, observation of true physiological processes.

Summary

Selecting the correct labeling technique crucially determines the outcome of an experiment. In particular, when it comes the location/precision of the label in relation to the protein of interest. However, the choice can be further complicated by the limited accessibility of the structure of interest.

The options for the label offer as wide of a range as the labeling techniques themselves. The number of photons emitted, the spectral response, overall characteristics (blinking, cycle time, quantum yield, sensitivity to the environment etc.), all features have to be carefully considered.

References

Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. (2007) **Multicolor super-resolution imaging with photo-switchable fluorescent probes.** *Science*. Sep 21;317(5845):1749-53.

Davis, L. & Chin, J. W. (2012) **Designer proteins: applications of genetic code expansion in cell biology.** *Nat Rev Mol Cell* Biol. Feb 15;13(3):168-82. doi: 10.1038/nrm3286.

Fernández-Suárez, M. & Ting, A. Y. (2008) **Fluorescent probes for super-resolution imaging in living cells.** *Nat Rev Mol Cell Biol.* Dec;9(12):929-43. doi: 10.1038/nrm2531.

Heilemann, M., van de Linde, S., Mukherjee, A. & Sauer, M. (2009) **Super-resolution imaging with small organic fluorophores.** *Angew Chem Int Ed Engl.* 48(37):6903-8. doi: 10.1002/anie.200902073.

Ishikawa-Ankerhold, H. C., Ankerhold, R. & Drummen, G. P. (2012) **Advanced fluorescence microscopy techniques--FRAP, FLIP, FLAP, FRET and FLIM.** *Molecules*. Apr 2;17(4):4047-132. doi: 10.3390/molecules17044047
Ichimura, T., Jin, T., Fujita, H., Higuchi, H. & Watanabe, T. M. (2014) **Nano-scale measurement of biomolecules by optical microscopy and semiconductor nanoparticles.** *Front Physiol.* Jul 29;5:273. doi: 10.3389/fphys.2014.00273

Romanini, D. W. & Cornish, V. W. (2012) **Protein labelling: Playing tag with proteins.** *Nat Chem.* Mar 22;4(4):248-50. doi: 10.1038/nchem.1325.

Stennett, E. M., Ciuba, M. A. & Levitus, M. (2014) **Photophysical processes in single molecule organic fluorescent probes.** *Chem Soc Rev.* Feb 21;43(4):1057-75. doi: 10.1039/c3cs60211g.

Tschesche, H. (2011) **Methods in Protein Biochemistry.** *De Gruyter.* ISBN-10: 9783110481501, ISBN-13: 978-3110481501