THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Shining a light on the dark secrets of the cell

OPEN ACCESS

Shining a light on the dark secrets of the cell: synthetic proteins for better fluorescence imaging

Curran Oi, Diana Tokarska and Lynne Regan

Abstract

The discovery of intrinsically fluorescent proteins revolutionized our ability to visualize proteins within living cells. Since that original observation, a plethora of fluorescent proteins with varied color and brightness have been obtained. For a number of imaging purposes, however, synthetic biology approaches have been required to create new labelling methodologies. Here we describe the protein engineering technologies that underlie some of those key designs and show how they have been used to great effect in different cell types.

Introduction

It is a grand challenge of biology to be able to see every molecule in the cell, and to track when and where they move. Proteins are responsible for the vast majority of functions in a cell, and consequently much research has focused on observing and tracking them. Ideally, we would be able to track multiple proteins, simultaneously, in live cells. Importantly, any method we use to label and track a protein must not change the latter's behavior.

The discovery of intrinsically fluorescent proteins (FPs) was game-changing and recognized by the 2008 Nobel Prize. Prior to their discovery, essentially the only way to observe a protein in a cell was to chemically fix (i.e. kill and immobilize) cells, then visualize specific proteins by immunofluorescence. FPs mature and fluoresce without the requirement for any other proteins or small molecule co-factors. Thus, they can be produced in virtually any cell and will spontaneously fold and fluoresce, in the live cell. Thankfully, many proteins tolerate being genetically fused to a FP without any impact on their function. Figure 1 (A, B) shows two beautiful examples of intracellular proteins directly fused, at the gene level, to FPs and visualized by fluorescence microscopy.

Sometimes, however, such fusions do perturb function, or are not compatible with the types of measurement researchers wish to make. In these situations, a battery of different protein engineering and synthetic biology approaches can be used. Here we describe several approaches of this kind and show how they have been applied to overcome a variety of limitations associated with direct fusions (Figure 2A).

The first FP was green fluorescent protein (GFP) but an abundance of new FPs with different colors and properties have subsequently been discovered in new species or created in the laboratory using mutagenesis combined with screening. Consequently, we now have a large repertoire of different FPs.
The minimum prerequisite for a FP to be genetically fused to a natural protein, without perturbing its function, is that the FP is monomeric. Many naturally occurring FPs tend to be dimers or tetramers, so mutagenesis was required to create synthetic, monomeric proteins. Without this intervention, FPs associate into higher order structures, dragging along the protein to which it is fused and change the localization and properties of the protein under investigation. Researchers have found that this phenomenon can be particularly severe when membrane proteins are fused to a FP that weakly self-associates. When such FPs are fused to soluble proteins, their weak propensity to associate can be insignificant, but when fused to a membrane protein the local concentration is much higher, and the 'weak' association becomes a serious issue.

How can creating new synthetic proteins expand the range of applications of FPs even more? Although direct fusion to a monomeric FP works well for many proteins, it does not work for all. For some proteins, especially membrane proteins, direct fusion to a FP prevents the protein from maturing and reaching its final functional location properly. In other situations, one wants to increase the signal from the protein of interest over the background fluorescence, or only track proteins that were labelled at a particular time.

Protein engineering/synthetic biology approaches have created novel solutions to these issues.

Post-translational labelling - covalent

Hinrichsen and colleagues invented a strategy to covalently label proteins, post translationally, with a FP. In this strategy, they fused the protein of interest to a short peptide (SpyTag) and they expressed the FP fused, at the gene level, to SpyCatcher (or a related protein, for example SpyoIPD). The SpyTag-SpyCatcher system is an excellent example of re-engineering a natural protein to create a synthetic one that can be used in a variety of different applications. In the cell wall of the bacterium *S. pyogenes*, a covalent isopeptide bond forms spontaneously between the side-chains of a Lys and an Asp residue between two β strands in the CnaB2 domain of the protein FbaB.

The SpyTag-SpyCatcher system was engineered by Howarth and colleagues, who showed that one of the β strands (SpyTag) could be completely removed from the protein FbaB, expressed separately, yet when it comes into contact with the remaining protein (SpyCatcher) it inserts and the covalent bond forms. Thus, any proteins that are fused to SpyCatcher and SpyTag become covalently linked (Figure 2C). This synthetic protein is a powerful addition to our *in vivo* engineering repertoire, because there are few other ways to covalently link two proteins, especially within living cells.

Making use of this SpyTag-SpyCatcher system, Hinrichsen and colleagues expressed SpyCatcher-FP from an inducible promoter in *Saccharomyces cerevisiae*. They could thus control the timing and the amount of this protein produced. Using this strategy, they were able to post-translationally label a variety of proteins. Of particular note was the labelling of the membrane protein plasma membrane ATPase 1 (Pma1) - the main proton pump of the plasma membrane of *S. cerevisiae*. Direct fusions of Pma1 to a FP result in accumulation of the Pma1-FP in the vacuole (Figure 3B), with a concomitant growth defect. By labelling Pma1 post-translationally, both the vacuolar accumulation and growth defect were avoided. Thus, the true behavior of Pma1, in live cells, could be observed for the first time (Figure 3C). Moreover, because the labelling is covalent, by switching off production of SpyCatcher-FP, then spatiotemporally tracking the Pma1 already covalently labelled with a FP, they were able to measure the lifetime of the protein in the membrane.

How could the applications of such covalent post-translational labelling be expanded? Perhaps the most obvious extension of this method would be to label multiple proteins simultaneously. To accomplish such 'multiplexed' labelling, non-cross-reacting (also referred to as orthogonal) SpyCatcher/SpyTag pairs are needed. Veggiani and colleagues reported a different, naturally occurring protein-peptide pair, which they named SnoopCatcher/SnoopTag. SnoopCatcher and SpyCatcher are orthogonal, so this might have been the first such pair that could be used simultaneously with SpyTag/SpyCatcher in live cells. Unfortunately, SnoopCatcher spontaneously accumulates in the nucleus, which precludes its use in the desired fashion.

It would also be useful if a set of different strength promoters were available, so that the level of expression of the labelling SpyCatcher-FP could be modulated. Dueber and colleagues described the systematic characterization of the strengths of different constitutive yeast promoters (11), which will be useful for post-translational *in vivo* labelling applications. To date, however, the repertoire of controllable (inducible) yeast promoters available is much more limited. Hinrichsen and colleagues used a GAL1 promoter to control expression of SpyCatcher-FP, which makes the transcriptional response to galactose linear allowing a more gradual change in expression levels in response to different levels of induction. This work was done in a strain in which the GAL2 gene was deleted (gal2Δ). In a gal2Δ strain, the amount of protein expressed via the GAL1 promoter is linear with respect to the concentration of galactose in the growth media.

Covalent post-translational labelling uniquely enables additional types of experiment to be performed. For example, if a Spy-Tagged protein is covalently labelled with SpyCatcher-FP expressed from a GAL1 promoter, expression from that promoter can be switched off by the addition of glucose. One can then spatiotemporally follow the fate of proteins that were labelled before the switch to glucose, for example measuring their lifetime without having to perturb the cell by adding cycloheximide (the translational inhibitor typically used in lifetime measurements).

Enhancing Signal:Background

In any fluorescent labelling and imaging, it is desirable to minimize background fluorescence. This background noise is a result of freely diffusing FPs, which are not bound to the protein of interest.

Kamiyama and colleagues used a protein engineering strategy to optimize the signal to background, successfully reducing background fluorescence by ensuring that only the bound form of the FP was fluorescent. They accomplished this feat by creating a split version of super-folder GFP (sfGFP) in which one of the eleven β strands (GFP_{11}) was expressed separately from the other ten (GFP_{1-10}). When these two parts of the FP are separate, the protein is not fluorescent. When GFP_{11} binds to GFP_{1-10}, the structure of the protein is complete, and is fluorescent (Figure 2D). Their ingenious solution was to fuse GFP_{11} to the target protein they wished to visualize and separately expressed GFP_{1-10}.

Because the non-bound form of the FP (GFP_{1-10}) is not fluorescent, this method of labelling is quite forgiving with respect to the level of expression of the FP: High expression levels do not significantly increase background fluorescence.

They were able to use this method to image several different proteins in mammalian cells and were able to expand this technique to multicolor imaging in a reasonably straightforward fashion by making the point mutations Y66W and T203Y in GFP_{1-10}, thus creating CFP_{1-10} (cyan) and YFP_{1-10} (yellow), respectively.

It is highly desirable to be able to perform multicolor imaging by simultaneously tagging multiple proteins in the cell with different colored fluorescent probes. Green and red are commonly used colors, because they are spectrally distinct, thus minimising bleed-through across different excitation and emission filters used on microscopes. Kamiyama and colleagues therefore sought to expand their method to red FPs by applying the approach described above to mCherry (monomeric cherry) and sfCherry (super-folder cherry). The technique was not nearly so straightforward with these proteins as with sfGFP and sfGFP-derived FPs.

Although co-expressing mCherry$_{11}$ with mCherry$_{1-10}$ resulted in increased fluorescence compared with expressing mCherry$_{1-10}$ alone, the reconstituted fluorescent signal was far less than that of the original, un-split mCherry. Thus, additional engineering was required. The researchers first tried switching their efforts from mCherry to sfCherry. However, despite splitting and reconstituting sfCherry resulting in a higher intensity fluorescence than was observed for mCherry, it was not sufficiently bright to be useful in imaging applications. Similarly, the researchers achieved improved fluorescence intensity using tandem repeats of sfCherry$_{11}$ fused, at the gene level, to a protein of interest, but the method was still not ideal for imaging applications (Figure 4).

Feng and colleagues therefore developed a novel screening strategy for engineering new split FP with improved signal to background fluorescence levels. They performed multiple rounds of random mutagenesis and DNA shuffling, which propagates beneficial mutations using repeated cycles of DNA fragmentation and overlap polymerase chain reactions to reamplify the gene of interest. After each round of mutagenesis or DNA shuffling, they selected the brightest colonies as the source of template DNA for the next round.

Using this approach, they were able to create two new split FPs: sfCherry2$_{1-10/11}$, which is 10-fold brighter than sfCherry$_{1-10/11}$ (Figure 5) and can now be used in cell imaging applications in the same way that was so powerful for GFP$_{1-10/11}$ and mNeonGreen2$_{1-10/11}$, which is a yellow-green split FP that offers improved signal to background over GFP$_{1-10/11}$ (Figure 6).

Post-translational labelling - non-covalent

For some applications, it is desirable to label the protein of interest neither by a direct fusion, nor by covalent, post-translational labelling, but via a non-covalent peptide: peptide-binding module-FP interaction (Figure 2B).

There are several requirements for optimal operation of such a labelling strategy. First, the peptide-binding module-FP must not interact with other cellular proteins; it must bind the peptide that is fused to the protein of interest with high specificity. It is extremely hard to foresee whether a protein will have issues with promiscuous interactions. Highly positively-charged proteins, which may interact with DNA or the cell membrane, is one thing to avoid, but beyond such extremes, it is not predictable.

Speltz and colleagues used a module that binds only to peptides with a free C-terminus. Thus, they were able to narrow down the possible cross-reactivity of their peptide-binding modules by assessing the 'C-terminome' of the cell for sequences with few miss-matches to the peptide with which they were tagging the protein of interest. Even so, they experimentally tested their designs, assessing the ability of cell extracts to compete with the protein of interest in 'pull down' assays. They were able to identify three peptide:peptide-binding module pairs whose interaction is not abrogated by the interaction of either partner with cellular components.

Additionally, if one seeks to label two proteins at once, the peptide:peptide-binding-module pairs must not cross-react, they must be orthogonal to each other. Speltz et al. were able to create such pairs, using a combination of rational design and chemical intuition, combined with experimental assessment at every stage.

Finally, if one is labelling a protein *via* a non-covalent interaction, it is vital to control the expression level of the peptide-binding module-FP relative to the dissociation constant of the peptide:peptide-binding module complex (Figure 7). If the expression level is too high, there will be excess, unbound binding module-FP and the background will be high. Conversely, if the expression level is too low, then not all the protein of interest will be bound to binding module-FP.

One can use either constitutive promoters of different strengths, or well-controlled inducible promoters. In *E. coli*, the use of the arabinose promoter, from which transcription is linearly proportional to the concentration of inducing arabinose is one strategy. Similarly, in *S. cerevisiae* one can use a GAL1 promoter, in a gal2Δ background, to obtain a linear response of transcription versus concentration of galactose.

New gene editing technologies will allow creative imaging experiments, such as those described above, to be implemented in mammalian cells. Currently, far fewer well-characterized promoters are available for use in mammalian cells, though finding and testing more is a topic of much current investigation.

Summary

By utilising synthetic biology and protein engineering approaches, researchers have devised clever new ways to fluorescently label proteins for live cell imaging. Using different types of post-translational labelling, they have circumvented the perturbation of function sometimes associated with directly fusing a protein to a FP. They have also devised novel strategies for spatiotemporal tracking and for enhancing signal:background. In the future, we can anticipate expanding such approaches still further, as synthetic biology expands the range of constitutive and controllable promoters available to use in different cell types. In addition, we also envision new methods being developed to enable the researcher to track multiple proteins simultaneously, over time, and at ever higher resolution. Thus, the combination of synthetic biology and protein

engineering provides new tools that enable us to better understand how cells work and move us towards meeting a grand challenge of biology.

Figure Legends
Figure 1
Fluorescence microscopy images of two different intracellular proteins directly fused, at the gene level, to FPs. Images are pseudocolored to match the emission profiles of the different FPs. (A) Actin with FP mTFP1 (a cyan FP) fused to its C-terminus. (B) Microtubule associated protein EB3 with FP YPet (a yellow FP) fused to its N-terminus. Adapted and reproduced with permission from Shaner NC, Patterson GH, Davidson MW. Advances in fluorescent protein technology. Journal of Cell Science. 2007;120(24):4247-60.

Figure 2
Comparison of different methods for fluorescently labelling a protein of interest. In all parts of the figure, the protein of interest is shown as a gray rectangle and the FP is shown as a bright green rectangle.
(A) Direct fusion, at the gene level, of a FP to the protein of interest.
(B) Non-covalent labelling of a protein of interest with a FP. The protein of interest is fused, at the gene level, to a short peptide tag (dark purple) and the FP is fused, at the gene level, to a peptide-binding module (light purple). Thus, the protein of interest is labelled by the non-covalent interaction of the peptide with the binding module.
(C) Covalent labelling of a protein of interest with a FP.
The protein of interest is fused, at the gene level, to the SpyTag peptide (yellow) and the FP is fused, at the gene level, to the SpyCatcher protein (orange). Thus, the protein of interest is labelled by the covalent interaction of the peptide with the binding module. The covalent bond between SpyTag and Spycatcher is indicated by the red lock. Far right: ribbon representation of the x-ray crystal structure of the SpyTag/SpyCatcher complex (4MLI). The covalent isopeptide bond is indicated by red sticks.
D) Labelling of a protein of interest by reassembly of a FP. The protein of interest is fused to a peptide that corresponds to a β strand (dark green bar) of a FP. The rest of the FP (light green) is expressed separately. Only when these two parts of the FP assemble is the protein fluorescent (bright green). Far right, a guide to the eye, based on the x-ray crystal structure of sfGFP (6DQ1), showing the β strand (dark green) that inserts into the remainder of the protein (pale green) to reconstitute a complete FP.

Figure 3
Comparison of different methods to visualize the plasma membrane protein Pma1.
(A) Immuno-staining of fixed *S. cerevisiae* cells (using anti-HA antibodies, which target a peptide epitope derived from the human influenza hemagglutinin protein) in a strain expressing Pma1 fused, at the gene level, to the HA peptide. This 'native' Pma1 localizes exclusively to the plasma membrane. None is evident in the vacuole. Reproduced with permission from Mason AB, Allen KE, Slayman CW. C-terminal truncations of the Saccharomyces cerevisiae PMA1 H+-ATPase have major impacts on protein conformation, trafficking, quality control, and function. Eukaryot Cell. 2014;13(1):43-52.
(B) Live cell fluorescence imaging of *S. cerevisiae* expressing, from the endogenous Pma1 promoter, a Pma1-EGFP (Enhanced GFP) fusion protein. A significant amount of fluorescence is observed in the vacuole, in addition to that at the plasma membrane.

(C) Post-translational labelling of Pma1. Live cell fluorescent imaging of *S. cerevisiae* expressing Pma1-SpyTag from the endogenous Pma1 promoter. Pma1 was labelled by the covalent interaction of SpyTag with SpyCatcher-EGFP, which was expressed from the GAL1 promoter. In this image, a post-labelling period of incubation with glucose was used, which enhances the signal: background.

Reproduced with permission from Hinrichsen M, Lenz M, Edwards JM, Miller OK, Mochrie SGJ, Swain PS, et al. A new method for post-translationally labelling proteins in live cells for fluorescence imaging and tracking. Protein Eng Des Sel. 2017;30(12):771-80.

Figure 4

Increasing the fluorescence signal by using tandem repeats of $sfCherry_{11}$.

(A) Cartoon representation of the multiple tandem repeats method. The protein of interest (grey rectangle) is fused, at the gene level, to four tandem copies of β strand 11 of sfCherry ($sfCherry_{11}$) (dark green ribbon) separated by short linkers. Thus, four molecules of $sfCherry_{1-10}$ can bind to this array, giving four $sfCherry_{1-10/11}$ assembled proteins (bright green) attached to each protein of interest.

(B) Comparing fluorescent images of β-actin labelled with single or multiple copies of $sfCherry_{1-10/11}$. Left: β-actin fused to a single $sfCherry_{11}$. Right: β-actin fused to four tandem repeats of $sfCherry_{11}$. In both scenarios, $sfCherry_{1-10}$ is overexpressed from the constitutive cytomegalovirus (CMV) promoter. Scale bars are 5 μm.

Reproduced with permission from Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, et al. Versatile protein tagging in cells with split fluorescent protein. Nat Commun. 2016;7:11046.

Figure 5

Comparing the whole cell fluorescence intensity (measured by flow cytometry) of HEK 293T cells expressing full length sfCherry, full length sfCherry2, assembled $sfCherry_{1-10/11}$ or assembled $sfCherry2_{1-10/11}$, as indicated.

Adapted and reproduced with permission from Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B. Improved split fluorescent proteins for endogenous protein labelling. Nat Commun. 2017;8(1):370.

Figure 6

Labelling histone H2B (HB2) or clathrin light chain A (CLTA) using assembled split FP. Top left: H2B fused, at the gene level, to $mNeonGreen2_{11}$ ($mNG2_{11}$) with $mNG2_{1-10}$ expressed in trans from the CMV promoter. Top right: CTLA fused, at the gene level, to $mNeonGreen2_{11}$ ($mNG2_{11}$) with with $mNG2_{1-10}$ expressed in trans from the CMV promoter. Bottom left: H2B fused, at the gene level, to GFP_{11} H2B with GFP_{1-10} expressed in trans from the CMV promoter. Bottom right: CLTA fused, at the gene level, to GFP_{11} with GFP_{1-10} expressed in trans from the CMV promoter. Scale bars are 10 μm. Reproduced with permission from Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B. Improved split fluorescent proteins for endogenous protein labelling. Nat Commun. 2017;8(1):370.

Figure 7

Schematic illustration of the effect of expression level of the FP on signal:background when using non-covalent labelling strategies. Here, cell division control protein 12 (Cdc12), a protein involved in formation of the bud neck in *S. cerevisiae* is fused, at the gene level, to a peptide. A peptide binding module, fused, at the gene level, to a FP, is expressed at different levels. The signal:background fluorescence level is indicated by the bell curve. The expression level of the peptide-binding module-FP is indicated by the green gradient. Signal is indicated by a green band localizing to the bud neck of *S. cerevisiae*, while background is indicated by the disperse

green in the cytoplasm. The black arrow points to Cdc12 under conditions where signal:background is optimal.

Further Reading

1.   Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. 2004;22(12):1567-72.
2.   Hinrichsen M, Lenz M, Edwards JM, Miller OK, Mochrie SGJ, Swain PS, et al. A new method for post-translationally labelling proteins in live cells for fluorescence imaging and tracking. Protein Eng. Des. Sel. 2017;30(12):771-80.
3.   Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN. Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure. Science. 2007;318(5856):1625.
4.   Keeble AH, Howarth M. Insider information on successful covalent protein coupling with help from SpyBank. Methods Enzymol. 2019;617:443-61.
5.   Veggiani G, Nakamura T, Brenner MD, Gayet RV, Yan J, Robinson CV, et al. Programmable polyproteams built using twin peptide superglues. Proc. Natl. Acad. Sci. U S A. 2016;113(5):1202-7.
6.   Lee ME, DeLoache WC, Cervantes B, Dueber JE. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth Biol. 2015;4(9):975-86.
7.   Hawkins KM, Smolke CD. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. J. Biol. Chem. 2006;281(19):13485-92.
8.   Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, et al. Versatile protein tagging in cells with split fluorescent protein. Nat Commun. 2016;7:11046.
9.   Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B. Improved split fluorescent proteins for endogenous protein labelling. Nat. Commun. 2017;8(1):370.
10.   Speltz EB, Nathan A, Regan L. Design of Protein-Peptide Interaction Modules for Assembling Supramolecular Structures in Vivo and in Vitro. ACS Chem. Biol. 2015;10(9):2108-15.
11.   Shaner NC, Patterson GH, Davidson MW. Advances in fluorescent protein technology. J. Cell Sci. 2007;120(24):4247-60.