Shining a light on the dark secrets of the cell

Citation for published version:
Regan, L. Oi, C & Tokarska, D 2019. 'Shining a light on the dark secrets of the cell: Synthetic proteins for better fluorescence imaging', The Biochemist, vol. 41, pp. 40-46. https://doi.org/10.1042/BIO04103040

Digital Object Identifier (DOI):
10.1042/BIO04103040

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
The Biochemist

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Shining a light on the dark secrets of the cell: synthetic proteins for better fluorescence imaging

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 colour 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 behaviour.

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 an 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 colours and properties, have subsequently been discovered in new species or created in the lab using mutagenesis combined with screening. Consequently, we now have a large repertoire of different FPs.

The minimum prerequisite for an 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 an 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 an FP prevents the protein from properly maturing and reaching its final functional location. 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 an 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 *Streptococcus 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 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 an FP result in accumulation of the Pma1-FP protein 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 behaviour 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 an 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. Lee and colleagues described the systematic characterization of the strengths of different constitutive yeast promoters, 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 the 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 carried out 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 superfolder GFP (sfGFP) in which one of the 11 β strands (GFP11) was expressed separately from the other 10 (GFP10). When these two parts of the FP are separate, the protein is not fluorescent. When GFP11 binds to GFP10, the structure of the protein is complete, and is fluorescent (Figure 2D). Their ingenious solution was to fuse GFP11 to the target protein they wished to visualize and separately expressed GFP10.

Because the non-bound form of the FP (GFP10) 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 expand this technique to multicolour imaging in a reasonably straightforward fashion by making point mutations Y66W and T203Y in GFP10, thus creating GFP10 (cyan) and YFP10 (yellow), respectively.

It is highly desirable to be able to perform multicolour imaging by simultaneously tagging multiple proteins in the cell with different coloured fluorescent probes. Green and red are commonly used colours, because they are spectrally distinct, thus minimizing 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 mCherry11 with mCherry10 resulted in increased fluorescence compared with expressing mCherry11, alone, the reconstituted fluorescent signal was far less than that of the original, unsplit 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 sfCherry11 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...
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 mismatches 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 the binding module-FP.

One can use either constitutive promoters of different strengths or well-controlled inducible promoters. In *Escherichia 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 gal2A 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 utilizing synthetic biology and protein engineering approaches, researchers have devised clever new ways to fluorescently label proteins for live cell imaging. Using
**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 the 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 an 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**

- Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E. and Tsien, R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nat. Biotechnol. **22**, 1567–1572
- Hinrichsen, M., Lenz, M., Edwards, J.M., et al. (2017) A new method for post-translationally labelling proteins in live cells for fluorescence imaging and tracking. Protein Eng. Des. Sel. **30**, 771–780
- Kang, H.J., Coulibaly, F., Clow, F., Proft, T. and Baker, E.N. (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. Science **318**, 1625
- Keeble, A.H. and Howarth, M. (2019) Insider information on successful covalent protein coupling with help from SpyBank. Methods Enzymol. **617**, 443–461
- Veggiani, G., Nakamura, T., Brenner, M.D., et al. (2016) Programmable polypeptides built using twin peptide superglues. Proc. Natl. Acad. Sci. USA **113**, 1202–1207
- Lee, M.E., DeLoache, W.C., Cervantes, B. and Dueber, J.E. (2015) A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth. Biol. **4**, 975–986
- Hawkins, K.M. and Smolke, C.D. (2006) The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in *Saccharomyces cerevisiae*. J. Biol. Chem. **281**, 13485–13492
- Kamiyama, D., Sekine, S., Bardi-Rhine, B., et al. (2016) Versatile protein tagging in cells with split fluorescent protein. Nat. Commun. **7**, 11046
- Feng, S., Sekine, S., Pessino, V., Li, H., Leonetti, M.D. and Huang, B. (2017) Improved split fluorescent proteins for endogenous protein labelling. Nat. Commun. **8**, 370
- Speltz, E.B., Nathan, A. and Regan, L. (2015) Design of protein-peptide interaction modules for assembling supramolecular structures *in vivo* and *in vitro*. ACS Chem. Biol. **10**, 2108–2115
- Shaner, N.C., Patterson, G.H. and Davidson, M.W. (2007) Advances in fluorescent protein technology. J. Cell Sci. **120**, 4247–4260

**Acknowledgements**

The authors thank Louise Holyoake, Ella Thornton and Jack Suitor for providing constructive feedback on this article.

Lynne Regan is Chair of Interdisciplinary Science, University of Edinburgh. She read Biochemistry at Oxford University, obtained her PhD at MIT and was a postdoc with Bill DeGrado at du Pont. For many years she was Professor of Molecular Biophysics & Biochemistry and of Chemistry at Yale University. Email: Lynne.Regan@ed.ac.uk

Curran Oi is a Molecular Biochemistry & Biophysics PhD student at Yale University, under the supervision of Lynne Regan. He previously obtained his BS in Physics and Nuclear Engineering from MIT. Email: curran.o@yale.edu

Diana Tokarska is a Biological Sciences undergraduate student at the University of Edinburgh. She joined the Regan lab where she completed her Biochemistry (Hons) project. Email: s1661238@sms.ed.ac.uk