Chapter 6

Method for Developing Optical Sensors Using a Synthetic Dye-Fluorescent Protein FRET Pair and Computational Modeling and Assessment

Joshua A. Mitchell, William H. Zhang, Michel K. Herde, Christian Henneberger, Harald Janovjak, Megan L. O'Mara, and Colin J. Jackson

Abstract

Biosensors that exploit Förster resonance energy transfer (FRET) can be used to visualize biological and physiological processes and are capable of providing detailed information in both spatial and temporal dimensions. In a FRET-based biosensor, substrate binding is associated with a change in the relative positions of two fluorophores, leading to a change in FRET efficiency that may be observed in the fluorescence spectrum. As a result, their design requires a ligand-binding protein that exhibits a conformational change upon binding. However, not all ligand-binding proteins produce responsive sensors upon conjugation to fluorescent proteins or dyes, and identifying the optimum locations for the fluorophores often involves labor-intensive iterative design or high-throughput screening. Combining the genetic fusion of a fluorescent protein to the ligand-binding protein with site-specific covalent attachment of a fluorescent dye can allow fine control over the positions of the two fluorophores, allowing the construction of very sensitive sensors. This relies upon the accurate prediction of the locations of the two fluorophores in bound and unbound states. In this chapter, we describe a method for computational identification of dye-attachment sites that allows the use of cysteine modification to attach synthetic dyes that can be paired with a fluorescent protein for the purposes of creating FRET sensors.

Key words Synthetic dye, Optical sensor, Computational modeling, Förster resonance energy transfer

1 Introduction

Optical sensors have allowed for the investigation of physiological processes such as neurotransmission with both spatial and temporal resolution. FRET-based optical sensors are particularly useful, as they are capable of giving quantitative recordings independent of sensor concentration due to the ratiometric signal output of the sensor and the concentration independence of FRET donor lifetimes [1, 2]. Contemporary sensors typically use fluorescent proteins (FPs). However, some have used synthetic fluorescent

Viktor Stein (ed.), *Synthetic Protein Switches: Methods and Protocols*, Methods in Molecular Biology, vol. 1596,
DOI 10.1007/978-1-4939-6940-1_6,

dyes rather than FPs as the signaling component of the sensor. As there is much greater control over the precise location of synthetic fluorophores, synthetic dyes can theoretically allow for even small conformational changes to produce a measurable FRET signal. While a large conformational change for a given protein is always desirable for sensor construction, it is often a necessity when using FPs, meaning that synthetic dyes are potentially applicable across a much larger range of proteins, rather than being restricted to those with large distance-based conformational changes.

Sensors that use synthetic dyes are either intensity-based sensors (non-FRET), which require multiple synthetic components, or must be developed through extensive high-throughput screening [3, 4]. For example, in addition to the use of two synthetic dyes, the Snifit-type sensor design requires the development and synthesis of a tethered competitive ligand that can occupy the binding active site, which necessarily introduces an extra design phase of engineering and screening [3]. On the other hand, the EOS-type sensor developed by Namiki et al. only requires a single synthetic component (a dye) [4], but still requires exhaustive screening of different residues to find a location that gives a strong signal. In addition, EOS sensors do not produce ratiometric output and are therefore not quantitative.

It is possible to improve on one or more of these design components when creating a synthetic dye-based sensor. Specifically, it has been shown that it is possible to create FRET sensors that are a combination of one FP and one synthetic dye [5], which is an improvement over both single fluorophore and two dye sensors as it is ratiometric and requires one fewer site-specific modification. Additionally, rather than using brute force high-throughput screening of all residue locations to identify a dye labeling site, we have expedited sensor development through the use of computational screening, which can reduce the number of possible dye labeling sites to a subset with a higher likelihood of yielding a functional sensor. We have used these computational techniques in tandem with synthetic dyes (via thiol-maleimide labeling of residues) to develop optical sensors with large dynamic ranges.

2 Materials

Whenever possible, prepare all stock solutions and buffers in ultrapure water (MilliQ). For reagents that have poor solubility in water, dissolve them with the smallest possible proportion of organic solvent (i.e., 5% DMSO would be preferable to 10% DMSO) as some proteins may have poor stability in organic solvent. All solutions and reagents should be prepared as fresh as possible as some reagents will have short lifetimes when in solution, even at −20 °C. All buffer solutions should be filtered through a membrane filter (pore size 0.45 μm or smaller) after preparation.

2.1 Software

1. Python 2.7.11 (www.python.org).
2. Bash 4.2.46 (www.gnu.org/software/bash/).
3. GROMACS 5.1.2 (www.gromacs.org) [6].
4. MARTINI 2.1 (http://md.chem.rug.nl/) [7].
5. PyMol 1.7.6.0 (https://sourceforge.net/projects/pymol).
6. GAWK 4.0.2 (www.gnu.org/software/gawk/).
7. DSSP 2.2.1 (http://swift.cmbi.ru.nl/gv/dssp/).
8. Grep 2.5.1 (https://www.gnu.org/software/grep/).
9. Curl 7.29.0 (https://curl.haxx.se/).

2.2 Cloning and Protein Purification Components

1. The gene of interest in an expression vector (*see* Subheading 3.3).
2. Primers containing the mutation of interest (*see* Subheading 3.3).
3. High-fidelity DNA polymerase kit.
4. Gibson assembly kit.
5. Thin-walled PCR tubes.
6. PCR thermocycler.
7. PCR purification kit.
8. Competent cells for cloning (Top10).
9. Competent cells for protein expression (BL21DE3).
10. Materials for colony PCR and analysis by gel electrophoresis:
    (a) PCR master mix.
    (b) T7 primers.
    (c) 1% agarose gel made with SB buffer (46 g/L boric acid, 8 g/L sodium hydroxide), with a visualizing stain.
    (d) DNA ladder mix.
    (e) Agarose gel electrophoresis apparatus.

2.3 Dye Labeling Components

1. Buffer solution, Phosphate buffer (*see* **Note 1**): 0.05 M Sodium phosphate, 0.2 M NaCl. Adjust the pH to what is appropriate for both the protein of interest and the chemistry that is needed to label the protein with the synthetic dye (using either hydrochloric acid or sodium hydroxide).
2. TCEP stock solution: dissolve 0.1437 g of Tris(2-carboxyethyl) phosphine hydrochloride (TCEP-HCl) in 1 mL of buffer solution (501 mM stock solution) (*see* **Note 2**).
3. Dye stock solution: Add a suitable solvent to the synthetic dye to achieve a stock solution with a final concentration of 10 mM. In the case of the Alexa Fluor 532 C5 Maleimide, 1 mg was dissolved in 123 μL of phosphate buffer (10 mM stock solution) (*see* **Note 3**).

4. Protein solutions: proteins should be concentrated as much as practical (ideally at least 500 μM) and exchanged or dialyzed into the same buffer solution as used to prepare the reagent stock solutions (*see* **Note 4**).

3 Methods

All experimental (noncomputational) work should be performed at 4 °C unless otherwise specified.

3.1 Computational Screening and Residue Selection

First, prepare models of the target fusion protein in both its bound and unbound conformations. This involves modeling both the solute-binding and fluorescent domains with the appropriate linker. Start with crystal structures or high-quality homology models of the desired binding core in both conformations and the fluorescent protein. Ensure no nonstandard residues exist in the models, as these are not parameterized in the MARTINI forcefield (*see* **Note 5**). Reconstruct any residues missing from the crystal structures at the termini by which they are fused. Then, construct the linker sequence as expressed experimentally (*see* **Note 6**). Complete the model by fusing the two domains (*see* **Note 7**). Save both SBP-linker-FP models as .PDB files.

Create a directory for each conformation, copy the appropriate .PDB file into each, and also copy simulations.sh into each (*see* **Note 8**). The script simulations.sh automatically prepares and runs a number of MARTINI simulations. It depends on all of the software in Subheading 2.1 except MARTINI, which is downloaded automatically by the script, provided all software dependencies are available in your $PATH (*see* **Note 9**). The script can be configured by making a copy of it in each conformation directory and editing the leading "CONFIG" portion. Most defaults should be appropriate; however, the input_file, system_name and linker_residues variables should be set for your system (*see* **Note 10**). Configure and run the script, read and follow the directions given, then repeat for the other conformation.

Run the second script "process-data.py," giving as the first two arguments the locations of each set of output files. The residue number of the central residue of the FP fluorophore and the range of residues that should be checked for sensor generation should also be given as arguments, respectively, with the –f and –r switches (*see* **Note 11**). process-data.py reads the given .PDB files, and predicts dynamic ranges for sensors formed by labeling each residue in the given range with a dye with configurable Forster distance. This output is stored by default as comma-separated values with appropriate headers in sensor_predict.csv and can be visualized graphically using a program such as Excel.

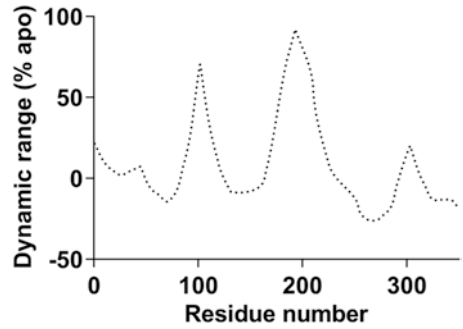

Fig. 1 A graphical representation of the script output. The data shown in this instance is hypothetical and is not based on a true set of calculations. The script determines the dynamic range for a hypothetical sensor that has been labeled with a fluorescent dye at a given residue. For example, for a construct with an ECFP fused at the N-terminus of the binding protein, a sensor construct that has been labeled at residue 50 is predicted to have a very small or nonexistent dynamic range. Conversely, for a sensor that has been labeled at residue 200, the dynamic range should be large, as it is approaching the theoretical maximum dynamic range possible for the construct

Select residues that are appropriate for cysteine mutagenesis (or mutagenesis to an appropriate residue). Note that this prediction does not account for any disruption of binding core function associated with chemical labeling. Therefore, a residue that yields the largest predicted dynamic range may not necessarily yield the best sensor, as the residue may have some structural or functional importance, which may be disrupted with mutagenesis. Residues with side chains oriented toward the solvent, or that are not a part of a structural motif should be selected preferentially. In the hypothetical data set example (Fig. 1), residue 200 is predicted to yield a large dynamic range upon labeling with a dye. Suppose, however, that for this hypothetical protein residue 200 is both not exposed to solvent and has its sidechain oriented toward the binding site of the protein (Fig. 2). Mutating and labeling this residue may abolish protein function, as the sterically bulky dye excludes the substrate from the active site. In contrast, although residue 100 has a smaller predicted dynamic range than residue 200, it is located on a flexible loop that does not have significant contact with other structural features of the binding protein. This makes this location preferable to residue 200, as labeling the residue is less likely to impact the correct function and dynamics of the binding domain, while still providing an excellent dynamic range.

3.2 Cloning and Purification of the Mutant Proteins

1. The sensor construct (SBP fused with the fluorescent protein) should be first cloned into an expression vector (with a T7 promoter). The sequence to be cloned should match the sequence used to model the sensor exactly, with the exception

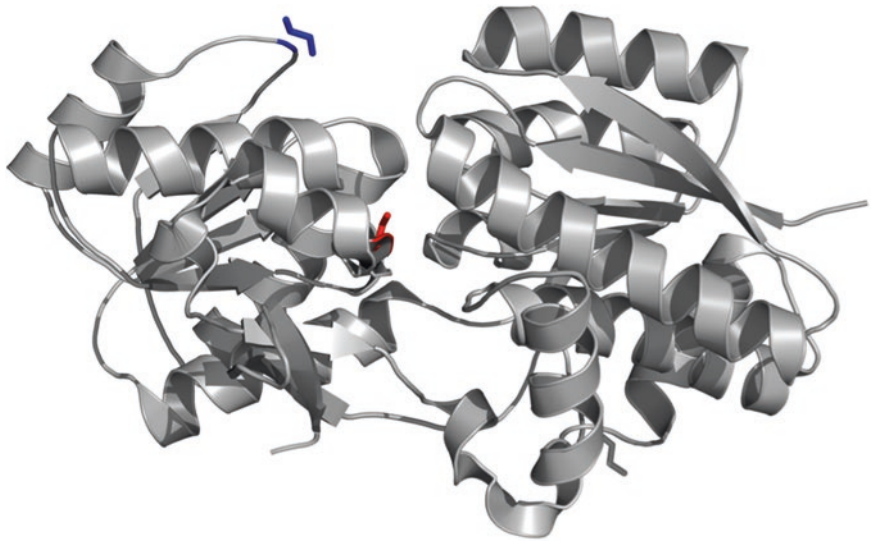

Fig. 2 A generic structure of an amino acid-binding protein (PDB 3IP9) used to illustrate that residue location and function needs to be considered along with the computational prediction. Residues 100 (*blue*) and 200 (*red*) are shown. Although labeling residue 200 would produce a sensor with the theoretically largest dynamic range (Fig. 1), in reality, as this would require the dye to occupy the ligand-binding site between the two domains, it would not result in a functional sensor. Alternatively, residue 100 is also predicted to yield a sensor with a significant dynamic range and is located on a flexible loop, where chemical modification and mutagenesis is less likely to negatively impact correct function of the binding protein

that there can be a histidine tag at the C-terminus of the fluorescent protein to facilitate protein purification.

2. Next, cysteine mutants of this sensor should be created at the residue locations identified by the computational screening (*see* **Note 12**). Any unwanted surface cysteines should be mutated to alternative residues to avoid nonspecific labeling (*see* **Note 13**). Although many cloning methods are suitable to introduce mutations, our preferred method is Gibson assembly.
3. In order to create the cysteine mutants through Gibson assembly, first synthesize or order a set of complementary primers (forward and reverse primers encoding the same sequence) that encompasses the residue of interest, with the total length of the primer between 30 and 50 nucleotides. The nucleotides coding the residue of interest should be changed to encode a cysteine residue, all other residues should match the template DNA exactly.
4. For the PCR, these primers will then be paired with the T7 promoter/terminator primers for PCR amplification. The T7 promoter forward primer is paired with the reverse mutagenic primer, while the T7 terminator reverse primer is paired with the mutagenic forward primer.

5. The PCRs using these primer sets should result in two fragments, with one fragment overlapping with the T7 promoter region and the mutagenic region, and the other fragment overlapping with the mutagenic region and the T7 terminator region. The mutagenic regions of these fragments will overlap and anneal during Gibson assembly. Gel purification is highly recommended to improve cloning efficiency, even if the agarose gel electrophoresis of the PCR product shows a single clear band.
6. Linear vector for use in the Gibson assembly reaction can be made through PCR using complementary primers of the T7 regions. Specifically, a T7 terminator forward primer and a T7 promoter reverse primer should be used. The linearized vector produced from this PCR reaction will have T7 regions that can overlap with the T7 regions of the gene fragments for the Gibson assembly reaction.
7. The PCR fragments (for both the sensor and vector) can then be combined into the Gibson master mix (in equimolar ratios). The typical incubation time for the master mix is 50 °C for 1 h.
8. The Gibson assembly mixture should then be transformed into competent cells that are designed for DNA cloning (e.g., TOP10 cells). If electrocompetent cells are used, the Gibson mix can be diluted with water to prevent arcing.
9. After confirming that the cloning is successful through sequencing, the gene should be transformed into an expression cell type (e.g., BL21DE3).
10. If the sensor construct contains a histidine tag, it can then be purified through nickel affinity chromatography.

3.3 Dye Labeling and Purification

1. To a volume of 849 μL buffer add 100 μL of the concentrated protein solution (assuming the concentration of the protein solution is 500 μM) and 1 μL of the TCEP solution. Wait approximately for 5 min to allow this solution to equilibrate to room temperature and for any disulfides to be reduced. This reaction can be performed in an Eppendorf tube or an equivalent (*see* **Note 14**).
2. To the reaction mixture add 50 μL of the dye stock. The final composition of the reaction mixture should contain approximately 50 μM of protein, 500 μM TCEP, and 500 μM dye. The reaction should be allowed to proceed overnight (16 h) at 4 °C in the dark with constant agitation (*see* **Note 15**).
3. After the reaction period, centrifuge the reaction mixture at high speed (18,000 × *g*, 5 min) to separate out any precipitated dye or protein (*see* **Note 16**).
4. The mixture should then be purified through gel filtration, with a desalting column usually being sufficient (*see* **Note 17**).

5. After one round of purification with the desalting columns, the protein mixture should be re-concentrated and buffer exchanged using centrifugal protein concentrators, which will remove trace TCEP and further remove unreacted and free dye.
6. The protein should then undergo a final desalting step to ensure that no free dye or TCEP remains and labeling efficiency can be evaluated using the following equation.
7. $$\text{Moles of } \textit{dye per} \text{ mole protein} = \frac{A}{\varepsilon \times C}$$

where for a given sample "A" is the absorbance at the peak excitation wavelength of the dye in use, "ε" is the extinction coefficient of the dye, and "C" is the concentration of protein.

4 Notes

1. Phosphate buffer might not be compatible with all proteins; substitute with an appropriate buffer if needed, but ensure that the pH range is compatible with the dye system in use. In the case of thiol-maleimide conjugations, the desired pH is between 7.0 and 7.5.
2. TCEP, or reducing agents in general, are typically only necessary for thiol based conjugation. TCEP is far preferable to other reducing agents such as DTT, as under normal conditions TCEP will not interfere with the maleimide-thiol reaction and also does not have issues with odor.
3. Sometimes there may be trace precipitate or undissolved dye in the stock solution. This is not a cause for concern so long as the precipitate is suspended homogenously prior to use. This trace precipitate will dissolve slowly over time as the reagent is consumed.
4. The protein should be purified to the highest degree possible, as other proteins could be labeled by the dye, which can affect the observed dynamic range of the protein sample. If additional purification is needed, size exclusion chromatography can be performed either before or after the labeling reaction.
5. For fluorescent proteins, this generally means mutating the fluorophore back to the three autocatalytic residues that form it; other proteins may incorporate nonstandard residues as a result of chemical modification for crystallization, etc. Unrecognized residues can be mutated to glycine by opening the .PDB file in a text editor, deleting the side chain atoms, and changing the residue names to GLY. They can then be mutated to the appropriate residue with PyMOL's mutagenesis wizard.

6. We found this easiest to do by working out from both domains and allowing the constructed extended linker model to meet in the middle. Residues can be added to an existing model in PyMOL in Editing mode, accessed by clicking the mouse key shortcuts box in the lower right corner of the viewer window while in the default Viewing mode. Once in Editing mode, select the N-terminal nitrogen or C-terminal carbonyl carbon by clicking on it, then add residues by holding Alt and typing the single-letter code associated with the desired amino acid.
7. In PyMOL's Editing mode, select both atoms that should be bonded in the product, then run PyMOL's fuse command. For example if a model with an ECFP on the N terminus of the protein is desired, start by selecting both the amine of the N terminus of the protein and the carbonyl carbon of the C terminus of the ECFP. Then, while both atoms remain selected, enter "fuse" as a command input, which will generate an approximate model of the fusion protein. The fuse command may sometimes orient the proteins poorly; make sure to rotate the proteins as such that they do not overlap in physical space and the linker is fully extended. This can be done in editing mode by holding shift and right-clicking on a bond to rotate the associated torsion angle.
8. For example, you may be working in a subfolder of your home directory called ~/dyes. You should create new directories for each conformation, perhaps ~/dyes/open and ~/dyes/closed. You add the script and starting structure to each directory, and are left with the following files:

```
~/dyes/open/open-start.pdb
~/dyes/open/simulations.sh
~/dyes/closed/closed-start.pdb
~/dyes/closed/simulations.sh
```

9. The script will then generate folders for setup, each run, and the trimmed, fitted trajectories as pdb files:

```
~/dyes/open/setup/
~/dyes/open/run1/, ~/dyes/open/run2/,
~/dyes/open/run3/ etc.
~/dyes/open/results/
```

10. The version numbers given in the materials are known to work; other versions will probably work as well but have not been tested. Note that MARTINIZE.py, which is downloaded and run by simulations.sh, is not compatible with Python 3, and thus Python 2 must be available on your system for the setup steps. The typical name for the DSSP executable varies from system to system; simulations.sh can be told the correct name for your system either by editing the dssp_name variable or by passing the correct name as an argument

with the –dssp_name switch. Finally, GROMACS versions prior to 5.0 are not supported.

11. By default, simulations.sh prepares, equilibrates, and runs all simulations it is asked to without taking a break. However, if it is terminated, it can restart from the beginning of the last step it finished successfully, so it may be used (with some modification) as a resubmit script for systems like PBS. If preferred, the -o switch can be supplied to the script, which will perform the setup steps and generate individual run folders, but will not perform the full-length production runs. The full-length simulations can then be run on whatever hardware is appropriate by simply running gmx grompp on the provided .MDP files and starting structures.

    The script first prepares the MARTINI coarse-grained force field topology, and converts the provided .PDB file to a coarse-grained model with the script Martinize. This includes construction of an elastic network around both protein domains, which keep their conformations constant and allow sampling of linker collapse. It then energy minimizes and solvates the model, including addition of neutralizing ions, antifreeze MARTINI water, and experimental salt concentration if desired. It performs a second energy minimization, and then equilibrates the system with thermostat and barostat in several rounds of progressively weaker backbone-restrained MD. Force constants of 1000, 500, 100, 50, and 10 are used by the script. The production run involves 30 replicate 200 ns simulations. New velocities are generated for each of these runs in a final unrestrained equilibration step. These simulations take approximately 90 min each for a 600 residue fusion model running on dual Tesla K40s with 32 cores. Finally, it outputs the last 500 ns of each simulation, sampled in 1 ns intervals, as a PDB file to the results directory.

    For instance, if **step 2** was performed in directories called ~/dyes/open and ~/dyes/closed, and the script is being run in ~/dyes, per the example in **Note 8**, a typical call for a fusion model with the binding protein occupying residues 242–586 and the fluorophore at residue 63 might be:

```
    python process-data.py open/results/*.
pdb closed/results/*.pdb -f 63 -r 242-586
```

12. While cysteine mutagenesis is a functional method of chemically labeling a protein with a synthetic dye, alternative chemistries are possible through the use of unnatural amino acid incorporation. This can allow for biorthogonal labeling of proteins in vivo, and in principle, it can allow for the sensor to be genetically encoded in an organism capable of utilizing the required unnatural amino acid.

13. There are two conserved cysteine residues in the GFP family; we have had success with the mutations C49S and C71V with minimal fluorescence loss, as described by Suzuki et al. [8].
14. The reaction volume and reagents should be scaled appropriately, relative to the concentration of the protein stock solution.
15. When agitating the solution, care should be made that the solution does not begin to form froth or foam as this can lead to precipitation of protein.
16. Filtration can be used instead of centrifugation; however, there is typically some volume/yield loss when filtering small volumes.
17. PD-10 columns (GE healthcare) are usually sufficient to separate free dye from the protein. If the purity of the protein is a concern, the protein should be first buffer exchanged to remove excess TCEP and then purified with SEC (GE healthcare, Hiload 26/600 superdex 200pg, adequate for most proteins). This should separate labeled protein from free dye and any contaminant proteins.

References

1. Scanziani M, Hausser M (2009) Electrophysiology in the age of light. Nature 461(7266):930–939
2. Hou B-H, Takanaga H, Grossmann G, Chen L-Q, Qu X-Q, Jones AM, Lalonde S, Schweissgut O, Wiechert W, Frommer WB (2011) Optical sensors for monitoring dynamic changes of intracellular metabolite levels in mammalian cells. Nat Protoc 6(11):1818–1833. http://www.nature.com/nprot/journal/v6/n11/abs/nprot.2011.392.html#supplementary-information
3. Masharina A, Reymond L, Maurel D, Umezawa K, Johnsson K (2012) A Fluorescent sensor for GABA and synthetic GABAB receptor ligands. J Am Chem Soc 134(46):19026–19034. doi:10.1021/ja306320s
4. Namiki S, Sakamoto H, Iinuma S, Iino M, Hirose K (2007) Optical glutamate sensor for spatiotemporal analysis of synaptic transmission. Eur J Neurosci 25(8):2249–2259. doi:10.1111/j.1460-9568.2007.05511.x
5. Suzuki M, Tanaka S, Ito Y, Inoue M, Sakai T, Nishigaki K (2012) Simple and tunable Förster resonance energy transfer-based bioprobes for high-throughput monitoring of caspase-3 activation in living cells by using flow cytometry. Biochim Biophys Acta 1823(2):215–226. doi:10.1016/j.bbamcr.2011.07.006
6. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4(3):435–447
7. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J (2008) The MARTINI coarse-grained force field: extension to proteins. J Chem Theory Comput 4(5):819–834
8. Suzuki T, Arai S, Takeuchi M, Sakurai C, Ebana H, Higashi T, Hashimoto H, Hatsuzawa K, Wada I (2012) Development of cysteine-free fluorescent proteins for the oxidative environment. PLoS One 7(5):e37551. doi:10.1371/journal.pone.0037551