Structural characterisation of novel GFP-like and phytochrome-derived fluorescent proteins

Hadrien Depernet

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 25 mai 2016

Présentée par

Hadrien DEPERNET

Thèse dirigée par **Antoine ROYANT**, HDR, DR CNRS, IBS, Grenoble, et
codirigée par **Gordon LEONARD**, SB Group Leader, HDR, ESRF, Grenoble

préparée au sein du **Structural Biology Group (SB), European Synchrotron Radiation Facility (ESRF)**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Caractérisation structurale de nouvelles protéines fluorescentes de type GFP ou dérivées de phytochromes

-

Structural characterisation of novel GFP-like and phytochrome-derived fluorescent proteins

Thèse soutenue publiquement le **30 Mars 2021**,
devant le jury composé de :

Rapporteur : Dr. Jean-Denis PÉDELACQ
CR CNRS, HDR, IPBS, Toulouse, France
Rapportrice : Dr. Nushin AGHAJARI
DR CNRS, HDR, IBCP, Lyon, France
Examinatrice : Dr. Gerlind SULZENBACHER
IR CNRS, AFMB, Marseille, France
Président : Dominique BOURGEOIS
DR CNRS, HDR, IBS, Grenoble, France
Directeur de thèse : Dr. Antoine ROYANT
DR CNRS, HDR, IBS, Grenoble, France

Acknowledgements

First, I would like to thank my thesis supervisors for having welcomed me in the Structural Biology group and let me pursue my thesis at the ESRF. I would like to thank Dr. Antoine Royant for the constant supervision during my thesis and the extensive knowledge he brought concerning fluorescent proteins, I learned a lot. I also thank Dr. Gordon Leonard for the monitoring of my thesis and both of them for the corrections and writing experience they brought for the completion of this manuscript.

I also thank the members of my thesis defence jury for their discerning remarks, comments and questions. Dr. Dominique Bourgeois as president of the jury, Dr. Jean-Denis Pédelacq and Dr. Nushin Aghajari as referees and Dr. Gerlind Sulzenbacher as examiners.

I would like to thank Dr. Pascal Arnoux and Dr. Carlos Contreras Martel for accepting to be part of my thesis monitoring committee and for their remarks and comments about my experiments and the scientific work of my thesis.

I would also like to especially thank Dr. Guillaume Gotthard for all the knowledge I learned from him about structural biology, data processing and the collection of crystals at a synchrotron beamline.

I also thank Dr. Sylvain Aumonier for this help in the lab and the advice he gave me in the beginning of my thesis concerning crystallogenesis and also for being my desk neighbour for most part of my PhD. And thank you, the advice you gave me when I was learning how to ski were very good.

I am grateful to Dr. Nathan Shaner from the University of California in Santa Diego, Dr. Xiaokun Shu from the University of California in San Francisco and Dr. Isabelle Navizet from the University Gustave Eiffel in Marne-la-Vallée for the rewarding scientific collaborations.

I warmly thank Fabien, Hugo and Thierry for their technical support on the beamlines and also Marcus, Gianluca, Nicolas, Daniele, Didier, Sasha, Romain, Max and David Flot for the good atmosphere and laugh during coffee breaks.

I also thank Jean Susini for this support to my PhD project and Christoph Mueller-Dieckmann for the support but also for the help and comments during the rehearsal of my thesis.

Many thanks to Dr. Montserrat Soler-Lopez for greeting me in the CIBB wet lab of the ESRF Structural Biology group and to Samira, Melissa, Lyna and Jennyfer for their help in the lab.

I also would like to thank Florent Bernaudat the Scientific Coordinator of the Partnership for Structural Biology (PSB) to let me be part of the student PSB committee and the other members of the student committee, Wikor Adamski (IBS), Nina Christou (IBS), Lukas Gajdos (ILL), Joanna Wandzik (EMBL) and Michael Adams (EMBL), the recurrent meetings we had for almost three years were really fun and I learned a lot about the structure and organisations of the EPN campus.

I would also like to thank all the PhD students and post-docs of room 115, I had some really good laughs. A special thanks to David, finding a person who had the exact same passions and hobbies as me was unexpected and helped me a lot during all the years of my thesis. I can tell without doubt you became my best friend and if you want to go back to visit Japan you can call me anytime. Thanks to Bart for the lengthy discussions we had in the lab when it was only the two of us, it was really good to have someone to talk during the never ending testing of crystallisation conditions. Thank you Igor for the stories about Russia you told to David and me, it was really fun and instructive. Thank you Maksim for the good mood, you lighten the atmosphere really fast and laugh all the time, it was really refreshing. Thank you Romain, I never understood your talk about soccer with David but at least it was fun to hear you debate about it. Thank you Gabriele for the jokes and good mood you always had (also thank you for your knowledge about the AKTA and purification procedures). Thanks to Anton for the good humour. Finally, thanks to Mark for all the talks we had with you and the other members of the “lunch group” and all the diners and invitations to your house it was really fun and relaxing.

Finalement, je voudrais également remercier mes parents pour leur soutien et patience tout au long de ces années d’études.

List of abbreviations

*Av*GFP	*Aequorea victoria* Green Fluorescent Protein
*Aaus*FP1	*Aequorea australis sp.* Fluorescent Protein 1
*Aaus*FP2	*Aequorea australis sp.* Fluorescent Protein 2
asFP	*Anemonia sulcata* Fluorescent Protein
BbFPs	Bilin-binding Fluorescent Proteins
BphP	Bacteriophytochrome
BphB	Bacteriophytochrome from *Bradyrhizobium*
BFP	Blue Fluorescent Protein
BR	Bilirubin
*Br*BphP	*Bradyrhizobium* Bacteriophytochrome
BV	Biliverdin
CBD	Chromophore Binding Domain
CCP4	Collaborative Computational Project, number 4
cDNA	complementary deoxyribonucleic acid
CFP	Cyan Fluorescent Protein
CP	Chromoprotein
CRIMS	Crystallization Information Management System
*GAF*FP4	*Clytia simplex* Fluorescent Protein 4
*Dr*CBD	*Deinococcus radiodurans* Chromophore Binding Domain
*Dr*BphP	*Deinococcus radiodurans* bacteriophytochrome
DsRed	*Discosoma sp.* Red
ECFP	Enhanced Cyan Fluorescent Protein
EGFP	Enhanced Green Fluorescent Protein
EMBL	European Molecular Biology Laboratory
eqFP	*Entacmaea quadricolor* Fluorescent Protein
ESRF	European Synchrotron Radiation Facility
FbFPs	Flavin-binding Fluorescent Proteins
FPs	Fluorescent Proteins
FRET	Förster (or Fluorescence) resonance energy transfer
GAF	cGMP phosphodiesterase/adenyl cyclase/FhlA
GFP	Green Fluorescent Protein
HEPES	4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid
HTX	High Throughput Crystallization
IC	Internal conversion
IFP	Infrared Fluorescent Proteins
iRFP	Near-infrared Fluorescent Protein

ISC	Intersystem crossing
kDa	Kilo Dalton
LOV	Light Oxygen Voltage
miniSOG	Minimal-sized Singlet Oxygen Generator
mKO	monomeric Kusabira Orange
MWCO	Molecular Weight Cut-Off
MX	Macromolecular crystallography
NIR FPs	Near-Infrared Fluorescent Proteins
OSER	Organized Smooth Endoplasmic Reticulum
PAFP	Photoactivable Fluorescent Protein
PAS	Per/ARNT/Sim
PCB	Phycocyanobilin
PCFP	Photoconvertible Fluorescent Protein
PDB	Protein Data Base
PEB	Phycoerythrobilin
PEG	Polyethylene Glycol
Pfr	Far-red absorbing state
pH	Potential of hydrogen
phiYFP	*Phialidium sp.* Yellow Fluorescent Protein
PHY	Phytochrome-specific
PISA	Proteins, Interfaces, Structures and Assemblies
pK	Dissociation constant
pKa	Acid dissociation constant
Pr	Red-absorbing state
PUB	Phycourobilin
PVB	Phycoviolobilin
PΦB	Phytochromobilin
r.m.s.d	root-mean-square deviation
REFMAC5	REFinement of MACromolecular structures, version 5
RFP	Red Fluorescent Protein
RpBphP	*Rhodopseudomonas palustris* Bacteriophytochrome
RSFP	Reversibly photoswitchable Fluorescent Protein
SDS-PAGE	Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis
TB	Terrific Broth
TEV	Tobacco Etch Virus
TRIS	2-amino-2-hydroxyméthylpropane-1,3-diol
VR	Vibrational relaxation
XDS	X-ray Detector Software
YFP	Yellow Fluorescent Protein

List of figures and tables

Table of Contents

Chapter 1

INTRODUCTION

Fluorescent proteins and the structure diversity of their chromophores

1.1 Why do we need fluorescent proteins?

While electron microscopy progressively became the favourite technique to look at details of continuously increasing resolution in fixed (dead) cells (McIntosh, 2001), fluorescence microscopy became an exciting tool to look at cellular inner morphology processes in live cells in the 1970s (Renz, 2013). This triggered the development of a new field in Chemistry, that is the synthesis of small fluorescent molecules which could traverse biological membranes and target specific cellular compartments to probe physiological parameters including pH and [Ca^{2+}] (Terai & Nagano, 2013). However, these small molecules lacked addressability to specific protein targets. The discovery of the Green Fluorescent Protein (see **Paragraphs 1.2** and **1.3**) opened the possibility of having genetically encoding fluorescence within cells, obviating the need to supply exogenous fluorophores and creating the possibility of fusing a fluorescent tag to any given protein by molecular biology (Chalfie *et al*, 1994). The general properties of fluorescent proteins (FPs) will be described in **Paragraph 1.4**. Given the scope of the PhD subject, the huge diversity of fluorescent proteins will then be described with a focus on the variety of chromophores. FPs homologous to GFP will be described in **Paragraph 1.5**. FPs derived from bilin-binding proteins (BbFPs), essentially fluorescing in the near-infrared part of the light spectrum, will be described in **Paragraph 1.6**. FPs derived from flavoproteins (FbFPs), fluorescing in the cyan-green part of the light spectrum, will be described in **Paragraph 1.7**. A general overview of the various applications made possible by FPs will be presented in **Paragraph 1.8**. Finally, the precise objectives of the thesis work will be presented in **Paragraph 1.9**.

1.2 Historical background

At the beginning of the 1960s, nobody could have predicted the importance of fluorescent proteins (FPs) in Biology. It all began in 1962 with the publication of Osamu Shimomura and Frank H. Johnson (Shimomura, *et al* 1962). Their study on the protein aequorin and the bioluminescence of the jellyfish *Aequorea victoria* (**Figure 1**) permitted the discovery of a new protein that eluted before aequorin during the column chromatography (Morise *et al*, 1974) and gave the jellyfish squeezates a "greenish luminescence" (reproduced from Shimomura *et al*, 1962). This protein would be later called Green Fluorescent Protein (GFP).

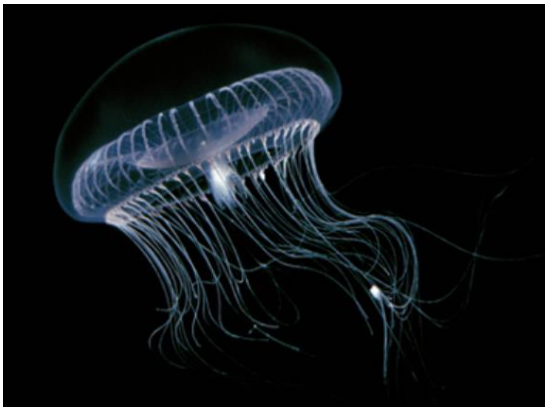

Figure 1. Photograph of a jellyfish from the *Aequorea Victoria* species (reproduced from https://www.nobelprize.org/prizes/chemistry/2008/illustrated-information, photo credit: Kevin Raskoff).

As Dr. Osamu Shimomura explained in his Nobel lecture of 8 December 2008 (Shimomura, 2008), GFP was originally considered as an uninteresting by-product of the Ca-sensitive photo protein *aequorin* and the protein remained for almost 30 years without any real application. The next significant step in the history of the GFP occurred in 1992 when Douglas Prasher successfully cloned and sequenced the GFP gene (Prasher *et al*, 1992) and made available its cDNA sequence for the rest of the scientific community. In 1994, Chalfie used GFP as a cell marker for the first time, expressing the fluorescent protein in the prokaryote *E. coli* and in the neurons of the eukaryote *Caenorhabditis elegans* (Chalfie *et al*, 1994) (**Figure 2**) This publication showed that GFP can be used to track protein localisation and gene expression inside living organisms without any substrates other than the protein itself and without any cofactors.

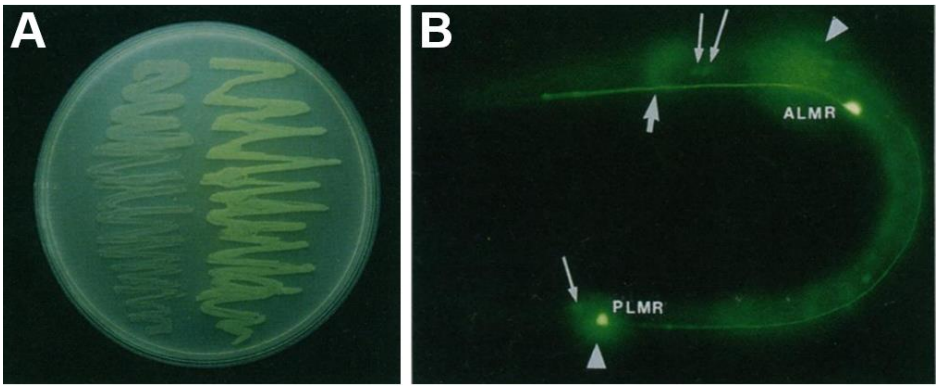

Figure 2. GFP used as a cell marker for the first time (**A**) Colony of *E. coli* expressing for the first time the *Av*GFP gene (right streak) compared to normal *E. coli* colony (left streak). (**B**) *Av*GFP expressed inside the neurons of *Caenorhabditis elegans* (reproduced from Chalfie *et al*, 1994).

Roger Tsien responded to this publication by pointing at a lack of spectral diversity and between 1994 and 1998 worked with collaborators to produce a number of GFP mutants, some which were brighter and more stable, and some whose fluorescence exhibiting the yellow, cyan and blue colours complementary to the original green (Tsien, 1998).

In 2008, Shimomura, Chalfie and Tsien (**Figure 3**) received the Nobel Prize in Chemistry for their work on GFP and the advances in the scientific fields. At this occasion, the work of Douglas Prasher was overlooked by the committee but not by the three recipients who acknowledge that the work of Prasher was critical for their work on GFP (https://www.mediatheque.lindau-nobel.org/laureates/chalfie).

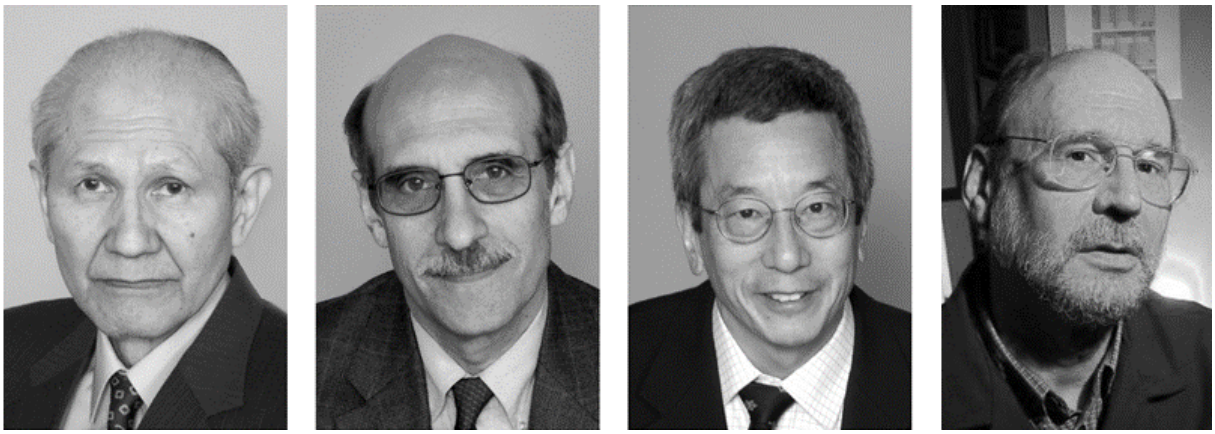

Figure 3. Pictures from left to right of Osamu Shimomura, Martin Chalfie, Roger Tsien and Douglas Prasher, Shimomura, Chalfie and Tsien obtain the Nobel prize of chemistry in 2008 for their work on GFP (reproduced from https://www.nobelprize.org/prizes/chemistry/2008/summary/ and https://www.sandiegouniontribune.com/lifestyle/people/sdut-cientist-turned-shuttle-van-driver-2013apr13-story.html).

Since the discovery of GFP in 1962 enormous progress have been made in biology thanks to this protein, its mutants and homologues. Indeed, FPs has been discovered in other animals such as corals, sea anemones, copepods (small crustacean) and lancelets. FPs have also been obtained by mutagenizing coloured proteins of different functions (phytochrome, fatty-acid binding protein, phototropin, allophycocyanin) in order to favour radiative de-excitation pathways after light absorption by the chromophore.

1.3 The Green Fluorescent Protein from *Aequorea victoria*: *Av*GFP

*Av*GFP has been extensively characterized both structurally and spectroscopically in the 1990s (Tsien, 1998). *Av*GFP is a weak dimer (dissociation constant ~100 µM), whose arrangement was determined in the first crystallographic structure of *Av*GFP (Yang *et al.* 1996) (**Figure 4**). Each monomer of the dimer is formed of a β-barrel composed of eleven β-strands. The barrel has dimensions of ~24 Å in diameter and ~42 Å in length. The chromophore belongs to the α-helix that traverses the barrel, and is located approximately halfway through, providing it with a rigid environment and protecting the chromophore from exposure to the agitation of solvent molecules. These two features explain why the small-molecule analogue of the GFP chromophore is not fluorescent in solution. Both N- and C-termini form disordered coils in the bulk solvent, providing the possibility to fuse proteins of interest at both positions.

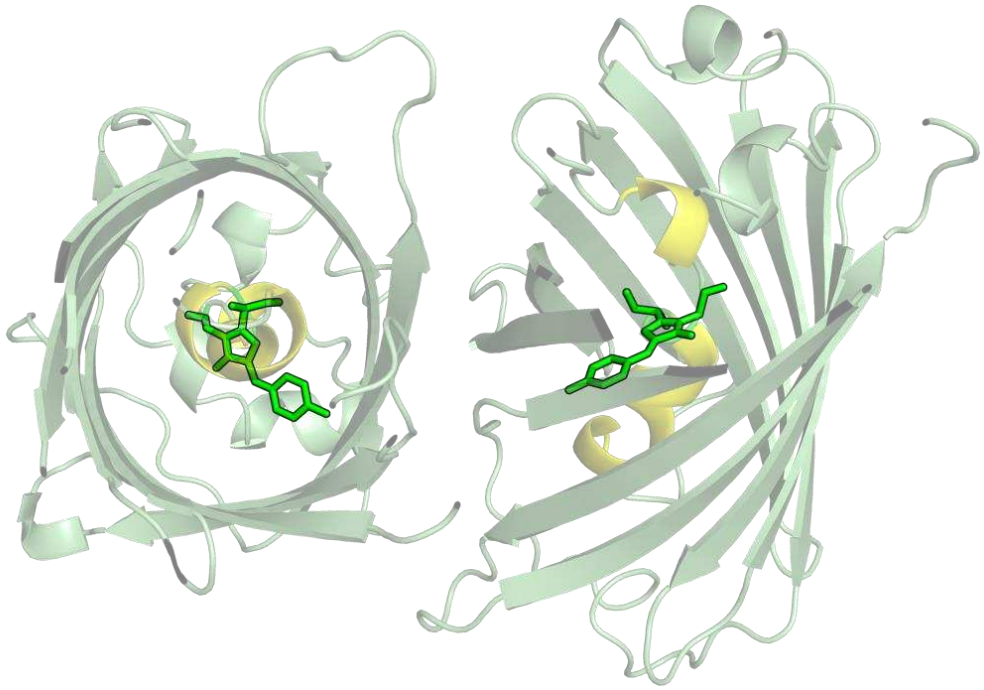

Figure 4. Crystallographic structure of the *Av*GFP dimer (PDB entry code: 1GFL) (Yang et al, 1996a). The chromophore (bright green) is located in the middle of the central α-helix (yellow), which traverses the β-barrel formed of eleven β strands (pale green).

The fluorescent chromophore is formed by the autocatalytic cyclisation of the three consecutive amino acid residues Ser65-Tyr66-Gly67, which only requires molecular oxygen as a cofactor (**Figure 5**). The two residues Arg96 and Glu222 (quasi-conserved in all

sequences of later discovered homologues) act as catalysers in chromophore maturation, Glu222 as a general base and Arg96 as an electrophile. The precise mechanism after the cyclisation step is still disputed (**Figure 5B**), but results in the formation of a conjugated electron cloud extended on both the phenol(ate) ring and the newly-formed imidazolinone ring, which explains the properties of light absorption in the visible part of the light spectrum. The two bonds linking the two rings are called the methylene bridge.

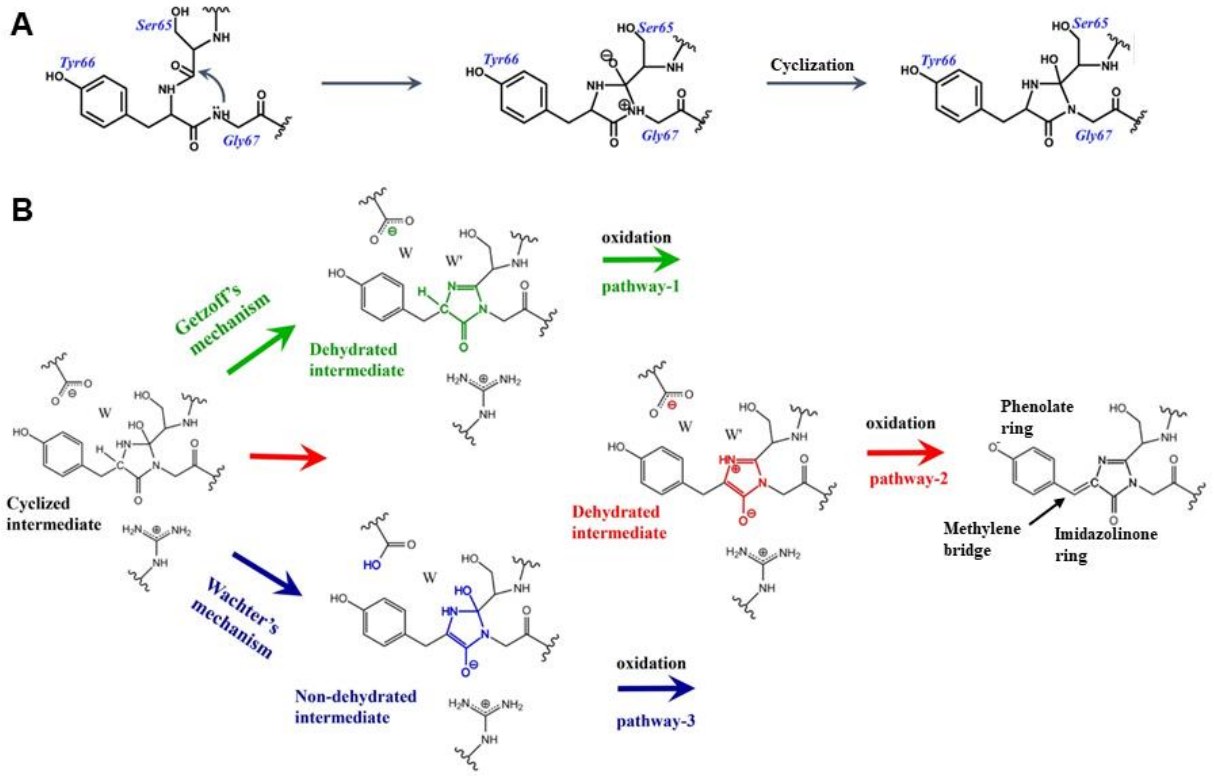

Figure 5. Proposed mechanisms for the formation of the GFP chromophore. (**A**) Cyclisation step. (**B**) Three proposed mechanisms leading to the mature chromophore from the cyclised intermediate, with possibly one first dehydration step then a final oxidation step (adapted from Grigorenko *et al*, 2017).

*Av*GFP presents a peculiar absorption spectrum, with two absorption maxima, one in the near UV at ~395 nm (major peak) and one in the blue-cyan at ~475 nm (minor peak), which were attributed to two spectroscopic states A and B, respectively (**Figure 6A**). Light excitation of both absorption bands leads to green fluorescence emission with a maximum at ~509 nm. This is explained by the fact state A has a phenol group (protonated, neutral) in the chromophore, while state B has a phenolate (deprotonated, negative charge) instead (**Figures 6B & 6C**). While state B fluoresces with a relatively small Stokes shift of ~30 nm (difference between the absorption/fluorescence excitation and fluorescence emission peaks), state A gets deprotonated in the excited state A* and fluorescence occurs from an electronic state close to that of B*, leading to a large Stokes shift of ~110 nm.

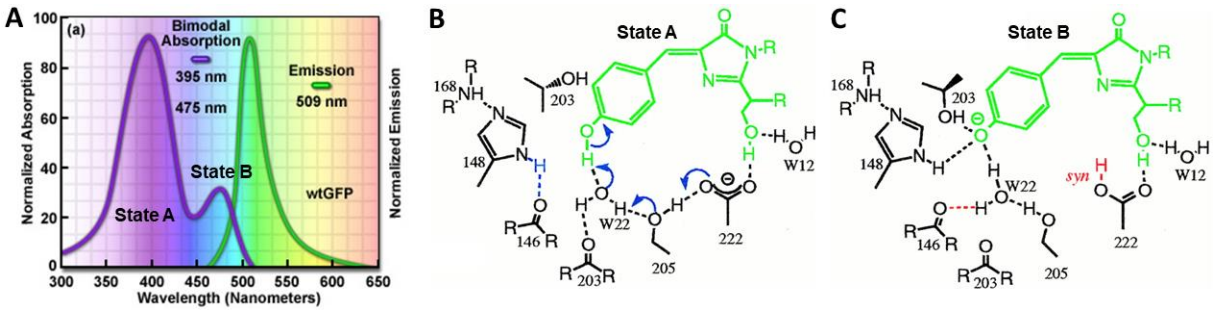

Figure 6. Peculiar absorption spectrum of *Av*GFP and chromophores structures. (A) Absorption and emission spectra of *Av*GFP (reproduced from http://zeiss-campus.magnet.fsu.edu/articles/probes/fpintroduction.html) (**B**) Chemical structure of the *Av*GFP chromophore and its protein environment in its state A. (**C**) Chemical structure of the *Av*GFP chromophore and its protein environment in its state B (adapted from Brejc *et al*, 1997).

While GFP has been extensively used in molecular biology, biochemistry and cell biology and thus have gained a strong biotechnological utility, it is striking to realize that its function in nature is still uncertain. Some have hypothesized that light emission by GFP is an attempt to attract secondary predators when the jellyfish is attacked so that the primary predator gets distracted (Abrahams *et al*. 1993). Others have postulated that GFP homologues play a photoprotective role in coral reefs (Salih *et al*. 2000).

1.4 Photophysical and structural properties of Fluorescent Proteins

1.4.1 Simplified mechanism of fluorescence

Fluorescence (**F**) occurs when a molecule is brought from its ground state $\mathbf{S_0}$ to an excited state $\mathbf{S_1}$ by absorption of a light quantum (**A**) and relaxes back to the ground state with emission of a light quantum of lower energy than the exciting one: this relaxation pathway is called **radiative**. There are other relaxation pathways as described on the simplified Jablonski diagram, which does not consider sub-states (**Figure 7**). First, the molecule can be brought into a higher excited $\mathbf{S_2}$, which rapidly converts to the $\mathbf{S_1}$ state by internal conversion (**IC**) (by heat generation). $\mathbf{S_1}$ can also relax to $\mathbf{S_0}$ by vibrational relaxation (**VR**) (also by heat generation), a **non-radiative** relaxation pathway. A transition can occur infrequently from the singlet state $\mathbf{S_1}$ to the triplet state $\mathbf{T_1}$, which will then relax back to the ground state either by **VR** or by emission of a light quantum (**P**, phosphorescence). The time scales of fluorescence and phosphoresence are typically nanoseconds and microseconds to milliseconds, respectively.

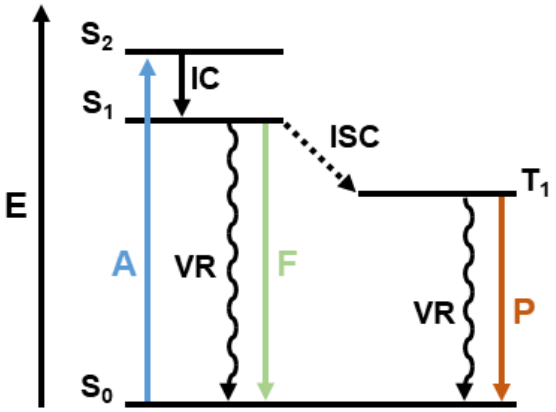

Figure 7. Simplified version of the Jablonski diagram. **A**: Absorption, **F**: Fluorescence, **P**: Phosphorescence, **IC**: Internal conversion, **ISC**: Intersystem crossing, **VR**: Vibrational relaxation, **E**: Energy levels.

1.4.2 Extinction coefficient, quantum yield, brightness

The extinction coefficient (EC) of the chromophore of a FP is a measure of its capacity to absorb light efficiently at a given wavelength. It is expressed in $M^{-1}.cm^{-1}$ or $mM^{-1}cm^{-1}$.

The quantum yield (QY) is a measure of how efficiently the chromophore of a FP emits fluorescence. It consists in the ratio of the number of emitted photons over the number of absorbed photons, and is thus greater than (or equal to) 0 and lower than (or equal to) 1.

The brightness is the product of EC and QY and represents a relative measure of the fluorescence signal emitted by a given chromophore upon a constant number of exciting photons. The EC and QY values of a given chromophore (and thus its brightness) are strongly modulated by the protein environment.

1.4.3 Configuration and conformation of the chromophore

The *Av*GFP chromophore, and by extension any similar chromophore, has two configurations with respect to the methylene bridge which consists of one single bond and one double bond in resonance: a *cis* and a *trans* configuration (**Figure 8A**). Most FPs with a *cis* chromophore are fluorescent whereas most FPs with a *trans* chromophore are non-fluorescent (Day & Davidson, 2009; Seward & Bagshaw, 2009). Noteworthy, the Trp-based chromophore of cyan fluorescent proteins derived from *Av*GFP have four distinct configurations due to the asymmetry of the indole ring (a phenolate group is symetric).

The aromatic character of the chromophore promotes the coplanarity of the two conjugated rings *in vacuo*. However, within the FP, interactions with neighbouring residues (hydrogen bonds, van der Waals interaction, steric hindrance, covalent bonds) may distort the chromophore away from coplanarity. The tilt dihedral angle τ and the twist dihedral angle Φ provides a quantitative estimate of the distortion (**Figure 8B**). Usually, the more distorted a chromophore is, the less fluorescent it becomes.

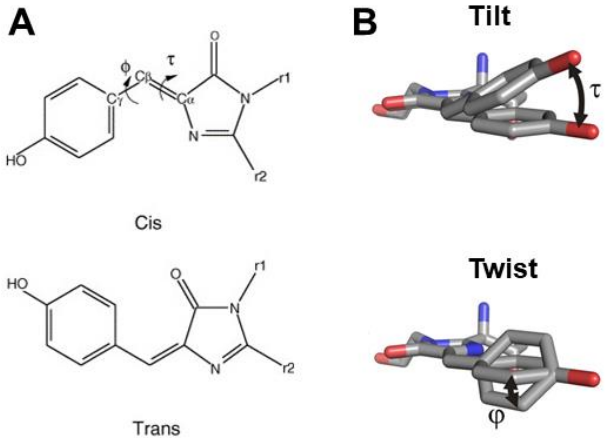

Figure 8. Configuration and conformation of the chromophore (**A**) cis and trans configurations of the Tyr-based chromophore in GFP homologues (adapted from Nifosì & Tozzini, 2006) (**B**) Illustration of coplanar and non-coplanar conformations of Tyr-based chromophore. A coplanar chromophore has both values of τ and ϕ close to 0°, a non-planar chromophore has τ and/or ϕ significantly different from 0° (adapted from Pakhomov & Martynov, 2008).

1.4.4 Fluorescence pKa

When pH is changed from physiological pH, the fluorescence signal usually falls sharply to zero at some point during acidification (but also during alkalization, a phenomenon that is far less investigated in live cell imaging). The fluorescence pKa of a FP is defined as the pH for which half the maximum fluorescence signal is lost (**Figure 9**). The pKa thus gives a lower limit of the pH range of general usability of the FP. For GFP-like FPs with a Tyr-based chromophore, the loss of fluorescence in most cases is due to the protonation of the phenolate group. For non-ionizable chromophores with low pKa, the loss of fluorescence can be attributed to acidification-induced denaturation of the protein.

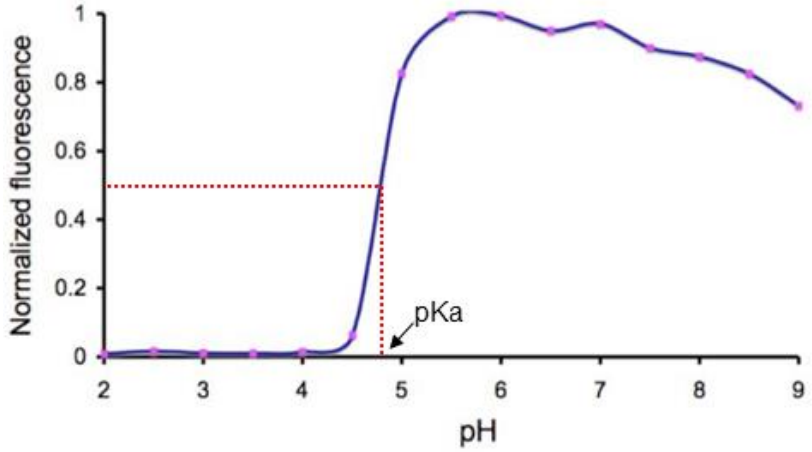

Figure 9. pKa determination of IFP1.4. For each pH, an emission spectrum was obtained upon excitation at 640 nm, integrated between 700 and 800 nm and the result normalized to the highest signal in the whole range (Shu *et al*, 2009).

1.4.5 Oligomeric state

Most of wild-type FPs are not monomeric, and form obligate oligomers such as dimers or tetramers (**Figure 10**). The oligomeric state appears to be key to the fluorescence properties, as monomerization usually collapses or even cancels fluorescence, and further engineering is required to restore an acceptable level of fluorescence (Shaner *et al*, 2013; Clavel *et al*, 2016). The oligomeric state of a FP needs to be assessed prior to its use in cell imaging as any unwanted affinity between two copies of the same FP fused to a protein of interest (POI) could create an artefactual interaction between the two POIs (Swulius & Jensen, 2012). The OSER (Organized Smooth Endoplasmic Reticulum) assay has been developed to assess the degree of monomer character of a given FP by evaluating the capacity of the FP to induce a restructuration of the endoplasmic reticulum (Costantini *et al*, 2012)

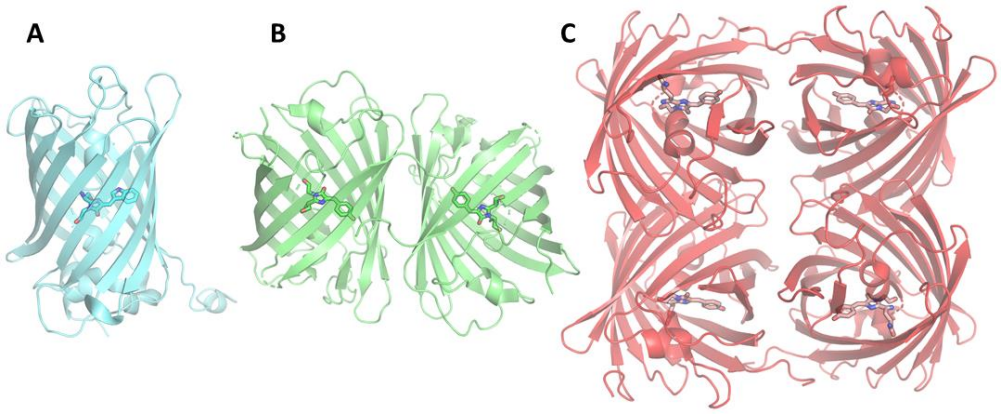

Figure 10. Oligomeric state of various FPs. (**A**) Monomer of mTurquoise2 (PDB entry code: 3ZTF) (Goedhart *et al*, 2012). (**B**) Dimer of rrGFP (PDB entry code: 2RH7) (Crystal Structures of the Luciferase and Green Fluorescent Protein from Renilla reniformis, 2007). (**C**) Tetramer of DsRed (PDB entry code: 1G7K) (Yarbrough *et al*, 2001).

1.4.6 Förster Resonance Energy Transfer (FRET)

The FRET mechanism consists of a **non-radiative** energy transfer between one donor molecule in the excited state and one acceptor molecule in the ground state, which occurs through dipole-dipole coupling. The FRET efficiency decreases with the sixth power of the distance between the donor and the acceptor, but is also dependent on the respective orientation of dipoles. For FRET to occur between two FPs there must be a spectral overlap between the donor emission spectrum and the acceptor absorption spectrum. In fact, the acceptor does not need to be fluorescent, as efficient FRET can be seen by the decrease in fluorescence of the donor. FRET has been a favourite tool to estimate distances between probes and to detect interactions between proteins. A number of biosensors rely on the disruption or build-up of a FRET signal.

1.4.7 Photobleaching

Each time a FP absorbs a photon, there is a possibility that this leads to the loss of fluorescence capability, either reversibly or irreversibly. This phenomenon is called **photobleaching**. **Irreversible photobleaching** may be ascribed to a chemical degradation of the chromophore or its environment. **Reversible photobleaching** may occur when the configuration (isomerisation), conformation or even chemical nature (e.g. (de)protonation) of the chromophore is changed after light absorption. FPs may photobleach at various rates, and

a precise control of light intensity and experiment duration is necessary to perform a successful cell imaging experiment.

1.4.8 Phototransformable FPs

An oversimplified view of a FP is to consider that, upon light excitation, it emits a single type of fluorescence signal, progressively losing intensity by irreversible photobleaching. However, some FPs exhibit peculiar properties. A **photoactivable** FP (**PAFP**) is a protein that is initially not fluorescent and becomes irreversibly fluorescent upon absorption of light of a certain wavelength. In some other type of FPs, the fluorescence of a **reversibly photoswitchable** FP (**RSFP**) can be reversibly switched on and off depending on the excitation wavelength. Finally, a **photoconvertible** FP (**PCFP**) can have its emission maximum changed thanks to a specific excitation wavelength. The mechanism of their phototransformation is presented in **Figure 11**.

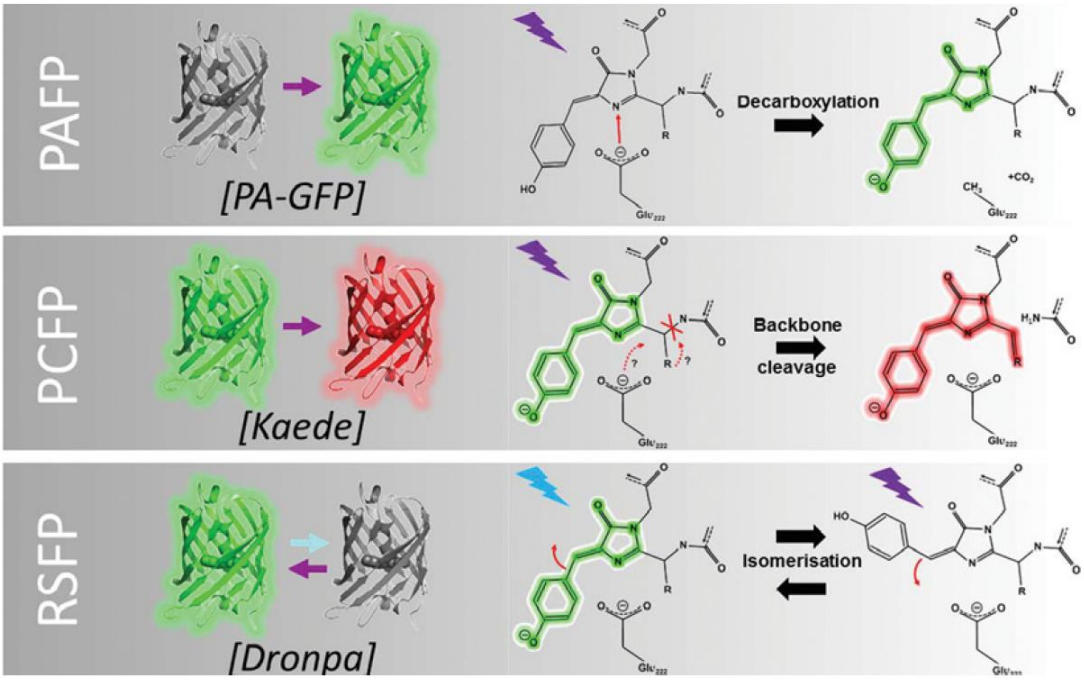

Figure 11. Three types of phototransformables fluorescent proteins. PA-GFP (Patterson & Lippincott-Schwartz, 2002), Kaede (Ando *et al*, 2002) and Dronpa (Ando *et al*, 2004) are prototypal members of each class (reproduced from (Bourgeois *et al*, 2012).

1.5 Variety of fluorescent proteins homologous to GFP ('GFP-like FPs')

GFP-like proteins consist of GFP from *Aequorea victoria*, *Av*GFP mutants and all homologues (and their subsequent mutants) that have been discovered in a number of marine organisms. The main spectroscopic properties of all GFP homologues presented in this paragraph (§ **1.5**) are gathered in **Table 1**. The family of GFP-like fluorescent proteins is the most well-known and most diverse of all FP families. Their colour ranges the entire visible light spectrum from deep blue to far red. One of its main drawbacks is its requirement for molecular oxygen for chromophore maturation, thus limiting its application to aerobic conditions.

1.5.1 Tuning the brightness, folding and oligomeric state of *Av*GFP by mutagenesis

The main drawback of *Av*GFP is the location of its main absorption band in the UV part of the light spectrum, which is associated to the protonated form of the chromophore (**Figure 12A**), as it requires UV light for fluorescence excitation, which is potentially phototoxic. Mutation of the first residue of the chromophore, Ser65, into a threonine suppresses the 395 nm excitation peak and promotes ionisation of the chromophore through the neighbouring Glu222 (Yang *et al*, 1996b) (**Figure 12B**). Addition of the F64L mutation just before the chromophore improves protein folding at 37 °C (Palm *et al*, 1997), facilitating cell imaging. The resulting protein has been dubbed EGFP, for Enhanced Green Fluorescent Protein. Mutation S30R for this part induces the formation of an unusual electrostatic network between β-strands S1, S2, S5 and S6. Additional mutations at the surface of the protein (N105T and A206V) resulted in a version of GFP, dubbed superfolder GFP that folds fast and well, even when fused with a poorly folding protein, and is resistant to chemical denaturation (Pédelacq *et al*, 2006).

Conversely, GFPuv was obtained from *Av*GFP with the idea of favouring the protonated state of the chromophore, thus providing a GFP with only near-UV excitable green fluorescence (Crameri *et al*, 1996). While of less practical application than other variants, it shows how the properties of the *Av*GFP chromophore can be tuned to each of its spectroscopic species (protonated/anionic).

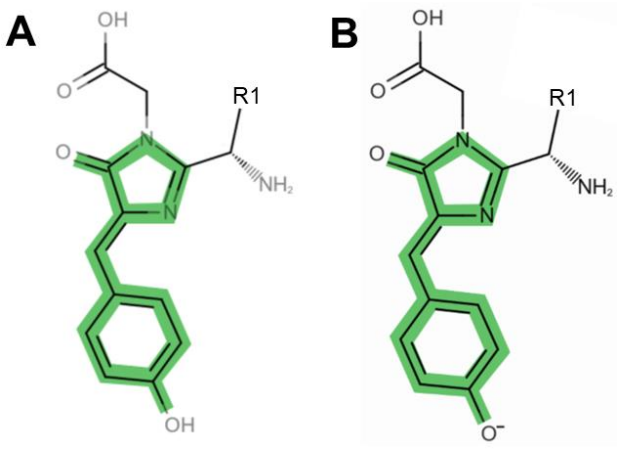

Figure 12. Protonated and deprotonated form of the chromophore. (**A**) Deprotonated chromophore (for *Av*GFP and EGFP). (**B**) Protonated chromophore (for *Av*GFP and GFPuv).

Table 1. Basic photophysical parameters of GPF-like FPs mentioned in this chapter. Values are reproduced from the FPbase database https://www.fpbase.org/ , except where indicated (Lambert, 2019).

Fluorescent Protein	Max. exc. (nm)	Max. em. (nm)	Ext. Coeff. (mM^{-1} cm^{-1})	QY	Brightness	pK_a
Sirius	355	424	15.0	0.24	3.6	3.0
EBFP	380	440	31.5	0.20	6.3	n.d.
Azurite	383	447	26.2	0.55	14.4	5.0
EBFP2	383	448	32.0	0.56	17.9	5.3
SBFP2	380	446	34.0	0.47	16.0	5.5
mTagBFP	402	457	52.0	0.63	32.8	2.7
ECFP*	435	474	30.0	0.30	9.0	4.7
Cerulean*	434	474	30.7	0.44	13.7	4.7
SCFP3A	433	474	30.0	0.56	16.8	4.5
mTurquoise	434	474	30	0.84	25.2	4.5
mTurquoise2	434	474	30	0.93	27.9	3.1
GFPuv	397	506	30.0	0.79	23.7	n.d.
EGFP	488	507	55.9	0.6	33.5	6.0
***Av*GFP**	395/475	509	25.0	0.79	19.8	4.5
Superfolder GFP	485	510	83.3	0.65	54.2	5.9
mNeonGreen	506	517	116	0.8	92.8	5.7
Kaede (green form)	508	518	98.8	0.88	86.4	5.6
LanYFP	513	524	150.0	0.95	142.5	3.5
EYFP	513	527	67.0	0.67	44.9	6.9
SYFP2	515	527	101	0.68	68.7	6.0
Venus	515	528	92.2	0.57	52.6	6.0
Citrine	516	529	77.0	0.76	58.5	5.7
phiYFP	525	537	115.0	0.60	69.0	n.d.
zFP538	528	538	20.2	0.42	8.5	n.d.
mBanana	540	553	6.0	0.70	4.2	6.7
mHoneydew	478	536/562	17.0	0.12	2.0	4.0
mOrange	549	565	58.0	0.60	34.8	6.5
mKO	551	565	63.8	0.62	39.6	5.5
eqFP578	552	578	102.0	0.54	55.1	n.d.
Kaede (red form)	572	580	60.4	0.33	19.9	5.6
tdTomato	554	581	138.0	0.69	95.2	4.7
DsRed	558	583	72.5	0.68	49.3	n.d.
TagRFP	555	584	100.0	0.48	48.0	3.8
mTangerine	568	585	38.0	0.30	11.4	5.7
LaRFP	521	592	71.0	0.10	7.1	4.7
mScarlet	569	594	100.0	0.70	70.0	5.3
mStrawberry	574	596	90.0	0.29	26.1	4.5
mCherry	587	610	72.0	0.22	15.8	4.5
eqFP611	559	611	78.0	0.45	35.1	3.8
mRaspberry	598	625	86.0	0.15	12.9	n.d.
mKate	588	635	45.0	0.33	14.9	6.2
mPlum	590	649	41.0	0.10	4.1	4.5

*Data (except pKa values) as remeasured in (Lelimousin *et al*, 2009)

1.5.2 Tuning the colour of *Av*GFP by mutagenesis

The initial attempts consisted in modifying the tyrosine of the chromophore with other residues with an aromatic character (phenylalanine, histidine, tryptophan).

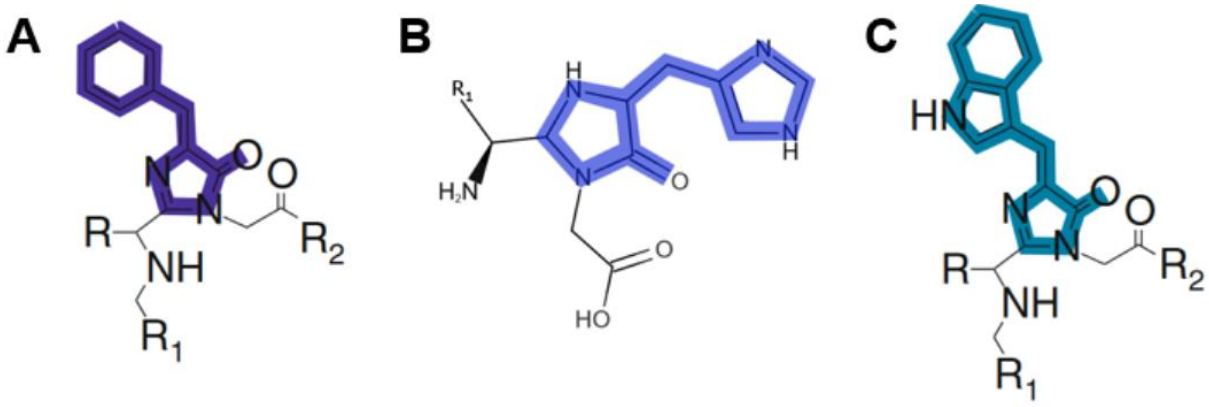

Figure 13. Chromophore of GFP-like proteins containing (**A**) a phenylalanine (**B**) an histidine or (**C**) a tryptophan (adapted from Merola *et al*, 2010).

1.5.2.a Effects of replacing Tyr66 with Phe

Replacing Tyr66 by a phenylalanine (**Figure 13A**) produces a mature chromophore that is maximally excitable at 360 nm, but with very weak blue fluorescence peaking at 442 nm (Cubitt *et al*, 1995). *Av*GFP-Y66F was difficult to evolve, and it took a cyan fluorescent mutant (see paragraph 1.5.2.c) to obtain a relatively bright variant whose fluorescence peaks at 424 nm, corresponding to a deep-blue, or ultramarine colour, helped cover the whole UV-visible spectrum with fluorescent proteins. This protein was dubbed Sirius and finds application in the design of biosensors used with excitation and emission filters in only the blue-cyan part of the spectrum, leaving room for the use of another biosensor with excitation/emission filters in the green to red part of the spectrum (Tomosugi *et al*, 2009).

1.5.2.b Effects of replacing Tyr66 with His

GFP-Y66H (**Figure 13B**) is weakly fluorescent around 450 nm (Heim *et al*. 1994). It took a couple of mutations, including Y145F (Heim *et al*. 1996), to obtain a useful blue fluorescent protein, EBFP (for Enhance Blue Fluorescent Protein) (Yang *et al*, 1998). Because of a limited brightness, EBFP was subject to multiple evolution efforts who resulted in Azurite (Mena *et al*, 2006), EBFP2 (Ai *et al*, 2007), SBFP2 (Kremers *et al*, 2007).

1.5.2.c Effects of replacing Tyr66 with Trp

Similarly to GFP-Y66F and GFP-Y66H, GFP-Y66W (**Figure 13C**) is weakly fluorescent around 480 nm (Heim *et al*. 1994). It took mutations on the seventh and eight strands of the protein to obtain a variant with an acceptable level of fluorescence: ECFP (for Enhanced Cyan Fluorescent protein), presumably to accommodate for the increased bulk of the tryptophan side chain (Heim *et al*. 1996) (Cubitt *et al*, 1999). This hypothesis was later confirmed by the structure determination of ECFP at physiological pH (Lelimousin *et al*, 2009), as it revealed an outward rotation of part of the seventh strand. ECFP became the most used FP as a donor in a yellow FP based FRET sensors (Zhang *et al*, 2002). However, the low brightness of ECFP and complex photophysics resulted in a poor dynamic range of the FRET, which triggered many efforts to improve its low fluorescence QY of 0.36. A second generation of cyan fluorescent proteins emerged in the mid 2000s with a QY around 0.5: Cerulean (Rizzo *et al*, 2004), CyPet (Nguyen & Daugherty, 2005) and SCFP3A (Kremers *et al*. 2006). One crucial mutation in this improvement appeared to be the H148D mutation on the seventh strand. Finally, two key subsequent mutations resulted in significant improvements of the QY: T65S in SCFP3A leading to mTurquoise (Goedhart *et al*, 2010) with a QY of 0.84 (mutation identified by fluorescence lifetime screening), then I146F in mTurquoise leading to mTurquoise2 (Goedhart *et al*, 2012) with a QY close to perfection (0.93).

1.5.2.d Effects of introduction a Tyr in π-stacking with Tyr66

The structure determination of the single-point mutant S65T of *Av*GFP revealed the precise location of all residues around the chromophore (Ormö *et al*, 1996). This led the authors to postulate that insertion of a histidine or tyrosine residue (with a polarisable pi electron system) at position 203 would lower the energy of the excited state of the adjacent chromophore and thus induce a red-shift of the emission maximum, paving the way for the design of the yellow fluorescent protein (E)YFP containing the T203Y mutation. Indeed the crystallographic structure of YFP showed a π-π stacking interaction between the phenolate group of the chromophore and the side chain of Tyr203 (Wachter *et al*, 1998) (**Figure 14**). EYFP was later improved by mutagenesis (for its brightness and pKa) into the widely used yellow FPs Citrine (Griesbeck *et al*, 2001), Venus (Nagai *et al*, 2002) and SYFP2 (Kremers *et al*. 2006), which all have improved folding and exhibit higher brightness, with a lower pKa and a lower sensitivity to chloride ion binding in the chromophore cavity (halide ion binding

induces a loss of fluorescence in FPs). These yellow FPs are primarily used to play the role of an acceptor in FRET experiments in conjunction with a cyan FP as a donor.

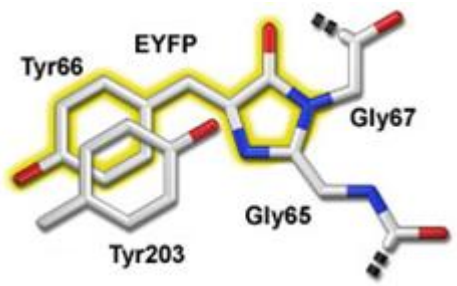

Figure 14. Representation of the π-π stacking interaction leading to yellow fluorescence in EYFP.

The T203Y mutation, originally found by rational design, was later rediscovered in the wild-type fluorescent protein PhiYFP of the jellyfish *Phialidium* (Pletneva *et al*, 2013b), which constitutes a glaring example of how exploiting the genetic diversity of GFP homologues in marine organisms can give leads to modify the spectroscopic properties of GFP-like proteins.

1.5.2.e Effects of the V68N mutation -> elimination of Ser65's side chain -> conjugation further on the main chain (DsRed-like)

*Av*GFP resisted evolution towards variants emitting in the orange and red parts of the visible spectrum long after the discovery of *Ds*Red in 1999. However, mutation of Val68 (the first residue after those of the chromophore) into an asparagine appears to catalyse the elimination of the Ser65 side chain, which is accompanied by the formation of an acylimine bond between the N and Calpha atoms of the main chain, which resembles the DsRed chromophore, in which the first residue would be an alanine (Mishin *et al*, 2008) (**Figure 15**). However, the elimination reaction is not complete, and while the resulting variant, R10-3, shows red fluorescence, it retains also a green fluorescence component, limiting its usefulness. Nevertheless, this result shows that mutagenesis of one single gene, that of *Av*GFP, can result in fluorescence maxima covering almost the whole visible spectrum, from deep blue to red.

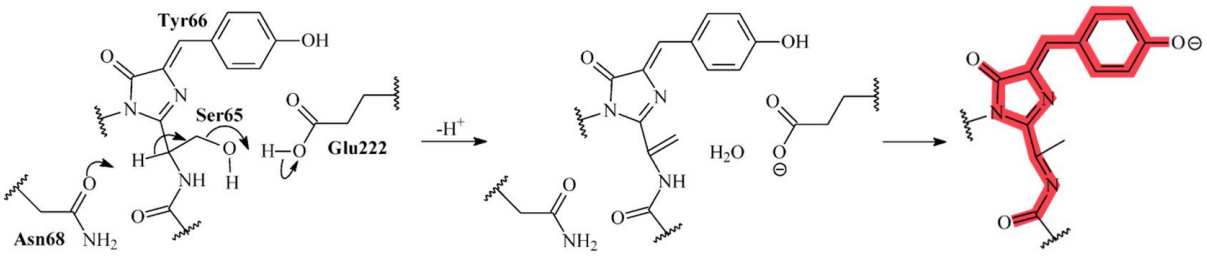

Figure 15. Formation mechanism of DsRed-like chromophore in a *Av*GFP variant (adapted from Mishin *et al*, 2008).

1.5.3 The discovery of GFP homologues in other marine organisms

*Av*GFP (originating from an organism of the Hydrozoa class of the Cnidaria phylum) long resisted tuning of its emission maximum above the 527 nm of EYFP, when additional red-shifted variants over the whole visible light spectrum would allow multicolour labelling of cells. Fortunately the team of Sergey Lukyanov discovered in 1999 the first natural red FP in a coral, *Discosoma sp.*, belonging to the Anthozoa class (Matz *et al*., 1999). Following up on this discovery, a number of teams looked for FPs in marine organisms of other classes, for instance in *Branchiostoma lanceolatum* (Baumann *et al*, 2008), an organism from the Laptocardii class of the Chordata phylum, and in *Pontella mimocerami* (Hunt *et al.*, 2010), an organism of the Copepoda subclass in the Arthropoda phylum. This search for new FPs with unknown photophysical properties still exists today and with the advancement of modern sequencing especially RNA sequencing there is a new interest in the study of GFP homolog especially in non-bioluminescent organism (Bork *et al*, 2015). This thesis is the continuation of this field and show that even today new mechanism of chromophore formation can be found.

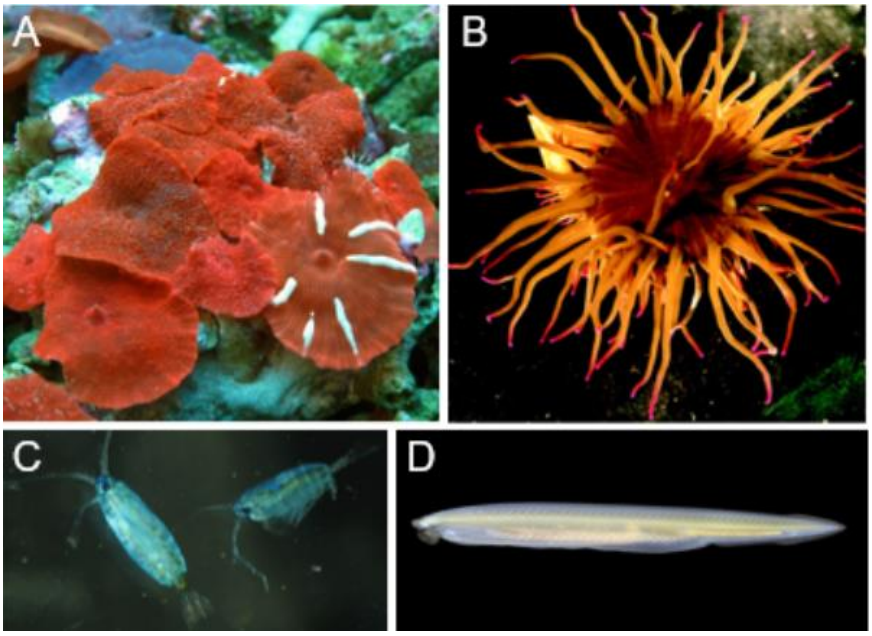

Figure 16. Marine organisms in which homologous of *Av*GFP have been found. (**A**) A colony of corals from the Discosoma species (reproduced from https://www.seahorseaquariums.com/Discosoma-sp/7496). (**B**) A sea anemone from the species Anemonia sulcata (reproduced from Lukyanov et al, 2000). (**C**) A copepod from the species Pontella mimocerami observed under visible light. (reproduced from Hunt et al., 2010). (**D**) A lancelet from the species Branchiostoma lanceolatum) (reproduced from https://naturephotographers.photoshelter.com/image/I0000UF4W8HcPqCQ).

In particular, most *Av*GFP homologues found in chordates have a Gly-Ser-Gly chromophore whose structure and spectroscopic properties are similar to the deprotonated chromophore of *Av*GFP, such as the tetrameric yellow fluorescent protein LanYFP found in *Branchiostoma lanceolatum*. The green FP mNeongreen was obtained by directed evolution from LanYFP, is to date the brightest monomeric FP developed so far (Shaner *et al*, 2013).

1.5.4 The canonical red fluorescent DsRed chromophore

The bright red fluorescence of the DsRed chromophore (QY = 0.68) was demonstrated by mass spectrometry to be attributed to a chromophore whose chemical structure is similar to that of a deprotonated *Av*GFP chromophore, in which the N-Ca bond of residue 66 is dehydrogenated. This dehydrogenation extends the electron conjugation of the GFP chromophore up to the carbonyl group of residue 65 (Gross *et al*, 2000) (**Figure 17**). However, it appears that some DsRed molecules have chromophores which do not proceed beyond the green-emitting form (state B, phenolate group), corresponding to a dead-end pathway. In fact, the phenol group needs to stay protonated for the chromophore to fully

mature to the red form, and there is a blue fluorescent intermediate on the maturation pathway, in which the methylene bridge is not formed yet (Strack *et al*, 2010).

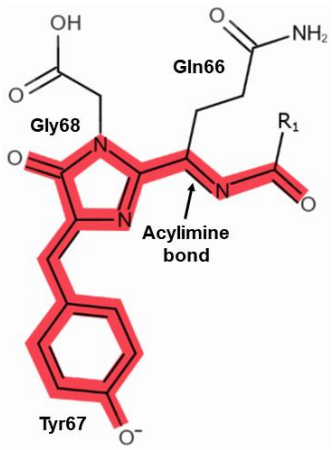

Figure 17. Chemical structure of the DsRed chromophore, in which the delocalized electron from the phenolate and imidazolinone rings is extended to the main chain, on which chromophore maturation has led to the formation of an acylimine bond.

1.5.5 The mFruit series derived from DsRed

DsRed was first monomerized into mRFP1 (Campbell *et al*, 2002). A comprehensive effort was thus attempted by (partial) saturation, random mutagenesis and FACS screening to identify mutants of various hues (Shaner *et al*, 2004). This resulted in the so-called mFruit series of FPs: mBanana (yellow), mHoneyDew (yellow), mOrange (orange), tdTomato (red), mTangerine (red), mStrawberry (red) and mCherry (red) (**Figure 18A**). Some of them show peculiar characteristics: mHoneydew has a Trp-based chromophore and an acylimine bond; mOrange has an additional modification of the chromophore (Shu *et al*, 2006) (**Figure 18B**), tdTomato is composed of a fused dimer of FPs to form a very bright non-aggregating RFP.

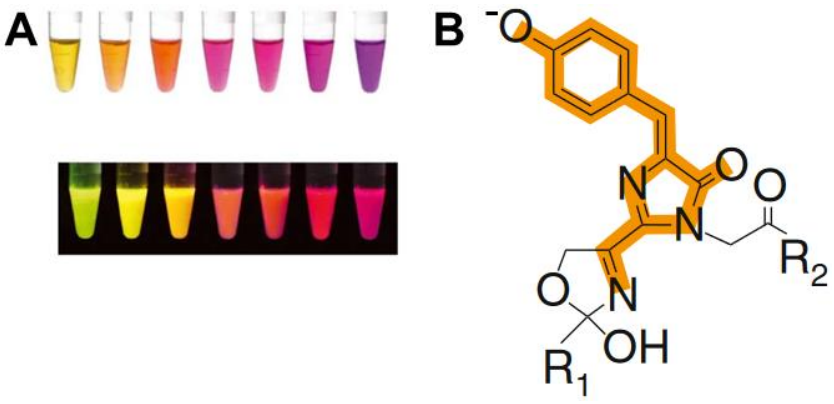

Figure 18. mFruit series of FPs. (**A**) From left to right: mHoneydew, mBanana, mOrange, tdTomato, mTangerine, mStrawberry and mCherry. Top: absorption colour, Bottom: fluorescence emission colour (reproduced from (Shaner *et al*, 2004) (**B**) Chemical structure of the mOrange chromophore, where a third heterocycle is formed upon the reaction of the first residue of the chromophore (a threonine) with the protein mainchain (adapted from Merola *et al*, 2010).

Further red-shifting of the emission maximum could be achieved by using somatic hypermutation, the mechanism by which B lymphocytes optimize immunoglobulins. This method is faster than the usual point or random mutagenesis and the successive screening, the limiting step being the transfection of each mutants into an individual cell. In the case of somatic hypermutation method, 23 rounds of FACS sorting led to the design of the FPs mRaspberry and mPlum, with emission maxima above 620 nm to qualify as in the **far-red** class, 625 and 649 nm, respectively (Wang *et al*, 2004) (**Table 1**).

1.5.6 eqFP578 and derivatives

eqFP578 is bright dimeric RFP from the sea anemone *Entacmaea quadricolor*, which was evolved into the bright monomeric variant TagRFP with complete chromophore maturation (Merzlyak *et al*, 2007). TagRFP was then evolved into the far-red FP mKate to find applications in deep tissue imaging (Shcherbo *et al*, 2007).

TagRFP was then mutated to trap the maturation of the chromophore to its blue emitting form (see § **1.5.4**) in order to generate blue FPs with a Tyr-based chromophore, which could be brighter than *Av*GFP mutants with a His-based chromophore. The resulting protein mTagBFP has a chromophore whose Cβ atom is protonated, thus reducing the conjugation of the chromophore to the imidazolinone ring and the acylimine bond (Subach *et al*, 2008, 2010) (**Figure 19**).

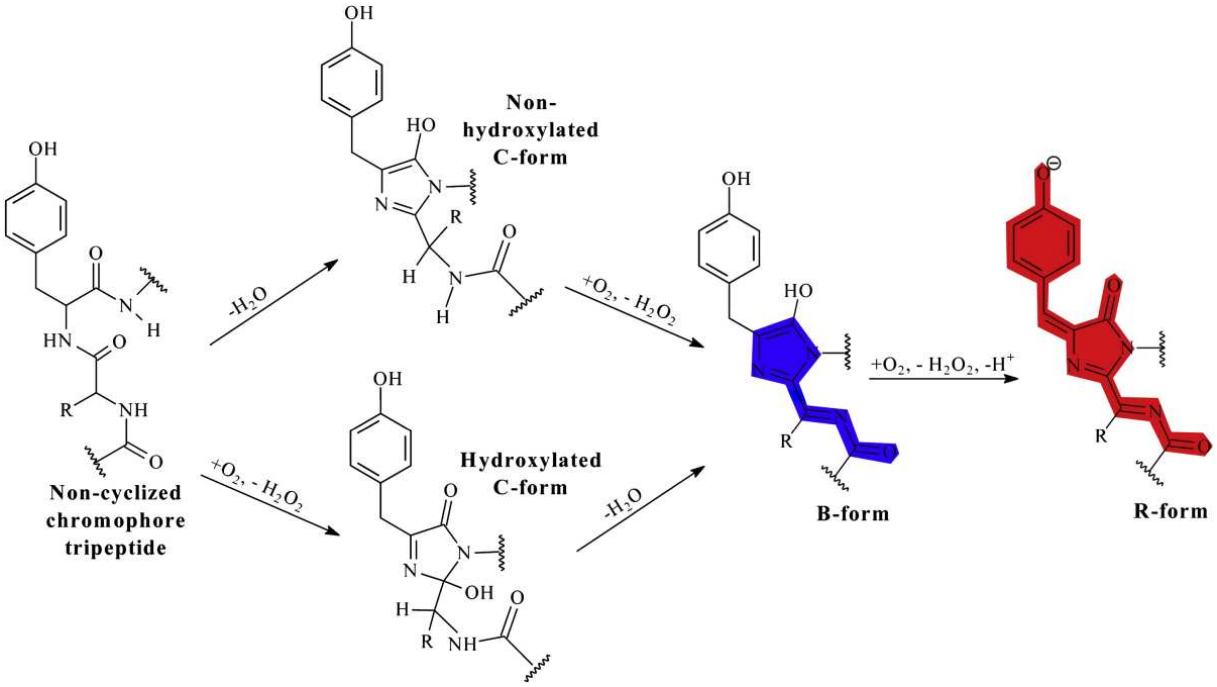

Figure 19. Mechanism of chromophore maturation for the blue FP mTagBFP and the red FP TagRFP. The blue fluorophore is a maturation intermediate that is stabilized in mTagBFP (reproduced from (Subach *et al*, 2010).

Most RFPs have been evolved from either DsRed or eqFP578. mScarlet, the brightest RFP so far, is an exception as it was evolved from a synthetic gene template derived from a sequence alignment of various naturally occurring RFPs and chromoproteins (Bindels *et al*, 2017).

1.5.7 A different fluorescent isomer of the DsRed chromophore: eqFP611

eqFP611 is a red FP which has the singular property of having its DsRed-like chromophore in its *trans* configuration (**Figure 20**) (Petersen *et al*, 2003), when the vast majority of GFP homologues with a trans chromophore are non-fluorescent (Day & Davidson, 2009; Seward & Bagshaw, 2009). The conjunction of a co-planar chromophore and a different stabilising environment can explain the spectroscopic properties of eqFP611 when compared to the fluorescent *cis*-configuration of DsRed and the noncoplanar *trans* configuration of the Rtms5 chromophore. Noteworthy, eqFP611 exhibits a long Stokes shift of 52 nm, this long shift can be explained by the presence of His197 that enters in a π-staking interaction with the phenolate ring of the chromophore. The presence of a different charge network surrounding the chromophore of eqFP611 when compared to DsRed can also partly explain the fluorescence properties of eqFP611 (Petersen *et al*, 2003).

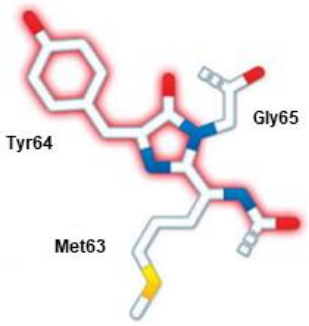

Figure 20. Chemical structure of the eqFP611 chromophore (adapted from (Day & Davidson, 2009).

1.5.8 The canonical red fluorescent Kaede chromophore

Kaede is a FP from the stony coral *Trachyphyllia geoffroyi* whose His-Tyr-Gly chromophore matures as a green fluorophore. Exposure to UV or violet light converts the chromophore to a red fluorophore (Ando *et al*, 2002). The mechanism of photoconversion induces a cleavage between the N and C_α atoms of the histidine residue of the chromophore via a formal β-elimination reaction. The subsequent formation of a C_α=C_β double bond on the histidine side chain extends the conjugation of the chromophore to the imidazole ring of the histidine (Mizuno *et al*, 2003) (**Figure 21**). Other FPs with a similar mechanism include EosFP (Wiedenmann *et al*, 2004), Dendra & Dendra2 (Gurskaya *et al*, 2006), KikGR1 (Tsutsui *et al*, 2005) and IrisFP (Adam *et al*, 2008).

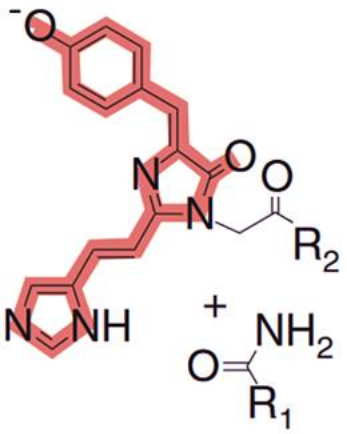

Figure 21. Chemical structure of the matured (red fluorescent) Kaede chromophore. (adapted from Merola *et al*, 2010).

1.5.9 Peculiar chromophores naturally found in other organisms

zFP538 is a yellow FP found in the button polyp *Zoanthus sp.* (Matz *et al*, 1999). The determination of the crystallographic structure of this yellow fluorescent chromophore revealed a uniquely different structure than in the YFPs derived from *Av*GFP or found in lancelets. The chromophore results from a transamination reaction, in which a DsRed-like acylimine bond is attacked by the amino group of the Lysine 66 side chain (the first residue of the chromophore) leading to the formation of a six-membered heterocycle and to the cleavage of the protein main chain between the chromophore and the previous residue 65. This amounts to the addition of a further double bond compared to the classical GFP chromophore (Remington *et al*, 2005) (**Figure 22A**).

mKO (monomeric Kusabira Orange) is a monomerized version of the wild-type orange KO from the plate coral *Fungia concinna* (Karasawa *et al*, 2004). The structure of its chromophore is quasi identical to that of mOrange, with a third heterocycle, except that the oxygen atom in the ring is replaced by a sulphur atom, as the first residue of the chromophore is a threonine in mOrange and a cysteine in mKO (Kikuchi *et al*, 2008) (**Figure 22B**). This shows that the chromophore of the peculiar FP mOrange, obtained by genetic engineering, had been previously invented by nature.

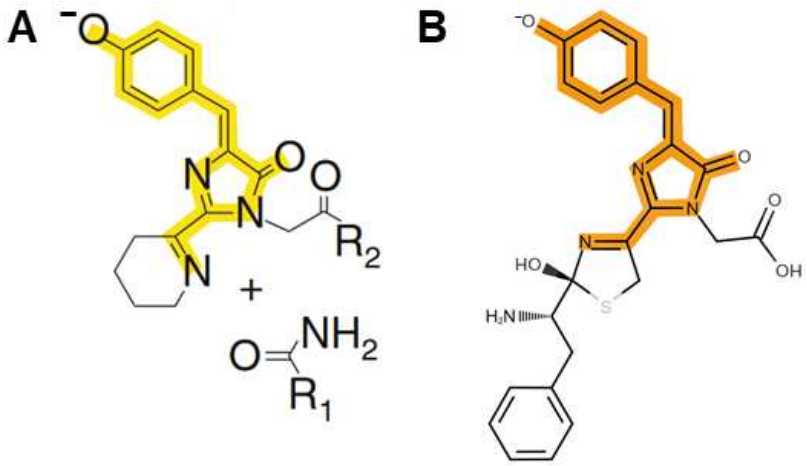

Figure 22. Peculiar chromophore structures in FPs naturally found in nature. (**A**) The chemical structure of the chromophore of the yellow FP zFP538 (adapted from Merola *et al*, 2010). (**B**) Chemical structure of the chromophore of the orange FP mKO.

1.5.10 The third type of red fluorescent chromophore: LanRFP

Apart from the two distinct pathways to generate red fluorescent chromophores, DsRed and Kaede, a third type has been discovered in *Branchiostoma lanceolatum*, with a very distinct mechanism of formation (Pletnev *et al*, 2013). The red fluorescent chromophore is composed of a classical green fluorescent GYG chromophore (very similar to that of LanYFP), but to which the phenol oxygen of a nearby tyrosine is covalently attached in the middle of the methylene bridge (**Figure 23**).

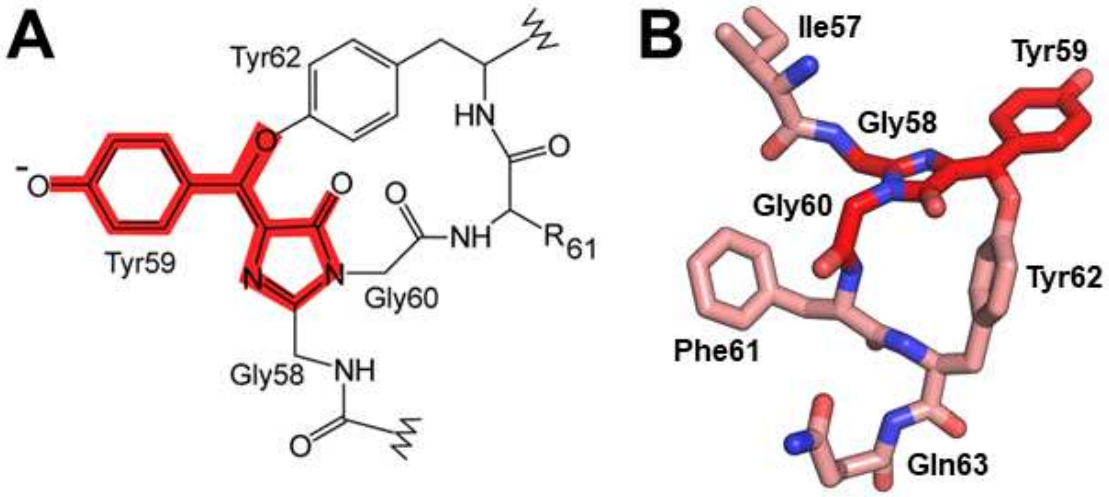

Figure 23. Red fluorescent chromophore in LaRFP. (**A**) Chemical structure (adapted from (Pletnev *et al*, 2013). (**B**) Three-dimensional structure of the chromophore derived from the crystallographic structure of LaRFP (PDB entry code: 4JF9).

1.5.11 Chromoproteins

A chromoprotein (CP) is a homologue of GFP whose chromophore is not fluorescent at all. The physiological function of CPs in marine organisms is unclear, and could be that of photoprotection of its organism. CPs find applications in imaging, as they can be used in photoacoustic imaging and as FRET acceptors (Li *et al*, 2016).

The pocilloporin Rtms5 from the reef-building coral *Montipora efflorescens* is a blue-coloured GFP-like protein, which qualifies as a CP. Its absorption maximum indicates that it contains a Ds-Red type of chromophore with an acylimine bond on the protein main chain next to the chromophore. The chromophore exhibits a *trans* configuration in a noncoplanar conformation (**Figure 24**).

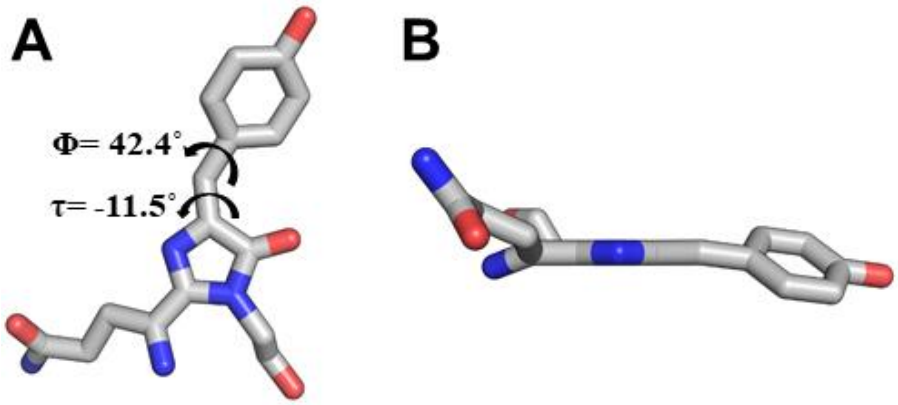

Figure 24. Chromophore conformation and configuration in the crystallographic structure of a single-point mutant of the CP Rtms5 from the reef-building coral *Montipora efflorescens*. (**A**) Orthogonal view showing the *trans* configuration. (**B**) Side view showing the noncoplanarity. (PDB entry code: 1MOV) (Prescott *et al*, 2003).

Interestingly, CPs can serve as starting material to generate new FPs. For instance, it has been shown that a single mutation in CPs from Anthozoa species can convert them into far-red FPs (Gurskaya *et al*, 2001).

1.6 Bilin-binding fluorescent proteins (BbFPs)

Bilins are biological pigments formed from the breakdown of porphyrins, constituting linear arrangements of four pyrrole rings. A key porphyrin is protoporphyrin IX, which is the precursor of chlorophylls and hemes. Hemes chelate ferrous iron (Fe^{2+}) and constitute, for instance, the functional site of the O_2-transporting proteins haemoglobin and myoglobin. Biliverdin IXa (**BV**) is the first breakdown product of heme catabolism, whose production is catalysed by haem oxygenase, while bilirubin (**BR**) is a second breakdown product in vertebrates, whose production is catalysed by biliverdin reductase (**Figure 25**). Phycobilins are obtained from BV in cyanobacteria and in the chloroplasts of red algae and cover widely the light spectrum: phycoerythrobilin (**PEB**, red, λ_{max} ~ 550 nm), phycourobilin (**PUB**, yellow-orange, λ_{max} ~ 495 nm), phycoviolobilin (**PVB**, purple) and phycocyanobilin (**PCB**, blue). Noteworthy, phycobilins are the pigments of phycobiliproteins (such as allophycocyanin, phycocyanin, phycoerythrin or phycoerythrocyanin) which are soluble proteins associated to the photosynthetic apparatus of these organisms, and absorb green, yellow, orange and red light, whose energy is then transferred to the chlorophyll molecules of the photosystems by energy transfer. Phytochromes are photosensors that converts between two states (Pr and Pfr for red-absorbing and far-red absorbing states, with two distinct configurations of the chromophore) and activating or deactivating a physiological signalling cascade, which are found in higher plants, cyanobacteria, bacteria and fungi (Rockwell *et al*, 2006). Phytochromobilin (**PΦB**) is the chromophore of plant phytochromes, PCB that of cyanobacterial phytochromes and BV. Two isomers of PΦB are formed from BV in chloroplasts (Terry *et al*, 1995). Bilins are usually bound to the biliprotein thanks to a thioether bond to a cysteine residue.

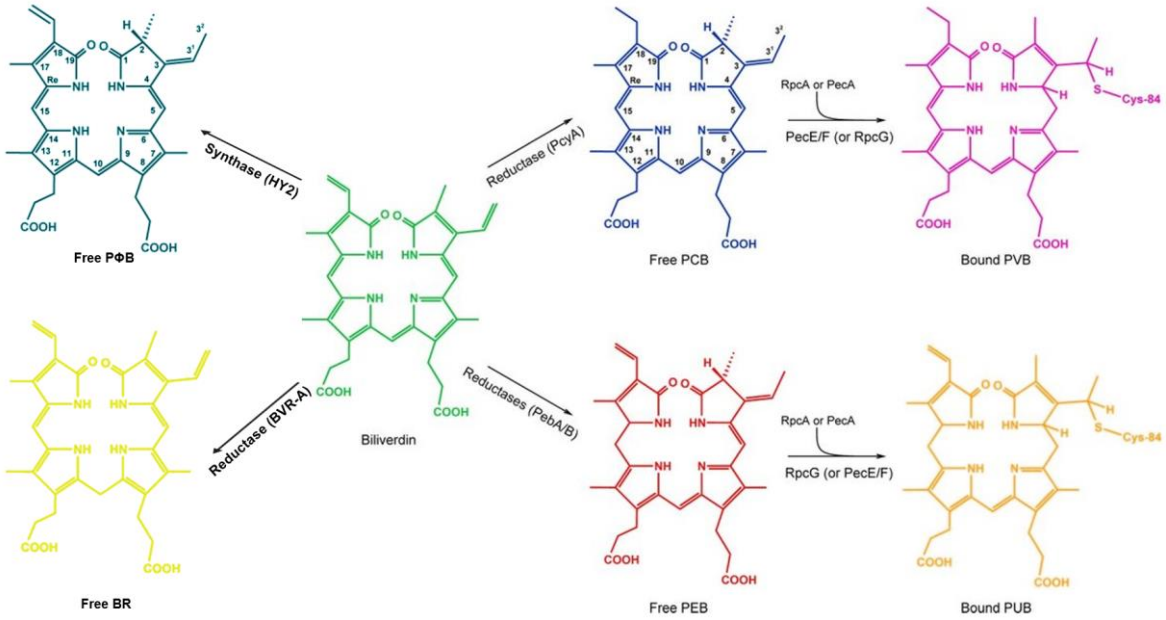

Figure 25. Overview of the various bioavailable bilins. BV is ubiquitous, BR is found in vertebrates, PΦB is found in plants, and the four phycobilins PCB, PVB, PEB and PUB are found in cyanobacteria and in the chloroplast of red algae (adapted from Blot *et al*, 2009).

The potential of BbFPs to be used as bright fluorescence probes has been recognized as early as 1984, when phycobiliproteins were dubbed as phycofluors (Glazer & Stryer, 1984). It was later demonstrated that replacing *in vivo* the natural chromophore PΦB of a plant phytochrome by PEB resulted in a bright orange fluorescent probe; this class of fluorescent proteins were correspondingly called phytofluors (Murphy & Lagarias, 1997).

1.6.1 Near-infrared BbFPs developed from the *D. radiodurans* phytochrome

The bacteriophytochrome from the poly-extremophilic bacterium *Deinococcus radiodurans* is a red/far-red photoreceptor mediating phototaxis and pigmentation. It is formed of a photosensory module composed of the three domains PAS, GAF and PHY (**Figure 26**) and of a carboxy-terminal regulatory module. The chromophore, a BV molecule, autocatalytically binds to a conserved cysteine residue of the PAS-domain via a thioether linkage and is located within a pocket of the GAF-domain. The ensemble composed of the PAS and GAF-domains is called the chromophore-binding domain (*Dr*CBD) and has a size of ~35 kDa. Light absorption induces an isomerisation of C_{15}=C_{16} double bond between the C and D rings, from the $5Z_{syn}$, $10Z_{syn}$, $15Z_{anti}$ configuration in the red-light-absorbing state Pr to the $15E_{anti}$ configuration in the far-red-light-absorbing state Pfr (**Figure 27**).

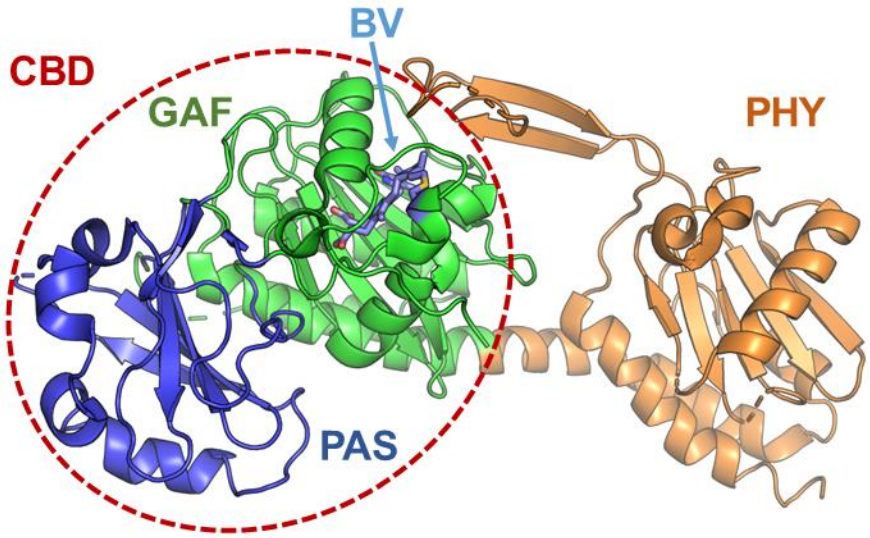

Figure 26. Structure of the photosensory module of the bacteriophytochrome from Deinococcus radiodurans, composed of the three domains PAS (Per/ARNT/Sim), GAF (cGMP phosphodiesterase/adenyl cyclase/FhlA) and PHY (phytochrome-specific) domains. The BV chromophore is located within a pocket of the GAF-domain. The PAS and GAF-domains constitutes the chromophore-binding domain (CBD) (PDB entry code: 4O0P) (Takala et al, 2014).

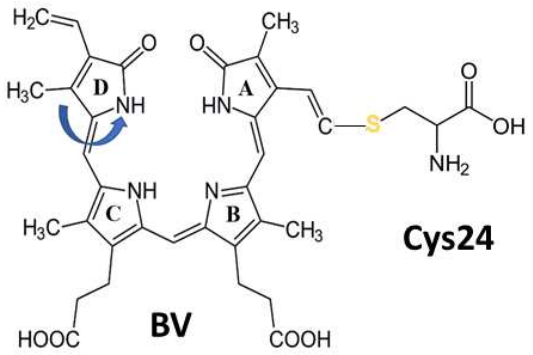

Figure 27. Chemical structure of the biliverdin chromophore in wild-type DrCBD. BV is composed of four pyrrole rings A, B, C and D, with two propionate groups on rings B and C. BV is attached to a cysteine residue of the protein via its vinyl group of ring A. Light-activation leads to isomerisation of the C15=C16 double bond (blue arrow).

The ground-breaking step in the development of BbFPs came from the observation that one given single-point mutant *Dr*CBD of the dimeric bacteriophytochrome exhibited a fluorescence peak at 690 nm after reconstitution with BV (Wagner *et al*, 2008). This specific mutant was subject to several rounds of evolution with saturation mutagenesis at selected locations (Shu *et al*, 2009). The first variant (*Dr*CBD-D207H, a.k.a IFP1.0) was showing light- and dark-adaptation behaviour (that is the loss of absorption properties upon light excitation, and subsequent recovery in the dark) suggesting that the chromophore was

undergoing configuration or conformation change. It was also dimeric, as inherited from the structural properties of the bacteriophytochrome. The second variant, IFP1.1, already showed a significant increase of the fluorescence QY, at the cost of a 9 nm blue-shift of its emission maximum (**Table 2**). The ultimate variant, IFP1.4, is monomeric, has a higher fluorescence quantum yield, a 14 nm blue-shifted emission maximum and has the light-adaptation behaviour abolished suggesting that mutations prevent isomerisation of the chromophore (a non-radiative decay pathway of the excited state) and favour fluorescence (a radiative decay pathway of the excited state) with emission maxima above 700 nm, *i.e.* in the near-infrared region of the light spectrum. Very conveniently, this wavelength lies well within the so-called optical window of tissues, in which absorption by haemoglobin, water, and lipids and light scattering is minimized. This result thus paved the way for the development of near-infrared fluorescent proteins for applications in deep tissue imaging based on the chromophore-binding domain of a phytochrome. This deletion followed by mutagenesis steps induced the monomerisation of the protein and the stabilisation of the chromophore and prevent the light-induced isomerisation of the chromophore. However, the formation of these fluorescent probes relies on the bioavailability of the chromophore biliverdin that needs to be produced by the degradation of haems.

During the evolution leading IFP1.4, it was shown that the monomerization process occurring between its predecessors IFP1.2 and IFP1.3 led to a decrease in cell brightness. After combining the beneficial mutations of IFP1.2 to those of IFP1.4, the directed evolution process led to reverse some of the mutations from IFP1.4 in addition to new mutations. IFP2.0 was eventually obtained with higher cellular brightness, which could potentially be explained by higher increased protein solubility, protein stability and BV-binding affinity (Yu *et al*, 2014).

A more conservative approach was used in the group that determined the first structure of *Dr*CBD at the University of Wisconsin in Madison. The weakly fluorescent *Dr*CBD-D207H variant was monomerized by inspection of the dimer interface and choice of three mutations destabilizing the interactions. Besides D207H, the mutant of a residue in the vicinity of the chromophore, Tyr263, into a phenylalanine was observed to become fluorescent as well, possibly by stabilizing the excited state, thus extending its lifetime. Interestingly, the monomerized D207H mutant was not fluorescent, and the addition of Y263F yields a monomeric fluorescent variant, which was called Wi-Phy for 'Wisconsin infrared phytofluor' (Auldridge *et al*, 2012).

Table 2. Basic photophysical parameters of BbFPs mentioned in this chapter. Values are reproduced from the FPbase database https://www.fpbase.org/ (Lambert, 2019), except where indicated.

Fluorescent Protein (chromophore)	Max. exc. (nm)	Max. em. (nm)	Ext. Coeff. (mM^{-1} cm^{-1})	QY	Brightness	pK_a
UnaG (BR)	498	527	77.3	0.51	39.4	4.0
smURFP* (PCB)	642	666	65.0	0.07	4.6	n.d.
BphP1-FP (BV)**	639	669	60.0	0.13	7.8	n.d.
smURFP (BV)	642	670	180.0	0.18	32.4	3.2
iBlueberry[¶] (BV)	644	667	38.0	0.07	2.6	<4
miRFP670 (BV)	642	670	87.4	0.14	12.2	4.5
iRFP670 (BV)	643	670	114.0	0.11	12.5	4.0
miRFP670nano (BV)	645	670	95.0	0.11	10.3	3.7
iRFP682 (BV)	663	682	90.0	0.11	9.9	4.5
iRFP702 (BV)	673	702	93.0	0.08	7.4	4.5
mIRFP (BV)	674	703	92.4	0.10	9.0	4.3
mIRFP703 (BV)	674	703	90.9	0.09	7.8	4.5
mIFP (BV)	683	704	82.0	0.08	6.6	3.5
IFP1.2[¶¶] (BV)	684	707	86.0	0.07	6.0	n.d.
IFP1.3[¶¶] (BV)	684	707	84.0	0.06	5.0	n.d.
IFP1.4[¶¶] (BV)	684	708	92.0	0.07	6.4	4.6
miRFP709 (BV)	683	709	78.4	0.05	4.2	4.5
IFP2.0 (BV)	690	711	86.0	0.08	6.9	n.d.
IFP1.1[¶¶] (BV)	686	713	86.0	0.05	4.3	n.d.
iRFP713 (BV)	690	713	105.0	0.06	6.3	4.0
Wi-Phy (BV)	701	719	93.0	0.05	4.4	n.d.
iRFP720 (BV)	702	720	96.0	0.06	5.8	4.5
miRFP720 (BV)	702	720	98.0	0.06	6.0	4.5
IFP1.0[¶¶] (*Dr*CBD) (BV)	699	722	60.0	0.03	1.8	n.d.

*Values from (Rodriguez *et al*, 2016) ** Values from (Shcherbakova *et al*, 2016), [¶]Value from (Yu *et al*, 2016). [¶¶]Values from (Shu *et al*, 2009).

1.6.2 Monomeric near-infrared BbFPs developed from a *Bradyrhizobium* phytochrome

After the development of IFP1.4 and IFP2.0 from a dimeric bacteriophytochrome as monomeric near-infrared FPs, it was realized that they still retained a dimeric character at high concentrations, with dissociation constants of 7.8 µM and 3.7 µM, respectively. In order to develop a truly monomeric FP, ~40 sequences of bacteriophytochrome from *Bradyrhizobium* (BphB) were analysed at the location of the dimer interface in DrBphP. Identification of a polar residue at this position in a sequence of a BphB from a *Bradyrhizobium* species (a symbiotic bacterium present in plants) indicated the possibility that this BphB was monomeric. Indeed, the CBD domain of this BphB was shown to be monomeric, and was then subjected to evolution by conjunction of saturation mutagenesis, DNA shuffling and random mutagenesis. The resulting protein was called mIFP (Yu *et al*, 2015). mIFP contains 19 mutations when compared to the parent BphB, five of them in the vicinity of the D ring of the biliverdin chromophore.

Based on the observation that a cyanobacterial phytochrome (Cph1 from *Synechocystis sp*. PCC 6803) could bind covalently a PCB molecule to a cysteine (Cys259) located in the GAF-domain, the equivalent residue in mIFP as inferred from structural alignment, Ile251, was mutated into a Cys. The resulting variant, iBlueberry, showed a ~40 nm blue-shift of its excitation and emission maxima, presumably upon a change in the binding mode of the chromophore, which can be used in conjunction with mIFP for two-colour labelling (Yu *et al*, 2016). A very similar approach was used to develop a blue-shifted NIR FP with a relatively high QY, BphP1-FP (Shcherbakova *et al*, 2015).

1.6.3 Near-infrared BbFPs developed from *R. palustris* bacteriophytochromes

After the development of IFP1.4, the observation that the full-length phytochrome *Rp*BphP2 from *Rhodopseudomonas palustris* was weakly fluorescent at 725 nm upon 710 nm light excitation (Giraud *et al*, 2005) triggered the search for more red-shifted near-infrared FPs (NIR FPs) with better affinity for BV. The fragment of *Rp*BphP2 composed of the PAS and GAF-domains, with the mutation D202H equivalent to D207H in *D. radiodurans*, were used as template for evolution through three rounds of random mutagenesis and one round of saturation mutagenesis at the position of the residues identified to play a role on the fluorescence properties. The brightest mutant contained 13 substitutions compared to the template and was called iRFP for infrared fluorescent protein (Filonov *et al*, 2011), later

renamed iRFP713 when other iRFPs became available. iRFP713 is dimeric, has a brightness similar to that of IFP1.4, is much more photostable, but has a 12-fold increased affinity for BV, obviating the use of haem oxygenase in cells or addition of exogenous BV.

In order to generate near-infrared FPs with blue or red-shifted emissions to do multi-colour imaging in the optical window of tissues, the template of another phytochrome from *R. palustris*, *Rp*BphP6, was used, since the absorption maximum of its Pr state is blue-shifted by 10 nm compared to that of *Rp*BphP2. An approach similar to that used to develop iRFP713 was used, with the addition of the chromophore-stabilizing mutation Y258F observed for the *Dr*BphP-derived NIR FP Wi-Phy (Auldridge *et al*, 2012), which resulted in the blue-shifted iRFP702 (Shcherbakova *et al*. 2013). Using random and saturation mutagenesis as well as FACS screening, red-shifted and blue-shifted mutants of iRFP702 and iRFP713 were identified. Eventually, three other variants were selected: iRFP670 deriving from *Rp*BphP6, and iRFP682 and iRFP720 both deriving from *Rp*BphP2. These five NIR FPs cover ~50 nm of the near-infrared region of the NIR light spectrum and allow for multicolour imaging with spectral unmixing in living mice (Shcherbakova *et al*. 2013).

The five aforementioned NIR FPs having the drawback of being dimeric, a new template for evolution was used from *Rp*BphP1, a dimeric phytochrome, but whose CBD domain is not implicated in dimerization. A particular effort was taken to mutagenize the C-terminal α-helix of the GAF-domain, which is known to be implicated in the dimerization of *Dr*BphP-derived NIR FPs. Eventually three bright, spectrally distinct monomeric NIR FPs were obtained: miRFP670, miRFP703 and miRFP709 (Shcherbakova *et al*, 2016). Finally, miRFP720 was obtained by a different approach, that is the monomerization of iRFP720 by mutagenizing five residues at the dimer interface inferred from the structure of *Rp*BphP2 with charge-changing, size-altering and hydrophobic-to-hydrophilic switching mutations, eventually leading to miRFP720 (Shcherbakova *et al*, 2018).

Finally a different group evolved miRFP from *Rp*BphP1, providing the brightest NIR FP with emission above 700 nm (Piatkevich *et al*, 2018). The particularity of this approach was to screen mutants using an automated robot that can control multiple parameters at the same time. This robot screened 300 000 cells expressing *Rp*BphP1 mutants in approximately four hours.

1.6.4 BbFPs derived from cyanobacterial PCB-binding proteins

One of the main drawbacks of the near-infrared BbFPs derived from phytochromes is the large size of the protein due to the necessity of both the PAS and GAF-domain for efficient binding of the bilin chromophore (biliverdin in the case of bacteriophytochromes). Alternate solutions to bacteriophytochromes are phycobiliproteins from cyanobacteria. One promising target was identified: the alpha subunit of allophycoyanin from *Trichodesmium erythraeum*: TeAPCα. Native APC is a highly fluorescent heterohexamer formed of three α/β dimers, which requires a lyase to incorporate PCB as chromophore. The 15 kDa TeAPCα was first evolved to obviate the need for a lyase and obtain autocatalytic binding of PCB. Because PCB is absent in mammals, the resulting BbFP was evolved to bind BV as chromophore as well. After twelve rounds of evolution, and the inserting of 20 mutations, the resulting protein has been called smURFP (for small ultra-redFP) (Rodriguez *et al*, 2016). smURFP is capable of autocatalytically binding BV but also PCB, which is more membrane-permeant than BV, providing an alternate way of producing smURFP holoproteins, yet with almost a 3-fold reduction of both the EC and QY, resulting in a 7-fold decrease in brightness (**Table 2**). Noteworthy, the smURFP dimer can sequentially bind one then a second copy of either BV or PCB (**Figure 28**), but the secondly bound chromophore partially quenches fluorescence affecting the overall quantum yield.

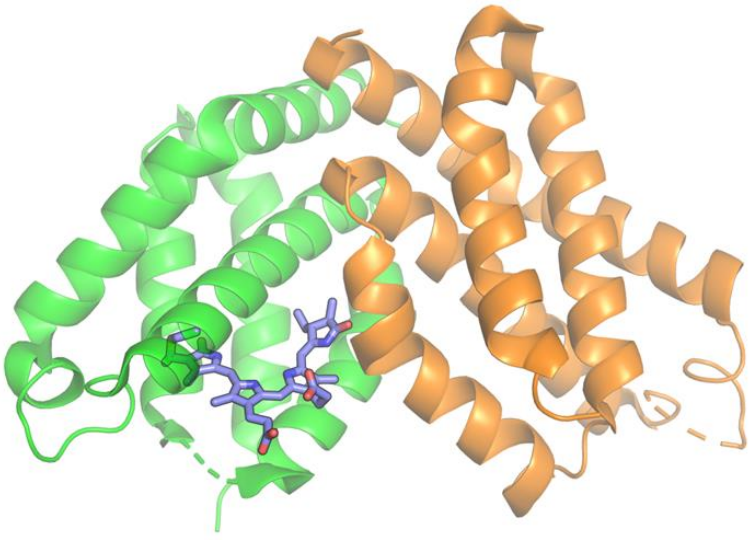

Figure 28. Structure of a dimer of the single-point mutant Y56R of smURFP loaded with one BV molecule (PDB entry code: 6FZN) (Fuenzalida-Werner *et al*, 2018).

Other potentially interesting cyanobacterial proteins are cyanobacteriochromes (CBCR). Unlike BphPs, the bilin chromophore can autocatalytically bind within a single GAF-domain (not requiring the presence of a PAS-domain for covalent linkage). Among other single domains, a 15 kDa GAF-domain of a CBCR from *Nostoc punctiforme* NpR3784 was shown to weakly bind BV in BV-producing *E. coli* bacteria. After 17 rounds of evolution, 18 mutations were introduced to produce a near-infrared FP, whose fluorescence peaks at 670 nm with a relatively high QY when compared to other FPs of the same colour class. Because of its particularly small size when compared to GFP-like and phytochrome-derived FPs, this protein was called miRFP670nano and is the smallest near infrared FP to date (Oliinyk *et al*, 2019). miRFP670nano is monomeric, a property inherited from its parent CBCR, and shows high affinity for BV.

1.6.5 A green fluorescent BbFP with bilirubin (BR) as chromophore

Green fluorescence has been observed in skeletal muscle of the *Anguilla japonica* eel species (Unagi in Japanese) leading to the purification of a green fluorescent protein of ~17 kDa (Hayashi & Toda, 2009). Samples of juvenile (glass, or transparent) eels showing green fluorescence were collected to generate cDNA. The protein responsible for the fluorescence was identified by sequencing using degenerate primers to be a 139-residue protein of the fatty acid binding protein (FABP) family and dubbed UnaG (Kumagai *et al*, 2013). As UnaG did not fluorescence when expressed in bacteria, it was postulated that the fluorescent chromophore was a mammalian cofactor. The conjunct use of mass spectrometry and of the screening with the apoprotein of a large number of biological material eventually yielding green fluorescence led to the identification of a 585.3 Da molecule sensitive to oxidation, which could be proved to be BR after reconstitution with the apoprotein. When heterologously expressed in cell lines or mouse brain, the protein produces oxygen-independent fluorescence. The structure of UnaG reveals a well-stabilized biplanar conformation of BR, which explains its high affinity, high specificity and fluorescence efficiency (**Figure 29**). UnaG constitutes the prototype of a new family of FPs, whose fluorescence is triggered upon binding of a specific ligand.

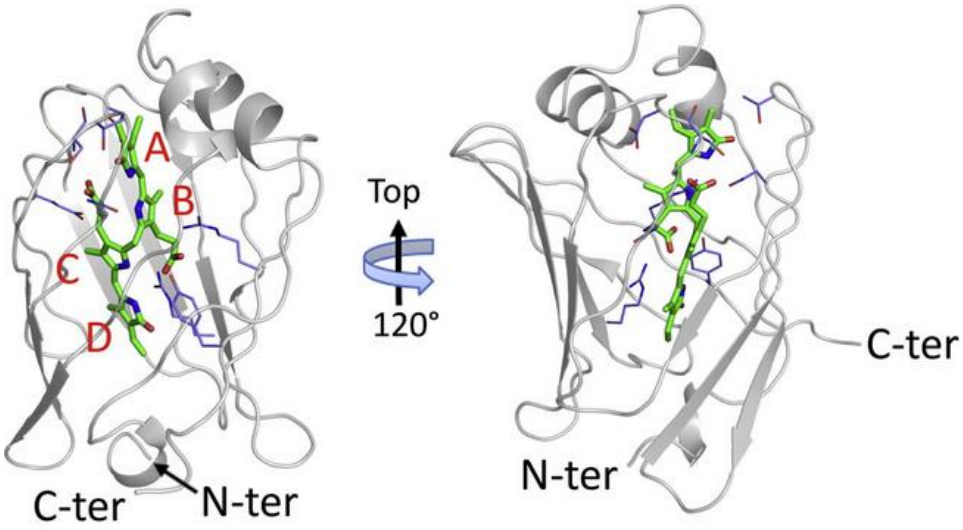

Figure 29. Structure of UnaG from two different angles with the chromophore bilirubin coloured in green (adapted from Kumagai et al, 2013) (PDB entry code: 4I3B).

1.7 FMN-binding fluorescent proteins (FbFPs)

One major drawback of using GFP-like FPs as biomarker is their requirement of molecular oxygen to obtain a fully matured fluorescent chromophore. This is also an issue, yet to a lesser extent, for BbFPs as the high-level production of biliverdin, which is the chromophore, or the precursor of the chromophore, requires the presence of heme oxygenase using O_2 for catalysis. The low blue autofluorescence of plant phototropins, coming from short wavelength-light excitation of the flavin mononucleotide (FMN) chromophore of their LOV (Light-oxygen-sensing) domain, suggested testing them as starting material for O_2-independent biomarkers.

Phototropins are plant photoreceptors which sense blue light to mediate various processes, in particular phototropism, which leads to the optimisation of photosynthesis upon variation in light intensity. Phototropins function as a light-activated protein kinase, whose autophosphorylation triggers a signalling cascade (Christie, 2007) (**Figure 30A**). They contain two LOV domains, which undergo a photocycle in which a covalent adduct between the FMN and a cysteine residue is formed in microseconds and lasts for tens to hundreds of seconds (**Figure 30B**). By mutating the cysteine into an alanine of a LOV domain from an oat phototropin, it has been shown that the photocycle is inhibited, and FMN fluorescence is increased (Swartz *et al*, 2001), which amounts to decreasing non-radiative decay pathways of the excited state in favour of a radiative decay pathway.

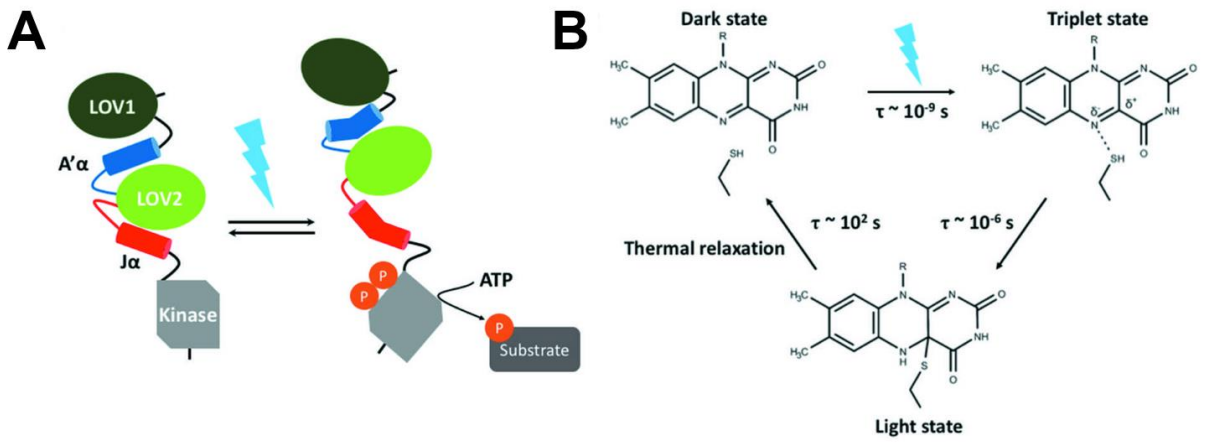

Figure 30. Phototropin model and photocycle. (**A**) Schematic model of phototropin activation upon blue light illumination. (**B**) Typicial photocycle of a LOV domain (adapted from Aumonier *et al*, 2020).

This approach was applied to bacterial blue-light photoreceptors from *Pseudomonas putida* and *Bacillus subtilis*, which were eventually restricted to their LOV domains. The photoactive cysteine was mutated into an alanine, resulting in the FPs PpFbFP (later renamed Pp2FbFP) and EcFbFP with an emission maximum at 495 nm for both and fluorescence quantum yields of 0.17 and 0.39, respectively (Drepper *et al*, 2007). Recently, the LOV domain of a histidine kinase from the thermophilic bacterium *Chloroflexus aggregans*, in which the functional cysteine is replaced with an alanine, has been shown to be the shortest FbFP (CagFbFP, 107 residues, 11.6 kDa) available to date (Nazarenko *et al*, 2019).

While the fluorescence of an FbFP is not genetically-encoded as in GFP, it relies on the incorporation of FMN, which is a key and abundant cellular cofactor serving as an electron shuttle in many enzymes. It has the advantages of not requiring molecular oxygen for formation of the fluorescent state, and of a reduced size (12-17 kDa) (Wingen *et al*, 2017) compared to GFP-like (~27 kDa) and phytochrome-derived BpFPs (~35 kDa). An obvious drawback however is the limitation of their fluorescence emission maximum in a narrow band of the visible spectrum next to 500 nm.

The LOV2 domain of phototropin-2 from *Arabidopsis thaliana* has attracted continuous interested for functional evolution. Cancelation of the photoreduction was obtained by the C426A mutation and in order to increase fluorescence, DNA shuffling was performed with the sequence of the other three LOV domains of *A. thaliana* phototropins using high fidelity PCR. The eventual fluorescent protein with a QY of 0.44 (a ~40 % improvement compared to the single mutant C426) was obtained with four mutations of the initial LOV2 domain and dubbed iLOV (Chapman *et al*, 2008) (**Table 3**).

Table 3. Basic photophysical parameters of FbFPs mentioned in this chapter. Values are reproduced from the FPbase database https://www.fpbase.org/ (Lambert, 2019).

Fluorescent Protein (chromophore)	Max. exc. (nm)	Max. em. (nm)	Ext. Coeff. (mM^{-1} cm^{-1})	QY	Brightness	pK_a
Pp2FbFP	449	495	14.2	0.22	3.1	n.d.
EcFbFP	448	496	13.9	0.44	6.4	n.d.
CagFbFP	447	497	16.3	0.36	5.5	n.d.
iLOV	447	497	n.d.	0.44	n.d.	n.d.
miniSOG	447	501	14.1	0.43	6.1	n.d.

This LOV domain was subjected to an evolution process aimed at exploiting the dual function of FMN, (i) its fluorescence and (ii) and its type II photosensitizing properties, i.e. its capacity of converting molecular oxygen ($^3O_2$, which is in a triplet state) into the very potent oxidizing molecule singlet oxygen $^1O_2$. A protein capable of these two functions would be extremely useful in correlative light and electron microscopy (CLEM) experiments to provide very high-resolution images (Gaietta *et al*, 2002). To this end, saturation mutagenesis was performed at selected locations around the chromophore, including Cys42, and mutants were screened by their capacity of bleaching a fused molecule of the near infrared FP IFP1.4 upon light irradiation. Eventually a 6-point mutant was obtained with a glycine at position 426 (instead of an alanine in other FbFPs), which was called miniSOG (**Figure 31**). miniSOG was shown to be an efficient tool in CLEM experiments to map gap junction channels formed by connexins between cells (Shu *et al*, 2011).

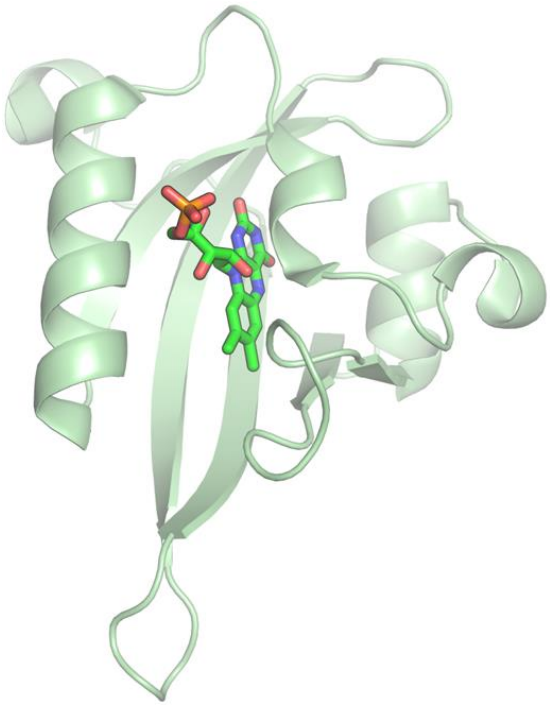

Figure 31. Structure of the labelling protein miniSOG with dual functions: fluorescence and singlet oxygen generation (PDB entry code: 6GPU) (Torra *et al*, 2019)

1.8 Applications of FPs

1.8.1 Protein colocalization

The first application of *Av*GFP, has been to be fused to a protein of interest so that the fluorescence signal is a reporter of colocalization of the POI within the cell, confined to membranes or cell compartments. In the first application, GFP was fused to β-tubulin, a protein abundant in touch receptors of *C. elegans* (Chalfie *et al*, 1994). This application is still the primary one nowadays, yet in more sophisticated ways. In particular, super-resolution microscopy was developed by taking advantage of reversibly photoswitchable FPs, allowing the determination of protein colocalization at resolutions below the diffraction limit of optical microscopy (Sahl *et al*, 2017, **Figure 32**). This major breakthrough led to the award of the 2014 Nobel Prize in Chemistry "for the development of super-resolution fluorescence microscopy" to Eric Betzig, Stefan W. Hell and William E. Moerner.

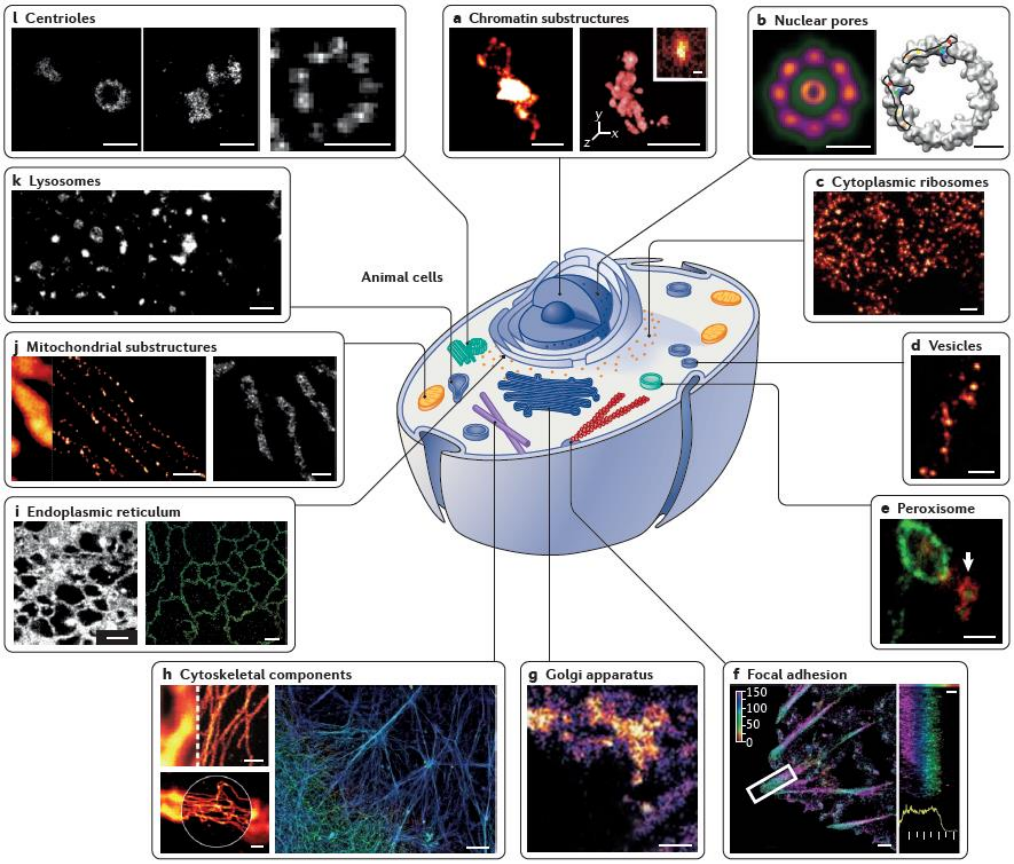

Figure 32. Localisation of the principal constituents of cell organelles (Nucleus, Endoplasmic reticulum, Golgi apparatus, peroxisome, lysosomes). (**a**) Chromatin structures, (**b**) nuclear pores, (**c**) cytoplasmic ribosomes, (**d**) vesicles, (**e**) peroxisome, (**f**) focal adhesion, (**g**) Golgi apparatus, (**h**) cytoskeletal components, (**i**) endoplasmic reticulum, (**j**) mitochondrial substructures, (**k**) lysosomes, (**l**) centrioles (adapted from Sahl *et al*, 2017).

1.8.2 Principle of a biosensor

Engineering FPs into active biosensors has followed two distinct pathways. The first one is based on the FRET technique between two complementary FPs. The donor and the acceptor are linked at each terminus of a protein fragment, whose structure is sensitive to the presence of a given small molecule. The binding of the molecules to the external protein sequence (e.g. a receptor domain) may enhance or decrease FRET efficiency and the modulation in emission ratio can be used to report the evolution of the concentration of the molecule of interest, while largely preserving normal cellular processes.

The second approach is to engineer an FP to render its fluorescence properties sensitive to the binding of a given molecule. The best-case scenario is when a given FP already shows sensitivity to pH, chloride or other molecules and can be exploited as it is. But it can also be engineered by rational design, usually in close vicinity of the chromophore through saturation mutagenesis at selected locations.

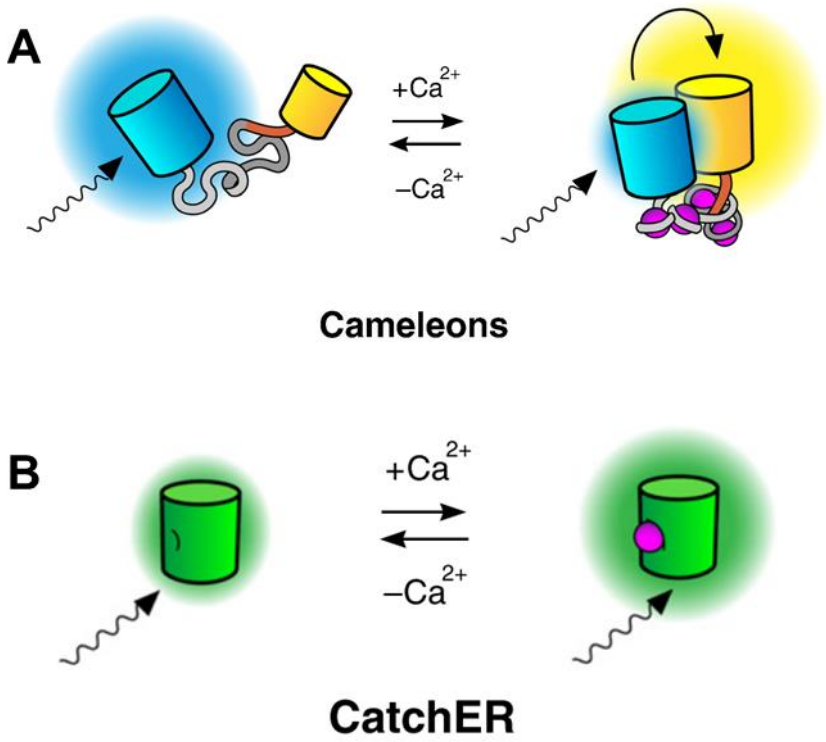

Figure 33. Two different types of biosensors. (**A**) FRET based calcium sensor Cameleon. Fluorescent proteins are represented as coloured cylinders while the CaM-binding peptide is represented as a grey linker and M13 as an orange linker (**B**) Single fluorescent protein calcium sensor, the binding of calcium changes the fluorescent intensities (CatchER) (adapted from Pérez Koldenkova & Nagai, 2013).

1.8.3 pH sensor

pH regulation plays an important role in protein function, metabolic reactions, autophagy, and other cellular processes. Mapping pH spatial and temporal gradients is thus important to help decipher cellular processes.

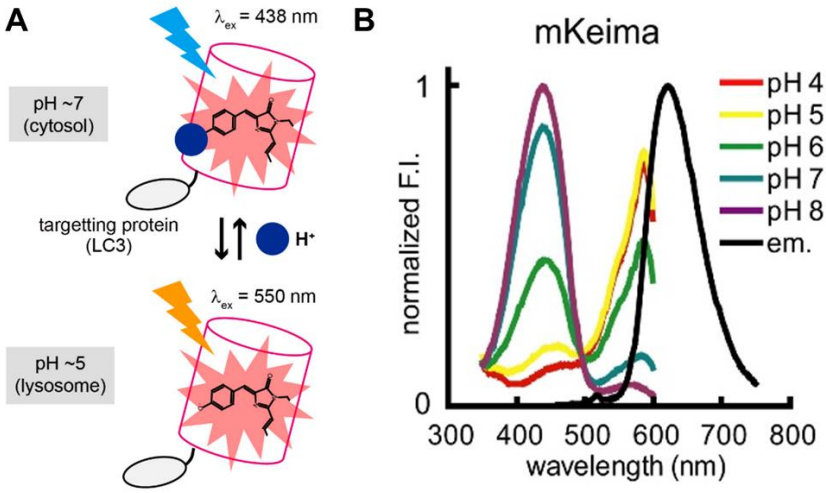

Figure 34. Ratiometric pH sensor (mKeima). (**A**) Schematic representation of mKeima (**B**) mKeima possesses dual excitation peaks at 438 and 550 nm dependent of the pH and a single emission peak at 620 nm (adapted from Shinoda *et al*, 2018 and Katayama *et al*, 2011)

FP-based pH sensors have been engineered according to three different types (Shinoda *et al*, 2018). The first one consists of a single FP, whose fluorescence intensity is affected by pH change. However, the fluorescence signal will vary according to the expression level of the FP and to the rate of its photobleaching, which will depend itself on excitation light intensity. Thus, pH values cannot be estimated precisely from changes in the fluorescence. This first type is not much used anymore.

The second type has been created to solve the drawback of the intensity-based sensor by using a ratiometric approach. It consists of a single FP which has either a dual excitation (D_{ex}) or a dual emission (D_{em}) mode. The ratio between the two fluorescence intensities at a given wavelength upon excitation at two different wavelengths (or at two different wavelengths when excited at a fixed wavelength) will cancel any drift of the absolute fluorescence intensity.

The third type is based on the FRET technique between two FPs, which gives again a relative signal corrected for variation in absolute intensities. However, the second type is more frequently used do its smaller size (~27 *vs* ~54 kDa).

mKeima is a large Stokes-shift monomeric RFP with two excitation peaks at 438 nm and 550 nm, corresponding to the protonated and deprotonated forms of the chromophore, respectively (Kogure *et al*, 2006). The intensity ratio between the fluorescence emitted by the two forms (each excited by blue and orange light, respectively) is dependent on pH (**Figure 34**). The pKa of 6.5 gives the centre of the usable pH range of detection.

1.8.4 Ion sensors

Calcium (Ca^{2+}) plays an important role in the brain neuronal transmission, muscle contraction and apoptosis. The first FP-based efficient Ca^{2+} sensors have been the series cameleons (Miyawaki *et al*, 1997, **Figure 33**). A cameleon is a FRET-based sensor composed of either a CFP/YFP or BFP/GFP pair, with each pair separated by a protein sequence composed of calmodulin (CaM) and M13, the 26-residue CaM-binding peptide of myosin light-chain kinase, which is disordered when unbound from CaM. Upon Ca^{2+} binding to CaM, M13 orders and binds CaM, inducing a structural change, which brings the two FPs closer to each other. The resulting FRET signal indicates the presence of Ca^{2+}.

Other types of Ca^{2+} sensors are based on single FPs, for example GCaMP (Nakai *et al*, 2001), GECO (Zhao *et al*, 2011), and Pericam (Nakai *et al*, 2001). In these cases, CaM and M13 are fused to the N- and C- termini of the FP. When Ca^{2+} is not present, the FP is not fluorescent due to the instability of the structure, whereas when Ca^{2+} binds to CaM, M13 binds CaM and stabilizes the whole structure, which restores the fluorescence. Red variants of this type of Ca^{2+} sensor based on mCherry or mApple have also been developped (Zhao *et al*. 2011; Carlson *et al*. 2013).

Chloride (Cl^-) is one of the main ions in the body as it account for about two thirds of the total anions present in the extracellular fluid and is an integral part of diverse cellular functions such as cell homeostasis or action potential transmission in neurons. The ratiometric sensor Clomeleon has been designed based on a CFP/YFP pair (Kuner & Augustine, 2000, **Figure 35**). While CFP is insensitive to Cl^-, the known sensitivity of YFP to Cl^- was exacerbated by mutations H79R and L68V.

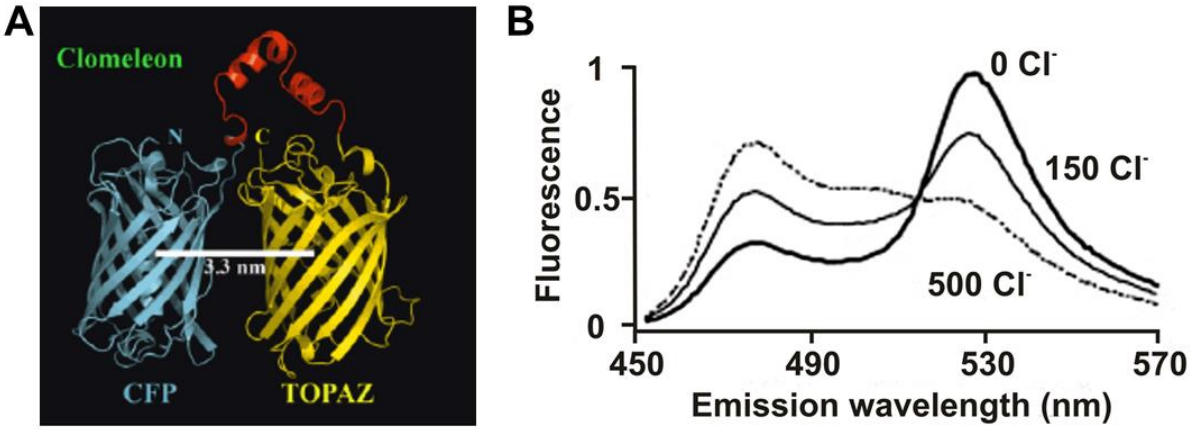

Figure 35. Chloride ion sensor Clomeleon. (A) 3-dimensional representation of Clomeleon composed of a cyan (CFP) protein and a yellow (Topaz) protein. The polypeptide linker of 24 aminoacids that connect CFP and Topaz is represented in red (B) Emission spectra of Clomeleon at various concentrations of chloride ions (mM) (adapted from Bregestovski *et al*, 2009).

Sensors for other types for ions have been developed, such as for Zn^{2+} (Qiao *et al*, 2006), Mg^{2+} (Lindenburg *et al*, 2013), Cu^{+} (Koay *et al*, 2013), inositol trisphosphate (IP3) (Tanimura *et al*, 2004) and ATP (Imamura *et al*, 2009).

1.8.5 Other types of sensors

Biosensors can monitor complex physiological and cellular events such as redox potential (roGFP1 and roGFP2 (Hanson *et al*, 2004)), hydrogen peroxide (HyPer (Belousov *et al*, 2006)), membrane potential (FlaSh (Siegel & Isacoff, 1997)), temperature (Donner *et al*, 2012), hydrostatic pressure (Watanabe *et al*, 2013) or molecular crowding (Boersma *et al*, 2015).

1.9 Objectives of the thesis work

The general objective of my thesis is to solve and analyse the structure of several new fluorescent proteins in order to understand what structural determinants control their oligomeric state and photophysical parameters, so that mutations can be proposed to improve them, for instance their oligomeric state (ideally monomeric) or their fluorescence quantum yield (as close to 1 as possible). The proteins I have been working on belong to two of the aforementioned families of fluorescent proteins.

1.9.1 GFP-like fluorescent proteins

In 2015, a NIH-funded collaboration was established between Dr. Nathan Shaner (Scintillon Institute, then University of California San Diego) as lead scientist, Dr. Anya Salih (The University of Western Sydney) and Dr. Antoine Royant (Institut de Biologie Structurale), one of my PhD supervisors, to look for, identify and characterize novel fluorescent proteins of the GFP family from diverse genetic material to develop new probes for various imaging applications. To this end, Dr. Shaner and his colleague Gerry Lambert went on an expedition to Heron Island, a research station on the Great Barrier Reef in Australia. They dived and snorkelled to collect plankton and invertebrates floating in the water. In particular, they collected two types of jellyfish, from which they eventually extracted RNA and sequenced once back home in Australia. From their sequences, they identified a number of genes homologous to that of GFP, and selected three of them for structural characterization. Chapter 3 of this manuscript describes the structure determination and analysis of a bright green fluorescent protein (*Aaus*FP1) and a blue chromoprotein (*Aaus*FP2) from *Aequorea australis*, while Chapter 4 focuses on those of a weakly red fluorescent protein (*Csi*FP4) from *Clytia simplex*.

1.9.2 Phytochrome-derived near-infrared fluorescent proteins

Dr. Royant has a long-standing collaboration with Prof. Xiaokun Shu from the University of California – San Francisco on the development of various BbFPs and FbFPs. In particular they have worked together on the development of phytochrome-base near-infrared FPs, first on IFP1.4 (Shu *et al*, 2009; Feliks *et al*, 2016) and IFP2.0 (Yu *et al*, 2014), which are both derived from the chromophore-binding domain of the bacteriophytochrome from the

extremophilic radio-resistant bacterium *Deinococcus radiodurans*. In the course of the development of iBlueberry (Yu *et al*, 2016), a near-infrared FP derived from a *Bradyrhizobium* species, crystallographic data had been obtained, for which the limited resolution could not allow for the unambiguous attribution of the configuration of the chromophore. As iBlueberry was an attempt to change the attachment site of the chromophore compared to mIFP (Yu *et al*, 2015), it was postulated that the right configurations of the chromophore had to be inferred from the comparison of the structures of mIFP, iBlueberry and a single-point mutant of iBlueberry. Chapter 5 describes the structure determination of these three near-infrared FPs, the identification of the various configurations and conformations of the chromophore in each structure and the consequences on the spectroscopic properties.

Chapter 2

MATERIALS AND METHODS

Proteins expression and purification

2.1 Phytochrome-derived near-infrared fluorescent proteins

2.1.1 Plasmids construct

Expression plasmids for the NIR FPs mIFP, iBlueberry and iBlueberry-C18I were provided by the the team of Professor Xiaokun Shu, University of California, San Francisco (Yu *et al*., 2015; Yu *et al*., 2016).

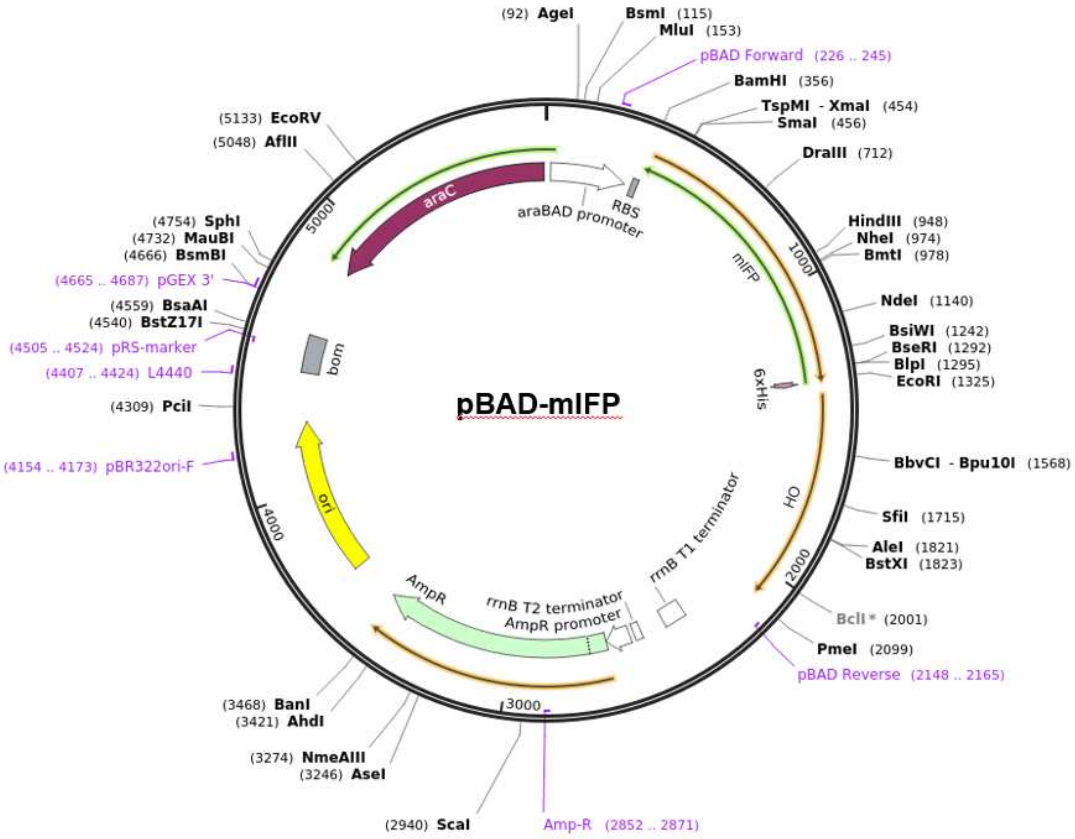

Figure 36. pBAD, expression vector containing mIFP. As noted in the main text the plasmid includes a haem oxygenase-1 gene from cyanobacteria, useful for the production of biliverdin inside bacterial expression host, and an araBAD allowing the induction of protein production via the addition of arabinose. The expression vector places a his_6 tag at the C-terminus of mIFP and contains an ampicillin resistance gene to allow for colony selection following cell transformation.

To create these plasmids (**Figure 36**) the BrBphP gene from *Bradyrhizobium* (synthesised by GenScript) was cloned into a pBAD expression vector (Invitrogen) containing a haem oxygenase-1 gene from cyanobacteria (useful for the production of biliverdin inside bacterial expression host (Migita *et al.*, 2003) and an araBAD allowing the induction of protein production via the addition of arabinose (Khlebnikov *et al.*, 2000). The expression vector

places a hexahistidine (his_6) tag at the C-terminus of the expressed protein and contains an ampicillin resistance gene to allow for colony selection.

As outlined in chapter 5, *BrBphP* was subject to site specific saturation mutagenesis to stabilise the binding of its biliverdin chromophore. This was followed by several rounds of random mutagenesis to produce the fluorescent protein (mIFP). The mutants iBlueberry and iBlueberry-C18I were produced by site-directed mutagenesis from mIFP. All proteins are expressed with a codon usage modified to accommodate *E. coli* expression.

2.1.2 Expression and purification

Upon receipt of the plasmids, expression tests were carried out to determine – in our hands - the best conditions for protein production (**Figure 37**). Following transformation, one colony was picked from each of the two cell lines tested (*E. coli* BL21, Rosetta 2^{TM}) to start a 4 mL pre-culture in LB medium (Sigma, 20 g/L LB in H_2O) which was grown overnight at 37°C at 150 rpm. 4 mL of different expression media (LB, 2YT, TB, ZYM 5052 and ZYM 505, see **Table 4** expression media composition) supplemented with ampicillin (100 µg/mL) were inoculated with the pre-culture (1/1000 dilution) and incubated at 37°C and 150 rpm until the optical density (OD) reached to between 0.5 to 0.8. Protein production was then induced by arabinose (0.2% v/v final) overnight at 17°C. 200 µL of culture were centrifuged, the supernatant discarded and the pellet resuspended in 20 µL of Buffer C (**Table 5**). Samples were then loaded on a 15% SDS-PAGE gel (3 min at 95°C) which were analysed using ImageJ (**Figure 37**, Schneider *et al.*, 2012).

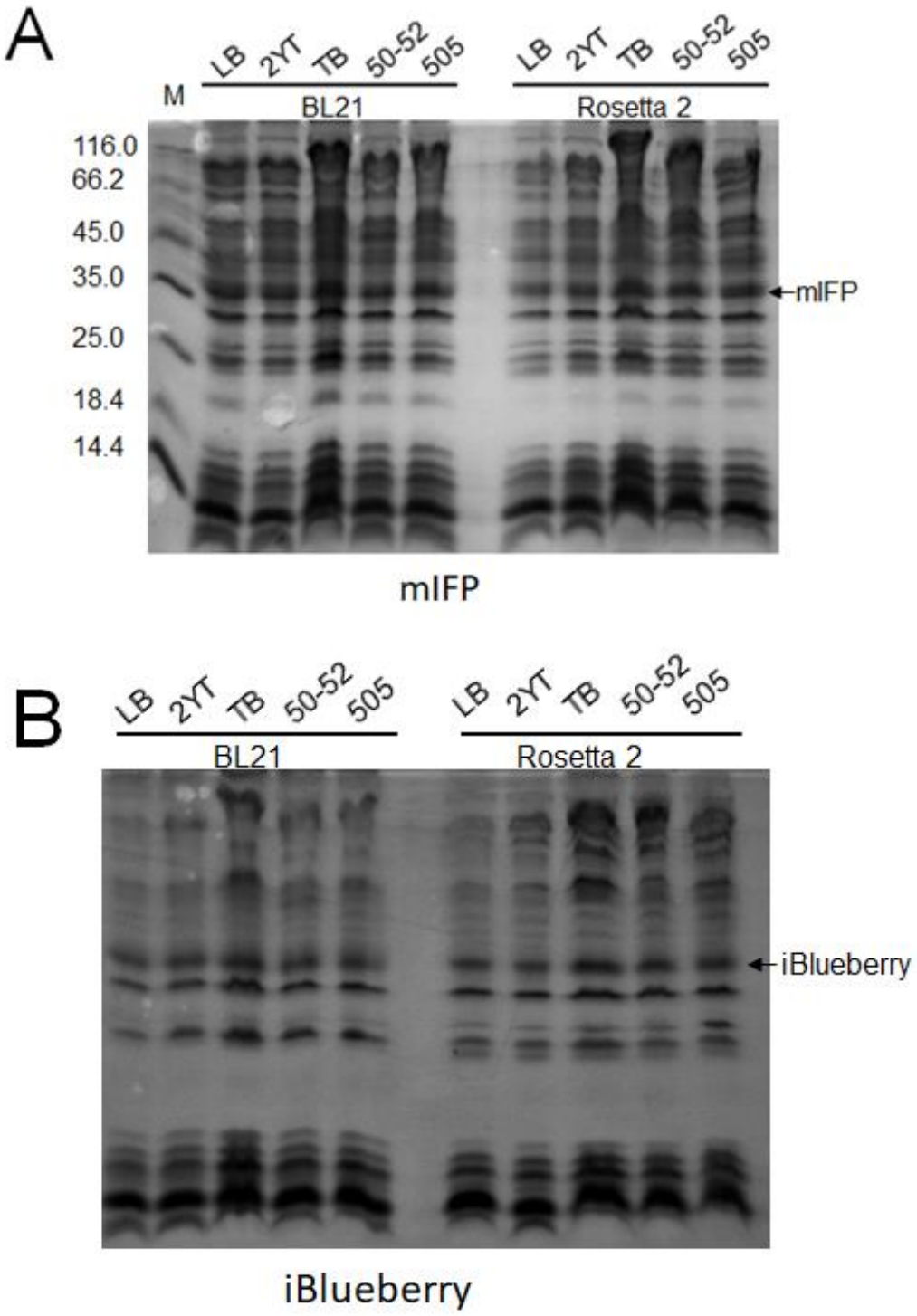

Figure 37. Results of initial expression tests for mIFP (A) and iBlueberry (B). Proteins were expressed in both *E. coli* BL21 and Rosetta 2™ cells and five different media were tested for each strain. Given the higher growth of the culture and higher protein expression, TB media was chosen with Rosetta 2™ cells as expression host.

Table 4. Composition of the various media used for the expression test and protein production.

Solution	Composition	Quantity (w/v)
LB	LB Broth (Sigma)	20 g/L
2YT	Tryptone (Bacto BD)	16 g/L
	Yeast Extract (Bacto BD)	10 g/L
	NaCl	5 g/L
TB	Tryptone (Bacto BD)	12 g/L
	Yeast Extract (Bacto BD)	24 g/L
	Glycerol	4 mL
	Potassium phosphate (0.17 M, KH_2PO_4 / 0.72 M, K_2HPO4)	100 mL
ZY	Tryptone (Bacto BD)	10 g/L
	Yeast Extract (Bacto BD)	5 g/L
NPS (20X)	$(NH_4)_2SO_4$	66 g/L
	KH_2PO_4	136 g/L
	Na_2PO_4	142 g/L
50 52 (50X)	Glucose	25 g/L
	α-Lactose	100 g/L
	Glycerol	250 g/L
505 (50X)	Glucose	25 g/L
	Glycerol	250 g/L

Table 5. Composition and function of the buffers used in protein purification.

Buffer	Function	Composition
L	Lysis Buffer	50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.25 $mg.mL^{-1}$ lysozyme, 400 $\mu g.mL^{-1}$ DNAse I, 20mM $MgSO_4$, Anti-protease Complete EDTA-free
A	Binding buffer	50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole
B	Elution buffer	50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole
C	Gel filtration buffer	20 mM Tris pH 8.0
D	Dialysis Buffer	50 mM Tris pH 8.0, NaCl 50 mM

As for all three proteins tested expression in Rosetta 2™ cells appeared to give the best results, for full-scale protein production plasmids were transformed in *E. coli* Rosetta 2™ cells (Novagen) on agar plates which were incubated overnight at 37°C with ampicillin (100 µg/mL) for selection. Protein expression was then carried out using the same protocol described above, but scaled up to 4 L of TB medium (http://www.thelabrat.com/protocols/TerrificBroth.shtml).

All three NIR FPs studied in this work were purified in a very similar fashion. After growth/expression, cells were harvested by 15 min centrifugation (4500 rpm) (**Figure 38A**) at 4°C and the resulting cell pellets suspended in 35 mL of buffer L (**Table 5**). The suspension was then flash cooled with liquid nitrogen and stored at -80°C until purification. Here, the pellet was first thawed in a water bath (40°C) and the cells lysed with 6 cycles of sonication (30 sec ON, 30 sec OFF, 60 % amplitude). The lysate was centrifuged 45 min at 14500 rpm, 4°C and the soluble fraction used in a further purification step comprising Ni-affinity chromatography on columns cleaned with degassed water and equilibrated with buffer A (**Table 5**). The lysates were loaded on the column (GE Healthcare HisTrap HP, 1mL)(**Figure 38B**) and thoroughly washed with buffer A to eliminate most of the contaminant and residual cell components. Proteins were then eluted with different concentrations of imidazole (gradient from 50 mM to 250 mM at 2 mL/min) with column fractions analysed with 15% SDS-PAGE gels to determine proteins purity and location in the elution peak (**Figure 39**).

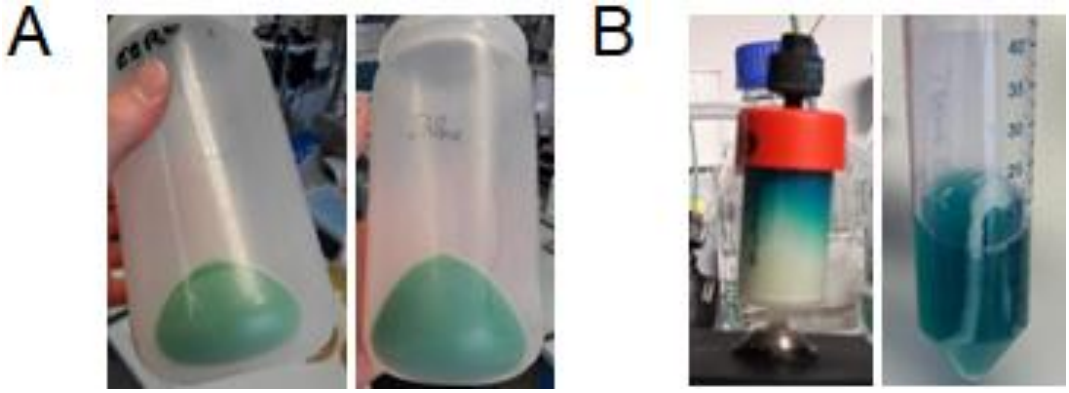

Figure 38. The production and purification of NIR FPs. (**A**) Pellets of cell containing iBlueberry-C18I (left) and iBlueberry (right) after centrifugation. (**B**) The loading of iBlueberry-C18I onto an Ni-affinity chromatography column and the resulting solution after the elution process.

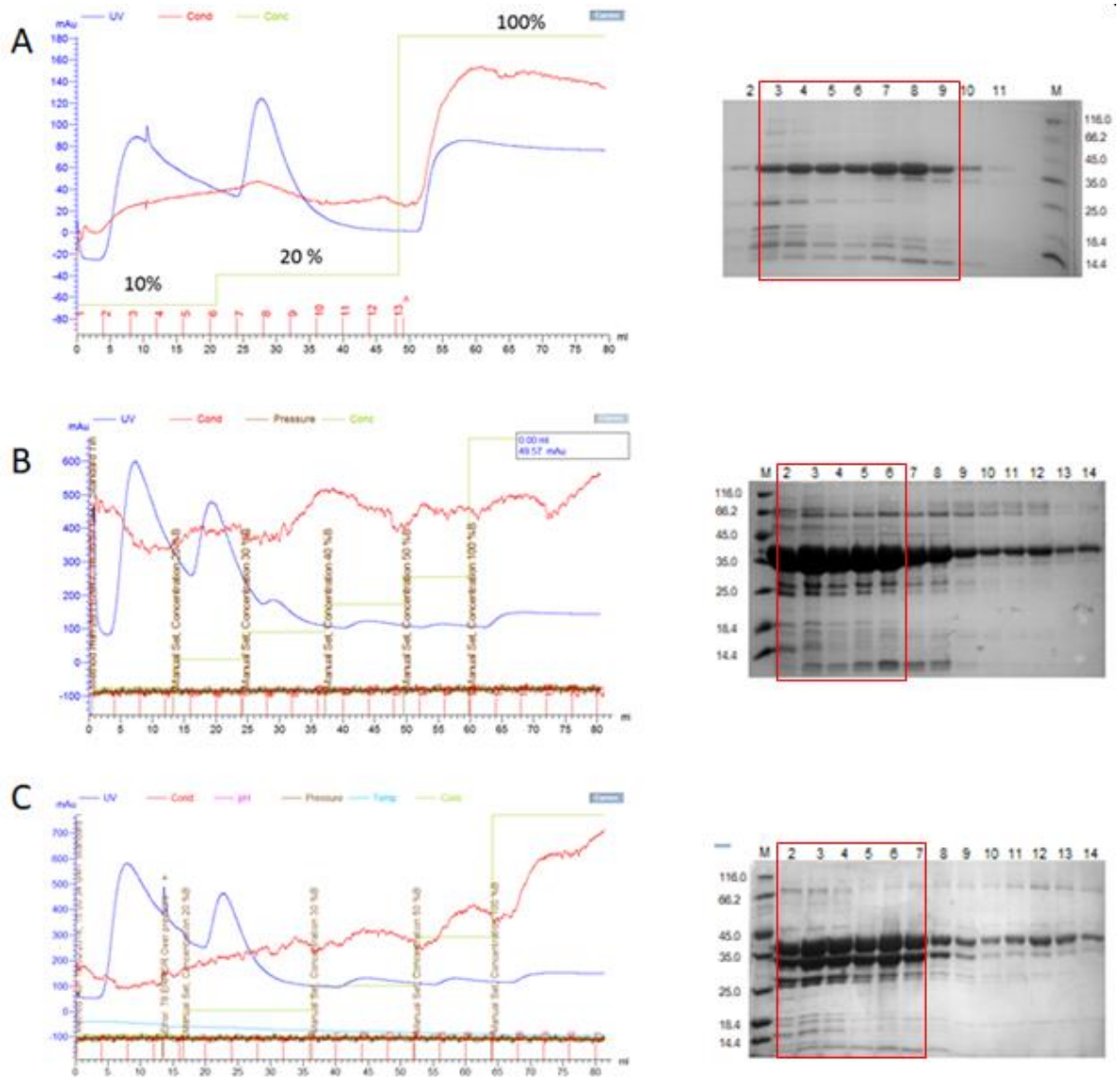

Figure 39. Elution curves of mIFP, iBlueberry-C18I and iBlueberry following nickel affinity chromatography and resulting 15% SDS-PAGE gel analysis. (**A**) mIFP, (**B**) iBlueberry-C18I, (**C**) iBlueberry. Absorbance at 280 nm (A280) is shown in blue, the conductance (mS.cm-1) is shown in red. The green line tracks the concentration of buffer B (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole (**Table 2**)) loaded on the column. M denotes the molecular weight marker also run on the gels.

Fractions containing the desired protein (3 to 9 for mIFP, 2 to 6 for iBlueberry-C18I and 2 to 7 for iBlueberry; red boxes in **Figure 39**) were pooled and dialysed (Spectra/Por, Dialysis Membrane, Standard RC Tubing, MWCO: 6-8 kD) against buffer D (50 mM Tris pH 8.0, NaCl 50 mM; **Table 5**) to eliminate imidazole. The buffer was replaced twice after 1h and the protein was left overnight at 4°C to equilibrate. The resulting protein solutions were then concentrated using an Amicon ultra concentrator (Merck Millipore, molecular weight cut off: 10 KDa).

The purification procedure continued by using gel filtration step (Superdex 75-10/300 GL columns, GE HealthCare) on a GE HealthCare AKTApure system and buffer C (20 mM Tris pH 8.0, **Table 5**) as the eluent. All gel filtration runs were performed with a flow rate of 0.8 mL/min and 0.5 mL fractionation. Fractions were analysed with 15% SDS-PAGE

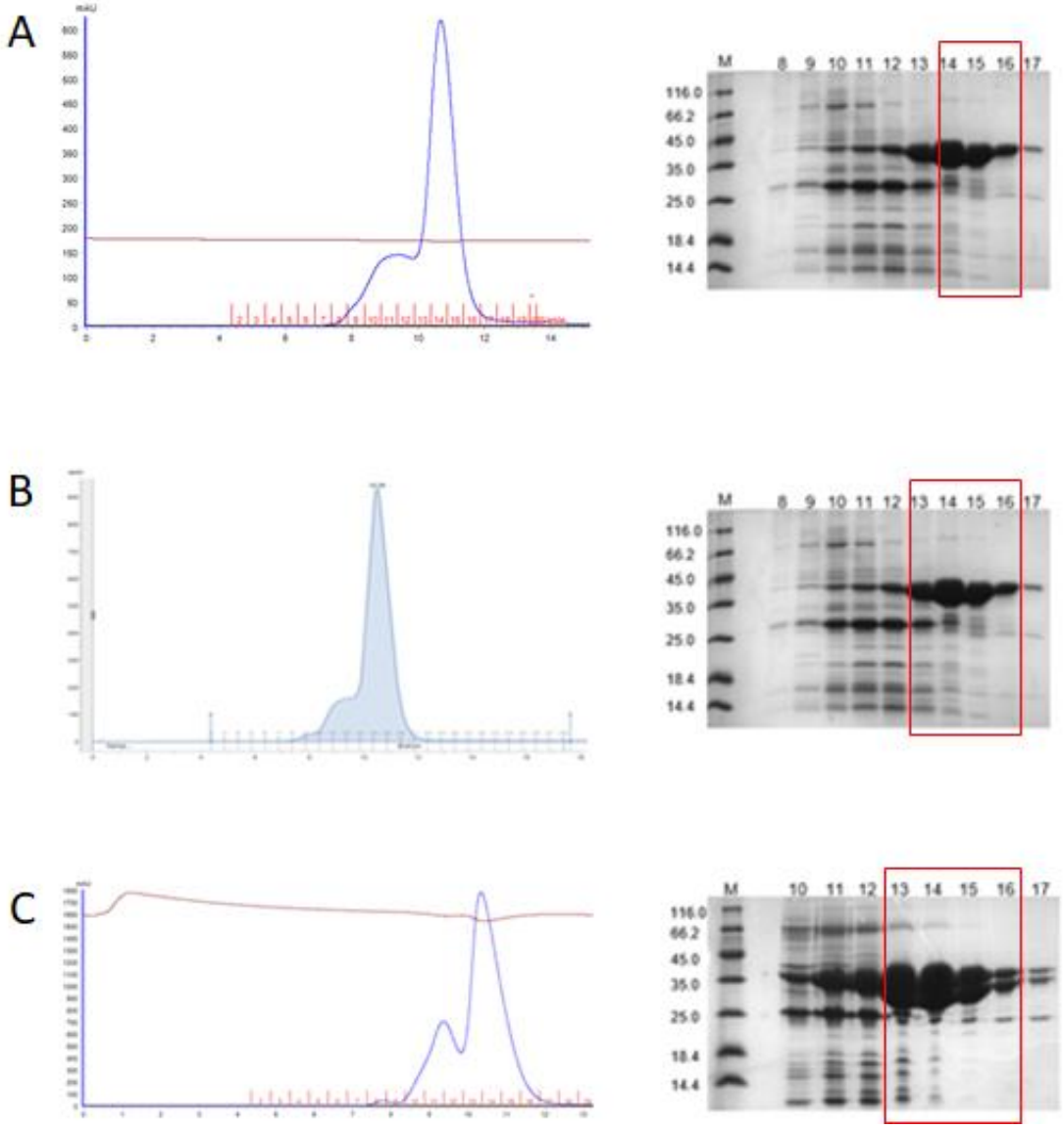

Figure 40. Elution curves of mIFP, iBlueberry-C18I and iBlueberry following gel filtration chromatography and resulting 15% SDS-PAGE gel analysis. (**A**) mIFP, (**B**) iBlueberry-C18I, (**C**) iBlueberry. A_{280} is shown in blue, the conductance ($mS.cm^{-1}$) in red. M denotes the molecular weight marker also run on the gels.

The fractions containing the proteins of interest were concentrated (14 to 16 for mIFP, 13 to 18 for iBlueberry-C18I and 13 to 16 for iBlueberry; red boxes in **Figure 40**), split into 50 µL aliquots, flash cooled with liquid nitrogen and stored at -80°C. A NanoDrop ND-1000 was used to quantify protein concentration (UV-vis / A_{280nm}).

Table 6. Concentration of NIR FPs purified in this study, concentration determination were done on a NanoDrop ND-1000 (UV-vis / A_{280nm})

Proteins	**Concentration (mg.mL^{-1})**
mIFP	18.3
iBlueberry-C18I	21.2
iBlueberry	21.4

2.1.3 His-tag cleavage and limited proteolysis

Previous studies (from Fanny Marzocca and Franck Borel) on a mutant of iBlueberry resulted in preliminary crystallisation conditions for this NIR FP. From these experiments, the his_6-tag was not specifically cleaved from the protein and, instead, the protein was subject to limited proteolysis to remove all the flexible regions that might hinder crystallisation (Dong *et al.*, 2007).

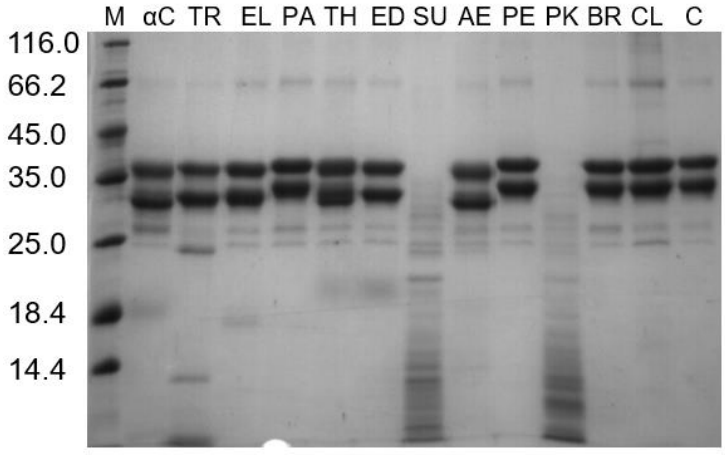

Figure 41. Limited proteolysis tests with iBlueberry..M= marker, αC= α-Chymotrypsin, TR=Trypsin, EL= Elastase, PA= Papain, TH= Thermolysin, ED= Endoproteinase, SU= Subtilisin (Na acetate trihydrate pH 7.5), AE= Actinase E, PE= Pepsin, PK= Proteinase K (Tris hydrochloride pH 7.5), BR= Bromelain, CL= Clostripain, C= Control.

To confirm that limited proteolysis would be an effective strategy for the NIR FPs studied during this thesis work, a test was conducted on iBlueberry using the Proti-Ace 2Kit (Hampton research). This kit includes 12 different proteases, including trypsin, and in the experiments carried out here, 10 µL proteinase at 0.01 mg/mL was mixed with 10 µL of protein stock solution at 5 mg/mL at room temperature for 1h and results visualised on a 15% SDS-PAGE gel (**Figure 41**). Treatment with α-chymotrypsin, trypsin, elastase, thermolysin, endoproteinase and actinase E produced a small shift in weight representative of limited proteolysis. Subtilisin and proteinase K for their parts completely digest the protein. From the six proteases that produce a positive results, trypsin was chosen as it is readily available in the lab for crystallisation trials and thus to produce protein solutions for crystallisation of the NIR FPs studied here, a 1/500 dilution from a 50 µg/µL trypsin stock solution was prepared and 10 µL of this dilution was used with 50 µL of protein stock at 20 mg/mL. The mixtures were kept for 2h at room temperature.

2.1.4 Initial crystallisation trials

Previous work from Fanny Marzocca and Franck Borel provided preliminary crystallisation conditions for the NIR FPs studied here (Yu *et al.*, 2015; Yu *et al.*, 2016). These conditions were optimised during the thesis work and the conditions obtained are reported here and were obtained at 293 K using either the hanging drop vapour diffusion method (Kam *et al.*, 1978; Martins *et al.*, 2008) for iBlueberry-C18I and mIFP or, in order to limit protein consumption, batch methods in PCR tube for iBlueberry (**Table 7**). Two protein stock solution : precipitant solution ratios (2:1; 1:1) were used in hanging drop experiments. Batch method experiments for iBlueberry used conditions comprising 20 µL of precipitant solution and 3 µL of protein at 21 $mg.mL^{-1}$. All plates were wrapped in aluminium foil to preserve the crystals from light and photobleaching. The size, shape and nucleation rate of the crystals were controlled by microseeding (see below).

2.1.5 Crystals optimisation

Table 7. Optimised crystallisation conditions for mIFP, iBlueberry-C18l and iBlueberry. These were further improved using microseeding and trypsin digestion. Protein concentrations were determined via absorbance at 280 nm using standard extinction coefficients.

Protein	Protein concentration (mg/mL)	pH	Conditions
mIFP	13	7.3 – 8.2	14-24% PEG3350 (w/v) 0.15 M malic acid 2:1 ratio (protein/condition)
iBlueberry-C18I	18	7.3 – 8.2	14-24% PEG3350 (w/v) 0.15 M malic acid 2:1 ratio (protein/condition)
iBlueberry	21	7.9	20-40% PEG3350 (w/v) 0.15 M malic acid 20 µL of condition and 3 µL of protein (batch)

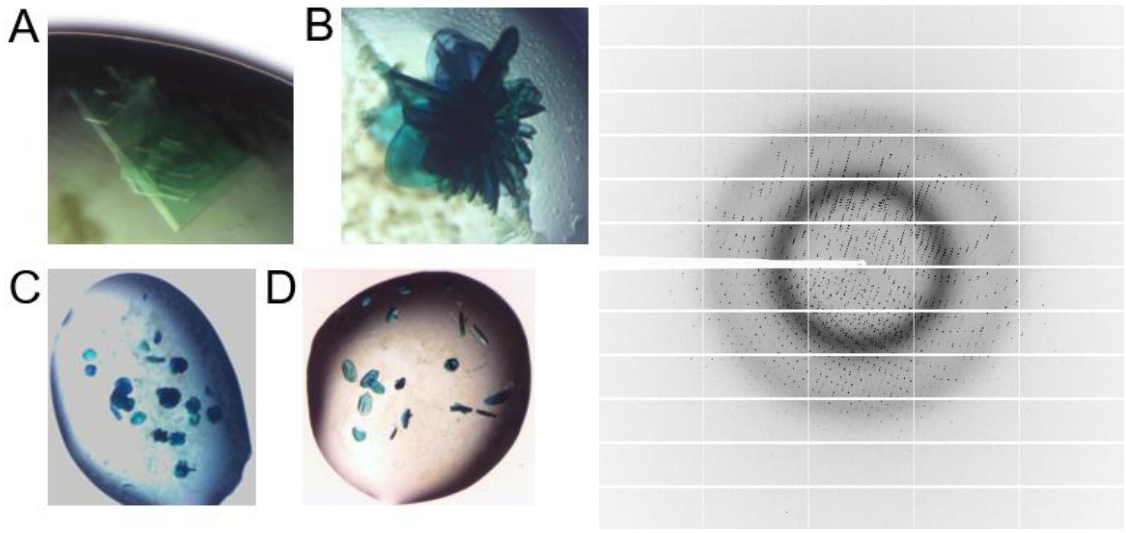

Figure 42. Crystals of iBlueberry-C18I and resulting diffraction pattern. (**A**) Crystals of iBlueberry-C18I after 2 months of crystallisation (first crystallisation assay, conditions shown in **Table 7**), this crystal was fished and used for diffraction data collection after which it was crushed to produce seeds for further crystals improvement. (**B**) First seeding cycle. (**C**) Second seeding cycle. (**D**) Third seeding cycle. Final crystallisation conditions are as shown in **Table 7** (14-24% PEG3350 (w/v), 0.15 M malic acid, pH 7.3 to 8.2, 2:1 ratio). A typical diffraction from the best crystals obtained is shown on the right

The first protein NIR FP to be crystallised in this work was iBlueberry-C18I which 2 months after setting up crystallisation drops produced multiple, interleaved flat-plate shaped crystals (**Figure 42A**). While these crystals were used for diffraction data collection, the diffraction pattern was not ideal. This crystal was used for data collection and then crushed to produce seeds. Several consecutive steps of seeding were carried out (**Figures 42B - 42D**) producing crystals with a much improved appearance and form.

Microseeding was used to improve the quality of the crystals obtained of mIFP and iBlueberry. For the microseeding procedure, crystals were harvested with 10 µL of the solution well and crushed using a 0.1 mL potter (Wheaton), 90 µL of precipitant solutions 1 to 2 % more concentrated than the original condition to avoid crystals dissolution was added. The crystals used for seeding solutions were chosen the smallest possible to avoid the accumulation of defects when crystals grow bigger. Seeding solutions were diluted at 1/10, 1/100, 1/1000, 1/10000 and 1/100000, each dilution was mixed with protein with a 1/10 ratio and used directly for hanging drop vapour dilution or batch methods.

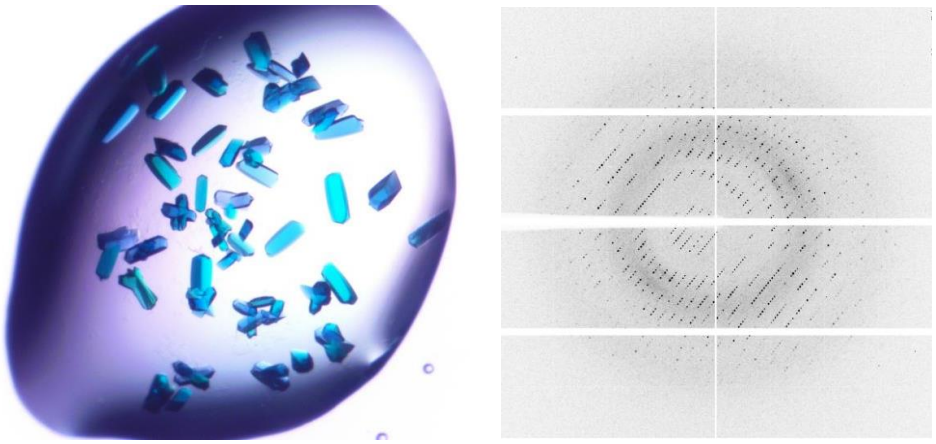

Figure 43. Crystals of mIFP using seeds from crystals obtained from the third seeding cycle of iBlueberry-C18I. Final crystallisation conditions are as shown in **Table 7**. A typical diffraction obtained from these crystals is shown on the right.

Crystals of mIFP diffracting at an acceptable resolution (**Figure 43**) were obtained from a single hetero-seeding cycle using crushed crystals of iBlueberry-C18I with the same crystallisation condition as for iBlueberry-C18I (**Table 7**).

A similar protocol was used in an attempt to produce crystals of iBlueberry using the hanging drop method. However, all attempts (changing the concentration of the protein stock, the pH, the concentration of the precipitant, testing different precipitants, a condition rescreen

of the protein or adding different additive to the condition) results in spherical or odd shaped crystals that did not diffract well (**Figure 44**, left).

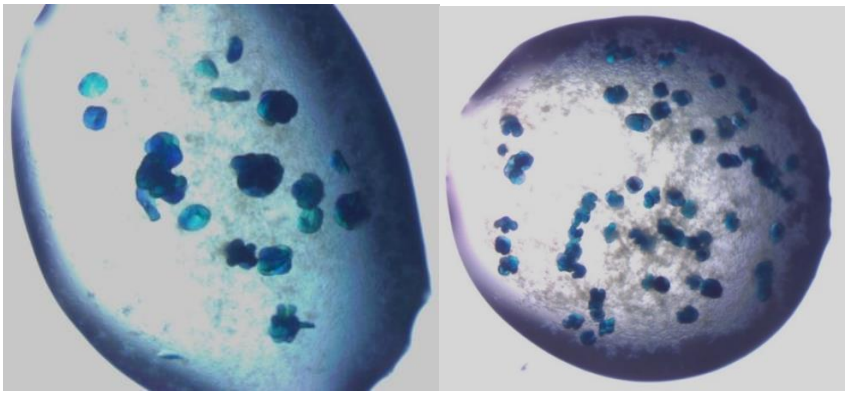

Figure 44. Left: Crystals obtained for iBlueberry using crushed crystals of iBlueberry-C18I as hetero seeds. Crystallisation protocol: PEG 3350 18%, malic acid 0.15 M pH 7.9, [Proteins] 21.4 mg/mL, hanging drop, vapour diffusion method. Right: Crystals of iBlueberry-C18I obtained in a crystallisation condition consisting of HEPES 0.15 M, PEG 3350 18%, pH 7.6, protein at 13 mg/mL, ratio 2:1, seeding 1/10000 from iBlueberry-C18I.

In addition to the above experiments, it was also checked whether microseeding would help produce crystals of iBlueberry-C18I in a different crystallization condition. Unfortunately, although crystals could be obtained (**Figure 44**, right), these showed very poor diffraction.

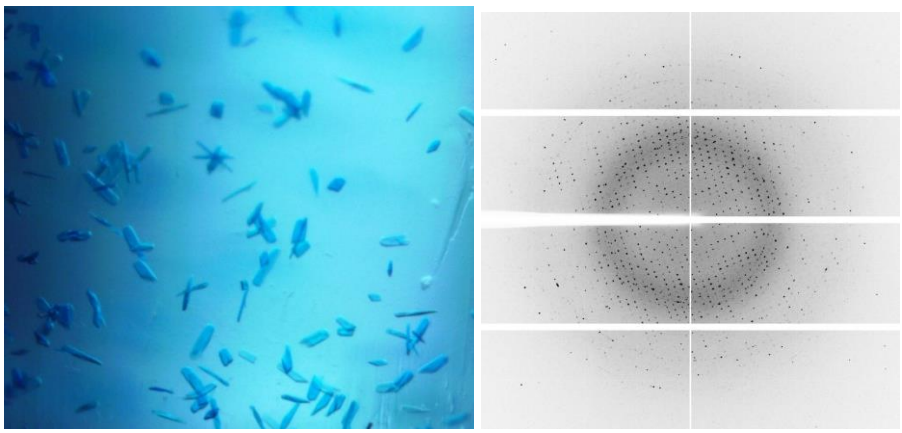

Figure 45. Crystals of iBlueberry obtained using batch method (3 months; See **Table 7** for crystallisation conditions). Microseeding (crushed crystals of iBlueberry seeds) and protein digestion by trypsin were required. A typical diffraction obtained from these crystals is shown on the right.

While the above hanging drop-based microseeding experiments yielded diffraction quality crystals for both iBlueberry-C18I and mIFP, this was not the case for iBlueberry. To solve this problem, different crystallisation methods were tested (free interface diffusion or batch

methods). Batch methods were successful for crystallisation of iBlueberry resulting in plate shaped crystals (**Table 7, Figure 45**). Here again though, microseeding and trypsin digestion of the protein were required to obtain the best crystals.

2.1.6 X-ray data collection

For diffraction data collection crystals of mIFP, iBlueberry and iBlueberry-C18I (**Figure 42, Figure 43, Figure 45**) were harvested and, for cryoprotection, soaked in a solution of the crystallisation precipitant supplemented with 20% (v/v) glycerol. All diffraction data were recorded at 100 K using the rotation method and exploiting monochromatic X-rays. Diffraction data from crystals of mIFP, iBlueberry and iBlueberry-C18I, were collected on beamline ID30A-3 (Von Stetten *et al.*, 2020) of the European Synchrotron Radiation Facility (ESRF, Grenoble). In particular, mIFP was collected using Helical Data collection (Flot *et al.*, 2010) to minimise radiation damage around the chromophore covalent bond to the protein.

X-ray diffraction intensities were integrated, scaled and merged using the *XDS* program package (Kabsch, 2010). The reduction of intensities to structure factors was carried out using *XDSCONV* (Kabsch, 2010) to convert reflection data into a format required by the CCP4 package (Winn *et al.*, 2011). The resolution limit for all data sets was determined using a combination of $CC_{1/2}$ (Evans and Murshudov, 2013) and $\langle I/\sigma(I)\rangle$ with the highest resolution bin chosen having $CC_{1/2}$ of ~0.70 and $\langle I/\sigma(I)\rangle$ close to 2.0.

2.1.7 Structures solution

The crystal structures of mIFP, iBlueberry and iBlueberry-C18I were solved by the molecular-replacement (MR) method using *Phaser* (McCoy *et al.*, 2007) with the crystal structure of *Rp*BphP1 from *Rhodopseudomonas palustris* ((4GW9 (Bellini *et al.*, 2012), ~ 60% of sequence identity) used as the search model. Residue replacement for the construction of the initial post-MR model was carried out with *ARP/WARP Classic* (Perrakis *et al.*, 1999). Waters molecules were placed with *ARP/WARP Solvent*. Initial refinement of the models obtained was carried out using alternating rounds of manual rebuilding with *Coot* (Emsley *et al.*, 2010) and *REFMAC5* (Murshudov *et al.*, 2011). Chromophores of each FPs were placed and initially fitted manually with *Coot* after a few rounds of refinement of the protein and link restrains were adjusted by manually modifying geometric restrains of the CIF files of the chromophores.

Structure validation was carried out using *Coot* (Emsley *et al*., 2010) and included analysis of Ramachandran plot, root mean square deviations from ideal bond lengths and angles, real space correlation coefficients, chiral volumes, unmodelled blobs, etc. The statistics for all final models can be found in **table 17** in chapter 5.

2.1.8 Recording of fluorescence excitation and emission spectra

UV-visible absorption spectra were recorded on a JASCO UV-vis spectrophotometer to estimate protein concentrations. Dilutions were performed in a buffer containing 20 mM TRIS pH 7.50 and 50 mM NaCl to reach an absorbance of circa 0.1 A.U. at the absorption peak of each chromophore.

Fluorescence excitation and emission spectra were recorded on a JASCO spectrofluorometer with an emission and excitation bandwidth of 5 nm, a scan speed of 100 nm per minute and a data interval of 0.5 nm. Each spectrum corresponds to an accumulation of five spectra. Emission spectra were measured using respectively 610 nm excitation wavelength for iBlueberry and iBlueberry C18I and 640 nm for mIFP. Excitation spectra were recorded using 710 nm for iBlueberry and iBlueberry C18I and 750 nm for mIFP.

2.2 GFP-like proteins

2.2.1 Plasmids construct

The team of Nathan Shaner from the Scintillon Institute for Biomedical and Bioenergy Research, San Digo, USA provided *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 genes (all from the Australian jellyfish (*Aequorea australis*)) cloned into pNCS expression vectors (Shaner *et al*., 2013, **Figure 46**) and sited between BamHI and EcoRI restriction sites. The pNCS vector is a derivative of a constitutively expressed plasmid in *E.coli*. It carries a SV40 replication origin to allow high expression levels in 293T cells, an N-terminal his_6-tag with an enterokinase site for cleavage, an X-press tag peptide for immuno-detection (Invitrogen), a linker with an enterokinase cleavage site and an ampicillin resistance gene to selection of colonies. As for the NIR FPs above, the codon usage of all three genes was to accommodate *E. coli* expression.

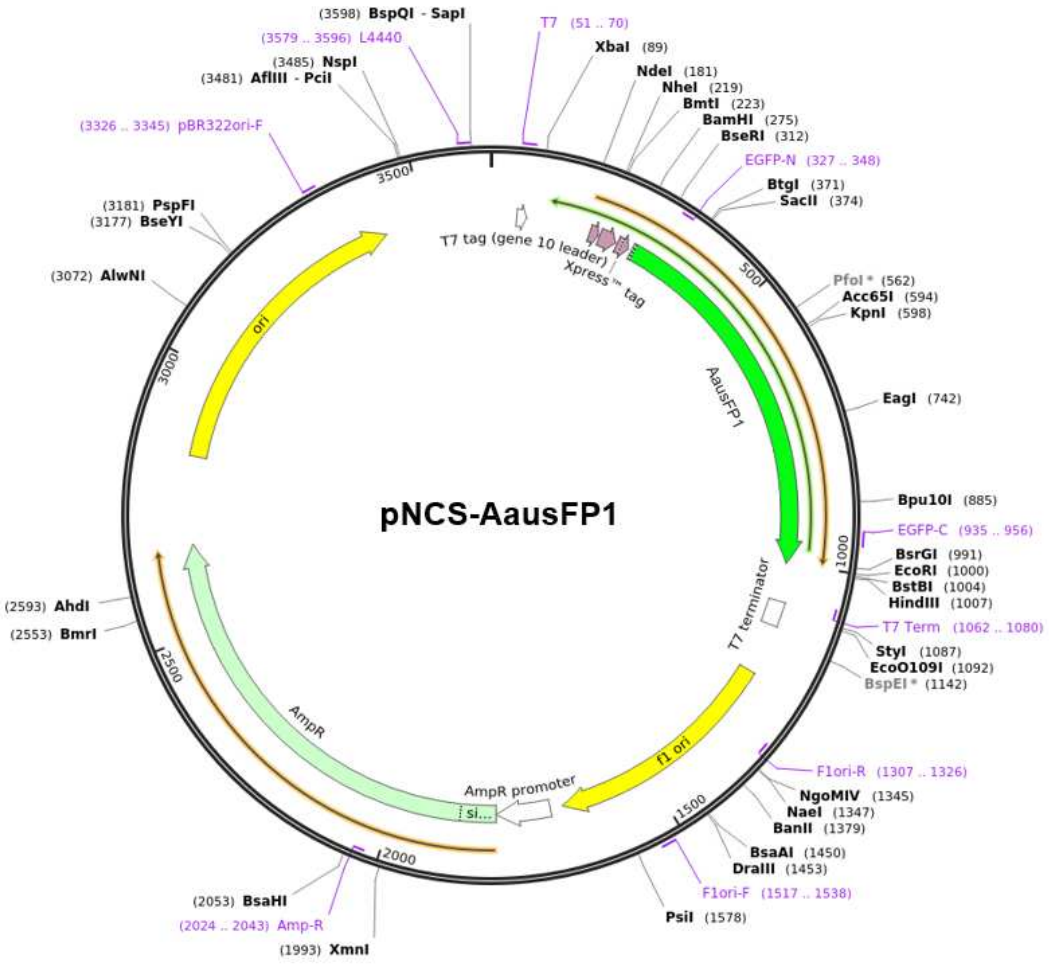

Figure 46. A pNCS expression vector, a constitutive plasmid containing *Aaus*FP1, an ampicillin resistance gene to select cells who integrated the plasmid, a 6xHis tag, an enterokinase cleavage site and an X-press tag (for immuno-detection) on the N-terminus of the protein.

2.2.2 Expression and purification

Plasmids containing *AausFP*1, *Aaus*FP2 and *Csi*FP4 were transformed in *E. coli* BL21 cells (Invitrogen) on agar plates and incubated overnight at 37°C with ampicillin for selection (100 $\mu g.mL^{-1}$). Protein expression was then carried out in 2 L of TB medium using a similar protocol as described for NIR FPs (§ 1.1.2). A major difference here was that arabinose was not added due to the constitutive expression of proteins by the pNCS vector. For *Aaus*FP1 and *Csi*FP4 cells were grown overnight at 37°C while for *Aaus*FP2 cells were incubated overnight at 17°C due to poor expression levels at 37°C (**Figure 47A**, middle flasks).

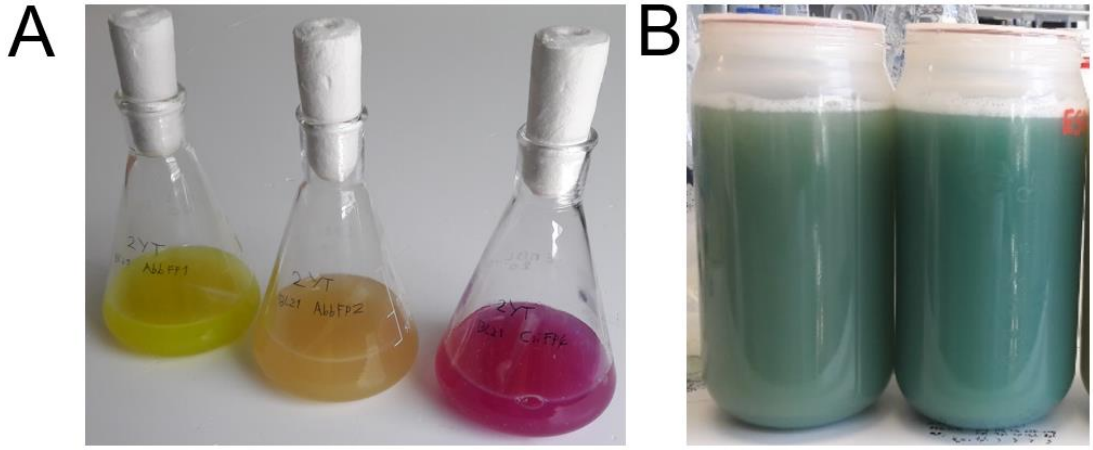

Figure 47. Expression of *Aaus*FP1, *Aaus*FP2 and *Csi*FP4. (**A**) Pre-cultures (2YT media) incubated at 37ºC of, from left to right, *Aaus*FP1, *Aaus*FP2 and *Csi*FP4. *Aaus*FP2 (middle flask) did not express at 37ºC (**B**) Culture of *Aaus*FP2 after testing different growth conditions were grown at 17°C overnight.

Very similar protocols as for the NIR FPs studied here were also used for the purification of *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 (**Figure 48**). However, a problem encountered during purification was that the high affinity of *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 on the nickel column which necessitated the use of 1M of imidazole for elution of the proteins directly followed by a step of dialysis to eliminate it.

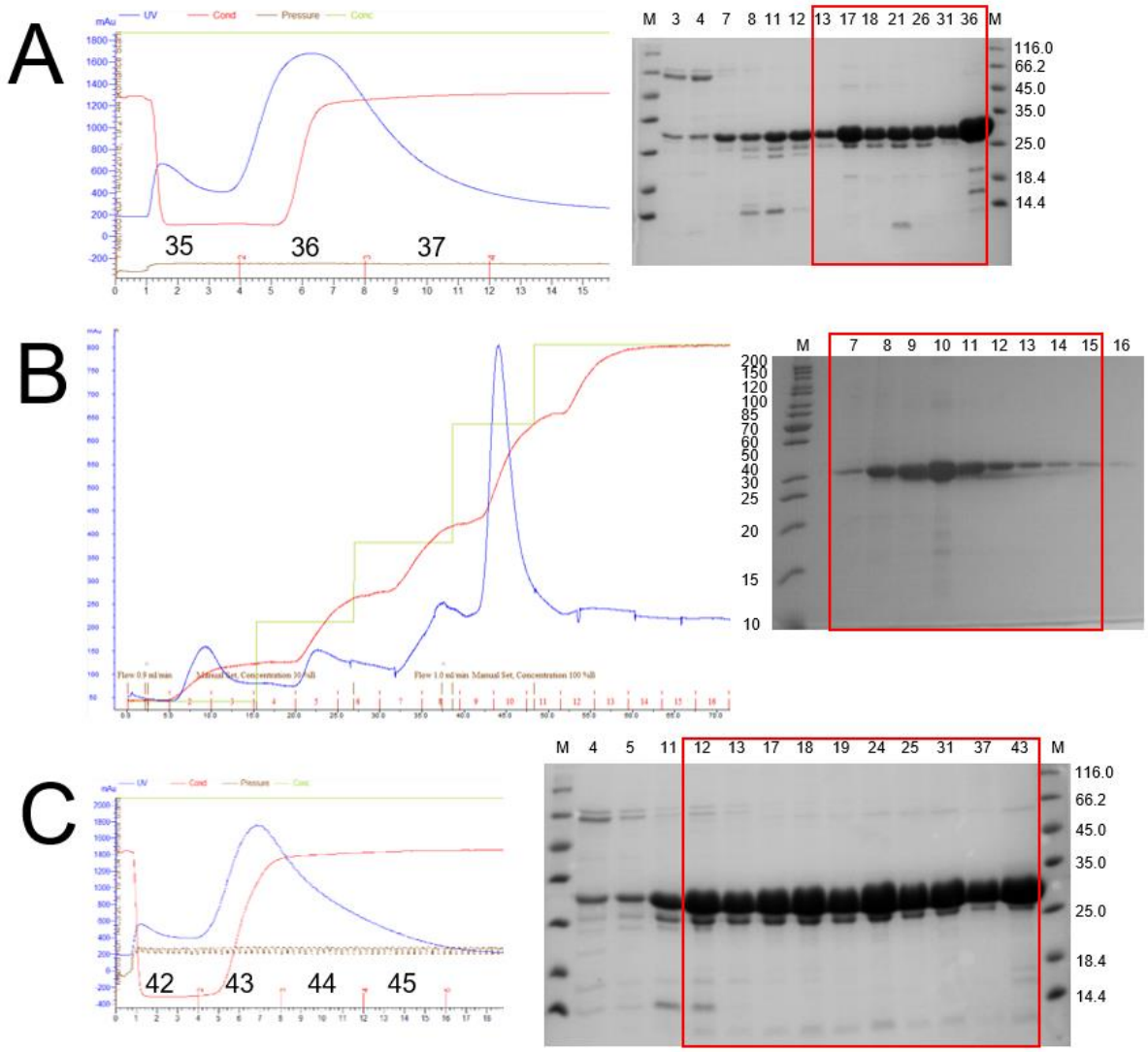

Figure 48. Elution curves of *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 following nickel affinity chromatography and resulting 15% SDS-PAGE gel analysis. (**A**) *Aaus*FP1, (**B**) *Aaus*FP2, (**C**) *Csi*FP4. Absorbance at 280 nm (A_{280}) is shown in blue, the conductance (mS.cm-1) is shown in red. The green line tracks the concentration of buffer B (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole (**Table 5**)) loaded on the column. M denotes the molecular weight marker also run on the gels. Note that for *Aaus*FP1 and *Csi*FP4 only the latter part of the elution profiles are shown due to the high concentrations of imidazole required for the elution of the proteins. For both these samples the proteins did not completely elute from the column.

Fractions containing the desired protein (13 to 16 for *Aaus*FP1, 7 to 15 for *Aaus*FP2, and 12 to 45 for *Csi*FP4; red boxes in **Figure 48**) were pooled and dialysed (Spectra/Por, Dialysis Membrane, Standard RC Tubing, MWCO: 6-8 kD) against buffer D (50 mM Tris pH 8.0, NaCl 50 mM; **Table 5**) to eliminate imidazole. The buffer was replaced twice after 1h and the proteins was left overnight at 4°C to equilibrate. The resulting protein solutions were then concentrated using an Amicon ultra concentrator (Merck Millipore, molecular weight cut off: 10 kDa).

The purification procedure continued by a gel filtration step (Superdex 75-10/300 GL columns, GE HealthCare) on a GE HealthCare AKTApure system and buffer C (20 mM Tris pH 8.0, **Table 5**) as the eluent. All gel filtration runs were performed with a flow rate of 0.8 mL/min and 0.5 mL fractionation. Fractions were analysed with 15% SDS-PAGE (**Figure 49**).

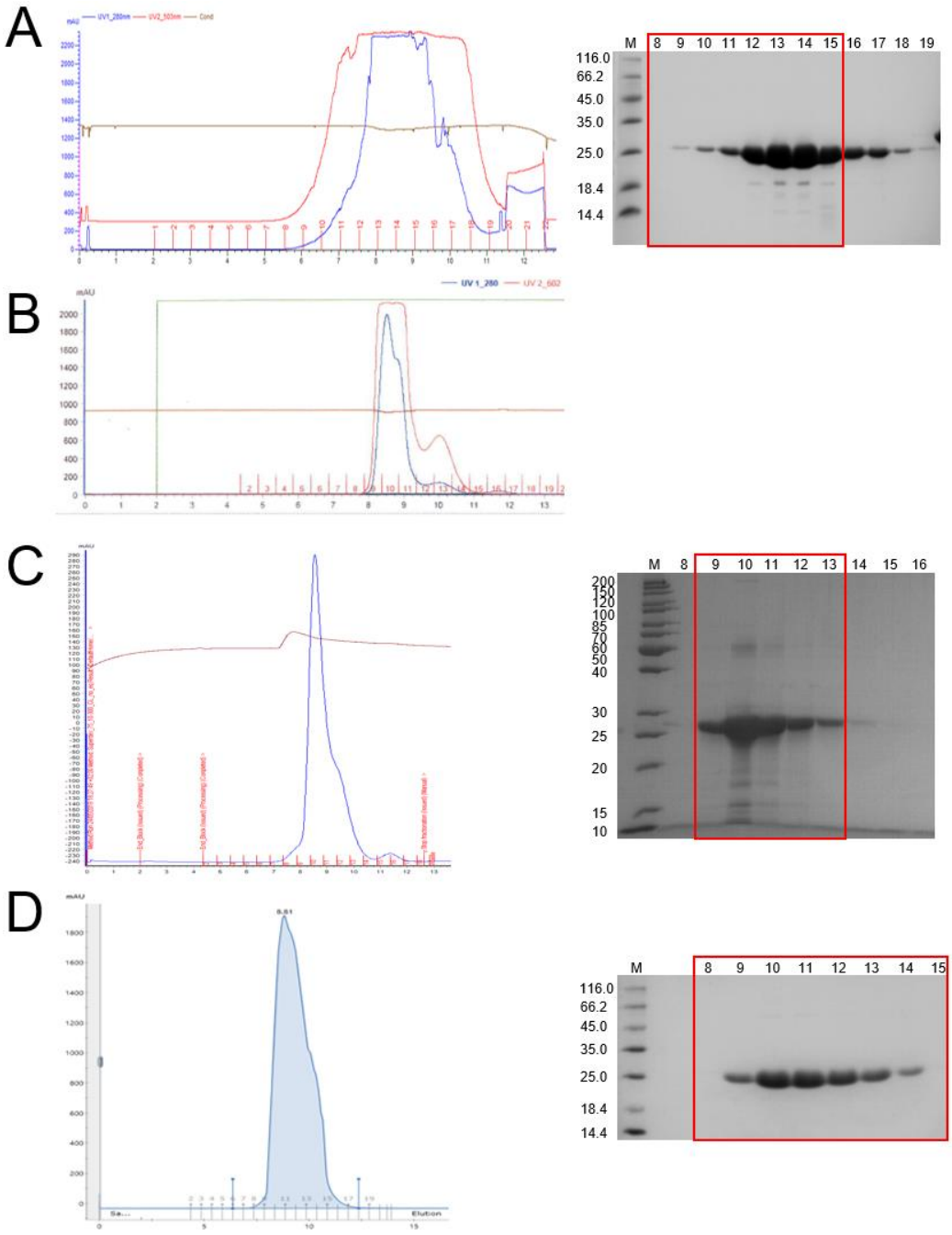

Figure 49. Elution curves of *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 following gel filtration chromatography and resulting 15% SDS-PAGE gel analysis. (**A, B**) *Aaus*FP1, (**C**) *Aaus*FP2, (**D**) *Csi*FP4. A280 is shown in blue, the conductance (mS.cm-1) in red. M denotes the molecular weight marker also run on the gels. Figure B is the same as figure A but with less protein injected and after flushing the AKTA system of the bubbles it contain.

The fractions containing the proteins of interest (8 to 15 for *Aaus*FP1, 9 to 13 for *Aaus*FP2 and 8 to 15 for *Csi*FP4) were combined, concentrated (see previous section), split into 50 µL aliquots, flash cooled with liquid nitrogen and stored at -80°C. A NanoDrop ND-1000 was used to quantify protein concentration (UV-vis / A_{280nm}).

Table 8. Concentration of GFP-like proteins purified in this study. Concentration determination were done on a NanoDrop ND-1000 (UV-vis / A_{280nm}).

Proteins	**Concentration (mg.mL^{-1})**
*Aaus*FP1	49
*Aaus*FP2	51
*Csi*FP4	39

2.2.3 His-tag cleavage

This step was carried out between the first nickel column chromatography (GE Healthcare HisTrap HP, 5mL) and the gel filtration chromatography (Superdex 75-10/300 GL columns, GE HealthCare). As proteins expressed via the pNCS vector contain a TEV protease site for the cleavage of the his_6 tag with which proteins are labelled, incubation with TEV protease was used in a first trial to remove the his_6 tag with which *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 were expressed. A 1: 100 (TEV : protein) ratio was used with incubation carried out overnight and at room temperature. As can be seen from **Figure 50**, TEV removal of the his_6 tag from *Csi*FP4 was successful. However, this was not the case for *Aaus*FP1. On another note, the small autocatalytic cleavage of the His-tag can be explained by either naturally occurring protease in the medium or by the fact that the his-tag is inherently unstable.

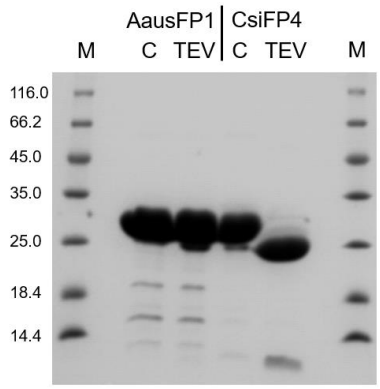

Figure 50. SDS-PAGE gel analysis of his6 tag removal (overnight incubation with TEV protease) for *Aaus*FP1 and *Csi*FP4. The protocol does not remove the his6 tag of *Aaus*FP1.

In a similar fashion to the experiments carried out on NIR FPs above (§ 1.1.3) the Proti-Ace 2Kit (Hampton research) kit was used to assay the best means of removing the his6 tag from *Aaus*FP1 depending on the cleavage site recognised by the protease. The size of the protease and its structure (location of the catalytic site) can also play a role in the recognition of the cleavage site. These suggested that his_6 tag removal could be affected using, in particular Subtilisin or Proteinase K (**Figure 51A, B**) and after a series of further trials *Aaus*FP1 his_6 tag removal was performed by incubation with Proteinase K for 3 h at 4 °C (**Figure 51C**).

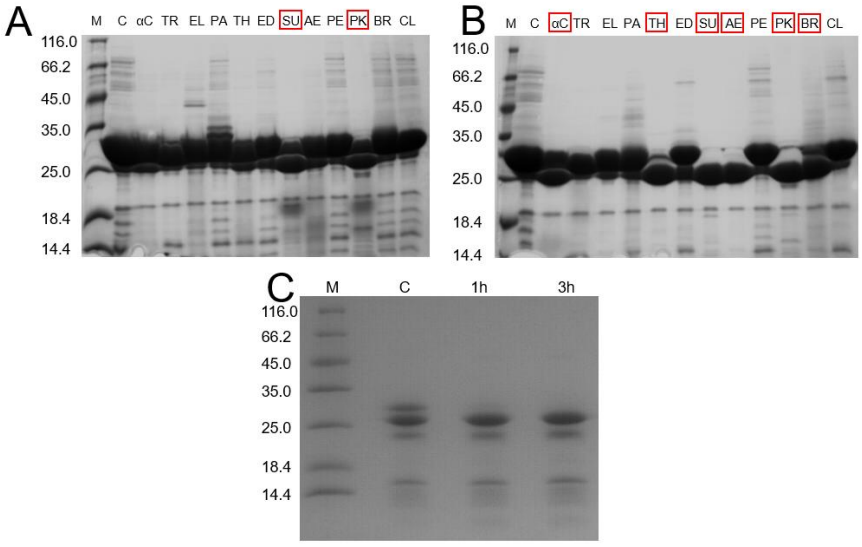

Figure 51. Limited proteolysis assay for his6 tag removal from *Aaus*FP1. (**A**) Assay after 1h digestion at room temperature (**B**) Assay after 70h digestion at room temperature (10 µL proteinase at 0.01 mg/mL + 10 µL protein at 5 mg/mL). C= Control, αC= α-Chymotrypsin, TR=Trypsin, EL= Elastase, PA= Papain, TH= Thermolysin, ED= Endoproteinase, SU= Subtilisin (Na acetate trihydrate pH 7.5), AE= Actinase E, PE= Pepsin, PK= Proteinase K (Tris hydrochloride pH 7.5), BR= Bromelain, CL= Clostripain. (**C**) Incubation of *Aaus*FP1 with Proteinase K after 1h and 3h at 4 °C (C= Control without digestion, which reveals an autocatalytic cleavage of the his6-tag.

*Aaus*FP2 is also resistant to TEV digestion and was subject to the same assay as *Aaus*FP1. Here again the best conditions for his_6 tag removal proved to be incubation with proteinase K after 3h at 4 C.

Following successful his_6 tag removal, the last purification step for *Aaus*FP1, *Aaus*FP2 was to load the protease-incubated solutions on a Ni-affinity column (50 mM Tris pH 8.0, 300 mM NaCl). The flow-through was collected and concentrated using an Amicon ultra concentrator (Merck Millipore, molecular weight cut off: 10 KDa) washing the protein solution in buffer C to remove the salt (20 mM Tris pH 8.0). Gel filtration followed this step (see previous section).

2.2.4 Initial crystallisation trials

Initial crystallisation conditions for the GFPs studied here were determined using robotic screening carried out at the HTX lab platform (EMBL Grenoble, Dimasi *et al.*, 2007, htxlab.embl.fr). Three different screens were used: *JCSG+* (Molecular Dimensions); *Wizard_I&II* (Rigaku Reagents) and *Classics-Suite* (Qiagen/Nextal, Jancarik and Kim 1991). Crystallisation drops were set up in Crystaldirect low profile crystallisation plates (CrystalDirect Plate, MiTeGen, Cipriani *et al*., 2012) which are highly transparent to UV and visible light (to allow for easy drop inspection) and designed to produce a low X-ray background in the case, not needed here, that *in situ* data collection is required. At the HTX platform, pictures of crystallisation drops are automatically taken after 1, 3, 7, 15, 33, 61 and 87 days after setup (visible or UV light) and the crystallisation plates can be stored at 20 ^{o}C or 4 ^{o}C.

An interesting feature available at the HTX platform is automated crystal harvesting and cryocooling using CrystalDirect (Cipriani *et al.*, 2012; Márquez and Cipriani 2014). Here, crystals to be mounted are selected through the HTX platform's CRIMS website (**Figure 52**, htxlab.embl.fr). Crystallisation conditions for *Aaus*FP1, *Aaus*FP2 and *Csi*FP4 can be found in **Table 9**. *Csi*FP4 proved difficult to crystallise, with only one of the conditions tested yielding crystals and this after two and a half months of growth and trying for three month to optimise the crystallisation conditions that yielded amorphous precipitate in the initial screens. The optimisation involved the testing of different pH and concentrations of precipitate and salts, the use of different additives kits and the testing of different crystallisation methods (hanging drop, siting drop, and batch methods). The crystals of the initial crystallisation screen were mounted onto pins for diffraction data collection using CrystalDirect (**Figure 52**). Attempts were made to manually optimise this condition but no crystals were obtained.

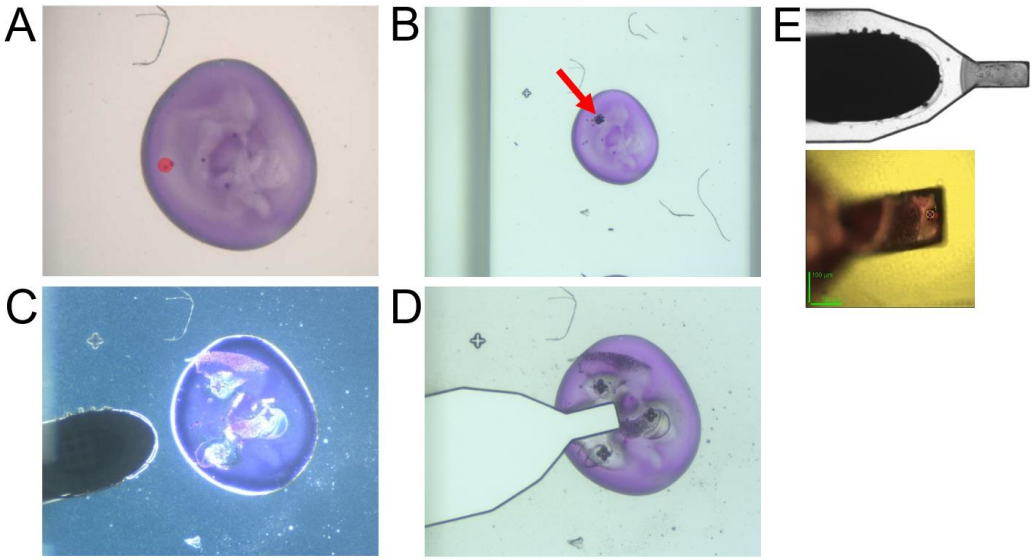

Figure 52. Automatic harvesting at the HTX platform of a *Csi*FP4 crystal after two and a half months of growth. (**A**) Picking of the crystals via the CRIMS software. (**B**) Removal of the crystallisation mother liquor by puncturing the crystallisation plate membrane with a syringe and aspirating the liquid. (**C**) Fixation of the pin on the membrane using a laser to melt the plastic. (**D**) Cutting of the shape to retrieve crystals. (**E**) Final pin, after automatic flash cooling in liquid nitrogen, with the crystal used to collect diffraction data for *Csi*FP4.

2.2.5 Crystals optimisation

Table 9. Crystallisations conditions of GFP-like proteins. Protein concentrations were determined at 280 nm with standard extinction coefficient.

GFP	Stock solution concentration ($mg.mL^{-1}$)	pH	Crystallisation Method	Conditions
***Csi*FP4**	39	8.5	Sitting drop	1 M lithium sulfate 0.1 M tris 0.01 M nickel chloride 3:1, 2:1 and 1:1 ratio
***Aaus*FP1**	40	6.5 – 8.0	Hanging drop	0.7-1.3 M trisodium citrate 0.2 M sodium chloride 0.1 M of Tris buffer 3:1, 2:1 and 1:1 ratio
***Aaus*FP2**	51	7.3 – 8.2	Hanging drop	14-24% PEG 3350 0.2 M NaCl 0.1 M Hepes 3:1 ratio

*Aaus*FP2 gave crystals in multiple conditions. However, most of these showed poor diffraction and manual optimisation was required to produce plates-shaped crystals showing diffraction patterns at a good resolution (**Figure 53**).

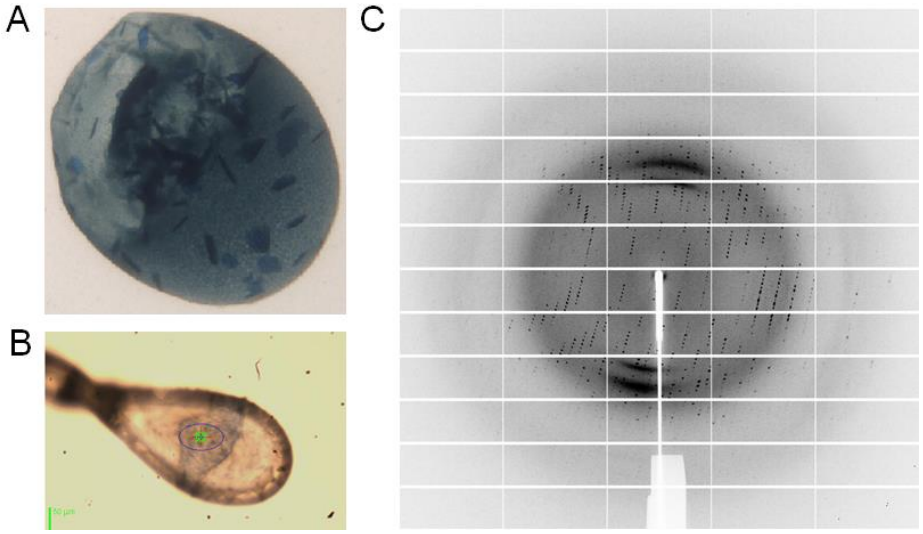

Figure 53. (**A**) Crystals of *Aaus*FP2 after condition optimisation (see **Table 9**) and one week of growth. (**B**) A single crystal mounted in a nylon loop and cryocooled was used for data collection. (**C**) An example of the diffraction images obtained.

Well diffracting crystals of *Aaus*FP1 were difficult to obtain. While the protein crystallised in less than one day in almost every condition tested, the crystals obtained were very soft and showed poor diffraction. Crystals exhibiting diffraction patterns at a good resolution were obtained only after manually optimising, including microseeding of one of the rare conditions that took one week to crystallise (**Table 9, Figure 54**).

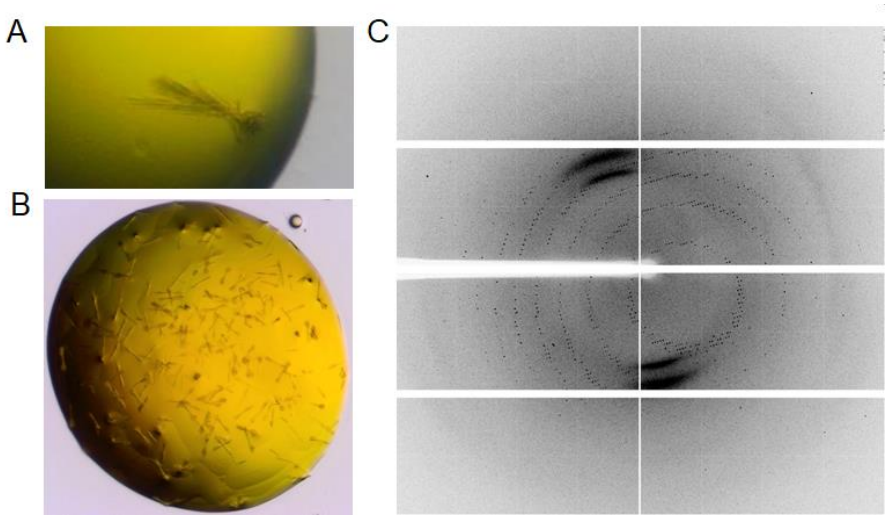

Figure 54. (**A**) Initial crystals obtained for *Aaus*FP1 after 1 week of growth. (**B**) Crystals after condition optimisation using seeding. (**C**) A typical diffraction pattern obtained from the crystals shown in **(B)**.

2.2.6 X-ray data collection

For diffraction data collection, crystals of *Aaus*FP1 were harvested and, for cryoprotection, soaked in a solution of the crystallisation precipitant supplemented with 20% (v/v) glycerol before cryocooling in liquid nitrogen. Crystals of *Aaus*FP2 were frozen directly from the crystallisation drop with no extra cryoprotectant added. Crystals of *Csi*FP4 were harvested using CrystalDirect (see above) and no cryoprotectant was used before the cryocooling of crystals. Diffraction data from crystals of *Aaus*FP1 were collected at ALBA beamline Xaloc (Juanhuix *et al*. 2014) and data for crystals of *Csi*FP4 were collected on beamline ID30A-3 of the European Synchrotron Radiation Facility (ESRF, Grenoble, Von Stetten *et al*., 2020). Diffraction data from crystals of *Aaus*FP2 were collected at ESRF beamline ID30B (McCarthy *et al.*, 2018).

As for the NIR FPs studied here (§1.1.5), X-ray diffraction intensities were integrated, scaled and merged using the *XDS* program package. The reduction of intensities to structure factors was carried out using *XDSCONV* (Kabsch, 2010) to converts reflexion data into a format required by the CCP4 package (Winn *et al.*, 2011). The resolution limit for all data sets was determined using a combination of $CC_{1/2}$ and $\langle I/\sigma(I)\rangle$ with the highest resolution bin chosen having $CC_{1/2}$ of ~0.70 and $\langle I/\sigma(I)\rangle$ close to 2.0 (see chapter 5 for statistics tables).

2.2.7 Structure solution

The crystal structures were solved by MR using *Phaser*. For *Aaus*FP1 and *Aaus*FP2, the crystal structure of yellow fluorescent protein from *Phialidium sp.* (phiYFP, 4HE4 (Pletneva *et al*., 2013) was used as the search model. For *Csi*FP4, the search model was the crystal structure of green fluorescent protein from *Clytia gregaria* (PDB entry code: 2HPW, Liu et al., "Crystal Structure of Green Fluorescent Protein from *Clytia Gregaria* at 1.55 Å resolution", to be published. For all three crystals structures, residue replacement for the construction of the initial model following MR was carried out in *Coot*. Structure improvement was then carried out with alternating rounds of manual rebuilding in *Coot* and refinement with *REFMAC5*. Waters were placed using *Coot:findwaters.* In order to improve the refinement process TLS (Translation/Libration/Screw) protocols were used (Schomaker & Trueblood, 1968; Painter & Merritt, 2006).

Structure validation was carried out using Coot (Emsley *et al*, 2010) and included analysis of Ramachandran plot, r.m.s. deviations from ideal bond lengths and angles, real space correlation coefficients, chiral volumes, unmodelled blobs, etc. Full details of the final models

obtained can be found in statistics tables in chapter 3 and 4. Estimation of buried surfaces was carried out using the *PISA* extension in *WinCoot* (Krissinel and Henrick 2005, 2007; Emsley *et al*., 2010).

Chapter 3

RESULTS

Structure determination of the bright green fluorescent protein *Aaus*FP1 and of the blue chromoprotein *Aaus*FP2 from *Aequorea australis*: discovery of a novel type of chromophore

3.1 Introduction

This chapter summarises the results contributing to the article '*Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing*' (Lambert *et al.*, 2020), the full text of which is appended at the end of the thesis manuscript. The work results from a large collaboration led by Prof. Nathan C. Shaner (from UC San Diego in La Jolla, (California), previously from the Scintillon Institute in San Diego (California)), including groups from the Birch Aquarium at Scripps in La Jolla (California), the Monterey Bay Aquarium Research Institute (California), Harvard Medical School in Boston (Massachusetts), the ESRF and the IBS in Grenoble (France), the Université Gustave Eiffel in Marne-la-Vallée (France) and the Western Sydney University in New South Wales (Australia).

This article reports the identification, cloning and characterization of five FPs homologous to GFP from *Aequorea Victoria* (see **§1.1** and **§1.2**) in the jellyfish *Aequorea* cf. *australis*, a blue jellyfish harvested at night during an expedition to Heron Island on the Great Barrier Reef in Australia, on which a marine research station is located. The presence of these FPs was hinted by the cyan-blue coloration of the radial canals of the collected specimen suggesting the presence of red-absorbing CPs (**Figures 55Aiii & 55Aiv**). The transcriptome of the animal was constructed after mRNA extraction, sequenced and the sequence of five *Av*GFP homologues were identified by BLAST homology searching.

The discovery of these homologues spurred the San Diego team to investigate a sample of *A. victoria* (**Figures 55Ai & 55Aii**) present at the Birch Aquarium at Scripps in La Jolla, in order to determine whether this species contained additional FPs. Indeed, the sequencing of the transcriptome revealed unsurprisingly the presence of *Av*GFP, then, as anticipated, those of the *Aaus*FPs orthologs, but surprisingly, also that of a distinct bright green-emitting FP with properties close to those of EGFP.

In brief, a total of nine homologues to *Av*GFP were discovered in *A. victoria* and *A.* cf. *australis* (*Avic*FP1/2/3/4, *Aaus*GFP and *Aaus*FP1/2/3/4) (**Figure 55B**) which belong to three broad classes: green-emitting FPs (*Avic*FP1, *Avic*FP4, *Aaus*GFP (a close equivalent to *Av*GFP with double peak excitation), *Aaus*FP1), long-wavelength-absorbing CPs (*Avic*FP2, *Avic*FP3, *Aaus*FP2, *Aaus*FP3) and a reversibly photoswitchable CP (*Aaus*FP4).

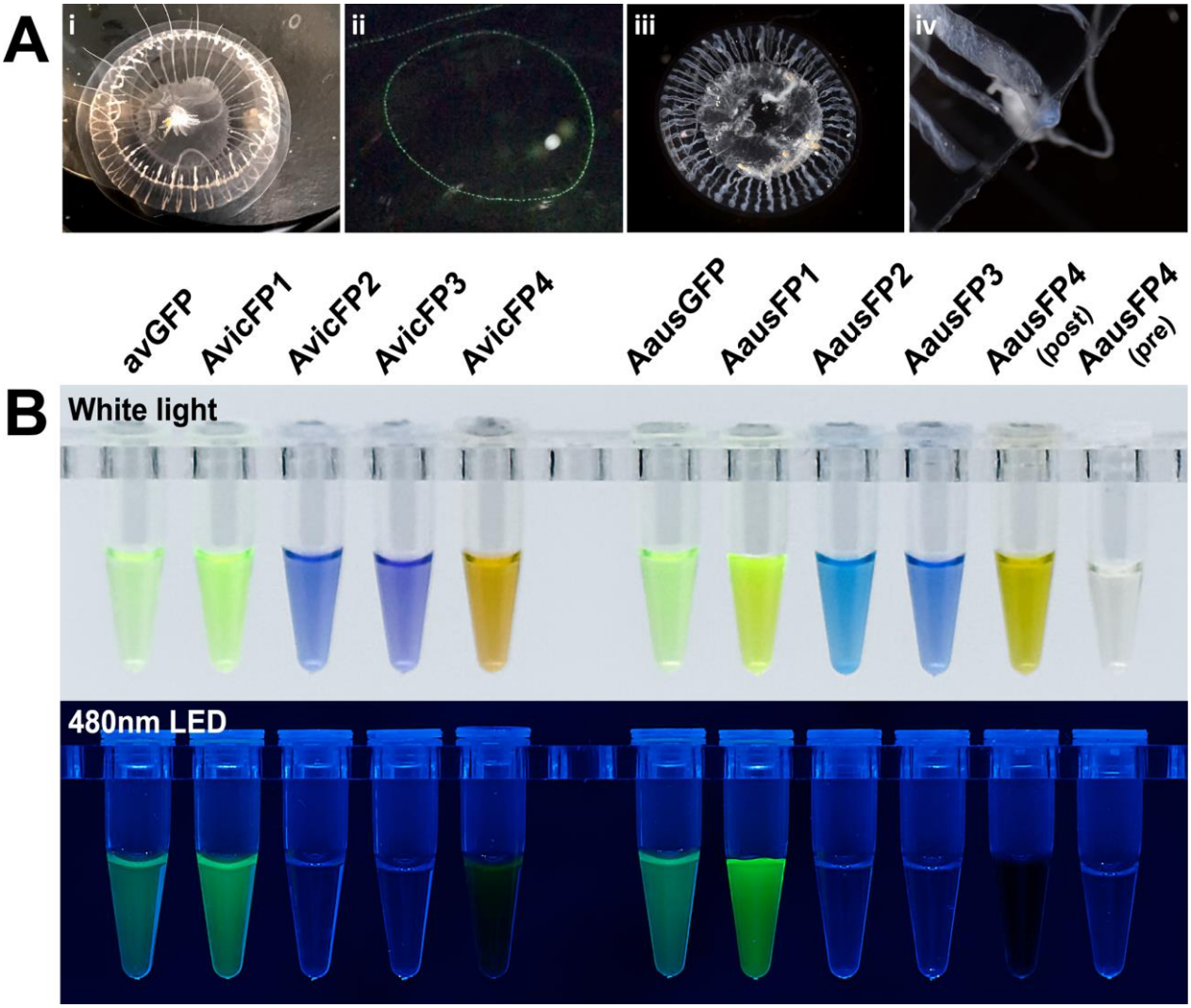

Figure 55. Identification of previously unknown FPs in the two jellyfish species *Aequorea victoria* and *Aequorea* cf. *australis*. (**A**) White light photographs of *A. Victoria* (**i**), *A.* cf. *australis* (**iii**) and a close-up on one tentacle bulb of the latter, showing its blue coloration (**iv**). A fluorescence photograph of A. Victoria (**ii**) was obtained using a violet LED (peaking at 400 nm) and showed green fluorescence in the bell margin. (**B**) Purified recombinant proteins from *Av*GFP the nine newly-discovered FPs shown under white light (top) and unfiltered blue LED (peaking at 480 nm) illumination. The fifth FP from *A.* cf. *australis* is shown (post) and before (pre) exposure to UV light (reproduced from (Lambert *et al.*, 2020)).

The goal of my work has been to determine the structure of two of these nine proteins, *Aaus*FP1 and *Aaus*FP2, because *Aaus*FP1 is the brightest FP of them all (approximately 5-fold brighter than EGFP and 2-fold brighter than mNeonGreen). *Aaus*FP2 is the CP with the most red-shifted absorption peak. The crystallographic structure of these two proteins would provide (1) the structural determinants of bright fluorescence, which could eventually be passed on to other FPs and (2) structural information on a close to far-red absorbing chromophore, which could lead to the design of a far-red FP. *Aaus*FP1 and *Aaus*FP2 share

54% sequence identity. Their sequences (**Figure 56**) indicate that *Aaus*FP1 has the TYG chromophore of EGFP (for the record the S65T mutation was engineered in *Av*GFP to make it brighter, so this is another observation that a man-designed mutation had already been produced by nature previously, as the T203Y mutation leading to the YFP class of FPs (see **§1.5.2.d**) and that *Aaus*FP2 possesses a rather infrequent AYG chromophore, which has only been found naturally in four anthozoan genera (Haddock *et al.* 2010) and in an hydrozoan, the jellyfish *Rhacostoma atlantica* (Tota *et al*, 2016).

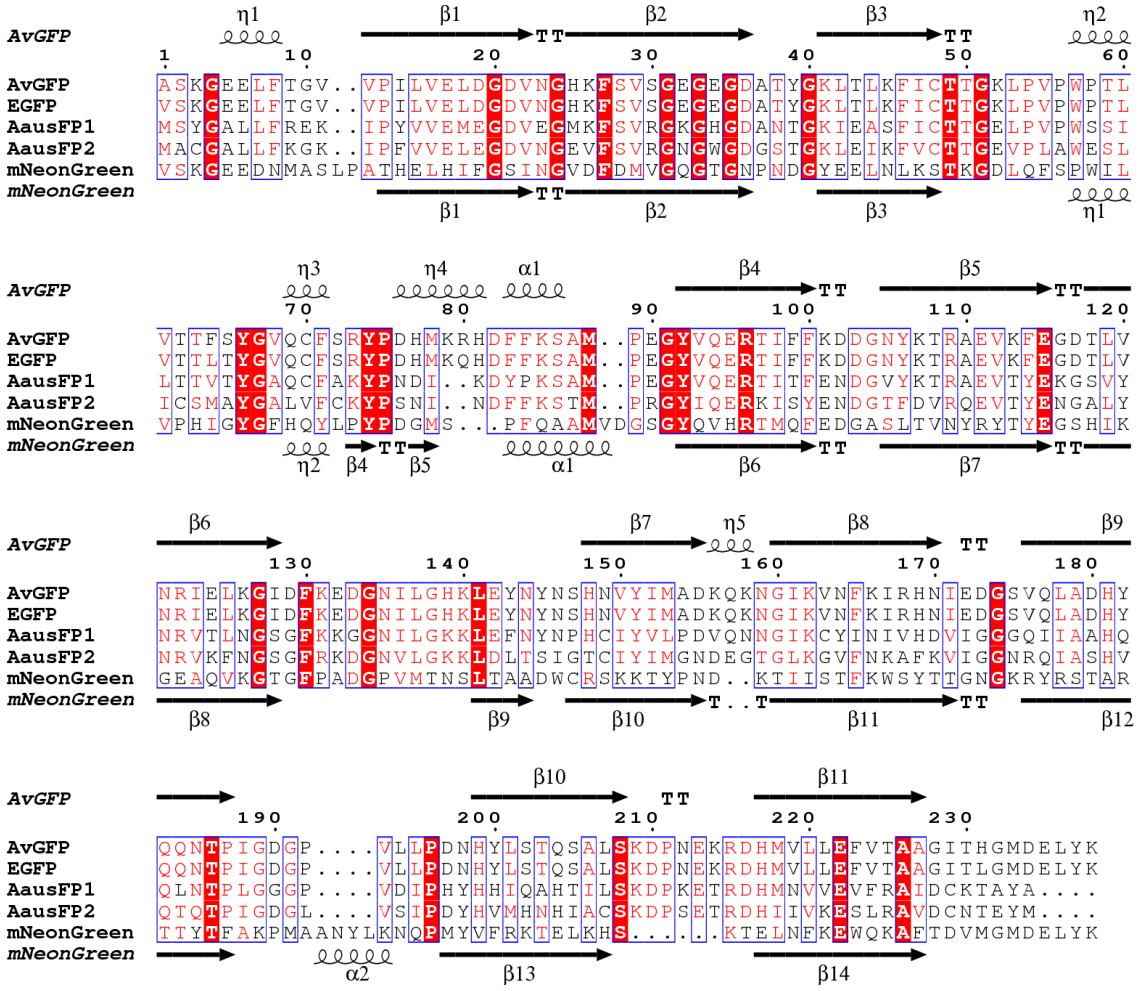

Figure 56. Sequence alignment of *Aaus*FP1 and *Aaus*FP2 with the very well-known *Av*GFP and EGFP and the very bright GFP mNeonGreen from *Branchiostoma Lanceolatum*. Secondary structure elements (α-helix, β-sheet, turn) are indicated for the *Av*GFP structure (top) (PDB entry code: 1GFL) and for the mNeonGreen structure (bottom) (PDB entry code: 5LTR). The sequence alignment was prepared with Clustal Omega (Madeira *et al*, 2019) and was prepared with ESPript (Robert & Gouet, 2014).

3.2 Spectroscopic properties

*Aaus*FP1 is a green fluorescent protein that is much brighter than other proteins from the same spectral class (**Table 10**). It presents very narrow excitation and emission peaks, as can be visualized from the comparison with the already bright EGFP (**Figure 57A**). These characteristics are particularly valuable when it comes to choose a set of narrow excitation and emission band-pass filters for the fluorescence microscope, possibly allowing for the increase of spectral multiplexing probes in a multicolour imaging experiment, by simultaneously using more fluorescent probes, five at least. The narrowness of the spectra has to be correlated to the optimal fluorescence properties of the chromophore, which includes a very short Stokes shift (6 nm *vs.* 11 nm for mNeonGreen and 19 nm for EGFP) illustrating that the ground state S_0 and excited state S_1 are very close in energy. However, a short Stokes shift is not ideal as the separation of excitation and emission signals require very high-quality filters or a decrease in the number of collected photons by moving away from the peaks. Finally, the pKa value of 4.4 makes it superior to other green FPs, as it allows probing fluorescence in a wider range of pH, for instance for the study of endosomes, secretory granules, lysosomes or plant vacuoles (Shinoda *et al.*, 2018).

Table 10. Basic photophysical parameters of GFP-like FPs mentioned in this chapter. Values are reproduced from the FPbase database https://www.fpbase.org/ , except where indicated (Lambert, 2019).

Protein	λ_{exc} or λ_{abs} (nm)	λ_{em} (nm)	EC (mM^{-1} cm^{-1})	QY	Brightness	pKa
***Av*GFP**	395 / 475	509	25.0	0.79	19.8	4.5
EGFP	488	507	55.9	0.60	33.5	6.0
***Aaus*FP1**	504	510	170.0	0.97	164.9	4.4
mNeonGreen	506	517	116.0	0.80	92.8	5.7
***Aaus*FP2**	609	-	52.0	< 0.001	-	6.0*

*pKa determined from pH titration of *Aaus*FP2 absorbance

*Aaus*FP2 is a chromoprotein, meaning that fluorescence of its chromophore is below the detection limit of standard fluorimeters. Its blue colour stems from a broad absorption peak in the red and far-red part of the visible spectrum (**Figure 57B**), suggesting a peculiar chromophore structure or environment, which triggered our structural study. Following up on our successful structure determination, the role of the nearby residue Cys62 was investigated by mutagenesis. Indeed, the single-point point mutant C62S appears yellow, with an absorption maximum at 465-470 nm (**Figure 57B**) demonstrating that it plays a major role in the red/far-red light absorption properties of the *Aaus*FP2 chromophore.

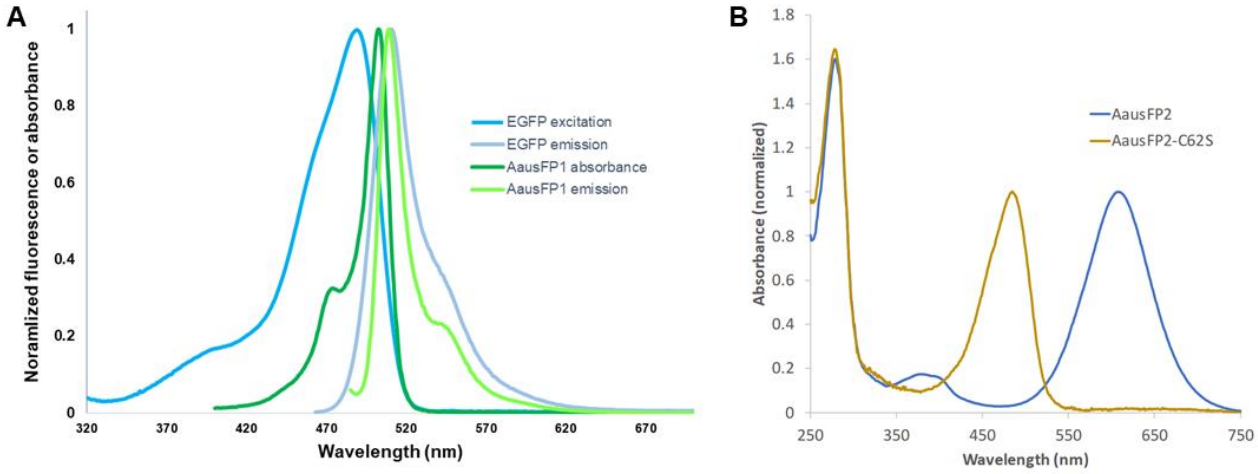

Figure 57. Absorption/excitation and emission spectra of EGFP, *Aaus*FP1 *Aaus*FP2 and *Aaus*FP2-C62S. (**A**) Normalized absorption, fluorescence excitation and emission spectra for *Aaus*FP1 and EGFP. EGFP data are reproduced from the FPbase database https://www.fpbase.org/, (Lambert, 2019). (B) Normalized absorption spectra of *Aaus*FP2 and of its single-point mutant *Aaus*FP2-C62S.

3.3 Structure determination

The crystal of *Aaus*FP1 used for structure determination was grown with the hanging drop method in 0.9 M trisodium citrate, 0.2 M sodium chloride in 0.1 M Tris buffer (pH 7.0) and was cryoprotected with 20 % glycerol. The crystal of *Aaus*FP2 used for structure determination was crystallized with the hanging drop method in 24% PEG 3350, 0.2 M sodium chloride and 0.1 M HEPES buffer (pH 7.9) and was not cryprotected to prevent melting of the crystal. Diffraction data were collected on beamline XALOC of the ALBA synchrotron in Spain for *Aaus*FP1 and on beamline ID30B of the ESRF in Grenoble for *Aaus*FP2. Both crystals were cryoprotected with 20% glycerol and both data collections were performed at 100 K. The structures of both proteins were solved by the molecular replacement method using the crystal structure of the yellow fluorescent protein phiYFP from *Phialidium sp.* (PDB entry code: 4HE4) as the search model, as the phiYFP structure came up as the top hit in a BLAST search for both proteins. Data reduction and refinement statistics are presented in **Table 11**.

Table 11. Data collection and structure refinement for *Aaus*FP1 and *Aaus*FP2. Values in parentheses are for the outer resolution shell, chosen using $CC_{1/2}$ (Evans and Murshudov, 2013).

	***Aaus*FP1**	***Aaus*FP2**
Data reduction		
Wavelength (Å)	0.979	0.976
Space group	$P2_12_12_1$	$C222_1$
Cell dimensions		
a, b, c (Å)	65.08, 101.41, 161.43	54.41, 75.13, 100.41
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	50.0 – 2.47 (2.53 – 2.47)	50.0 – 2.06 (2.11 – 2.06)
Wilson B-factor ($Å^2$)	37.9	38.0
No. of reflections	228,828 (16,791)	78,841 (5,174)
Unique reflections	39,086 (2845)	13,079 (976)
Multiplicity	5.9 (5.9)	6.0 (5.3)
Completeness (%)	99.9 (99.9)	99.8 (99.7)
$\langle I/\sigma(I) \rangle$	9.72 (2.07)	12.79 (2.16)
R_{meas} (%)	21.5 (120.3)	10.0 (85.1)
$CC_{1/2}$	0.993 (0.707)	0.998 (0.683)
Structure refinement		
Resolution (Å)	48.42 – 2.47 (2.50 – 2.47)	44.11 – 2.06 (2.09 – 2.06)
R_{work} (%)	18.1 (29.1)	16.6 (24.0)
R_{free} (%)	22.8 (35.0)	20.2 (32.0)
No. of atoms	7567	1937
Protein	7196	1788
Solvent	371	149
B-factors ($Å^2$)	37.2	24.1
Protein	37.4	22.4
Solvent	34.1	44.3
R.m.s. deviations		
Bond lengths (Å)	0.010	0.009
Bond angles (°)	1.73	1.69

3.4 Structural analysis of *Aaus*FP1

3.4.1 Oligomeric state

Determining the oligomeric nature of a wild-type FP is of a prime importance for evaluating its potential in cell imaging. If a monomer, its spectroscopic properties will be easily retained in a fusion version. If dimeric or tetrameric, the FP cannot be used as it is, and needs to be monomerized, which in general strongly affects negatively its brightness. The exact oligomeric nature of *Aaus*FP1 and *Aaus*FP2 could not be assessed from SEC MALS experiments, in which purified proteins co-eluted with aggregates (**Figure 58**). I thus performed a crystal packing analysis for *Aaus*FP1 and *Aaus*FP2 to identify potential oligomerisation interfaces as this method can differentiate between physiological interfaces and crystal packing interfaces depending on the area of contact (see explanation below).

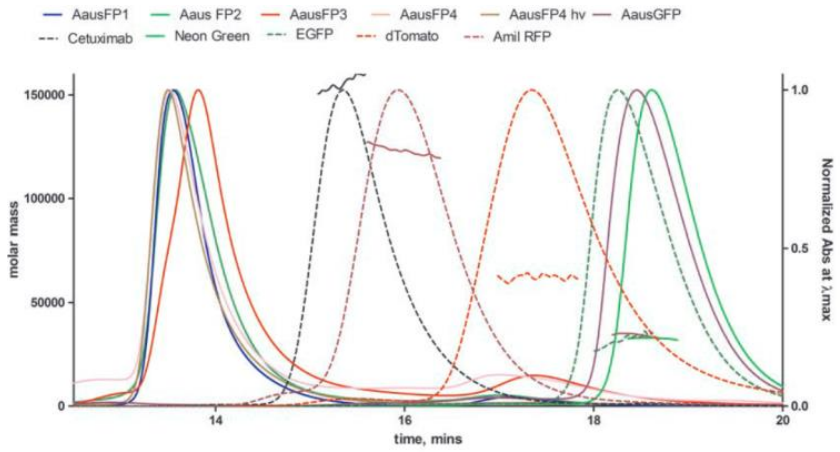

Figure 58. Normalised elution profiles in SEC-MALS experiments from the five different *A. australis* FPs and reference proteins (monomeric FPs: mNeonGreen and EGFP, dimeric FP: tdTomato, tetrameric FP: AmilFP593, Cetuximab: highly purified pharmaceutical antibody of MW= 153 kDa) (reproduced from Lambert *et al.*, 2020).

The asymmetric unit of the *Aaus*FP1 crystal consists of a tetramer (**Figure 59**, left and centre). A PISA analysis of monomer-monomer interfaces within the crystal packing reveals two large interfaces of 1210 Å^2 area between the A and B molecules of the asymmetric unit on one side, and the C and D molecules on the other side, respectively. The third and fourth larger interface areas are only of 360 Å^2 between the A and D molecules, and the B and C molecules, respectively. Previous crystallographic work performed in the group on tetrameric LanYFP and monomeric mNeonGreen crystallized in various space groups (Clavel *et al*, 2016) taught us that interface areas between FP molecules above 700 Å^2 were likely to be

physiological, while interface areas below 400 Å^2 were likely to be due mere crystal contacts. We thus conclude that *Aaus*FP1 is probably a dimer in solution (**Figure 59**, right).

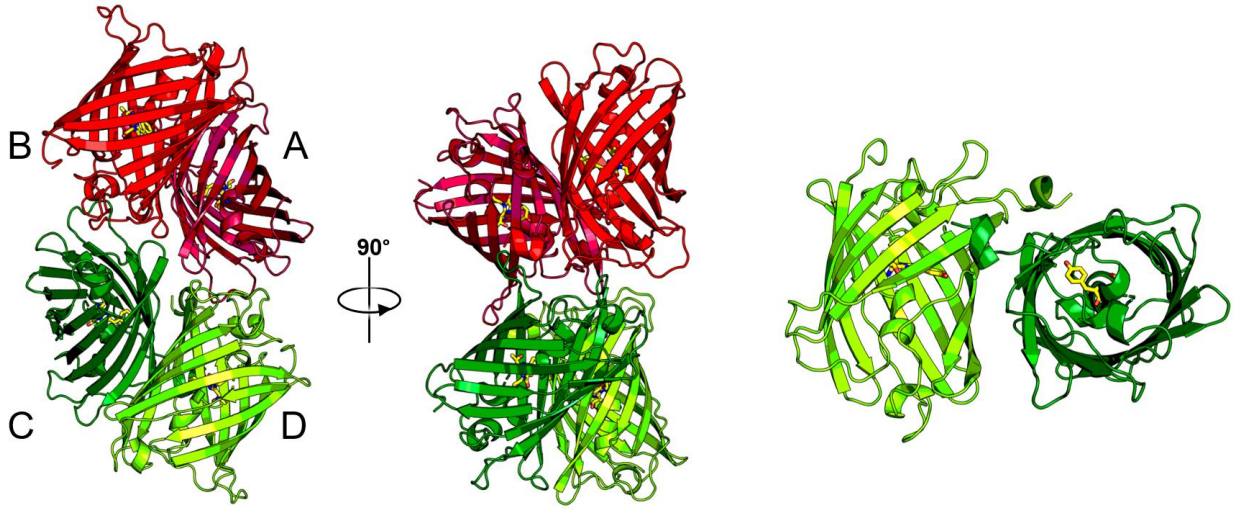

Figure 59. Oligomeric state of *Aaus*FP1. (Left) Tetramer of *Aaus*FP1 in the asymmetric unit. (Centre) Tetramer rotated by 90° along the vertical axis. (Right) Probable physiological dimer of *Aaus*FP1.

3.4.2 The chromophore and its environment

In order to understand the rationale for very bright fluorescence of *Aaus*FP1, we decided to compare its structure to that of another bright green FP from another jellyfish species, EGFP, a double-point mutant from the jellyfish *A. victoria*, with the relatively high fluorescence QY of 0.60, and not to that of the very bright mNeonGreen, which originates from an organism belonging to a different phylum.

EGFP is the double-point mutant F64L/S65T of *Av*GFP and thus shares with *Aaus*FP1 the exact same sequence for the chromophore. They also share the same chemical structure, with a deprotonated oxygen in the phenolate group. Both chromophores are almost perfectly planar, with dihedral angles of τ = -0.3° and ϕ= -1.8° for EGFP (PDB entry code: 2Y0G) and τ = 1.1° and ϕ= -4.8° for *Aaus*FP1 (monomer A). The stabilisation of the phenolate group oxygen in both proteins is very similar, except that the hydrogen bond to Thr203 does not have an equivalent is *Aaus*FP1 (**Figure 60A**). To the contrary, the whole set of interactions to a group of three key residues in EGFP and superfolder GFP (Leu220, Glu222, Val224) have their equivalent residues significantly displaced in *Aaus*FP1 (residues Val218, Glu220 and Phe222), but also in mNeonGreen (residues Phe208, Glu210 and Gln212) in relation to the chromophore (**Figure 60B**), which has implications on the stabilization mode of the bulk of

the chromophore. Interestingly, this difference appears to be due to a swap of residues of distinct side chain bulk volumes: at position (223, 224) in EGFP a Phe/Val motif becomes a Val/Phe motif at position (221, 222) in *Aaus*FP2.

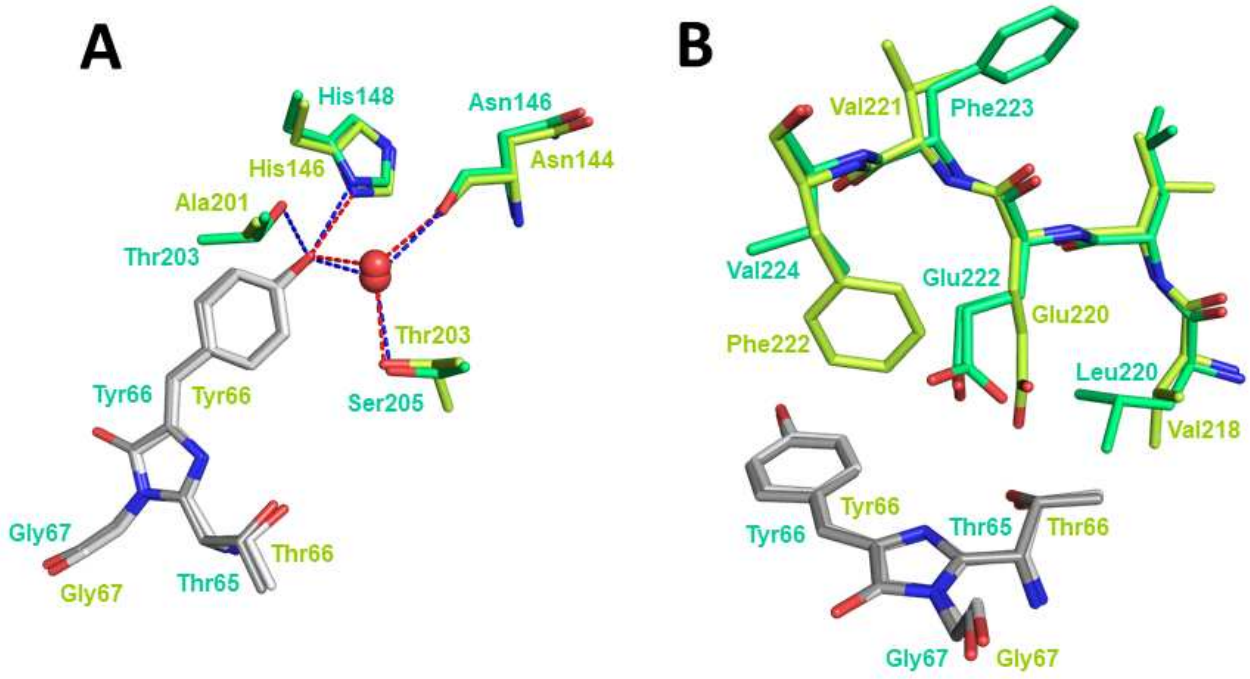

Figure 60. *Aaus*FP1 and EGFP hydrogen stabilisation mode and residues concerted displacement. (**A**) Comparison of the stabilisation mode by hydrogen bonds in *Aaus*FP1 (yellow-green, chromophore in light-grey) *vs.* EGFP (blue-green, chromophore in dark-grey). (**B**) Concerted displacement of three neighbouring side chains on one side of the chromophore in *Aaus*FP1 (residues Val218, Glu220 and Phe222) when compared to EGFP (residues Leu220, Glu222, Val224).

Analysis of residues within 4.0 Å of the *Aaus*FP1 and EGFP chromophores gives insights on the difference in structural determinants of chromophore stabilisation. The list of distances between can be found in the Supporting Information (S1 Text) of Lambert *et al.*, 2020 at: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000936. When looking to the side of the chromophore, stabilisation of the phenolate ring appears grouped in EGFP (**Figure 61B**, left) but separated in two rings of surrounding residues : one (Leu61, Thr62, Ile148, Ile163, Phe222) is centred on the C_{γ}, $C_{\delta 1}$ and $C_{\delta 2}$ atoms of the phenolate and one (Tyr143, Ile165, Ala201, Thr203) centred on the $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, C_{ζ}, and O_{η} atoms (**Figure 61A**, left), seemingly providing a wider van der Waals interaction along the axis of the chromophore. Moreover, when looking along the axis of the chromophore, the distribution of van der Waals interactions appear evenly distributed in *Aaus*FP1 (**Figure 61A**, right, purple circular arrow) but leaves a free angular volume of ~60° in EGFP (**Figure 61B**, right, purple

incomplete circular arrow). Our interpretation is that the *Aaus*FP1 chromophore appears evenly stabilized from all sides, thus favouring optimal radiative decay of the excited state, while the EGFP chromophore may find a free volume for vibrational relaxation from the excited state.

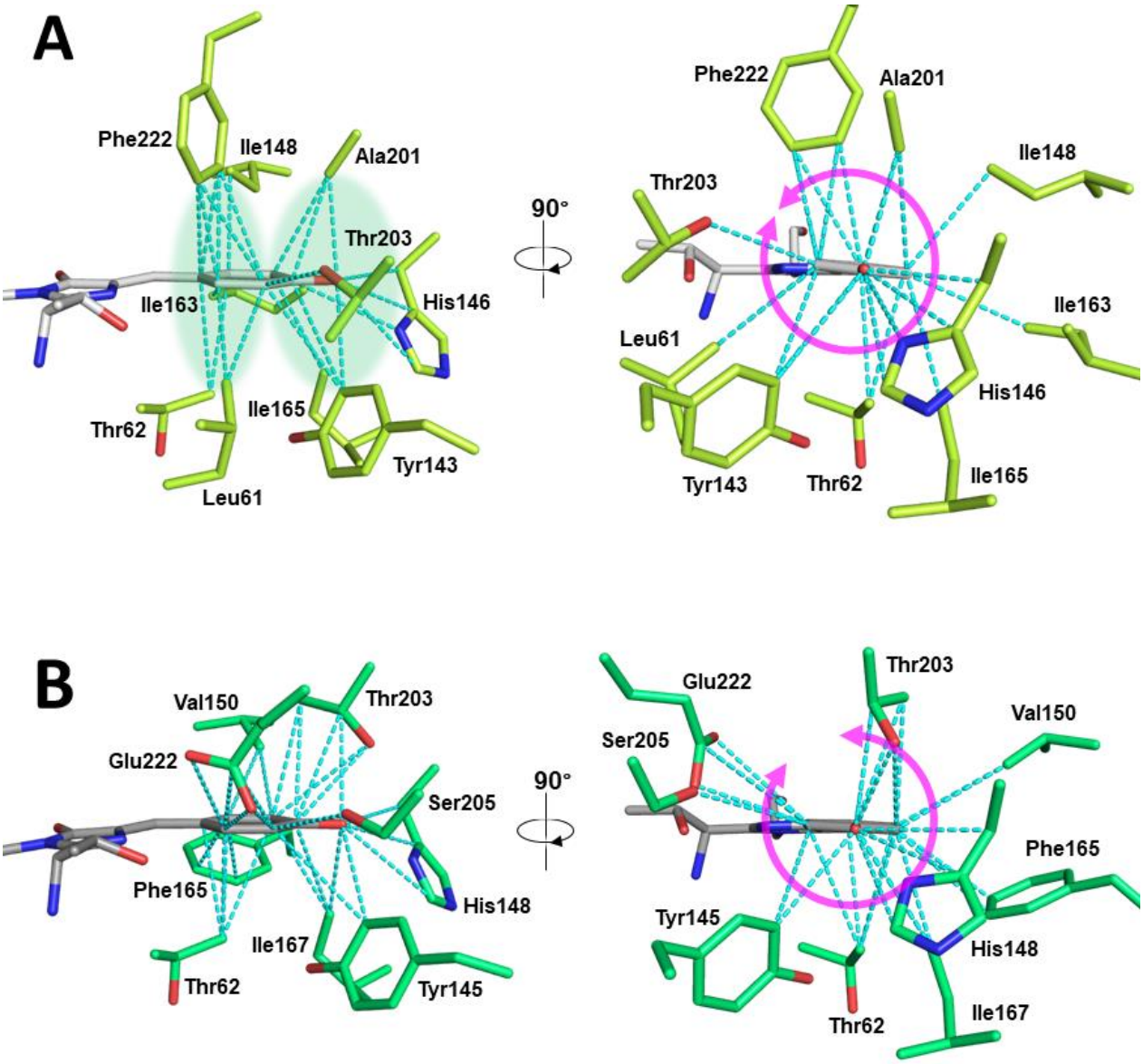

Figure 61. Two views (rotated by 90° from each other) of van der Waals interactions in the (**A**) *Aaus*FP1 and (**B**) EGFP chromophore environments.

3.5 Structural analysis of *Aaus*FP2

3.5.1 Oligomeric state

The asymmetric unit of the *Aaus*FP2 crystal consists of a monomer. A PISA analysis of monomer-monomer interfaces within the crystal packing reveals a single large interface of 1290 Å^2 area between the monomer and a symmetry-related molecule. The second, third and fourth largest areas range from 540 down to 310 Å^2, which are likely to stem from crystal contacts. We thus postulate that *Aaus*FP2 is a dimer in solution given the single large interface between the two monomers (**Figure 62**).

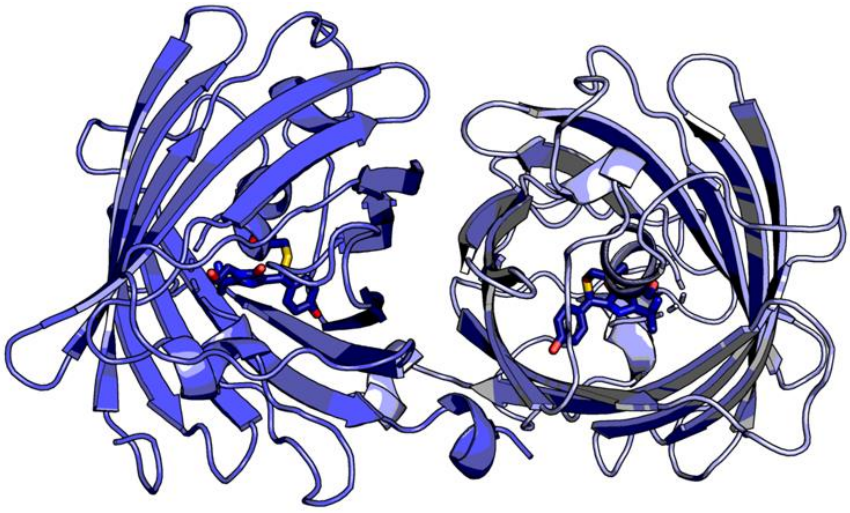

Figure 62. Putative dimer of *Aaus*FP2 formed by the monomer in the asymmetric unit and a symmetry-related molecule. in the asymmetric unit.

3.5.2 The peculiar structure of the chromophore

The most striking feature of the 2.06 Å structure of *Aaus*FP2 is the unusual environment of the chromophore (**Figure 63A**). Examination of the electron density map strongly suggests that Cys62 residue is covalently bound to the C_β atom at the centre of the methylene bridge. Indeed, spectroscopic characterization and a mutational study had strongly suggested that this residue controls the absorption properties of the chromophore. This covalent bond does not affect the sp^2 character of the C_β atom. Indeed, if the C_β atom had become sp^3, the electron conjugation system of the chromophore would have been interrupted at this position, and the absorption maximum would have blue-shifted compared to a GFP-like chromophore given the smaller size of the conjugated system of the chromophore. Hence, the cysteine covalent attachment amounts to a substitution of the C_β hydrogen. Besides, the chromophore is very

much distorted, as shown by the comparison with the flat chromophore of *Aaus*FP1 (**Figure 63B**).

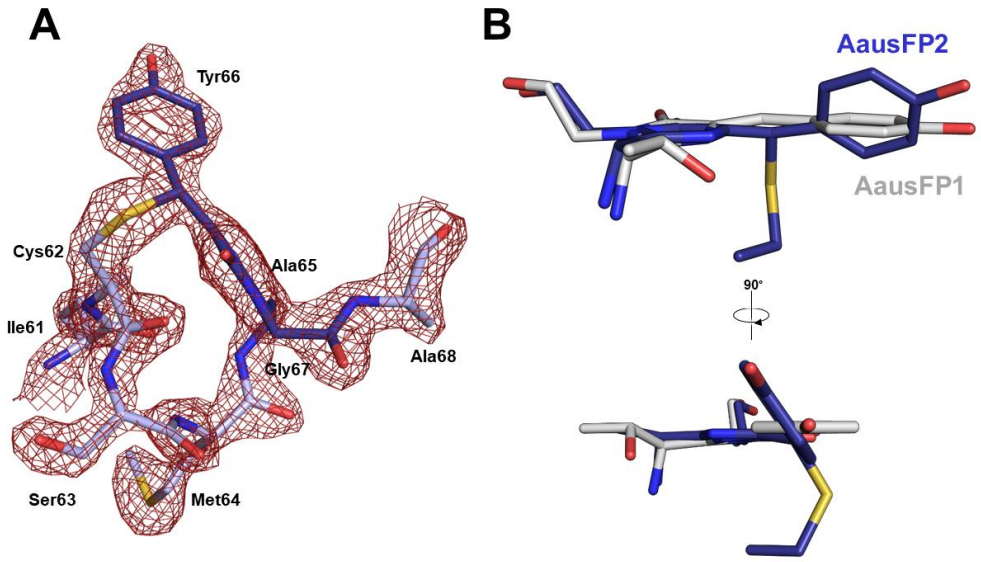

Figure 63. Peculiar structure of the *Aaus*FP2 chromophore (**A**) $2F_{obs} - F_{calc}$ electron density map, contoured at a contour level of 1.8 σ, represented on the chromophore region of *Aaus*FP2. (**B**) Two views, vertically rotated 90° from each other, of the superposition of the chromophores in *Aaus*FP1 and *Aaus*FP2.

Analysis of residues within 4.0 Å of the *Aaus*FP2 chromophore gives insights on the structural determinants of chromophore distortion. The list of these distances can be found in the Supporting Information (**Annex 1**) of Lambert *et al.*, 2020 at: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000936. The side chains of residues Asn201, Ile203 and Glu220 and the main chain of His202 appear to form a van der Waals clamp around the phenolate group of the chromophore, whose rotation around its symmetry axis is blocked *in fine* by the C_β atom of Thre146 (**Figure 64**). This asymmetric set of van der Waals interactions may play a role in the distortion of the chromophore, although one can observe that the sulphur atom of Cys62 is coplanar with the atoms of the phenolate group. There may be an electronic effect in play as well. Finally, the phenolate oxygen is stabilized by a water molecule and the amide nitrogen of the main chain of His202 (**Figure 65**).

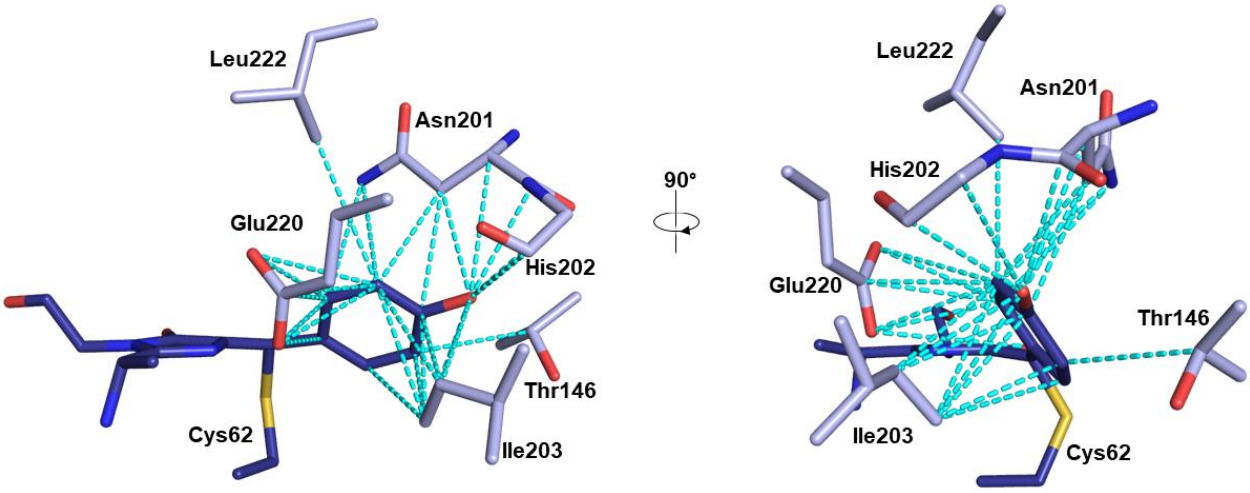

Figure 64. Two views (rotated by 90° from each other) of *Aaus*FP2 showing the asymmetry of van der Waals interactions between the protein and the phenolate group of the *Aaus*FP2 chromophore. One can note that the sulphur atom of Cys62 is in-plane with atoms of the phenolate group.

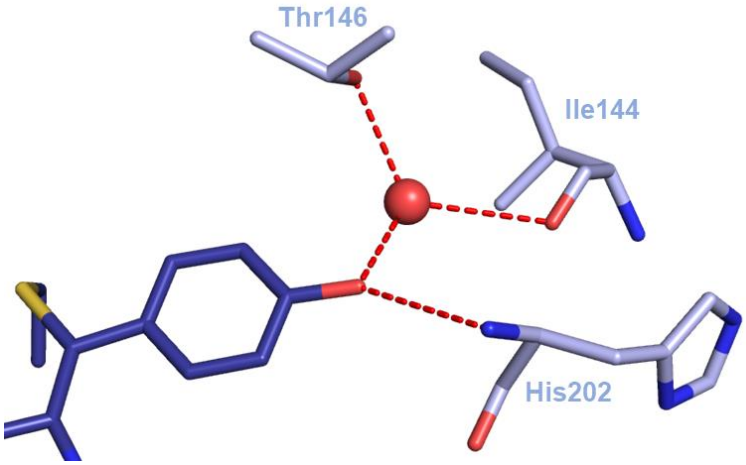

Figure 65. Stabilisation of the phenolate oxygen of the *Aaus*FP2 chromophore by hydrogen bonds.

3.5.3 Validation of the role of Cys62

The apparent presence of covalent bond on the methylene bridge of the chromophore of *Aaus*FP2 and the spectroscopic properties of the C62S mutant called for an investigation of the role of the sulphur atom in the absorption properties of the chromophore. First of all, the absorption peak of alkali-denatured *Aaus*FP2 shows a ~20 nm red-shift compared to the 445 nm absorption peak of an alkali- denatured FP with the classical Tyr-based GFP chromophore (**Figure 66A**). Further addition of the reducing agent β-mercaptoethanol cancels the red-shift, suggesting that the suspected covalent bond to the chromophore has been reduced.

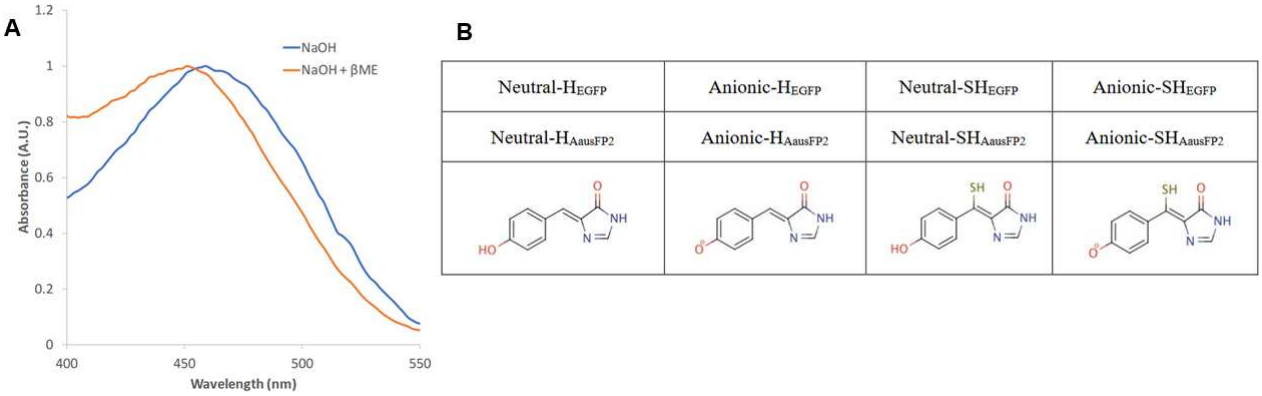

Figure 66. Demonstration of Cys-chromophore covalent bond and modelling of the chromophore absorption properties. (**A**) Absorption spectra of alkali-denatured *Aaus*FP2 in absence or presence of β-mercaptoethanol. (**B**) The eight chromophore models tested by QM calculations. Each of the models represented on the lower line by its chemical structure has either a planar (EGFP series) or a distorted (*Aaus*FP2) conformation.

Finally, quantum mechanical calculations were carried out by Prof. Isabelle Navizet, from the University Gustave Eiffel in Marne-la-Vallée (France), to test the effect of sulphur substitution at the C_β position of the *Av*GFP chromophore restricted to the atoms implicated in the conjugated electron cloud. This chromophore is tested with either a planar (as in *Av*GFP and GFP) or a distorted (as in *Aaus*FP2) conformation. In addition, the oxygen of the phenolate group is either protonated (neutral chromophore) or deprotonated (negatively charged chromophore), leading to eight possible chromophore models (**Figure 66B**). The time-dependent density functional theory (TD-DFT) calculations of the vertical transitions were performed for the first 6 singlet states.

The energies and wavelengths of the most intense transitions (which illustrate the bigger oscillator strength) (**Table 12**) only give a qualitative estimate of the absorption maximum of the various chromophores as they do not take into account the protein environment. However, the results give a clear trend, and allow identifying two models with a large red shift characteristic of *Aaus*FP2. Both models have a sulphur atom substituted on the methylene bridge C_β atom and adopt the distorted geometry, and they only differ by their protonation state. This suggests that both sulphur substitution and conformation distortion are required for the large red shift. Surprisingly, the mere sulphur substitution seems to hardly affect the absorption properties of a planar chromophore.

Table 12. Energies of the most intense vertical transition for the eight different chromophore models, corresponding wavelengths and oscillator strengths.

Model	Excited state	Transition energy (eV)	Wavelength (nm)	*f*
Anionic-H_{EGFP}	S2	2.66	466	0.81
Anionic-SH_{EGFP}	S2	2.66	467	0.64
Neutral-H_{EGFP}	S2	3.02	411	0.60
Neutral-SH_{EGFP}	S1	3.00	413	0.33
Anionic-$H_{AausFP2}$	S2	2.47	501	0.65
Anionic-$SH_{AausFP2}$	S1	1.93	641	0.25
Neutral-$H_{AausFP2}$	S2	2.94	421	0.50
Neutral-$SH_{AausFP2}$	S1	1.80	688	0.13

3.6 Discussion

Identifying that both *Aaus*FP1 and *Aaus*FP2 are dimeric in solutions given the crystal packing interfaces study poses the question of the similarity of this dimer to that of *Av*GFP, as these FPs come from two species of the same genus *Aequorea*. All three dimers are represented on **Figures 67A** to **67C**. The superposition of all three dimers (**Figure 67D**) shows that *Aaus*FP1 and *Aaus*FP2 shares a very similar dimer interface, which is very different to that of *Av*GFP. However, these results cannot rule out the fact that when *Aaus*FP1 and *Aaus*FP2 are present at high concentrations, higher oligomeric states can form (**Figure 58**).

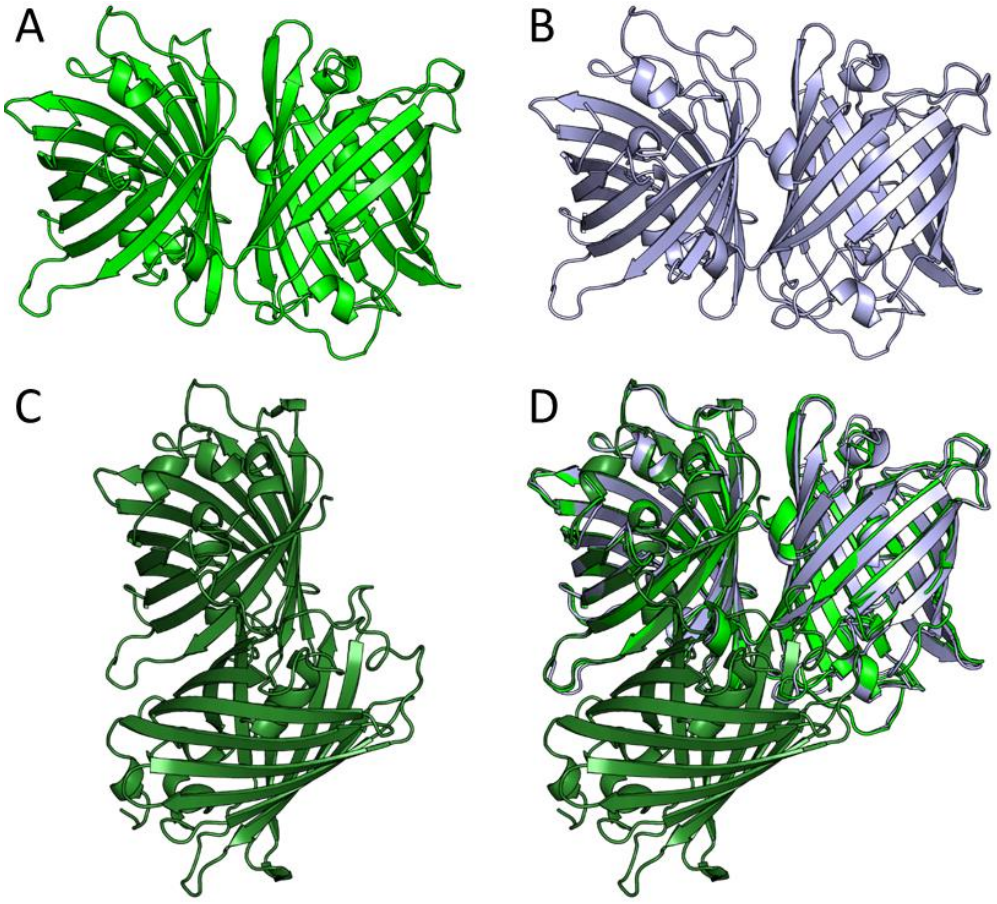

Figure 67. Comparison of *Aaus*FP1, *Aaus*FP2 and *Av*GFP dimeric interface. (**A**) *Aaus*FP1 putative physiological dimer. (**B**) *Aaus*FP2 putative physiological dimer. (**C**) *Av*GFP physiological dimer at high concentration (PDB entry code: 1GFL) (Yang *et al*, 1996a) (**D**) Superposition of all three dimers relative to monomer A.

We have shown that the very high brightness of *Aaus*FP1 was likely due to an optimized stabilisation of van der Waals interactions homogenously surrounding the phenolate group of the chromophore, first ensuring its planarity, but also minimizing the possibility for vibrational relaxation of the excited state by environment crowding. The lesson learned here will be difficult to propagate to other FPs as van der Waals interactions are weak and thus require to come from a large number of side chains, contrarily to stabilisation with hydrogen bonds, which is usually ensured by a couple of well-positioned residues. However, correcting the imperfection of certain incomplete van der Waals interaction sphere of residue should be possible by considering saturation mutagenesis at the specific location of the imperfection and should include the 2-3-4 neighbouring residues of the interaction sphere.

The unexpected discovery brought by the structure determination of the chromoprotein *Aaus*FP2 is the identification of a new type of chromophore, in which a nearby residue covalently binds to the methylene bridge of the chromophore. While this mechanism has been somewhat observed previously in the red FP LaRFP with the fragile binding of a tyrosine at the same location (Pletnev *et al*, 2013), the fact that this new chromophore involves a redox reaction involving a cysteine opens new possibilities for future new bioengineering ways of creating new chromophores, possibly leading to far-red fluorescence.

One caveat to the future use of *Aaus*FP1 as a very bright GFP is its oligomeric state. The dimer interface appears complex (**Figure 68A**), which probably requires to identify a small number of mutations (for the record the dimer interface of *Av*GFP could be completely disrupted with the single mutation A206K). Besides, there is a chance that the brightness of the protein will be affected, and additional mutagenesis rounds should be necessary to restore most of the brightness, as already seen for the monomerization of the very bright yellow FP LanYFP (Shaner *et al*, 2013). Would *Aaus*FP2 find application as a chromoprotein, the monomerization process appears to be as complex as *Aaus*FP1, if not slightly more (**Figure 68B**).

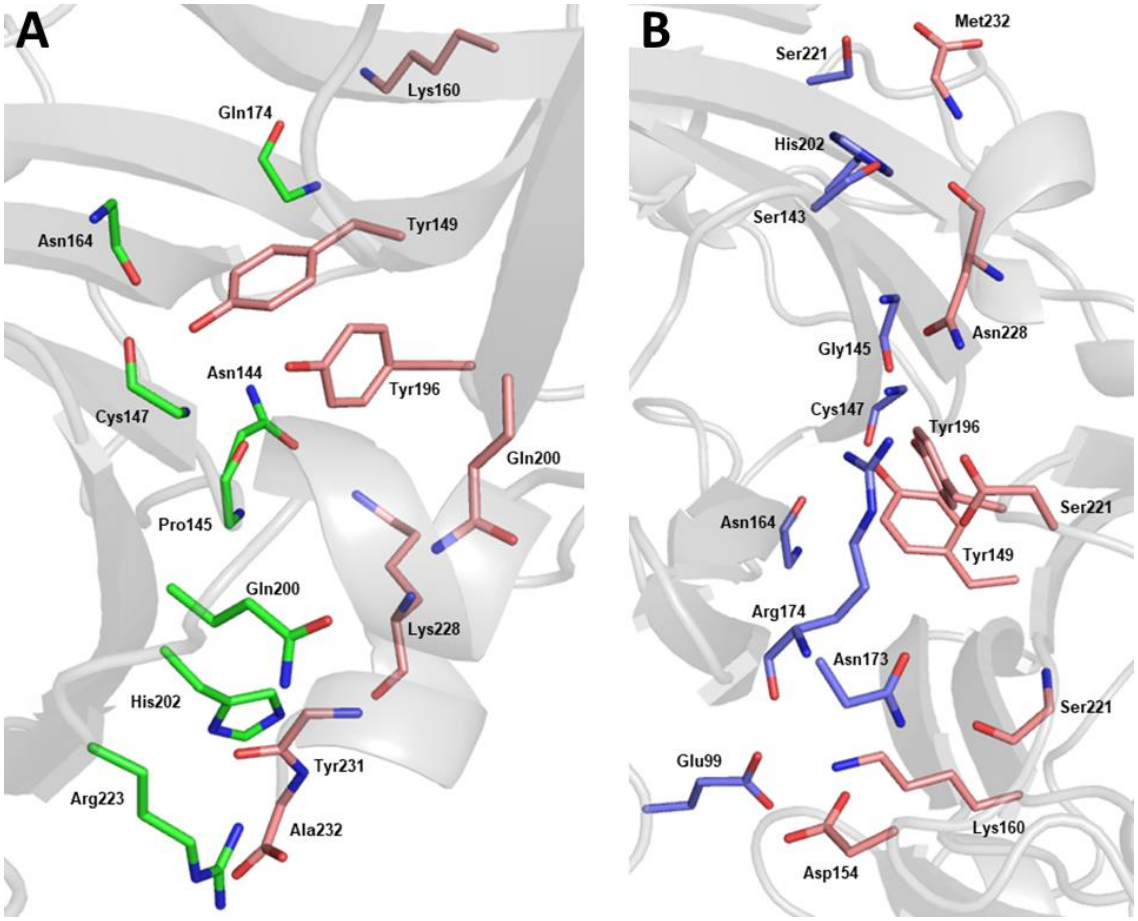

Figure 68. *Aaus*FP1 and *Aaus*FP2 dimeric interface. (**A**) Residues implicated in half of the *Aaus*FP1 dimer. (**B**) Residues implicated in half of the *Aaus*FP2 dimer. Note that the other halves are composed of the same residues, but belonging to the other monomer of each dimer.

Chapter 4

RESULTS

Structure determination of the weakly red fluorescent protein *Csi*FP4 from *Clytia simplex*: a fourth type of red fluorescent chromophore

4.1 Introduction

During their trip to the Great Barrier Reef in 2017, Nathan Shaner and Gerard Lambert collected a jellyfish using a hand-held plankton net in the scientific research zone surrounding Heron Island. The living animal displayed bright green fluorescence along with weak red fluorescence. Total RNA was isolated from the whole animal on site at Heron Island Research Station and then mRNA sequencing and *de novo* transcriptome assembly were performed to reconstruct the transcriptome of the animal, which was determined to be the hydrozoan jelly *Clytia simplex* (**Figure 69**). The sequences for FP homologs were identified from the total transcriptome via BLAST homology searching. Four FP homologs were identified from this animal, one weakly red fluorescent FP and three green and cyan FPs. The RFP, called *Csi*FP4, drew particular attention in the context of the search for new genetic material as starting scaffold for efficient FPs in the far-red region of the visible light spectrum.

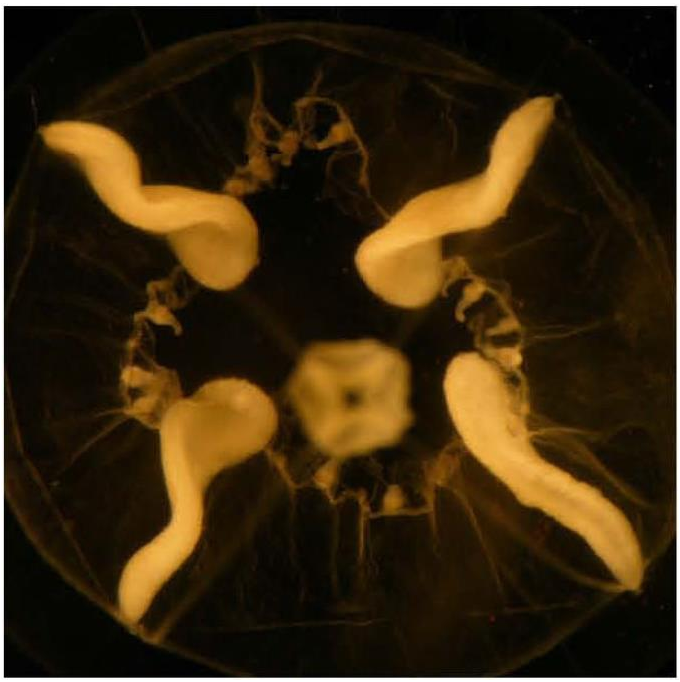

Figure 69. Photograph of a specimen hydrozoan jelly *Clytia simplex*. Sub-umbrellar view, showing gonads. The diameter of the bell is below 1 cm (reproduced from (Galea *et al*, 2007).

The sequence of *Csi*FP4 shares 40% and 36% identity with those of *Aaus*FP1 and *Aaus*FP2, respectively (**Figure 70**). *Csi*FP4 has a different chromophore, QYG, than *Aaus*FP1 (TYG) and *Aaus*FP2 (AYG). The QYG sequence is that of the *Ds*Red chromophore, posing the question of whether the red-shifted absorption spectrum of the chromophore is due to the formation of DsRed-like chromophore, *i.e.* with the formation an acylimine bond on the

protein backbone. However, the sequence alignment (**Figure 70**) reveals the presence of a cysteine (Cys64) at the same position as that covalently bound to the chromophore in *Aaus*FP2.

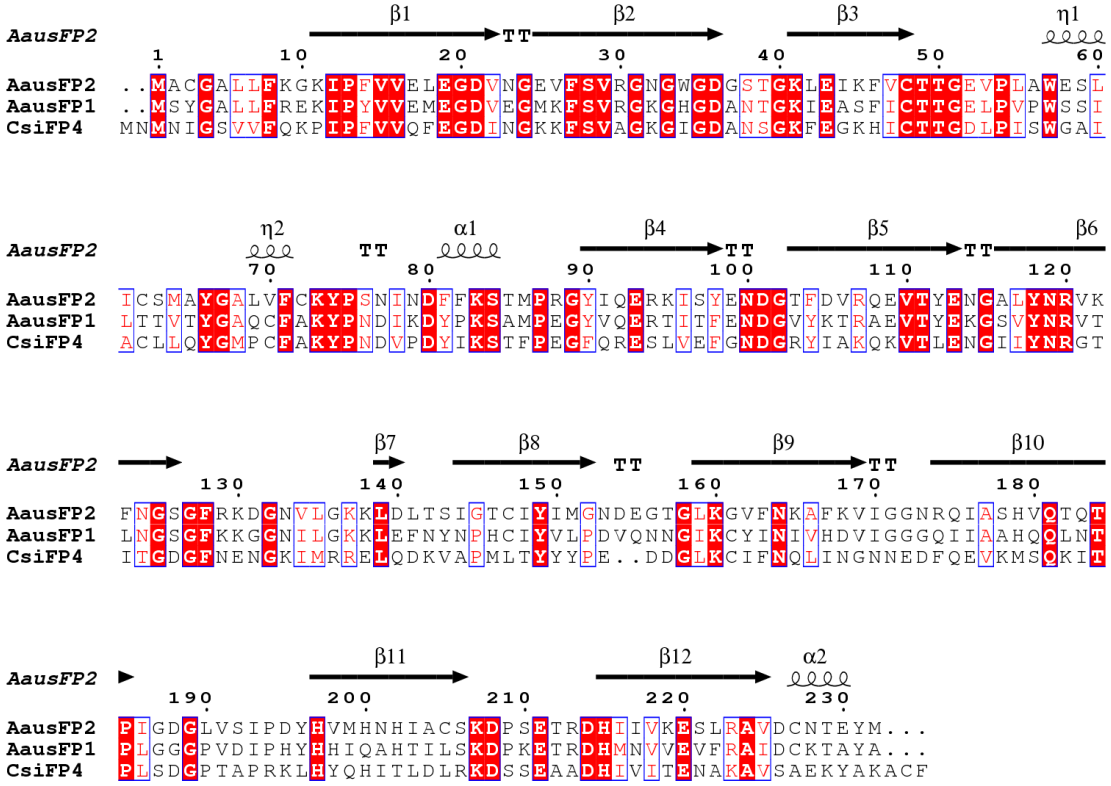

Figure 70. Sequence alignment of *Csi*FP4 with *Aaus*FP1 and *Aaus*FP2. Secondary structure elements (α-helix, β-sheet, turn) are indicated for the *Aaus*FP2 structure (PDB entry code: 6S68). The sequence alignment was prepared with Clustal Omega (Madeira *et al*, 2019) and was prepared with ESPript (Robert & Gouet, 2014)

4.2 Spectroscopic properties

*Csi*FP4 is weakly red fluorescent, with very good absorbance. The absorption peak is at 578 nm, while the fluorescence emission maximum is at 598 nm (**Figure 71**). The extinction coefficient has been calculated to be ~150.0 mM^{-1} cm^{-1} and the fluorescence QY of the protein to be 0.003 +/- 0.001 (**Table 13**).

Table 13. Basic photophysical parameters of *Csi*FP4 compared to *Aaus*FP1 and *Aaus*FP2. Values are reproduced from the FPbase database https://www.fpbase.org/ (Lambert, 2019), except for *Csi*FP4 (this chapter).

Protein	λ_{exc} or λ_{abs} (nm)	λ_{em} (nm)	EC (mM^{-1} cm^{-1})	QY	Brightness	pKa
***Aaus*FP1**	504	510	170.0	0.97	164.9	4.4
***Aaus*FP2**	609	-	52.0	< 0.001	-	6.0*
***Csi*FP4**	578	598	150.0	0.003	0.5	n.d.

*pKa determined from pH titration of *Aaus*FP2 absorbance

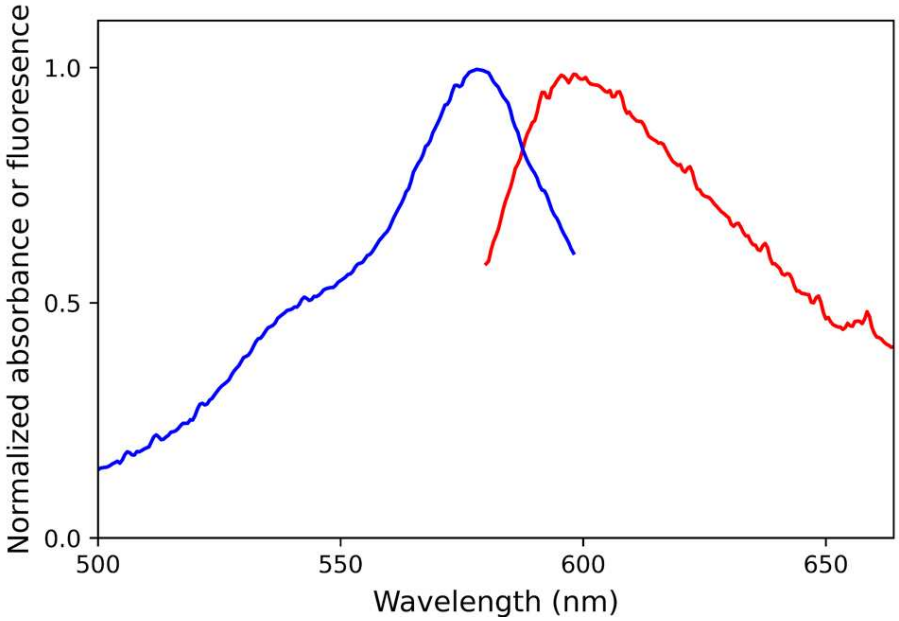

Figure 71. Normalized absorption (blue) and fluorescence (red) emission spectra for *Csi*FP4.

Because of the presence of a cysteine at position 64, a C64S mutant was produced as for *Aaus*FP2. *Csi*FP4-C64S absorbs around 513nm (data not shown), which demonstrates the significant role of this cysteine in the red-shifted absorption properties of *Csi*FP4.

4.3 Structure determination

The crystal of *Csi*FP4 used for structure determination was grown with the sitting drop method in 1.0 M lithium sulfate, 0.01 M nickel chloride in 0.1 M Tris buffer (pH 8.5) and was not cryoprotected. Diffraction data were collected at 100 K on beamline ID30A-3 of the ESRF in Grenoble. The structure of *Csi*FP4 was solved by the molecular replacement method using the crystal structure of the green fluorescent protein from *Clytia gregaria* (PDB entry code: 2HPW) as the search model, as this structure came up as the top hit in a BLAST search for the proteins. Data reduction and refinement statistics are presented in **Table 14**.

Table 14. Data collection and structure refinement for *Csi*FP4. Values in parentheses are for the outer resolution shell, chosen using $CC_{1/2}$ (Evans and Murshudov, 2013).

	***Csi*FP4**
Data reduction	
Wavelength (Å)	0.9677
Space group (Å)	$P6_5$
Cell dimensions	
A, b, c (Å)	182.9 182.9 155.9
A, *β*, ɣ (°)	90.0 90.0 120.0
Resolution range (Å)	100 – 3.04 (3.25 – 3.04)
Wilson B factor ($Å^2$)	35.4
Solvent (%)	82.9
No. of reflections	437,797 (80,631)
Unique reflexions	56,881 (10,271)
Multiplicity	7.7 (7.9)
Completeness (%)	99.9 (100)
$\langle I/\sigma(I) \rangle$	4.97 (1.37)
R_{meas} (%)	39.2 (146.3)
$CC_{1/2}$	96.3 (57.7)
Structure refinement	
Resolution (Å)	47.53 – 3.04 (3.12 – 3.04)
R_{work} (%)	17.5 (32.0)
R_{free} (%)	22.4 (36.0)
No. of atoms	7709
Protein	7309
Solvent	400
B-factors ($Å^2$)	24.3
Protein	24.6
Solvent	37.5
RMSD bond lengths (Å)	0.0152
RMSD bond angles (°)	2.1

4.4 Structural analysis

4.4.1 Crystal packing

The solvent content of the *Csi*FP4 crystals is particularly high at 83%, which can be visualised by the large solvent channels created by the crystal packing (**Figure 72**). This characteristic explains why crystals appeared soft when trying to fish them by hand, and fishing of the crystal used for successful data collection was made possible by the automated crystal collector robot of the *HTXLab* at the EMBL (Grenoble, France), which does not touch the crystal.

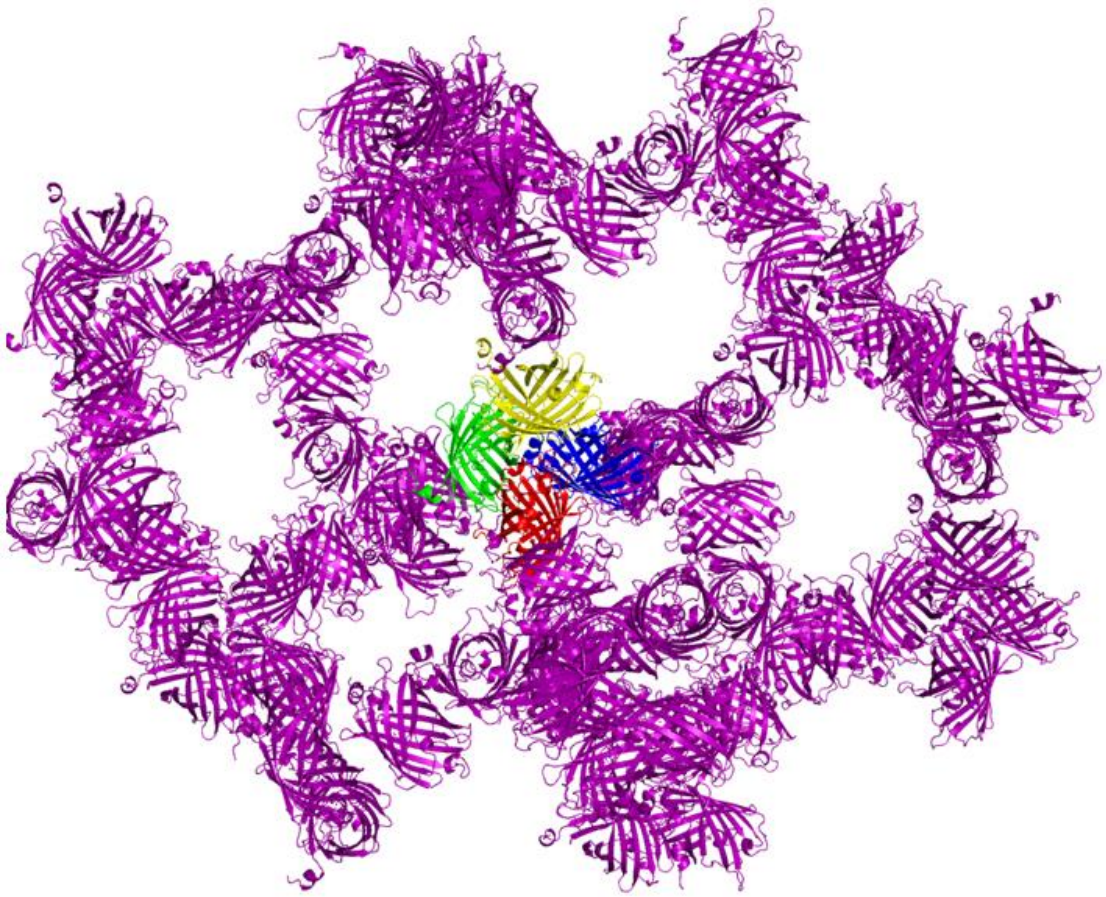

Figure 72. Crystal packing in the *Csi*FP4 crystals showing the presence of large solvent channels with diameters ranging from 58 to 70 Å. One tetramer, corresponding to the content of the asymmetric unit, is shown in green/yellow/blue/red and symmetry-related tetramers are shown in purple.

4.4.2 Oligomerisation state

*Csi*FP4 runs as a dimer on a gel in non-denaturating condition (data from Dr. Shaner's team). I performed a crystal packing analysis for *Csi*FP4 and to identify potential oligomerisation interfaces and verify the postulated oligomeric state. The asymmetric unit of the *Csi*FP4 crystal consists of a tetramer (**Figure 72**). A PISA analysis of monomer-monomer interfaces within the crystal packing reveals two large interfaces of 1510 and 1460 $Å^2$ area between the A and C molecules of the asymmetric unit on one side, and the B and D molecules on the other side, respectively. These surfaces are even larger than those identified in *Aaus*FP1 and *Aaus*FP2. The third, fourth, fifth and sixth larger interface areas range from 650 down to 440 $Å^2$ and correspond to the various cross-interactions between the four molecules of the asymmetric unit. While these areas are significant, they are still below the 700 $Å^2$ threshold that we use for FPs (Clavel *et al*, 2016), and we thus postulate that *Csi*FP4 is a dimer in solution (**Figure 73A**) confirming the non-denaturing gel conclusion.

The particularly large interface area in the postulated dimer compared to other FP dimers can be explained by the visually arresting observation that both C-termini of the dimer molecules adopt the shape of two hooks which are closely interacting with each other (**Figure 73B**). Monomerization of *Csi*FP4 would certainly require deleting the C-terminus, provided it does not affect protein folding and/or chromophore maturation.

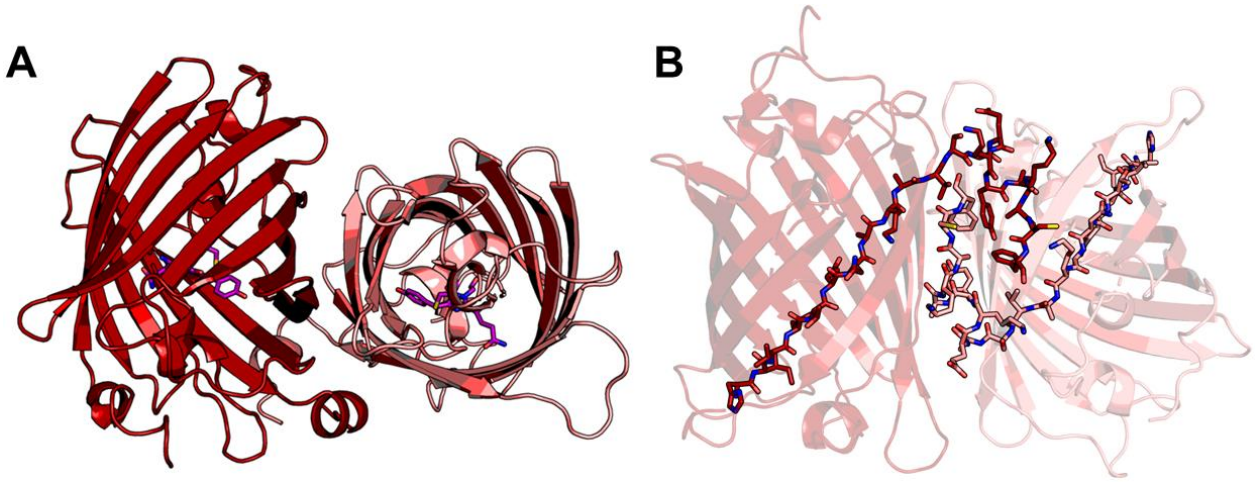

Figure 73. Oligomeric state of *Csi*FP4. (**A**) Putative dimer of *Csi*FP4 formed by the monomer in the asymmetric unit and a symmetry-related molecule in the asymmetric unit. (**B**) Visualisation of the interaction between the C-termini of both monomers of the dimer.

4.4.3 Chromophore configuration and environment

The striking feature of the *Csi*FP4 structure is the presence of a chromophore similar to that of *Aaus*FP2 (with the exception of Gln67 in *Csi*FP4 instead of Ala65 in *Aaus*FP2) with a covalent bond between the Cβ of the methylene bridge and a nearby cysteine residue (**Figure 74A**). However, if the cysteine is at the same position in the sequence compared to the chromophore, the thiol group attaches the methylene bridge from its other side, and when the configuration of the *Aaus*FP2 chromophore is *cis*, that of the *Csi*FP4 chromophore is *trans* (**Figure 74C**). Both chromophores are distorted with dihedral angles of τ = -23.3° and ϕ= -39.5° for *Aaus*FP2 and τ = 17.4° and ϕ= 63.0° for *Csi*FP4 (monomer A) (**Figure 74B**). Another notable difference is the absorption maximum at 570 nm, *i.e*. red-shifted by ~40 nm compared to the absorption maximum of *Aaus*FP2 (**Table 13**). Moreover, while the sulphur atom of Cys62 appeared in plane with the phenolate ring in the *Aaus*FP2 chromophore, that of Cys64 appears out of plane in the *Csi*FP4 chromophore. Finally, the most important difference is that the *Csi*FP4 chromophore is red fluorescent, although weakly.

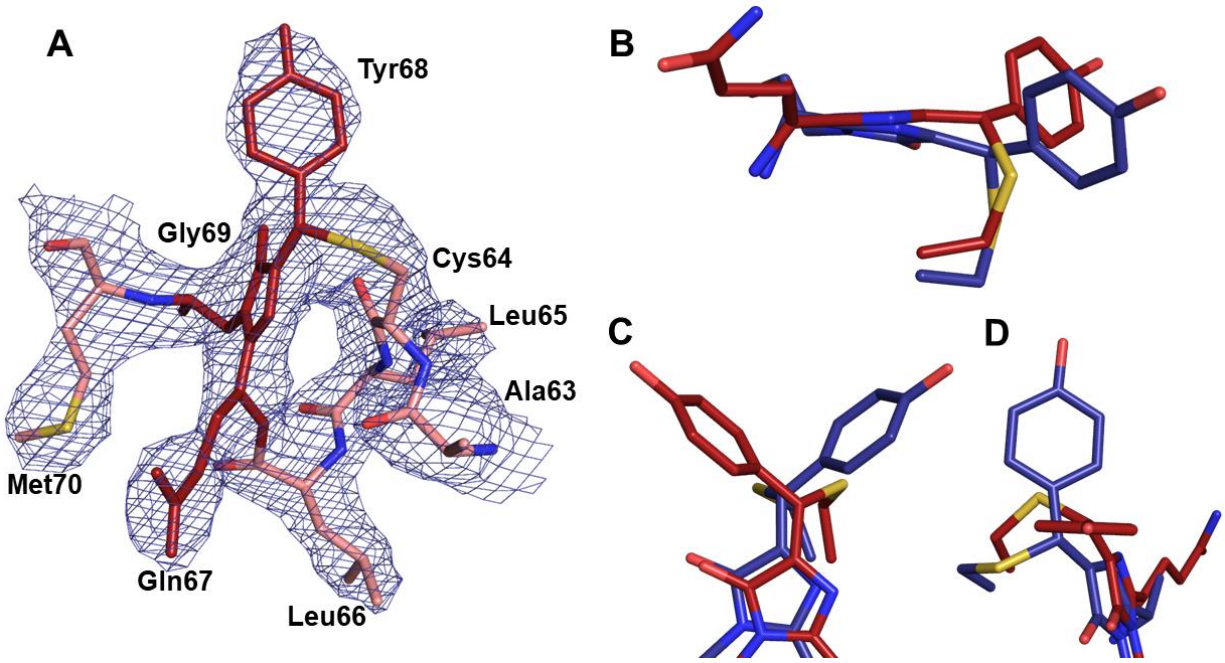

Figure 74. Peculiar structure of the *Csi*FP4 chromophore. (**A**) $2F_{obs} - F_{calc}$ electron density map, contoured at a contour level of 2.0 σ, represented on the chromophore region of *Csi*FP4. (**B**) Side view of the superposition of the *Aaus*FP2 (blue) and *Csi*FP4 (red) chromophores showing their distortion from planarity. (**C**) Top view showing the two distinct *cis* and *trans* configurations. (**D**) Front view for the *Csi*FP4 chromophore showing that the sulphur atom is not coplanar with the phenolate group.

Analysis of residues within 4.0 Å of the *Csi*FP4 chromophore gives insights on the structural determinants of chromophore stabilisation (**Figure 75**). Strikingly, the chromophore is well stabilized by van der Waals interaction on one side of the phenolate ring, while the other side is completely devoid of interactions, providing a straightforward explanation for the very low fluorescence QY.

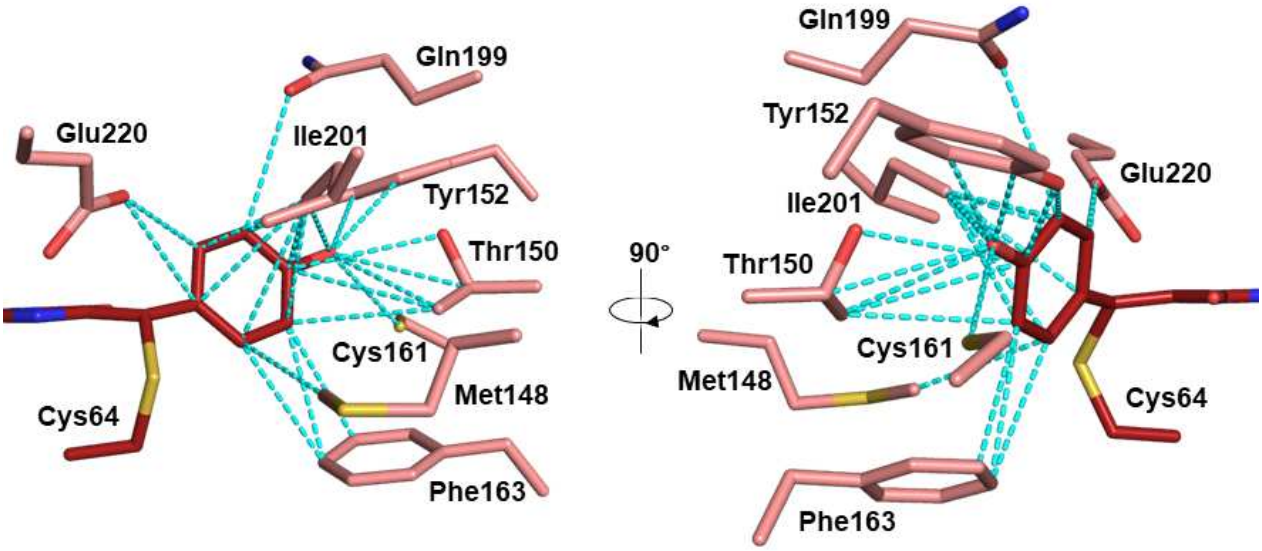

Figure 75. Two views (rotated by 90° from each other) of *Csi*FP4 showing the asymmetry of van der Waals interactions between the protein and the phenolate group of the *Csi*FP4 chromophore.

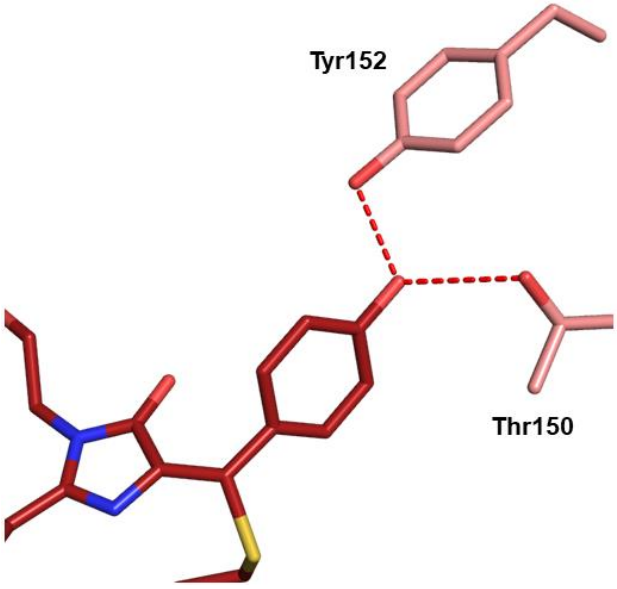

Figure 76. Stabilisation of the phenolate oxygen of the *Csi*FP4 chromophore by hydrogen bonds.

The different configuration in *Csi*FP4 compared to *Aaus*FP2 (**Figure 74B, C, D**) can be explained by a different stabilising environment of the phenolate of the chromophore. In *Aaus*FP2 the phenolate is stabilised by two hydrogen bonds, one with a neighbouring water molecule and another one to an amide nitrogen of the protein backbone. The phenolate is also stabilised with a network of Van der Waals (vdW) interactions with Ile203 and the backbone of His202 and Asn201 (**Figure 64** and **65**). In *Csi*FP4 the stabilising environment is almost completely different and the phenolate is stabilised by two hydrogen bond to neighbouring residues (Thr150 and Tyr 152) and a network of vdW interactions on one side of the chromophore (**Figures 75** and **76**) thus inducing an alternate configuration compared to *Aaus*FP2.

Finally, the difference between *Csi*FP4 and *Aaus*FP2 phenolate oxygen stabilisation can explain the different spectroscopic properties of the two proteins. It is conceivable that the water molecule stabilising *Aaus*FP2 phenolate ring is more displaceable than the side chain residues stabilising *Csi*FP4 phenolate ring, increasing further the possibility of vibrational relaxation of the excited state and thus leading to a seemingly non-fluorescent protein with *Aaus*FP2.

4.5 Discussion

Does the *Csi*FP4 dimer resembles that of *Av*GFP or that of *Aaus*FP2? Superposing the dimers of *Csi*FP4 and *Aaus*FP2 on one monomer only (**Figure 77**) shows that they both share a similar dimerization interface, different to that of *Av*GFP, which for the record, only has an interface of 850 $Å^2$.

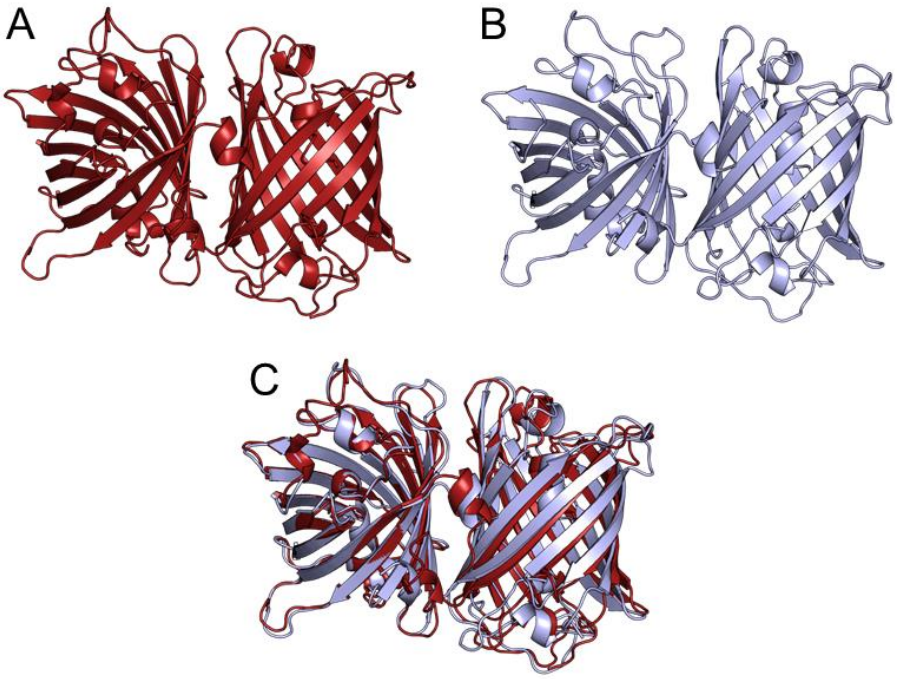

Figure 77. Comparison of *Aaus*FP2 and *Csi*FP4 dimeric interface. (**A**) *Csi*FP4 dimer. (**B**) *Aaus*FP2 dimer. (**C**) Superposition of the two dimers on monomer A only, showing the similarity of their dimerization interface.

Analysis of the interface dimer reveals that monomerization of *Csi*FP4 could be achieved with a limited number of steps. As previously mentioned, the hook interaction could be cancelled by deleting the C-termini, provided it does not affect fluorescence. Close inspection of the dimer interface shows that mutating a small number of residues could be sufficient to disrupt the interaction. There are four sets of interactions, which are actually symmetric so disrupting two of them will disrupt the two others. The first set of interactions (**Interaction 1** and **Interaction 1'**) involves Glu176 on one monomer and Tyr153 and Lys160 on the other one (**Figure 78**). The second set of interactions (**Interaction 2** and **Interaction 2'**) involves Tyr151 on one monomer and the main chain amide nitrogen and/or carbonyl oxygen of Leu149 and Asn164 on the other one. Mutation of Glu176 into a Lysine, for instance, would break the salt bridge interaction with Lys160 and the hydrogen bond to Tyr153. Mutation of Tyr151 into a phenylalanine, for instance, would break the three hydrogen bonds to the main chain atoms of Leu149 and Asn164. In brief, C-terminus truncation and two mutations could be sufficient to disrupt the large dimer interface in *Csi*FP4 and produce a monomeric version.

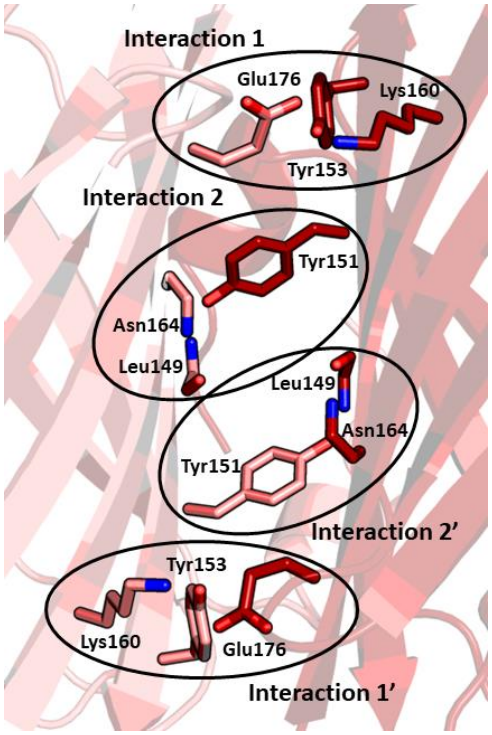

Figure 78. Interactions at the dimer interface in *Csi*FP4.

The structure of *Csi*FP4 has revealed that the two possible isomers *cis* and *trans* of a GFP chromophore covalently bound to a cysteine could be formed. Furthermore, the secondly discovered isomer has been shown to be fluorescent, albeit very weakly, which opens the possibility to improve this naturally found genetic material in order to obtain FPs with a significant level of fluorescence. This improvement will have to fully stabilize the phenolate ring of the chromophore and, maybe, force it into a less distorted conformation. For the record, the red FP eqFP611 is fluorescent with a trans configuration of its planar chromophore, which shows that the nature of the isomer should not be determinant for fluorescence.

Chapter 5

RESULTS

Structural characterization of a family of monomeric phytochrome-based near-infrared fluorescent proteins

5.1 Introduction

5.1.1 Monomeric near infrared FPs (NIR FPs) for multicolour labelling

Near infrared fluorescent proteins (NIR FPs) have been evolved from bacteriophytochromes to take advantage of the optical windows of tissues from living organisms, in which absorption of light is minimized. The first NIR FP, IFP1.4, has been developed from the chromophore-binding domain (CBD) of the bacteriophytochrome from *Deinococcus radiodurans* (Shu *et al*, 2009), where the chromophore is a biliverdin (BV) molecule, a breakdown product of haems. IFP1.4 was later improved into IFP2.0 (Yu *et al*, 2014). The CBD of *D. radiodurans*, *Dr*CBD, is a strong dimer (Wagner *et al*, 2005), and while mutations at the dimer interface have been introduced to obtain the monomeric variants IFP1.4 and IFP2.0, it has been shown that they dimerized at high concentration, with dissociation constants of 7.8 µM and 3.7 µM, respectively (Yu *et al*, 2015). As a consequence, a naturally monomeric bacteriphytochrome from *Badryrhizobium sp.*, a symbiotic bacterium, was restricted to its PAS and GAF-domains, and evolved into the NIR FP mIFP (Yu *et al*, 2015).

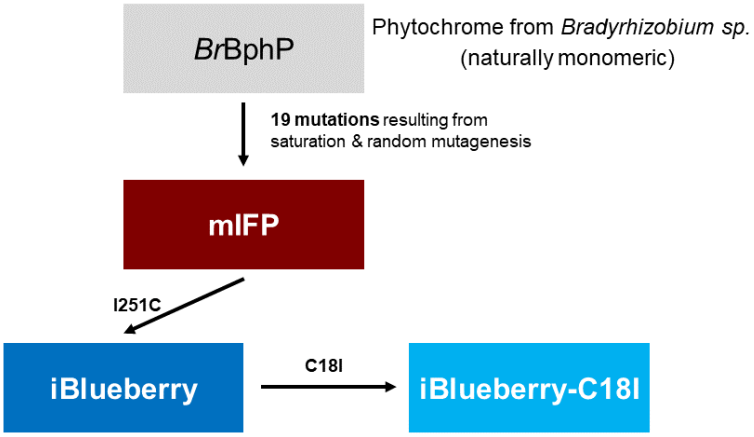

Figure 79. Evolution tree in the iBlueberry family

In order to perform two-colour imaging in the NIR region, attempts were made to blue-shift the excitation and emission peaks of mIFP by mutating residues around the A ring of BV. Reasoning that the cyanobacterial phytochrome Cph1 had attachment of its phycocyanobilin chromophore on a cysteine of the PAS-domain, superposition of the structure of Cph1 (PDB entry code: 2VEA) (Essen *et al*, 2008) and a model of mIFP derived from the IFP2.0 structure (PDB entry code: 4CQH) (Yu *et al*, 2015) indicated that isoleucine

251 was at the location of the PCB-binding cysteine of Cph1. Mutation of this isoleucine into a cysteine (**Figure 79**) induces a ~40 nm shift of the absorption, fluorescence excitation and emission maxima demonstrating that the binding of BV was affected by this mutation. mIFP-I251C was dubbed iBlueberry, and the iBlueberry-C18I mutant is still blue-shifted, indicated that covalent binding on Cys251 is indeed responsible for the blue shift. Excitation and emission spectra of all three proteins are represented on **Figure 80**, and the photophysical parameters can be found in **Table 15**.

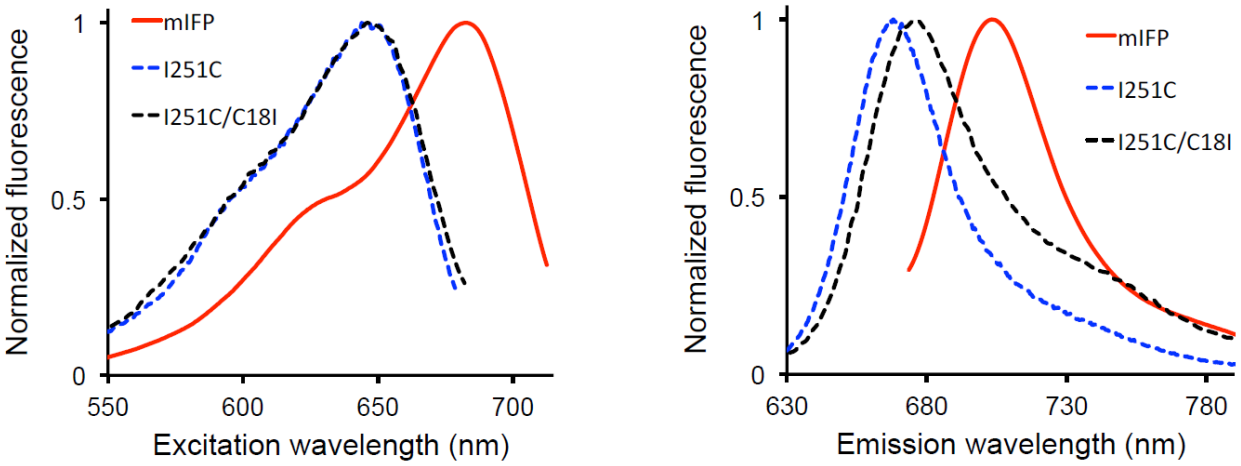

Figure 80. Excitation (left) and emission (right) spectra of mIFP, iBlueberry (mIFP-I251C) and iBluberry-C18I (mIFP-C18I/I251C).

Table 15. Basic photophysical parameters of *Bradyrhizobium* NIR FPs.

Fluorescent Protein (chromophore)	Max. exc. (nm)	Max. em. (nm)	Ext. Coeff. (mM^{-1} cm^{-1})	QY	Brightness	pK_a
iBlueberry[¶]	644	667	38.0	0.07	2.6	<4
iBlueberry-C18I[¶]	644	677	35.0	0.06	1.9	<4
mIFP[¶¶]	683	704	82.0	0.08	6.6	3.5

[¶]Value from (Yu *et al*, 2016), [¶¶]Values are reproduced from the FPbase database https://www.fpbase.org/ (Lambert, 2019)

A slightly different approach was used by the group of Prof. Verkhusha at the Albert Einstein College of Medicine in New York to come to similar conclusions. It had been realized that the first bacteriophytochrome from *Rhodopseudomonas palustris*, *Rp*BphP1, had the most blue-shifted fluorescence properties. Its truncation to the PAS-GAF-domain was subject to saturation mutagenesis at the key residue Asp201 (equivalent to Asp207 in *Dr*CBD) and the adjacent residue Ile202. Random mutagenesis then flow cytometry selection of blue-shifted mutants led to the identification of a mutant which had gained a cysteine in the GAF-domain, while retaining the original cysteine in the PAS-domain. This mutant, BphP-FP has excitation and emission maxima at 639 nm and 669 nm, significantly blue-shifted from other BphP-derived NIR FP from *Rhodopseudomonas palustris* (Shcherbakova *et al*, 2015)

5.1.2 Different binding modes of biliverdin (BV) in phytochromes and phytochrome-derived FPs

In the original 2.5 Å structure of DrCBD, the biliverdine chromophore appeared to be simply linked to Cys24 of the PAS-domain via a thioether bond without affecting the conjugated electron system of the chromophore (i.e. the precise locations of double bonds) (Wagner *et al*, 2005). However, the structure determination of a single-point mutant Y307S of DrCBD at much higher resolution (1.45 A) revealed that the C2 carbon atom becomes sp^3 and goes out of the plane of ring A, cancelling the C3=C2 double bond, and inverting the double and single-bond characters of the C3-$C3^1$ and $C3^1$-C^2 bonds (**Figure 81**) (Wagner *et al*, 2007).

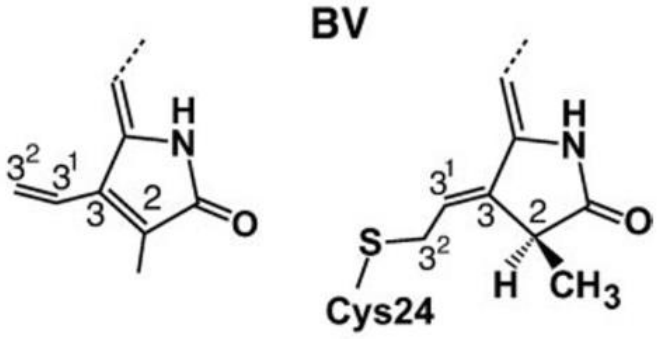

Figure 81. Chemical structure of free and bound biliverdin. (Left) Chemical structure of ring A of free BV. (Right) Chemical structure of ring A of BV bound to a Cysteine of the PAS-domain in a single-point mutant of *Dr*CBD. (adapted from (Wagner *et al*, 2007)

In 2017, the structure determination of the monomeric NIR FPs miRFP670, miRFP703 and miRFP709 derived from Bp revealed the three distinct binding modes of BV on a Cys residue from the PAS-domain, on a Cys residue from the GAF-domain and on two Cys from each of the PAS and GAF-domains (**Figure 82**, top line) (Baloban *et al*, 2017). The first one is slightly different from the binding mode in *Dr*CBD but confers similar spectral properties. The second and third ones are found as a mixture in the structure of miRFP670: the first species has a single covalent bond between the GAF-domain Cys and the $C3^2$ atom of BV chromophore, and the second species has two covalent bonds, one between the PAS-domain Cys and atom $C3^2$ of BV, and one between the GAF-domain Cys and atom $C3^1$ of BV, respectively (**Figure 82**, bottom line). In the last two configurations of BV, $C3^1$ is out of plane, indicating the sp3 character of atom C3 next to atom C2 with sp3 character, showing that ring A has lost conjugation with the rest of the chromophore, which is consistent with the blue shift of the chromophore in miRFP670. Noteworthy, binding to the GAF-domain Cys induces a movement of ~ 1 °Å of the ring A.

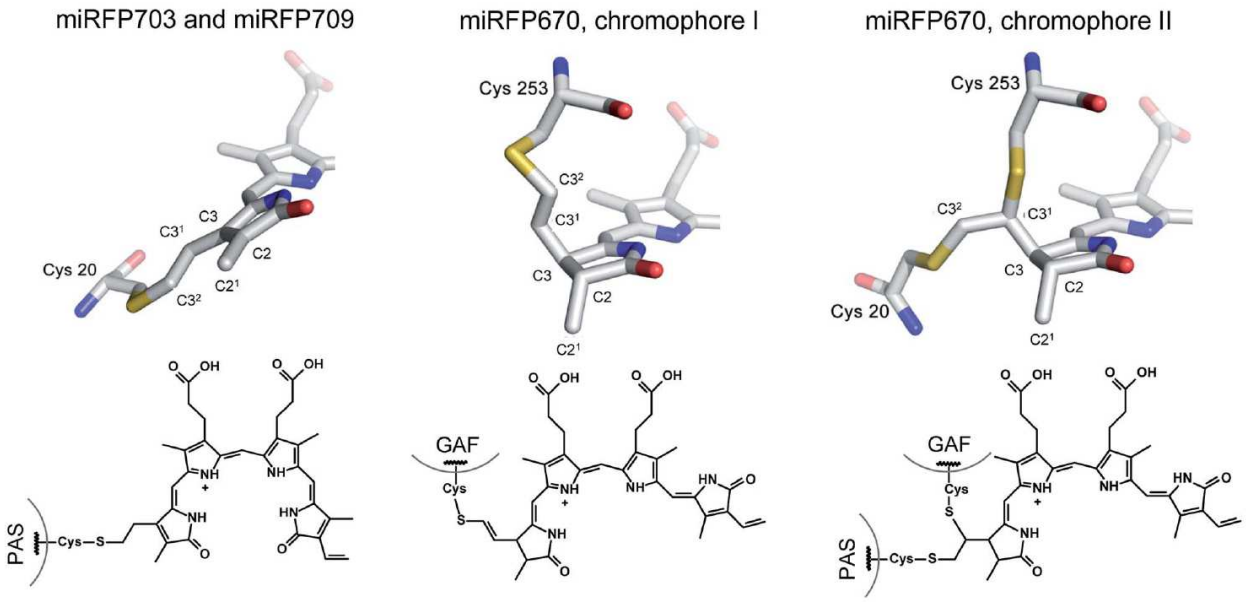

Figure 82. Distinct binding modes of BV in three structures of miRFPs. Top line: miRFP of origin and three-dimensional model. Bottom line: Chemical structures of the three covalently bound BV molecules. (adapted from (Baloban *et al*, 2017).

The chemical mechanism of formation of the three binding modes of BV has been tested by QM/MM simulations (Khrenova *et al*, 2018) and explains, in particular, the two alternate modes of binding of the chromophore in the presence of two Cys, one in the GAF-domain and the other one in the PAS domain. When atom $C3^2$ reacts with the PAS-domain Cys, the $C3^1$ atom can then react to the GAF-domain Cys (**Figure 83, path A**). However, when $C3^2$ reacts with the GAF-domain Cys, $C3^1$ is too far to be able to react with the PAS-domain (**Figure 83, path B**).

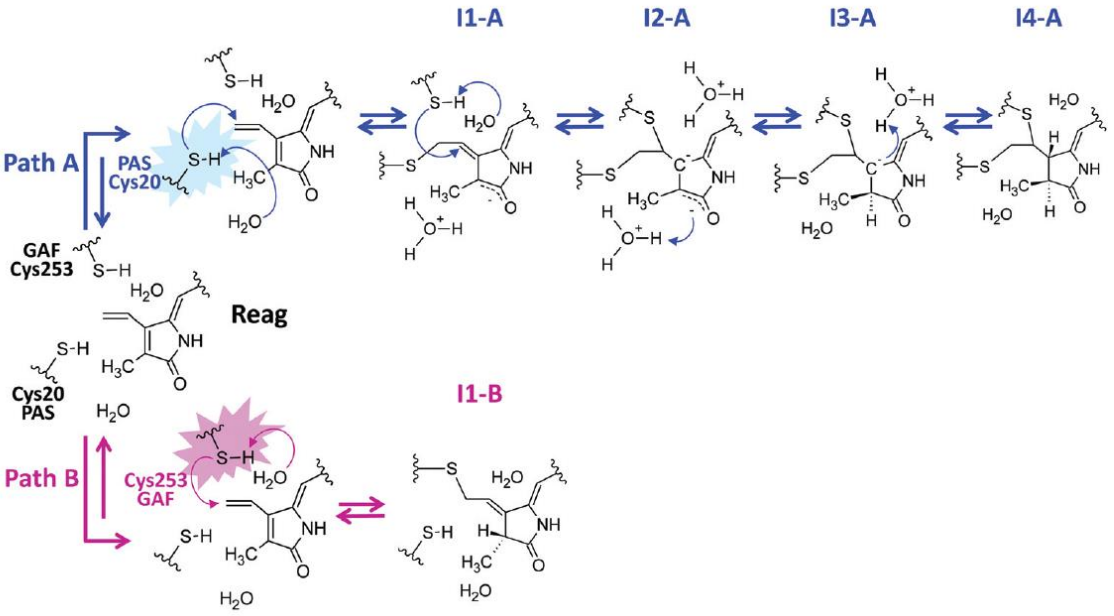

Figure 83. The two competitive paths of BV binding to the PAS and GAF-domain cysteines (adapated from (Khrenova *et al*, 2018).

Finally, a comprehensive study has envisaged all these possible binding modes of one or two cysteines on ring A (Buhrke *et al*, 2019) (**Figure 84**). The BV binding mode in *Dr*CBD is mode **B2**. The BV binding mode in RpBphP-derived NIR FPs with a PAS-domain Cys is mode **B1**. The BV binding modes in miRFP670 are **C1** and **D**.

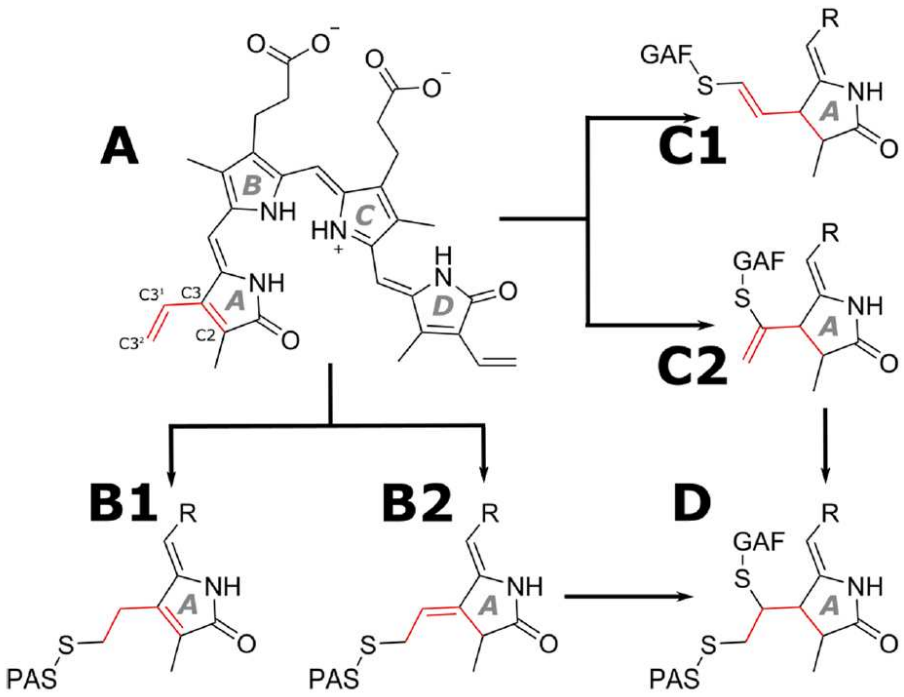

Figure 84. All possible binding modes in NIR FPs possessing cysteine residues on either PAS or GAF-domain, or on both domains. (reproduced from (Buhrke *et al*, 2019).

5.2 Spectroscopic properties

Table 16. Updated basic photophysical parameters of *Bradyrhizobium* NIR FPs.

Fluorescent Protein (chromophore)	Max. exc. (nm)	Max. em. (nm)	Ext. Coeff. (mM^{-1} cm^{-1})	QY	Brightness	pK_a
iBlueberry	647*	666*	38.0¶	0.07¶	2.6¶	<4¶
iBlueberry-C18I	647*	667*	35.0¶	0.06¶	1.9¶	<4¶
mIFP	682*	701*	82.0¶¶	0.08¶¶	6.6¶¶	3.5¶¶

*Values recorded for this work (see § **2.1.8**), ¶Value from (Yu *et al*, 2016), ¶¶Values are reproduced from the FPbase database https://www.fpbase.org/ (Lambert, 2019)

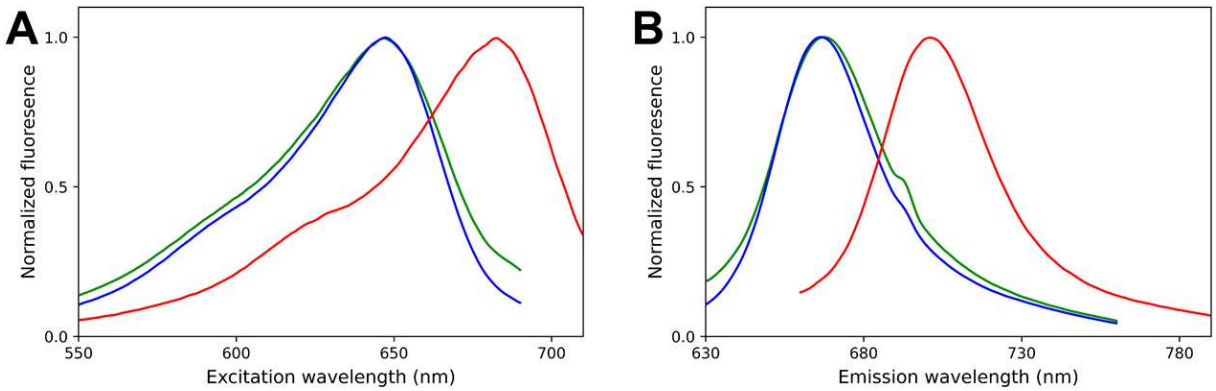

Figure 85. (**A**) Excitation and (**B**) emission spectra of mIFP (red), iBlueberry (mIFP-I251C, blue) and iBluberry-C18I (mIFP-C18I/I251C, green).

To confirm the maximum excitation and emission peak of mIFP, iBlueberry and iBlueberry C18I the fluorescence spectra were remeasured (previous data, **Figure 80**). These new spectra revealed that iBlueberry and iBlueberry C18I show similar spectroscopic properties (**Table 16**, **Figure 85**) contrary to previous results (§ **5.1.1**).

5.3 Structure determination

mIFP and iBlueberry-C18I were crystallized with the hanging drop method using 14% (w/v) PEG 3350 and 0.15 M malic acid (pH 7.3). iBlueberry was crystallized with the batch method using 30% PEG 3350, 0.15 M malic acid (pH 7.9). Diffraction data for all three proteins were collected at the ESRF on beamline ID30A-3. The crystal structure of each protein was solved by molecular replacement using the crystal structure of *Rp*BphP1 from *Rhodopseudomonas palustris* (PDB entry code: 4GW9) as the search model as this structure came up as the top hit in a BLAST search for the proteins. Data reduction and refinement statistics are presented in **Table 17**.

Table 17. Data collection and structure refinement for mIFP, iBlueberry-C18I and iBlueberry. Values in parentheses are for the outer resolution shell, chosen using $CC_{1/2}$ (Evans and Murshudov, 2013).

	mIFP	**iBlueberry C18I**	**iBlueberry**
Data reduction			
Wavelength (Å)	0.9677	0.9677	0.9677
Space group (Å)	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimension:			
A, b, c (Å)	46.4 58.9 98.8	46.5 58.7 98.7	46.6 58.6 98.3
A, β, ɣ (°)	90.0 90.0 90.0	90.0 90.0 90.0	90.0 90.0 90.0
Resolution range (Å)	80.0 – 2.15 (2.21– 2.15)	50.0 – 1.54 (1.64 – 1.54)	50.4 – 1.98 (2.03 – 1.98)
Wilson B factor ($Å^2$)	37.3	29.6	38.3
No. of reflections	47997 (3851)	169526 (29788)	105500 (7998)
Unique reflexions	14675 (1112)	40031 (6879)	19388 (1424)
Multiplicity	2.1 (3.5)	4.2 (4.3)	5.4 (5.6)
Completeness (%)	95.7 (98.7)	98.2 (99.8)	99.6 (99.7)
$\langle I/\sigma(I)\rangle$	9.65 (1.98)	12.87 (1.51)	16.25 (3.09)
R_{meas} (%)	10.1 (60.9)	5.9 (87.6)	6.7 (69.3)
CC1/2	99.7 (74.1)	99.9 (61.9)	99.9 (85.4)
Structure refinement			
Resolution (Å)	42.03– 2.15 (2.21 – 2.15)	29.34 – 1.54 (1.58 – 1.54)	34.21 – 1.98 (2.03 – 1.98)
R_{work} (%)	17.6 (23.0)	17.7 (29.0)	19.2 (26.0)
R_{free} (%)	23.7 (26.0)	21.3 (30.0)	23.2 (26.0)
No. of atoms	2775	2801	2632
Protein	2485	2522	2511
Solvent	290	279	121
B-factors ($Å^2$)	34.8	18.1	39.5
Protein	33.8	15.7	39.2
Solvent	43.5	40.6	46.2
R.m.s. deviations			
Bond lengths (Å)	0.0036	0.0137	0.0058
Bond angles (°)	1.59	2.31	1.94

5.4 Structural analysis

A PISA analysis was performed on all structures to verify the absence of large interaction interfaces. All three IFPs crystallize in the same space group with one molecule in the asymmetric unit and with the same crystal packing. The largest monomer-monomer interface area is 740 Å for mIFP, 760 Å in iBlueberry-C18I, and 860 Å in iBlueberry. For reference the dimer interface of the ancestor protein for IFP1.4 and IFP2.0, *Dr*CBD (PDB entry code: 1ZTU) (Wagner *et al*, 2005) has an area of 1090 Å, while the largest interface areas in the supposedly monomeric variants (at least at low concentration) IFP1.4 (PDB entry code: 5AJG) (Feliks *et al*, 2016) and IFP2.0 (PDB entry code: 4CQH) (Yu *et al*, 2014) are 690 Å and 710 Å, respectively. In brief, the crystal packing analysis does not reveal the presence of a large interface which could serve as dimerization interface.

5.4.1 Structure of mIFP

The structure of mIFP is classically organized as PAS-GAF protein, with a typical knot structure present in all phytochromes (the N-ter of the PAS-domain traverses a loop of the GAF-domain) (**Figure 86**). The chromophore is bound to Cys18 at the beginning of the PAS-domain but most of the stabilizing interactions come from the GAF-domain, which thus contains the chromophore binding pocket.

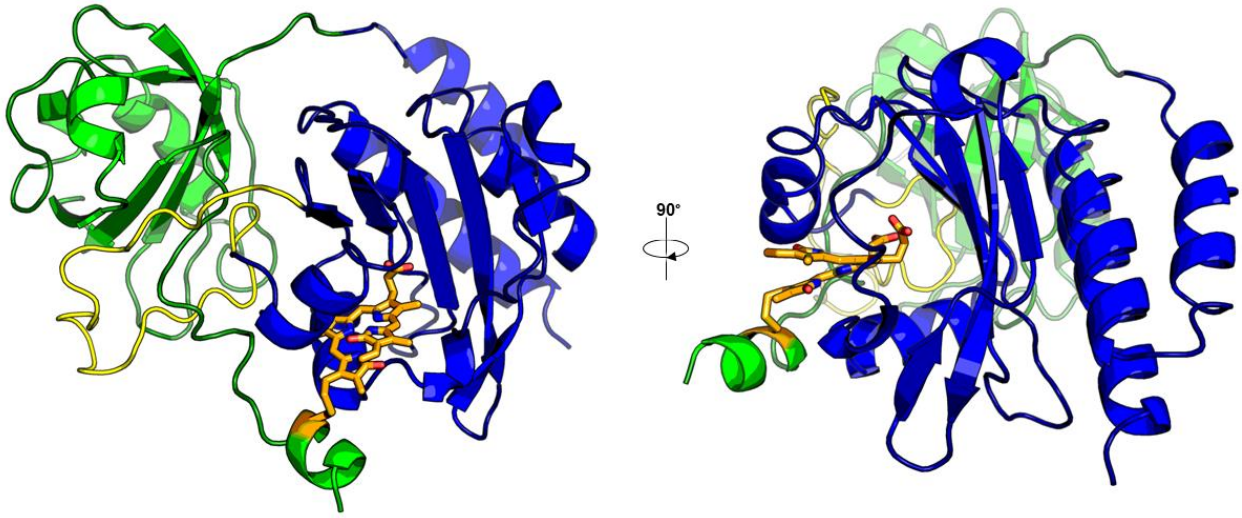

Figure 86. Secondary structure of mIFP. The PAS-domain is represented in green, the GAF-domain in blue, the knot structure in yellow and the chromophore BV in orange. The two views are approximately vertically rotated by 90°.

The chromophore of mIFP has a well-defined single conformation as shown by the unambiguous electron density map (**Figure 87A**). It adopts an unexpected conformation, which we decided to call 'compact', when compared to the structure of biliverdin in phytochromes, whose structure is known. In particular, the conformation of BV in mIFP is distinct from the 'extended' conformation of BV in IFP2.0 (**Figure 87B**).

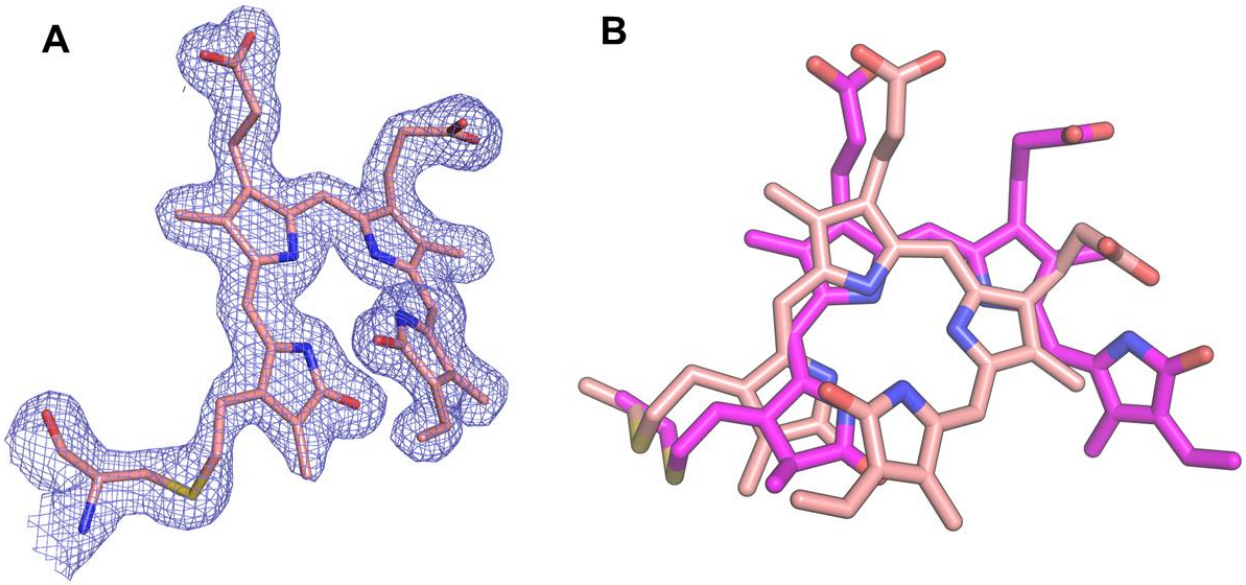

Figure 87. Structure of mIFP chromophore. (**A**) $2F_{obs} - F_{calc}$ electron density map superimposed on the mIFP chromophore contoured at a 1.0 σ level. (**B**) Superposition of the mIFP chromophore (salmon), in a 'compact' conformation and the IFP2.0 chromophore (magenta), in an 'extended' conformation.

The chromophore is bound to Cys18 of the PAS-domain through the vinyl group of its ring A (**Figure 88A**). As ring A is perfectly planar, the binding mode of BV to mIFP is mode **B1** (**Figure 84**). The propionate group of ring B is stabilized by two arginines, while that of ring C is stabilized by a histidine and a tyrosine. The 'compact' conformation of BV in mIFP is explained by two factors. First, ring D is loosely hydrogen-bonded to His252. More importantly ring D is stacked between ring A and the phenolate group of Tyr 255, with which a clear π-π stacking interaction is apparent (**Figure 88B**). Tyr255 seems to be additionally stabilized by the steric constraint of Met259.

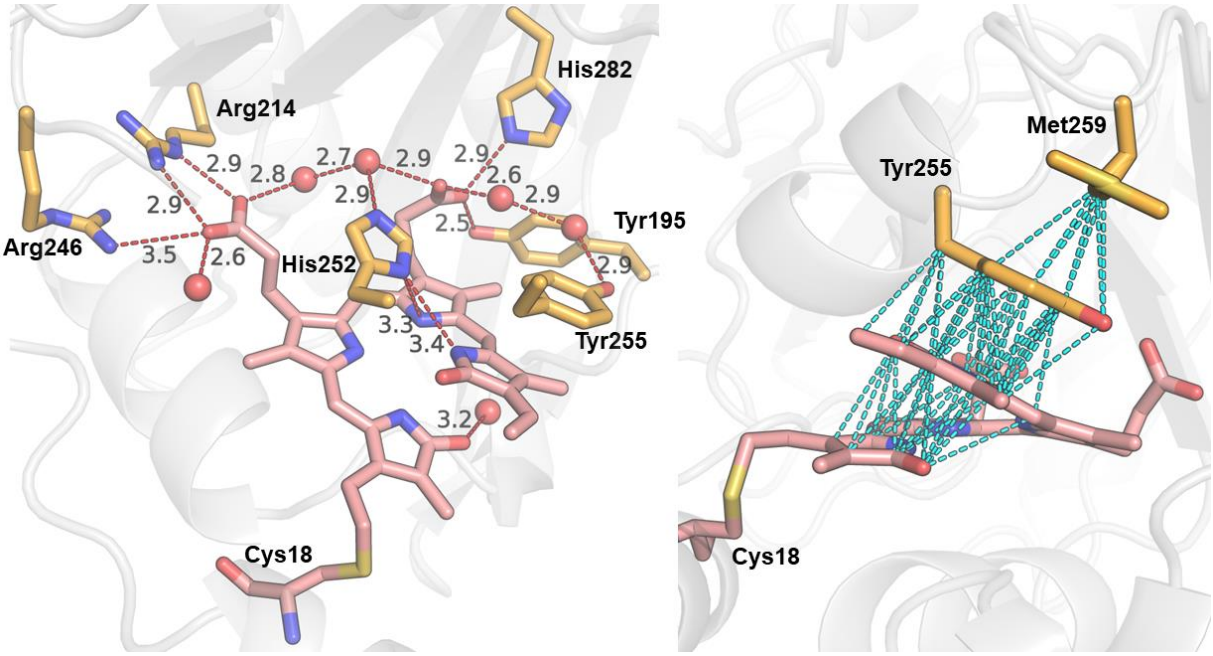

Figure 88. Chromophore environment in mIFP. (**A**) Interactions between the chromophore, water molecules and the protein in mIFP. (**B**) π-π stacking interactions stabilizing ring D of the chromophore in the 'compact' conformation (Structure not deposited yet).

5.4.2 Structure of iBlueberry-C18I

The chromophore of the double-point mutant C18I/I251C of mIFP, iBlueberry-C18I, can be connected only to Cys251, and shows two modes of binding **C1** and **C2** (**Figure 84**), in which the vinyl group attaches either by its $C3^2$ or $C3^1$ atom, respectively, which constitutes a minor structural perturbation. Both configurations adopt the 'compact' conformation first seen in mIFP (**Figure 89, 90**). In the two configurations the $C3^1$ atom is out of the plane of the ring A, as the methyl group on atom C2, which results in a reduced delocalized electron cloud, explaining the blue shift observed in iBlueberry-C18I when compared to mIFP. In other words, there is one less unsaturation on binding mode **C1** and **C2** compared to binding modes **B1** and **B2** resulting on a smaller chromophore conjugated system.

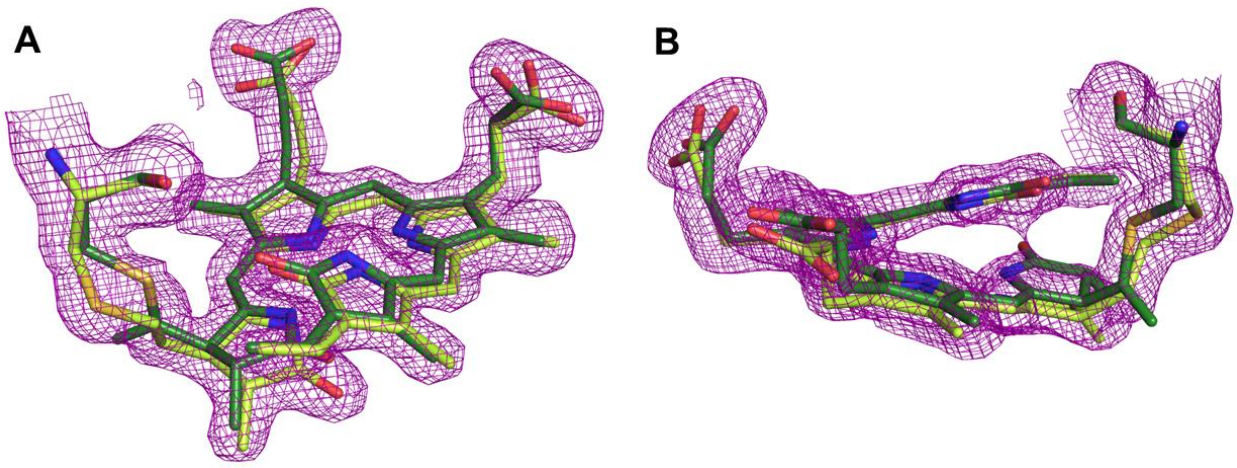

Figure 89. (**A**) $2F_{obs} - F_{calc}$ electron density contoured at a 0.8 σ level superimposed on the two configurations of the iBlueberry-C18I chromophore corresponding to binding modes **C1** and **C2**. (**B**) View from a different angle.

Apart from the change in binding mode, due to attachment to a different cysteine residue, the interactions with the protein are globally conserved. In particular, the 'compact' conformation is conserved with the stacking of the D ring between the A ring and Tyr255, itself blocked by Met259.

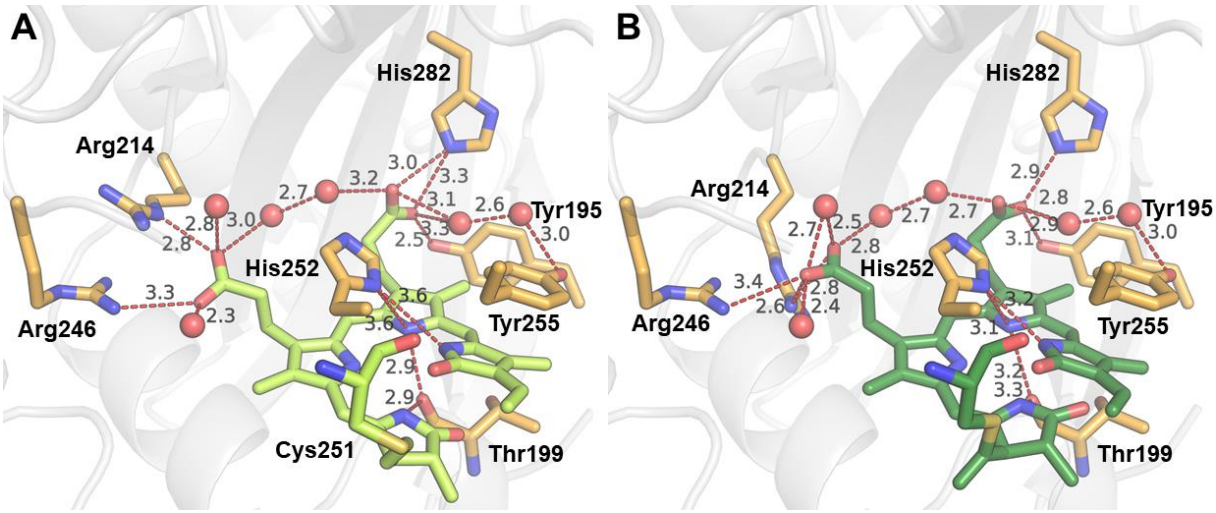

Figure 90. Chromophore environment in iBlueberry-C18I. (**A**) Interactions between the chromophore, water molecules and the protein in the first configuration of the BV in iBlueberry-C18I (binding mode **C1**) (**B**) Interactions between the chromophore, water molecules and the protein in the second configuration of the BV in iBlueberry-C18I (binding mode **C2**) (Structure not deposited yet).

5.4.3 Structure of iBlueberry

The chromophore of the single-point mutant I251C of mIFP, iBlueberry, can be connected to both Cys18 to Cys251. As shown in § **5.1.2**, if BV first binds to Cys18 by its $C3^2$ atom, the $C3^1$ atom can then react with Cys251 to form a doubly covalently bound chromophore. If BV binds to Cys251 by its $C3^2$ atom, the $C3^1$ atom is too far from Cys18 and cannot react with it. The doubly attached is evident from the electron density map. However, the conformation of the chromophore is not clear. I first calculated electron density maps without the model of the cysteines and of the chromophore (**Figure 91A**), which suggested the presence of an 'extended' conformation, which I built. In the corresponding electron density map (**Figure 91B**), there is a large residual density at the location of the D ring of a putative 'compact' conformation, which I then built in alternate conformation with the 'extended' one. The final map (**Figure 91C**) is consistent with partial binding of a BV molecule in binding mode **D** in the 'extended' conformation, and of a BV molecule in binding mode **C1** in the 'compact' conformation.

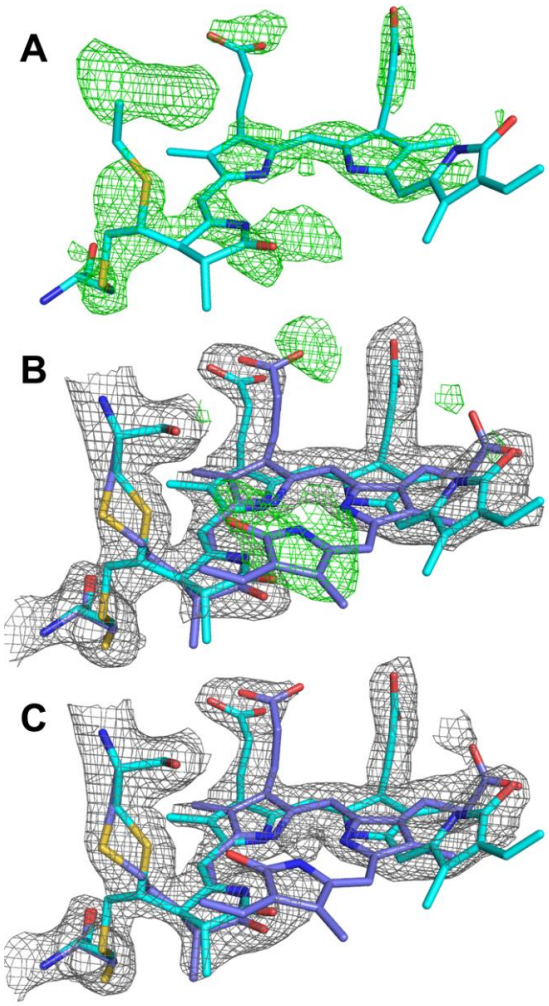

Figure 91. Model building of iBlueberry chromophore alternate configuration. (**A**) $F_{obs} - F_{calc}$ electron density map contoured at a 3.0 σ level calculated without the introduction of the chromophore and of the two cysteines. A doubly covalently bound chromophore (cyan) in an 'extended' conformation was built in the positive density map. (**B**) $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ electron density maps contoured at a 0.8 σ and 3.0 σ level, respectively., superimposed on the model of a doubly covalently bound chromophore in an 'extended' conformation. A chromophore in the 'compact' conformation singly bound to Cys251 (blue) was then modelled in the positive density map. (**C**) $2F_{obs} - F_{calc}$ electron density map contoured at a 0.8 σ level on the final model of the two configurations of the iBlueberry chromophore, which adopts two distinct conformations.

The 'compact' conformation of the chromophore is stabilized similarly to the other conformers in mIFP and iBlueberry-C18I (**Figure 92A**). The 'extended' conformation has clearly a different stabilization (**Figure 92B**), likely due to the strain on the linear chain of the bilin. In particular, there is in-plane rotation of rings B and C, which disrupts the interactions of the propionate groups. Arg214 switches its interaction with the ring B propionate to interact with the ring C propionate. Finally, His282 switches its interaction between the propionate of ring C to the carbonyl group of ring D through a strong hydrogen bond.

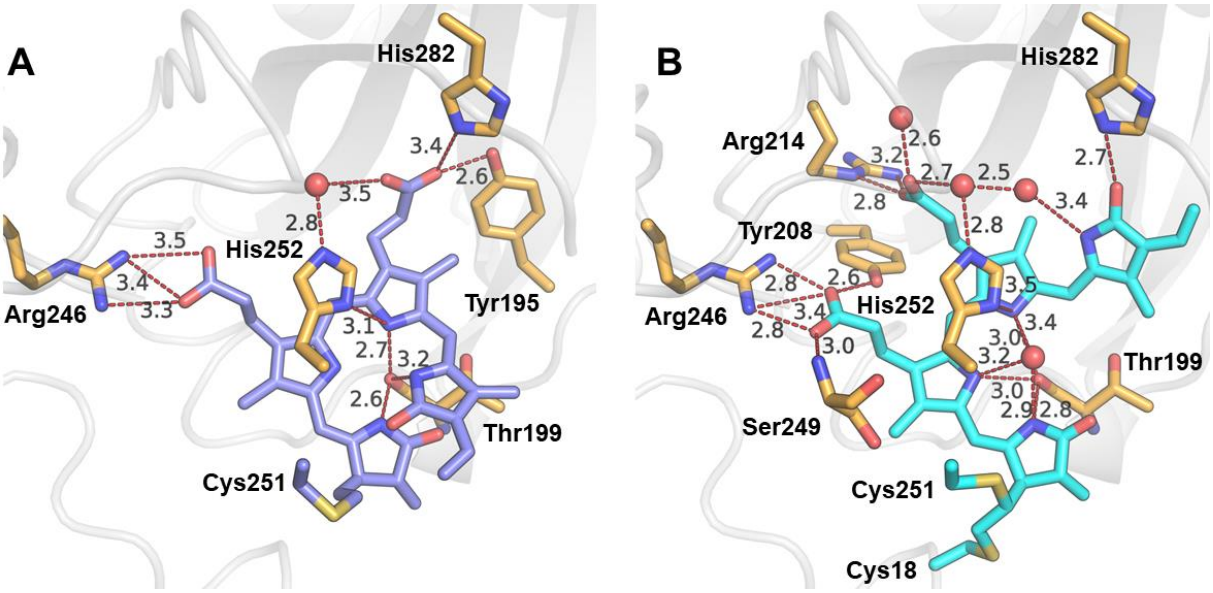

Figure 92. Chromophore environment in iBlueberry. (**A**) Interactions between the chromophore, water molecules and the protein in the first configuration of the BV iBlueberry (binding mode **C1**) (**B**) Interactions between the chromophore, water molecules and the protein in the second configuration of the BV iBlueberry (binding mode **D**) (Structure not deposited yet).

5.5 Discussion on chromophore binding mode and configuration

Structural analysis of all three proteins of the iBlueberry lineage have allowed us to visualize four of the five binding modes of BV to a set of two cysteine residues, one belonging to the PAS-domain, one to the GAF-domain. The fifth mode of binding (**B2**) had already been seen in the group for the structure determination of IFP1.4 and IFP2.0.

The unexpected discovery of this study is the observation of a ‘compact’ conformation of the chromophore, which has, to our knowledge, never been observed before, at least for BV-binding phytochromes and NIR FPs. However, the implications of this conformation are not obvious. Looking at the extinction coefficients of all NIR FPs (**Table 2**), it appears that that mIFP is the lowest of the ‘well-refined’ ones, which are terminal products of various evolution efforts, and that iBlueberry and iBlueberry-C18I have ECs smaller than half of that of mIFP. This suggests that a ‘compact’ conformation has lower chances of catching photons.

Finally, the slightly higher QY of iBlueberry compared to its C18I mutant is consistent with the demonstration that the doubly covalently bound chromophore has a higher QY than the singly covalently bound one (Buhrke *et al*, 2019), as this configuration is only present in iBlueberry.

Chapter 6

CONCLUSION
&
PERSPECTIVES

During this thesis, I have structurally characterized a very bright GFP-like green fluorescent protein, a weakly GFP-like red fluorescent protein and a GFP-like red-absorbing chromoprotein. These three structures have brought very surprising insights on the fluorescence properties of the protein.

The very bright GFP *Aaus*FP1 appears to benefit from an ideal stabilization by van der Waals interactions, a feature that has often been overlooked in rational design of improved FPs, due to the difficulty of rationalizing a cohesive effect of several residues at the same time. However, there have been a few examples of success (Goedhart *et al*, 2012) and this path of optimization is worth exploring further. *Aaus*FP1 thanks to its high brightness, low pKa and narrow excitation and emission peak, appears to be ideal for multiplexing colour imaging. However several obstacle need to be overcome. The first one is the monomerisation of the protein and the recovery of the fluorescence if brightness decreases due to the monomerisation process. The second obstacle is the increase of the Stokes shift of the protein. For now, the shift is only of 6 nm, which is not ideal for the selection of appropriate excitation and emission filters. The modification of the chromophore environment by mutagenesis needs to be pursued to increase the flexibility of the chromophore rings to extend the Stokes shift of the protein at the cost of a small drop in brightness (Piatkevich *et al*, 2013).

The structure of the red-absorbing chromoprotein *Aaus*FP2 revealed for the first time a chromophore with a unanticipated modification at the middle of the methylene bridge by a nearby cysteine residue, which brings sulphur redox chemistry in the realm of GFP-like chromophore biogenesis (apart from the peculiar mKO chromophore) as a strong nucleophile. Exploring oxygen chemistry has already been envisaged with a tyrosine as in LaRFP (Pletnev *et al*, 2013), but oxygen is a weaker nucleophile, the bond appears particularly sensitive and a tyrosine side chain is difficult to position and stabilize due to the steric hindrance of the side chain, although it has the length advantage compared to a cysteine and can interact at a greater distance when compared to a cysteine residue. *Aaus*FP2 is an interesting red absorbing chromoprotein thanks to the broad absorption peak with potential application as a FRET acceptor or for photoacoustic imaging (Li *et al*, 2016). Further mutagenesis need to be done to extend the extinction coefficient of this protein to improve its use in this type of applications. Moreover, *Aaus*FP2 revealed to be sensitive to the presence of oxygen inducing with the conjunction of light absorption the cleavage of the phenolate of the chromophore (data not shown). Depending on the results of an extensive characterisation, this observation could lead to the development of a potential new oxygen sensor.

The third structure that I solved of a GFP-like protein was that of a weakly red fluorescent, *Csi*FP4, which had, very surprisingly the same chromophore as *Aaus*FP2, although as a different isomer. The surprise also came from the fact that they originate from jellyfish species of different genii. FPs in the red to far-red part of the spectrum (RFPs and far-red RFPs) have so far resisted optimization of their fluorescence quantum yield, with the notable exception of mScarlet. Until now, three major types of chromophore had been identified in RFPs (**Figure 93**): (1) the further conjugation of the GFP chromophore electron conjugation system (phenolate and imidazolinone rings) to the protein main chain and the formation of a N-acylamine bond ('DsRed-like'), (2) the further conjugation to the indole ring of the first residue of the chromophore through photoinduced cleavage of the main chain ('Kaede-like') and (3) the covalent binding of a tyrosine residue to the central carbon of the chromophore methylene bridge ('LaRFP-like'). The *Aaus*FP2 and *Csi*FP4 chromophore thus constitutes a fourth type of potentially red-emitting chromophore. As the cysteine residue attaches to the methylene bridge of the (red fluorescent) chromophore, we can call it 'CysRFCh' or '*Aaus*FP2-like'. Given the apparent simplicity of cysteine chemistry (compared to that with a tyrosine), this intriguing chromophore requests its formation mechanism to be investigated so as to see if the fluorescence efficiency can be significantly improved, and even tuned by action of light.

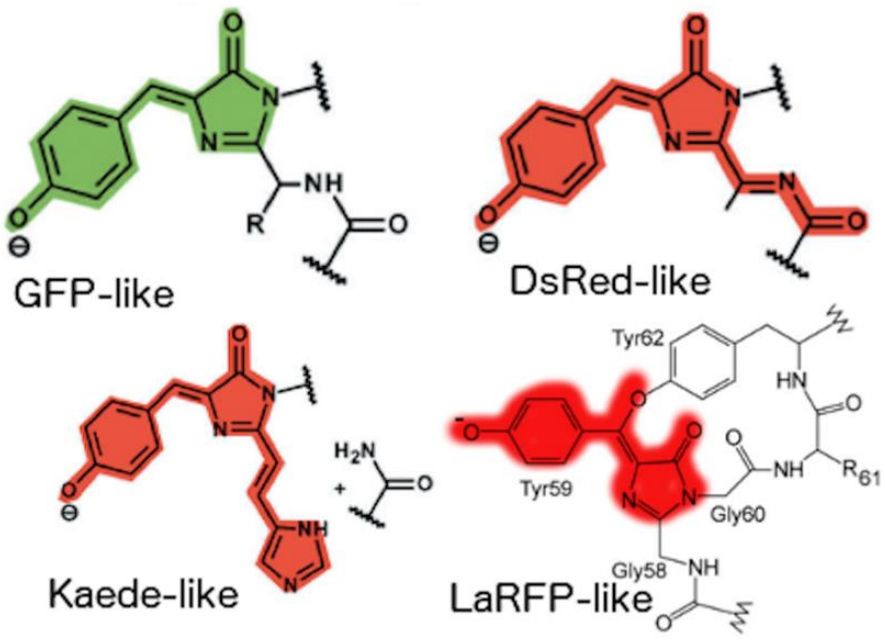

Figure 93. Prototypical chromophore in Green Fluorescent Protein (green) and three different types of chromophore in Red Fluorescent Proteins.

My structural investigation of the *Bradyrhizobium* NIR IFPs has allowed me to explore the complex chemistry of binding mode of a bilin to a cysteine. In particular, I have learned that rational design could lead to a double attachment of biliverdin to two different cysteines, however with only a ~50% efficiency, a single attachment chromophore being formed at the same time. It has been shown that this doubly attached chromophore is more fluorescent than the singly attached one (Buhrke *et al*, 2019). Consequently, it is desirable to explore whether it is possible to favour by mutagenesis the former mode of binding compared to the latter. Further work needs to be performed to understand the spectroscopic relevance of the "compact" conformation present in mIFP and iBlueberry-C18I. Preliminary observation of the extinction coefficient when compared to the other NIR FPs with an "extended" chromophore show a decrease in EC for mIFP and the two mutants. This fact shows that mIFP can be further improved by destabilising the current conformation by either mutating Tyr255 involved in the π-stacking of the chromophore or by trying to modify the direct environment of the chromophore and in particular the region between residue 194 and 201 (mIFP), a zone always disordered in iBlueberry-C18I and iBlueberry, to favour a stable "extended" conformation.

Abrahams M V. & Townsend LD (1993) Bioluminescence in Dinoflagellates: A Test of the Burgular Alarm Hypothesis. *Ecology* 74: 258–260

Adam V, Lelimousin M, Boehme S, Desfonds G, Nienhaus K, Field MJ, Wiedenmann J, McSweeney S, Nienhaus GU & Bourgeois D (2008) Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations. *Proc Natl Acad Sci U S A* 105: 18343–18348

Ai HW, Shaner NC, Cheng Z, Tsien RY & Campbell RE (2007) Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* 46: 5904–5910

Ando R, Hama H, Yamamoto-Hino M, Mizuno H & Miyawaki A (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci U S A* 99: 12651–12656

Ando R, Mizuno H & Miyawaki A (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science (80-)* 306: 1370–1373

Auldridge ME, Satyshur KA, Anstrom DM & Forest KT (2012) Structure-guided engineering enhances a phytochrome-based infrared fluorescent protein. *J Biol Chem* 287: 7000–9

Aumonier S, Santoni G, Gotthard G, Von Stetten D, Leonard GA & Royant A (2020) Millisecond time-resolved serial oscillation crystallography of a blue-light photoreceptor at a synchrotron. *IUCrJ* 7: 728–736

Baloban M, Shcherbakova DM, Pletnev S, Pletnev VZ, Lagarias JC & Verkhusha V V (2017) Designing brighter near-infrared fluorescent proteins: insights from structural and biochemical studies. *Chem Sci* 8: 4546–4557

Baumann D, Cook M, Ma L, Mushegian A, Sanders E, Schwartz J & Yu CR (2008) A family of GFP-like proteins with different spectral properties in lancelet Branchiostoma floridae. *Biol Direct* 3: 28

Bellini D & Papiz MZ (2012) Structure of a Bacteriophytochrome and Light-Stimulated Protomer Swapping with a Gene Repressor. *Structure* 20: 1436–1446

Belousov V V., Fradkov AF, Lukyanov KA, Staroverov DB, Shakhbazov KS, Terskikh A V. & Lukyanov S (2006) Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. *Nat Methods* 3: 281–286

Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, *et al* (2017) mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* 14: 53–56

Blot N, Wu XJ, Thomas JC, Zhang J, Garczarek L, Böhm S, Tu JM, Zhou M, Plöscher M, Eichacker L, *et al* (2009) Phycourobilinin trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *J Biol Chem* 284: 9290–9298

Boersma AJ, Zuhorn IS & Poolman B (2015) A sensor for quantification of macromolecular crowding in living cells. *Nat Methods* 12: 227–229

Bork P, Bowler C, De Vargas C, Gorsky G, Karsenti E & Wincker P (2015) Tara Oceans studies plankton at Planetary scale. *Science (80-)* 348: 873 doi:10.1126/science.aac5605

Bourgeois D, Regis-Faro A & Adam V (2012) Photoactivated structural dynamics of

fluorescent proteins. In *Biochemical Society Transactions* pp 531–538. Biochem Soc Trans

Bregestovski P, Waseem T & Mukhtarov M (2009) Genetically encoded optical sensors for monitoring of intracellular chloride and chloride-selective channel activity. *Front Mol Neurosci* 2: 15 doi:10.3389/neuro.02.015.2009

Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormö M & Remington SJ (1997) Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc Natl Acad Sci U S A* 94: 2306–2311

Buhrke D, Tavraz NN, Shcherbakova DM, Sauthof L, Moldenhauer M, Vélazquez Escobar F, Verkhusha V V., Hildebrandt P & Friedrich T (2019) Chromophore binding to two cysteines increases quantum yield of near-infrared fluorescent proteins. *Sci Rep* 9: 1866

Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA & Tsien RY (2002) A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* 99: 7877–82

Chalfie M, Tu Y, Euskirchen G, Ward WW & Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* 263: 802–5

Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, Roberts AG, Oparka KJ & Christie JM (2008) The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc Natl Acad Sci U S A* 105: 20038–43

Christie JM (2007) Phototropin blue-light receptors. *Annu Rev Plant Biol* 58: 21–45 doi:10.1146/annurev.arplant.58.032806.103951

Cipriani F, Röwer M, Landret C, Zander U, Felisaz F & Márquez JA (2012) CrystalDirect: A new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr Sect D Biol Crystallogr* 68: 1393–1399

Clavel D, Gotthard G, von Stetten D, De Sanctis D, Pasquier H, Lambert GG, Shaner NC & Royant A (2016) Structural analysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution. *Acta Crystallogr Sect D, Struct Biol* 72: 1298–1307

Costantini LM, Fossati M, Francolini M & Snapp EL (2012) Assessing the Tendency of Fluorescent Proteins to Oligomerize Under Physiologic Conditions. *Traffic* 13: 643–649

Crameri A, Whitehorn EA, Tate E & Stemmer WPC (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 14: 315–319

Crystal Structures of the Luciferase and Green Fluorescent Protein from Renilla reniformis (2007) *J Mol Biol* 374: 1017–1028

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA & Tsien RY (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* 20: 448–455

Cubitt AB, Woollenweber LA & Heim R (1999) Understanding structure - Function relationships in the Aequorea victoria green fluorescent protein. *Methods Cell Biol* 58: 19–30

Day RN & Davidson MW (2009) The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* 38: 2887–921

Dimasi N, Flot D, Dupeux F & Márquez JA (2007) Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory

receptor expressed on myeloid cells IREM-1. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 63: 204–208

Dong A, Xu X, Edwards AM, Chang C, Chruszcz M, Cuff M, Cymborowski M, Di Leo R, Egorova O, Evdokimova E, *et al* (2007) In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 4: 1019–1021

Donner JS, Thompson SA, Kreuzer MP, Baffou G & Quidant R (2012) Mapping intracellular temperature using green fluorescent protein. *Nano Lett* 12: 2107–2111

Drepper T, Eggert T, Circolone F, Heck A, Krauß U, Guterl J-K, Wendorff M, Losi A, Gärtner W & Jaeger K-E (2007) Reporter proteins for in vivo fluorescence without oxygen. *Nat Biotechnol* 25: 443–445

Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501

Essen LO, Mailliet J & Hughes J (2008) The structure of a complete phytochrome sensory module in the Pr ground state. *Proc Natl Acad Sci U S A* 105: 14709–14714

Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69: 1204–14

Feliks M, Lafaye C, Shu X, Royant A & Field M (2016) Structural Determinants of Improved Fluorescence in a Family of Bacteriophytochrome-Based Infrared Fluorescent Proteins: Insights from Continuum Electrostatic Calculations and Molecular Dynamics Simulations. *Biochemistry* 55: 4263–4274

Filonov GS, Piatkevich KD, Ting L-M, Zhang J, Kim K & Verkhusha V V (2011) Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat Biotechnol* 29: 757–761

Flot D, Mairs T, Giraud T, Guijarro M, Lesourd M, Rey V, Van Brussel D, Morawe C, Borel C, Hignette O, *et al* (2010) The ID23-2 structural biology microfocus beamline at the ESRF. *J Synchrotron Radiat* 17: 107–118

Fuenzalida-Werner JP, Janowski R, Mishra K, Weidenfeld I, Niessing D, Ntziachristos V & Stiel AC (2018) Crystal structure of a biliverdin-bound phycobiliprotein: Interdependence of oligomerization and chromophorylation. *J Struct Biol* 204: 519–522

Gaietta G, Deerinck TJ, Adams SR, Bouwer J, Tour O, Laird DW, Sosinsky GE, Tsien RY & Ellisman MH (2002) Multicolor and electron microscopic imaging of connexin trafficking. *Science (80-)* 296: 503–507

Galea HR, Häussermann V & Försterra G (2007) Hydrozoa, fjord Comau, Chile. *Check List* 3: 159

Gert-Jan Kremers, Joachim Goedhart, Erik B. van Munster and & Theodorus W. J. Gadella J. (2006) Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius†,‡.

Giraud E, Zappa S, Vuillet L, Adriano JM, Hannibal L, Fardoux J, Berthomieu C, Bouyer P, Pignol D & Verméglio A (2005) A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in Rhodopseudomonas palustris. *J Biol Chem* 280: 32389–32397

Glazer AN & Stryer L (1984) Phycofluor probes. *Trends Biochem Sci* 9: 423–427

Goedhart J, Von Stetten D, Noirclerc-Savoye M, Lelimousin M, Joosen L, Hink MA, Van

Weeren L, Gadella TWJ & Royant A (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat Commun* 3: 1–9

Goedhart J, van Weeren L, Hink MA, Vischer NOE, Jalink K & Gadella TWJ (2010) Bright cyan fluorescent protein variants identified by fluorescence lifetime screening. *Nat Methods* 7: 137–139

Griesbeck O, Baird GS, Campbell RE, Zacharias DA & Tsien RY (2001) Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J Biol Chem* 276: 29188–94

Grigorenko BL, Krylov AI & Nemukhin A V. (2017) Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *J Am Chem Soc* 139: 10239–10249

Gross LA, Baird GS, Hoffman RC, Baldridge KK & Tsien RY (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A* 97: 11990–11995

Gurskaya NG, Fradkov AF, Terskikh A, Matz M V, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA & Lukyanov SA (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett* 507: 16–20

Gurskaya NG, Verkhusha V V., Shcheglov AS, Staroverov DB, Chepurnykh T V., Fradkov AF, Lukyanov S & Lukyanov KA (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat Biotechnol* 24: 461–465

Haddock SHD, Mastroianni N & Christianson LM (2010) A photoactivatable green-fluorescent protein from the phylum Ctenophora. *Proc R Soc B Biol Sci* 277: 1155–1160

Hanson GT, Aggeler R, Oglesbee D, Cannon M, Capaldi RA, Tsien RY & Remington SJ (2004) Investigating Mitochondrial Redox Potential with Redox-sensitive Green Fluorescent Protein Indicators. *J Biol Chem* 279: 13044–13053

Hayashi S & Toda Y (2009) A novel fluorescent protein purified from eel muscle. *Fish Sci* 75: 1461–1469

Heim R, Prasher DC & Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci* 91: 12501–12504

Heim R & Tsien RY (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* 6: 178–82

Hunt ME, Scherrer MP, Ferrari FD & Matz M V. (2010) Very Bright Green Fluorescent Proteins from the Pontellid Copepod Pontella mimocerami. *PLoS One* 5: e11517

Imamura H, Huynh Nhat KP, Togawa H, Saito K, Iino R, Kato-Yamada Y, Nagai T & Noji H (2009) Visualization of ATP levels inside single living cells with fluorescence resonance energy transfer-based genetically encoded indicators. *Proc Natl Acad Sci U S A* 106: 15651–15656

Jancarik J & Kim SH (1991) Sparse matrix sampling. A screening method for crystallization of proteins. *J Appl Crystallogr* 24: 409–411

Juanhuix J, Gil-Ortiz F, Cuní G, Colldelram C, Nicolás J, Lidón J, Boter E, Ruget C, Ferrer S & Benach J (2014) Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron. *J Synchrotron Radiat*

21: 679–689

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66: 125–32

Kam Z, Shore HB & Feher G (1978) On the crystallization of proteins. *J Mol Biol* 123: 539–555

Karasawa S, Araki T, Nagai T, Mizuno H & Miyawaki A (2004) Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochem J* 381: 307–312

Katayama H, Kogure T, Mizushima N, Yoshimori T & Miyawaki A (2011) A sensitive and quantitative technique for detecting autophagic events based on lysosomal delivery. *Chem Biol* 18: 1042–1052

Khlebnikov A, Risa, Skaug T, Carrier TA & Keasling JD (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol* 182: 7029–7034

Khrenova MG, Kulakova AM & Nemukhin A V. (2018) Competition between two cysteines in covalent binding of biliverdin to phytochrome domains. *Org Biomol Chem*

Kikuchi A, Fukumura E, Karasawa S, Mizuno H, Miyawaki A & Shiro Y (2008) Structural characterization of a thiazoline-containing chromophore in an orange fluorescent protein, monomeric kusabira orange. *Biochemistry* 47: 11573–11580

Koay MS, Janssen BMG & Merkx M (2013) Tuning the metal binding site specificity of a fluorescent sensor protein: From copper to zinc and back. *Dalt Trans* 42: 3230–3232

Kogure T, Karasawa S, Araki T, Saito K, Kinjo M & Miyawaki A (2006) A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nat Biotechnol* 24: 577–581

Kremers GJ, Goedhart J, Van Den Heuvel DJ, Gerritsen HC & Gadella TWJ (2007) Improved green and blue fluorescent proteins for expression in bacteria and mammalian cells. *Biochemistry* 46: 3775–3783

Krissinel E & Henrick K (2005) Detection of protein assemblies in crystals. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* pp 163–174. Springer, Berlin, Heidelberg

Krissinel E & Henrick K (2007) Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* 372: 774–797

Kumagai A, Ando R, Miyatake H, Greimel P, Kobayashi T, Hirabayashi Y, Shimogori T & Miyawaki A (2013) A Bilirubin-Inducible Fluorescent Protein from Eel Muscle. *Cell* 153: 1602–1611

Kuner T & Augustine GJ (2000) A genetically encoded ratiometric indicator for chloride: Capturing chloride transients in cultured hippocampal neurons. *Neuron* 27: 447–459

Lambert GG, Depernet H, Gotthard G, Schultz DT, Navizet I, Lambert T, Adams SR, Torreblanca-Zanca A, Chu M, Bindels DS, *et al* (2020) Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. *PLOS Biol* 18: e3000936

Lambert TJ (2019) FPbase: a community-editable fluorescent protein database. *Nat Methods* 16: 277–278 doi:10.1038/s41592-019-0352-8

Lelimousin M, Noirclerc-Savoye M, Lazareno-Saez C, Paetzold B, Le Vot S, Chazal R, Macheboeuf P, Field MJ, Bourgeois D & Royant A (2009) Intrinsic dynamics in ECFP and cerulean control fluorescence quantum yield. *Biochemistry* 48: 10038–10046

Li Y, Forbrich A, Wu J, Shao P, Campbell RE & Zemp R (2016) Engineering Dark Chromoprotein Reporters for Photoacoustic Microscopy and FRET Imaging. *Sci Rep* 6: 22129

Lindenburg LH, Vinkenborg JL, Oortwijn J, Aper SJA & Merkx M (2013) MagFRET: The first genetically encoded fluorescent Mg2+ sensor. *PLoS One* 8

Lukyanov KA, Fradkov AF, Gurskaya NG, Matz M V., Labas YA, Savitsky AP, Markelov ML, Zaraisky AG, Zhao X, Fang Y, *et al* (2000) Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J Biol Chem* 275: 25879–25882

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, *et al* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47: W636–W641

Márquez JA & Cipriani F (2014) CrystalDirectTM: A novel approach for automated crystal harvesting based on photoablation of thin films. *Methods Mol Biol* 1091: 197–203

Martins PM, Rocha F & Damas AM (2008) Understanding water equilibration fundamentals as a step for rational protein crystallization. *PLoS One* 3

Matz M V., Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML & Lukyanov SA (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* 17: 969–973

McCarthy AA, Barrett R, Beteva A, Caserotto H, Dobias F, Felisaz F, Giraud T, Guijarro M, Janocha R, Khadrouche A, *et al* (2018) ID30B – a versatile beamline for macromolecular crystallography experiments at the ESRF. *J Synchrotron Radiat* 25: 1249–1260

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674

McIntosh JR (2001) Electron microscopy of cells: A new beginning for a new century. *J Cell Biol* 153: f25 doi:10.1083/jcb.153.6.F25

Mena MA, Treynor TP, Mayo SL & Daugherty PS (2006) Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat Biotechnol* 24: 1569–1571

Merola F, Levy B, Demachy I & Pasquier H (2010) Photophysics and Spectroscopy of Fluorophores in the Green Fluorescent Protein Family. In pp 347–383. Springer, Berlin, Heidelberg

Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Gadella TWJ, *et al* (2007) Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* 4: 555–557

Migita CT, Zhang X & Yoshida T (2003) Expression and characterization of cyanobacterium heme oxygenase, a key enzyme in the phycobilin synthesis: Properties of the heme complex of recombinant active enzyme. *Eur J Biochem* 270: 687–698

Mishin AS, Subach F V, Yampolsky I V, King W, Lukyanov KA & Verkhusha V V (2008)

The first mutant of the Aequorea victoria green fluorescent protein that forms a red chromophore. *Biochemistry* 47: 4666–73

Miyawaki A, Llopis J, Heim R, Michael McCaffery J, Adams JA, Ikura M & Tsien RY (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* 388: 882–887

Mizuno H, Mal TK, Tong KI, Ando R, Furuta T, Ikura M & Miyawaki A (2003) Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. *Mol Cell* 12: 1051–1058

Morise H, Shimomura O, Johnson FH & Winant J (1974) Intermolecular energy transfer in the bioluminescent system of aequorea. *Biochemistry* 13: 2656–2662

Murphy JT & Lagarias JC (1997) The phytofluors: A new class of fluorescent protein probes. *Curr Biol* 7: 870–876

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67: 355–67

Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K & Miyawaki A (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* 20: 87–90

Nakai J, Ohkura M & Imoto K (2001) A high signal-to-noise ca2+ probe composed of a single green fluorescent protein. *Nat Biotechnol* 19: 137–141

Nazarenko V V., Remeeva A, Yudenko A, Kovalev K, Dubenko A, Goncharov IM, Kuzmichev P, Rogachev A V., Buslaev P, Borshchevskiy V, *et al* (2019) A thermostable flavin-based fluorescent protein from: Chloroflexus aggregans: A framework for ultra-high resolution structural studies. *Photochem Photobiol Sci* 18: 1793–1805

Nguyen AW & Daugherty PS (2005) Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* 23: 355–360

Nifosì R & Tozzini V (2006) Cis-trans photoisomerization of the chromophore in the green fluorescent protein variant E2GFP: A molecular dynamics study. *Chem Phys* 323: 358–368

Oliinyk OS, Shemetov AA, Pletnev S, Shcherbakova DM & Verkhusha V V. (2019) Smallest near-infrared fluorescent protein evolved from cyanobacteriochrome as versatile tag for spectral multiplexing. *Nat Commun* 10: 279

Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY & Remington SJ (1996) Crystal structure of the Aequorea victoria green fluorescent protein. *Science* 273: 1392–5

Painter J & Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr* 39: 109–111

Pakhomov AA & Martynov VI (2008) GFP Family: Structural Insights into Spectral Tuning. *Chem Biol* 15: 755–764 doi:10.1016/j.chembiol.2008.07.009

Palm GJ, Zdanov A, Gaitanaris GA, Stauber R, Pavlakis GN & Wlodawer A (1997) The structural basis for spectral variations in green fluorescent protein. *Nat Struct Mol Biol* 4: 361–365

Patterson GH & Lippincott-Schwartz J (2002) A photoactivatable GFP for selective

photolabeling of proteins and cells. *Science (80-)* 297: 1873–1877

Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24: 79–88

Pérez Koldenkova V & Nagai T (2013) Genetically encoded Ca2+ indicators: Properties and evaluation. *Biochim Biophys Acta - Mol Cell Res* 1833: 1787–1797

Perrakis A, Morris R & Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6: 458–463

Petersen J, Wilmann PG, Beddoe T, Oakley AJ, Devenish RJ, Prescott M & Rossjohn J (2003) The 2.0-Å Crystal Structure of eqFP611, a Far Red Fluorescent Protein from the Sea Anemone Entacmaea quadricolor. *J Biol Chem* 278: 44626–44631

Piatkevich KD, Jung EE, Straub C, Linghu C, Park D, Suk H-J, Hochbaum DR, Goodwin D, Pnevmatikakis E, Pak N, *et al* (2018) A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat Chem Biol* 14: 352–360

Piatkevich KD, Malashkevich VN, Morozova KS, Nemkovich NA, Almo SC & Verkhusha V V. (2013) Extended stokes shift in fluorescent proteins: Chromophore-protein interactions in a near-infrared TagRFP675 variant. *Sci Rep* 3

Pletnev VZ, Pletneva N V, Lukyanov KA, Souslova EA, Fradkov AF, Chudakov DM, Chepurnykh T, Yampolsky I V, Wlodawer A, Dauter Z, *et al* (2013) Structure of the red fluorescent protein from a lancelet (Branchiostoma lanceolatum): a novel GYG chromophore covalently bound to a nearby tyrosine. *Acta Crystallogr D Biol Crystallogr* 69: 1850–60

Pletneva N V., Pletnev VZ, Souslova E, Chudakov DM, Lukyanov S, Martynov VI, Arhipova S, Artemyev I, Wlodawer A, Dauter Z, *et al* (2013a) Yellow fluorescent protein phiYFPv (*Phialidium*): structure and structure-based mutagenesis. *Acta Crystallogr Sect D Biol Crystallogr* 69: 1005–1012

Pletneva N V., Pletnev VZ, Souslova E, Chudakov DM, Lukyanov S, Martynov VI, Arhipova S, Artemyev I, Wlodawer A, Dauter Z, *et al* (2013b) Yellow fluorescent protein phiYFPv (Phialidium): Structure and structure-based mutagenesis. *Acta Crystallogr Sect D Biol Crystallogr* 69: 1005–1012

Prasher DC, Eckenrode VK, Ward WW, Prendergast FG & Cormier MJ (1992) Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* 111: 229–33

Prescott M, Ling M, Beddoe T, Oakley AJ, Dove S, Hoegh-Guldberg O, Devenish RJ & Rossjohn J (2003) The 2.2 Å crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation. *Structure* 11: 275–284

Qiao W, Mooney M, Bird AJ, Winge DR & Eide DJ (2006) Zinc binding to a regulatory zinc-sensing domain monitored in vivo by using FRET. *Proc Natl Acad Sci U S A* 103: 8674–8679

Remington SJ, Wachter RM, Yarbrough DK, Branchaud B, Anderson DC, Kallio K & Lukyanov KA (2005) zFP538, a yellow-fluorescent protein from Zoanthus, contains a novel three-ring chromophore. *Biochemistry* 44: 202–212

Renz M (2013) Fluorescence microscopy-A historical and technical perspective. *Cytom Part A* 83: 767–779

Rizzo MA, Springer GH, Granada B & Piston DW (2004) An improved cyan fluorescent protein variant useful for FRET. *Nat Biotechnol* 22: 445–449

Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42: W320–W324

Rockwell NC, Su YS & Lagarias JC (2006) Phytochrome structure and signaling mechanisms. *Annu Rev Plant Biol* 57: 837–858 doi:10.1146/annurev.arplant.56.032604.144208

Rodriguez EA, Tran GN, Gross LA, Crisp JL, Shu X, Lin JY & Tsien RY (2016) A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nat Methods* 13: 763–769

Sahl SJ, Hell SW & Jakobs S (2017) Fluorescence nanoscopy in cell biology. *Nat Rev Mol Cell Biol* 18: 685–701 doi:10.1038/nrm.2017.71

Salih A, Larkum A, Cox G, Kühl M & Hoegh-Guldberg O (2000) Fluorescent pigments in corals are photoprotective. *Nature* 408: 850–853

Schneider CA, Rasband WS & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9: 671–675 doi:10.1038/nmeth.2089

Schomaker V & Trueblood KN (1968) On the rigid-body motion of molecules in crystals. *Acta Crystallogr Sect B Struct Crystallogr Cryst Chem* 24: 63–76

Seward HE & Bagshaw CR (2009) The photochemistry of fluorescent proteins: implications for their biological applications. *Chem Soc Rev* 38: 2842

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE & Tsien RY (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 22: 1567–1572

Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M, *et al* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum.

Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M, *et al* (2013) A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat Methods* 10: 407–409

Shcherbakova DM, Baloban M, Emelyanov A V, Brenowitz M, Guo P & Verkhusha V V (2016) Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat Commun* 7

Shcherbakova DM, Baloban M, Pletnev S, Malashkevich VN, Xiao H, Dauter Z & Verkhusha V V (2015) Molecular Basis of Spectral Diversity in Near-Infrared Phytochrome-Based Fluorescent Proteins. *Chem Biol* 22: 1540–1551

Shcherbakova DM, Cox Cammer N, Huisman TM, Verkhusha V V. & Hodgson L (2018) Direct multiplex imaging and optogenetics of Rho GTPases enabled by near-infrared FRET. *Nat Chem Biol* 14: 591–600

Shcherbakova DM & Verkhusha V V (2013) Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nat Methods* 10: 751–754

Shcherbo D, Merzlyak EM, Chepurnykh T V, Fradkov AF, Ermakova G V, Solovieva EA, Lukyanov KA, Bogdanova EA, Zaraisky AG, Lukyanov S, *et al* (2007) Bright far-red

fluorescent protein for whole-body imaging. *Nat Methods* 4: 741–746

Shimomura O (2008) Discovery of Green Fluorescent Protein (GFP) (Nobel Lecture).

Shimomura O, Johnson FH & Saiga Y (1962) Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J Cell Comp Physiol* 59: 223–239

Shinoda H, Shannon M & Nagai T (2018) Fluorescent proteins for investigating biological events in acidic environments. *Int J Mol Sci* 19 doi:10.3390/ijms19061548

Shu X, Lev-Ram V, Deerinck TJ, Qi Y, Ramko EB, Davidson MW, Jin Y, Ellisman MH & Tsien RY (2011) A Genetically Encoded Tag for Correlated Light and Electron Microscopy of Intact Cells, Tissues, and Organisms. *PLoS Biol* 9: e1001041

Shu X, Royant A, Lin MZ, Aguilera TA, Lev-Ram V, Steinbach PA & Tsien RY (2009) Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome. *Science* 324: 804–7

Shu X, Shaner NC, Yarbrough CA, Tsien RY & Remington SJ (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45: 9639–9647

Siegel MS & Isacoff EY (1997) A genetically encoded optical probe of membrane voltage. *Neuron* 19: 735–741

Von Stetten D, Carpentier P, Flot D, Beteva A, Caserotto H, Dobias F, Guijarro M, Giraud T, Lentini M, McSweeney S, *et al* (2020) ID30A-3 (MASSIF-3) - A beamline for macromolecular crystallography at the ESRF with a small intense beam. *J Synchrotron Radiat* 27: 844–851

Strack RL, Strongin DE, Mets L, Glick BS & Keenan RJ (2010) Chromophore formation in DsRed occurs by a branched pathway. *J Am Chem Soc* 132: 8496–505

Subach OM, Gundorov IS, Yoshimura M, Subach F V., Zhang J, Grüenwald D, Souslova EA, Chudakov DM & Verkhusha V V. (2008) Conversion of Red Fluorescent Protein into a Bright Blue Probe. *Chem Biol* 15: 1116–1124

Subach OM, Malashkevich VN, Zencheck WD, Morozova KS, Piatkevich KD, Almo SC & Verkhusha V V. (2010) Structural Characterization of Acylimine-Containing Blue and Red Chromophores in mTagBFP and TagRFP Fluorescent Proteins. *Chem Biol* 17: 333–341

Swartz TE, Corchnoy SB, Christie JM, Lewis JW, Szundi I, Briggs WR & Bogomolni RA (2001) The Photocycle of a Flavin-binding Domain of the Blue Light Photoreceptor Phototropin. *J Biol Chem* 276: 36493–36500

Swulius MT & Jensen GJ (2012) The helical MreB cytoskeleton in Escherichia coli MC1000/pLE7 is an artifact of the N-Terminal yellow fluorescent protein tag. *J Bacteriol* 194: 6382–6

Takala H, Björling A, Berntsson O, Lehtivuori H, Niebling S, Hoernke M, Kosheleva I, Henning R, Menzel A, Ihalainen J a, *et al* (2014) Signal amplification and transduction in phytochrome photosensors. *Nature* 509: 245–8

Tanimura A, Nezu A, Morita T, Turner RJ & Tojyo Y (2004) Fluorescent biosensor for quantitative real-time measurements of inositol 1,4,5-trisphosphate in single living cells. *J Biol Chem* 279: 38095–38098

Terai T & Nagano T (2013) Small-molecule fluorophores and fluorescent probes for bioimaging. *Pflugers Arch Eur J Physiol* 465: 347–359 doi:10.1007/s00424-013-1234-z

Terry MJ, McDowell MT & Lagarias JC (1995) (3Z)- and (3E)-phytochromobilin are intermediates in the biosynthesis of the phytochrome chromophore. *J Biol Chem* 270: 11111–11118

Tomosugi W, Matsuda T, Tani T, Nemoto T, Kotera I, Saito K, Horikawa K & Nagai T (2009) An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nat Methods* 6: 351–353

Torra J, Lafaye C, Signor L, Aumonier S, Flors C, Shu X, Nonell S, Gotthard G & Royant A (2019) Tailing miniSOG: structural bases of the complex photophysics of a flavin-binding singlet oxygen photosensitizing protein. *Sci Rep* 9: 1–10

Tota MR, Allen JM, Karolin JO, Geddes CD & Ward WW (2016) Identification of a Fluorescent Protein from *Rhacostoma Atlantica*. *Photochem Photobiol* 92: 667–677

Tsien RY (1998) THE GREEN FLUORESCENT PROTEIN. *Annu Rev Biochem* 67: 509–544

Tsutsui H, Karasawa S, Shimizu H, Nukina N & Miyawaki A (2005) Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *EMBO Rep* 6: 233–238

Wachter RM, Elsliger M-A, Kallio K, Hanson GT & Remington SJ (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* 6: 1267–1277

Wagner JR, Brunzelle JS, Forest KT & Vierstra RD (2005) A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* 438: 325–331

Wagner JR, Zhang J, Brunzelle JS, Vierstra RD & Forest KT (2007) High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J Biol Chem* 282: 12298–12309

Wagner JR, Zhang J, Von Stetten D, Günther M, Murgida DH, Mroginski MA, Walker JM, Forest KT, Hildebrandt P & Vierstra RD (2008) Mutational analysis of Deinococcus radiodurans bacteriophytochrome reveals key amino acids necessary for the photochromicity and proton exchange cycle of phytochromes. *J Biol Chem* 283: 12212–12226

Wang L, Jackson WC, Steinbach PA & Tsien RY (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A* 101: 16745–16749

Watanabe TM, Imada K, Yoshizawa K, Nishiyama M, Kato C, Abe F, Morikawa TJ, Kinoshita M, Fujita H & Yanagida T (2013) Glycine Insertion Makes Yellow Fluorescent Protein Sensitive to Hydrostatic Pressure. *PLoS One* 8: e73212

Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Röcker C, Salih A, Spindler KD & Nienhaus GU (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc Natl Acad Sci U S A* 101: 15905–15910

Wingen M, Jaeger K-E, Gensch T & Drepper T (2017) Novel Thermostable Flavin-binding Fluorescent Proteins from Thermophilic Organisms. *Photochem Photobiol* 93: 849–856

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel

EB, Leslie AGW, McCoy A, *et al* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67: 235–242 doi:10.1107/S0907444910045749

Yang F, Moss LG & Phillips GN (1996a) The molecular structure of green fluorescent protein. *Nat Biotechnol* 14: 1246–1251

Yang TT, Cheng L & Kain SR (1996b) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res* 24: 4592–4593

Yang TT, Sinai P, Green G, Kitts PA, Chen YT, Lybarger L, Chervenak R, Patterson GH, Piston DW & Kain SR (1998) Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein. *J Biol Chem* 273: 8212–6

Yarbrough D, Wachter RM, Kallio K, Matz M V. & Remington SJ (2001) Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution. *Proc Natl Acad Sci U S A* 98: 462–467

Yu D, Baird MA, Allen JR, Howe ES, Klassen MP, Reade A, Makhijani K, Song Y, Liu S, Murthy Z, *et al* (2015) A naturally monomeric infrared fluorescent protein for protein labeling in vivo. *Nat Methods* 12: 763–5

Yu D, Dong Z, Gustafson WC, Ruiz-González R, Signor L, Marzocca F, Borel F, Klassen MP, Makhijani K, Royant A, *et al* (2016) Rational design of a monomeric and photostable far-red fluorescent protein for fluorescence imaging in vivo. *Protein Sci* 25: 308–15

Yu D, Gustafson WC, Han C, Lafaye C, Noirclerc-Savoye M, Ge W-P, Thayer DA, Huang H, Kornberg TB, Royant A, *et al* (2014) An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nat Commun* 5: 3626

Zhang J, Campbell RE, Ting AY & Tsien RY (2002) Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Biol* 3: 906–918 doi:10.1038/nrm976

Zhao Y, Araki S, Wu J, Teramoto T, Chang YF, Nakano M, Abdelfattah AS, Fujiwara M, Ishihara T, Nagai T, *et al* (2011) An expanded palette of genetically encoded Ca2+ indicators. *Science (80-)* 333: 1888–1891

Annex 1

ARTICLE

Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing

Lambert GG, Depernet H, Gotthard G, Schultz DT, Navizet I, Lambert T, Adams SR, Torreblanca-Zanca A, Chu M, Bindels DS, *et al* (2020) 'Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing'. *PLOS Biol* 18: e3000936

PLOS BIOLOGY

METHODS AND RESOURCES

Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing

Gerard G. Lambert[1], Hadrien Depernet[2], Guillaume Gotthard[2], Darrin T. Schultz[3,4], Isabelle Navizet[5], Talley Lambert[6,7], Stephen R. Adams[8], Albertina Torreblanca-Zanca[1], Meihua Chu[1], Daphne S. Bindels[9], Vincent Levesque[10], Jennifer Nero Moffatt[10], Anya Salih[11], Antoine Royant[2,12], Nathan C. Shaner[1]*

1 Department of Neurosciences, Center for Research in Biological Systems, University of California San Diego School of Medicine, La Jolla, California, United States of America, **2** Structural Biology Group, European Synchrotron Radiation Facility, Grenoble, France, **3** University of California Santa Cruz, Santa Cruz, California, United States of America, **4** Monterey Bay Aquarium Research Institute, Moss Landing, California, United States of America, **5** Laboratoire Modélisation et Simulation Multi-Echelle, Université Gustave Eiffel, Université Paris Est Creteil, Marne-la-Vallée, France, **6** Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, United States of America, **7** Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, United States of America, **8** Department of Pharmacology, University of California San Diego School of Medicine, La Jolla, California, United States of America, **9** Nikon Imaging Center, University of California San Diego, La Jolla, California, United States of America, **10** Birch Aquarium at Scripps, La Jolla, California, United States of America, **11** Confocal Facility, Western Sydney University, Penrith, New South Wales, Australia, **12** Institut de Biologie Structurale, Université Grenoble Alpes, CNRS, CEA, Grenoble, France

* ncshaner@ucsd.edu

OPEN ACCESS

Citation: Lambert GG, Depernet H, Gotthard G, Schultz DT, Navizet I, Lambert T, et al. (2020) *Aequorea*'s secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. PLoS Biol 18(11): e3000936. https://doi.org/10.1371/journal.pbio.3000936

Academic Editor: Ana J. Garcia-Saez, Universitat zu Koln, GERMANY

Received: October 31, 2019

Accepted: October 15, 2020

Published: November 2, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pbio.3000936

Data Availability Statement: A large portion of the relevant data are within the paper and its Supporting Information files. The native cDNA sequences for the coding region of each FP

Abstract

Using mRNA sequencing and de novo transcriptome assembly, we identified, cloned, and characterized 9 previously undiscovered fluorescent protein (FP) homologs from *Aequorea victoria* and a related *Aequorea* species, with most sequences highly divergent from *A. victoria* green fluorescent protein (avGFP). Among these FPs are the brightest green fluorescent protein (GFP) homolog yet characterized and a reversibly photochromic FP that responds to UV and blue light. Beyond green emitters, *Aequorea* species express purple- and blue-pigmented chromoproteins (CPs) with absorbances ranging from green to far-red, including 2 that are photoconvertible. X-ray crystallography revealed that *Aequorea* CPs contain a chemically novel chromophore with an unexpected crosslink to the main polypeptide chain. Because of the unique attributes of several of these newly discovered FPs, we expect that *Aequorea* will, once again, give rise to an entirely new generation of useful probes for bioimaging and biosensing.

Introduction

EGFP and other engineered variants of avGFP [1] have truly transformed biological imaging, allowing researchers to probe living cells in ways that were previously unthinkable [2–4]. The seemingly impossible task of producing a bright avGFP variant with an emission peak beyond

transcript described here have been deposited in GenBank, accession numbers MN114103 through MN114112. Raw Illumina RNA-Seq reads have been deposited in the NCBI Sequence Read Archive (SRA), accession numbers SRR9606756 through SRR9606760. Plasmids encoding the FPs described in this manuscript have been deposited with AddGene (plasmid numbers 129499 through 129512). The structures of AausFP1 and AausFP2 have been deposited in the Protein Data Bank under entry codes 6S67 and 6S68, respectively. Spectra from Fig 2 and photophysical characterization data from Table 1 are available on FPbase.org. Raw image data from all microscopy experiments are accessible from the DOI https://doi.org/10.26300/4x48-y393.

Funding: This work was supported by the following grant awards: NIH R01GM109984 (GGL, ATZ, MC, DSB, and NCS), NIH R01GM121944 (GGL, ATZ, MC, DSB, and NCS), NIH U01NS099709 (GGL, ATZ, MC, DSB, and NCS), NIH R21EY030716 (GGL, ATZ, MC, DSB, and NCS), NIH U01NS113294 (GGL, ATZ, MC, DSB, and NCS), NSF NeuroNex 1707352 (NCS), and NIH R01GM086197 (SRA). HD was supported by a PhD fellowship from the European Synchrotron Research Facility (https://www.esrf.eu). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. GGL, ATZ, MC, DSB, and NCS received salary support from the funding sources listed above.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: avGFP, *Aequorea victoria* green fluorescent protein; CP, chromoprotein; FP, fluorescent protein; FRET, Förster resonance energy transfer; GFP, green fluorescent protein; H2B, histone 2B; mRNA-Seq, mRNA sequencing; OSER, organized smooth endoplasmic reticulum.

yellow-green [5] drove many groups to explore other marine organisms as potential sources of fluorescent proteins (FPs) emitting at longer wavelengths. About 5 years after avGFP was cloned, FPs were discovered in corals [6,7], and since that time, FPs cloned from jellies, corals, and many other marine organisms have been reported (e.g., [8–10], among many others).

Despite this abundance of reported wild-type FPs, most FPs in widespread use as imaging tools are derived from only a handful of these organisms. Numerous avGFP variants with blue, cyan, green, and yellow-green emission remain the workhorses of live-cell imaging, and derivatives of red-emitting FPs from the soft coral *Discosoma* sp. [6,11,12] and the sea anemone *Entacmaea quadricolor* [10,13–15] make up the majority of commonly used FPs emitting at longer wavelengths. With the practical limitations of these particular FP scaffolds becoming more apparent as live-cell microscopy grows more complex and demanding, our group has focused on identifying, characterizing, and engineering FPs with low homology to these traditional choices. Our ongoing efforts include cloning new FPs from diverse sources as well as investigating previously un- or under-characterized FPs to identify new scaffolds with improved and/or novel spectral properties from which to launch new engineering efforts.

While searching for organisms expressing new and unusual FPs at Heron Island, a research station in the southern Great Barrier Reef, we collected a single individual of an unknown *Aequorea* species that we later determined was most similar to *A. australis*. This serendipitous encounter with a familiar genus led us to discover several novel FP homologs from 2 *Aequorea* species. Several of these proteins offer unique starting points for probe engineering.

Results and Discussion

The cyan-blue coloration of the radial canals of the *A.* cf. *australis* specimen we collected (Figs 1A, S1, and S2, and Figs K–N in S1 Text) suggested the potential presence of red-absorbing chromoproteins (CPs) and led us to reconstruct the transcriptome of the animal. In addition to transcripts encoding an FP clearly homologous to *A. victoria* green fluorescent protein (avGFP), as we expected, the *A.* cf. *australis* transcriptome also contains several transcripts encoding other, much more divergent avGFP homologs. Characterization of the recombinant proteins revealed that these avGFP homologs include a green-emitting FP that appears to be the brightest FP discovered to date, 2 long-wavelength-absorbing CPs, and a reversibly photochromic FP (Figs 1B and 2; Table 1).

Intrigued by the diversity of optical properties in the *A.* cf. *australis* FP homologs, we next investigated a sample of *A. victoria* from the Crystal Jelly exhibit at the Birch Aquarium at Scripps to determine whether this species also contained multiple diverse FPs. As we suspected, the *A. victoria* individual we sequenced expressed avGFP as well as orthologs of the bright green-emitting FP and the unusual CPs that we first identified in *A.* cf. *australis*. Surprisingly, this *A. victoria* also expressed a fairly close homolog of avGFP with surprisingly EGFP-like properties: a fully anionic chromophore, low pK_a, and much more efficient folding and maturation at 37˚C than wild-type avGFP. We found that only 2 mutations—1 to speed maturation at 37˚C and 1 to monomerize the protein—generated an FP with properties comparable to the commonly used avGFP variant mEGFP. *A. victoria*'s 2 CPs are distinct from those of *A.* cf. *australis*, undergoing an apparently unique mode of photoconversion from a green-emitting (green fluorescent protein [GFP]–like) state to a non-fluorescent, orange-red-absorbing state after exposure to blue light (Fig 2). While not characterized in depth during this study, this unusual property certainly warrants additional investigation of these CPs.

The 3 broad classes of FP homolog found in this study—green-emitting FPs, long-wavelength-absorbing CPs, and a reversibly photoswitchable CP—are discussed below. A phylogenetic tree of the FP homologs from this study is shown in Fig 3, and a sequence alignment is shown in Fig A in S1 Text.

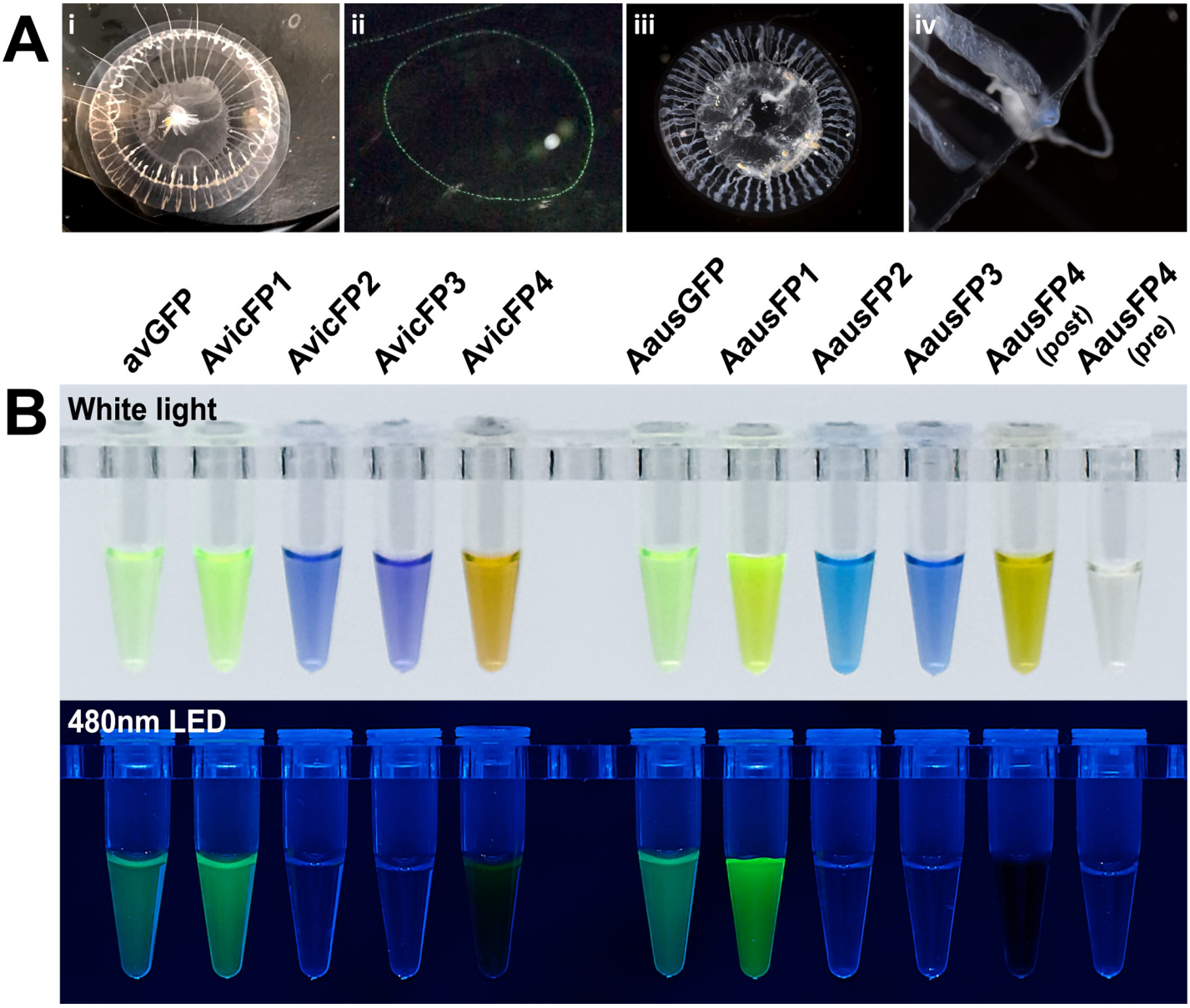

Fig 1. Photographs of *Aequorea* individuals from this study and purified fluorescent proteins cloned from these samples. (A) White-light (i) and fluorescence (400-nm LED illumination) (ii) photographs of *A. victoria* and white-light photographs of *A.* cf. *australis* (iii, iv). The blue coloration of *A.* cf. *australis* is shown in the higher magnification image of one of its tentacle bulbs (iv). (B) Purified recombinant proteins from *Aequorea* species, shown under white light and 480-nm LED without emission filters. Protein concentrations were adjusted to display similar optical density as judged by eye and were between 0.5 and 2 mg/ml for all samples.

https://doi.org/10.1371/journal.pbio.3000936.g001

Multiple, diverse *Aequorea* GFPs

As expected, both *Aequorea* species abundantly express close homologs of avGFP. Both AausGFP (*A.* cf. *australis* GFP) and the avGFP sequence identified in this work possess optical and biochemical properties similar to Prasher et al.'s original avGFP clone [1], characterized by an excitation spectrum with 2 peaks at 398 and 477 nm and an emission peak at 503 nm. The extinction coefficient and quantum yield of the 2 proteins are similar, and in our hands match closely with the literature values for avGFP [2].

We were surprised to discover a second green-emitting FP in *A.* cf. *australis*, AausFP1, that shares only 53% amino acid identity with avGFP (Fig A in S1 Text). AausFP1 is to our

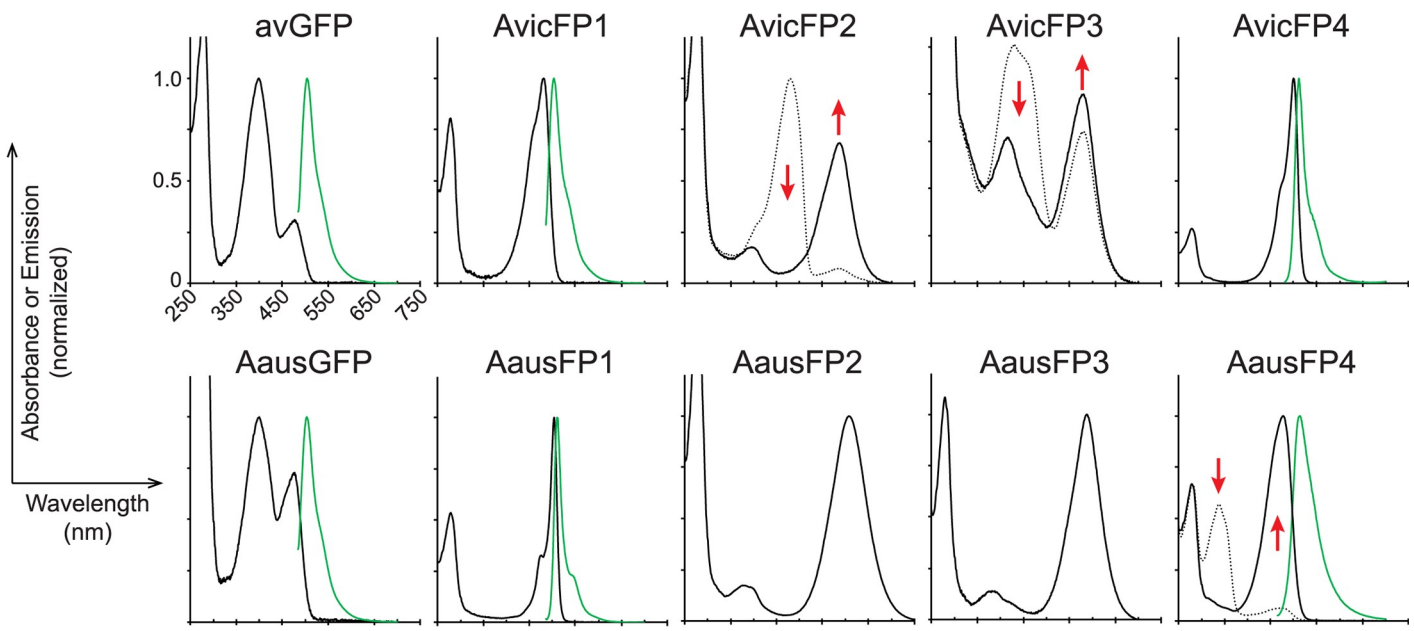

Fig 2. Absorbance and emission spectra (where measurable) for FP homologs in this study. Proteins from each species were designated AvicFP or AausFP and numbered in order of discovery, with chromoproteins retaining the "FP" nomenclature for consistency. For photoswitchable and photoconvertible proteins, pre-illumination absorbance spectra are shown as dotted lines, and post-illumination absorbance spectra as solid lines. Emission spectra are shown as green solid lines. The emission spectra for AvicFP2 and AvicFP3 were measured using 460-nm excitation prior to photoconversion. The emission spectrum of AausFP4 was measured using 440-nm excitation after photoswitching to the blue-absorbing state. Red arrows indicate peaks that increase or decrease upon photoconversion or switching. For ease of display, spectra are normalized to the maximum visible absorbance for non-photoactive proteins, and to the pre- (for AvicFP2) or post-illumination (for AvicFP3 and AausFP4) maximum for photoactive proteins. All plots share the same *x*-axis scale as shown for AausGFP. The data underlying this figure may be found at FPbase (https://www.fpbase.org). FP, fluorescent protein.

https://doi.org/10.1371/journal.pbio.3000936.g002

knowledge the brightest FP discovered to date, with a nearly perfect quantum yield (0.97) and a peak extinction coefficient of 170,000 $M^{-1}cm^{-1}$, making it nearly 5-fold brighter than EGFP on a per-molecule basis. These already extraordinary properties are further bolstered by a low fluorescence pK_a (4.4), unusually narrow excitation and emission peaks (see Fig 2; the emission peak of AausFP1 has a full width at half maximum [FWHM] of 19 nm, compared to 32 nm for EGFP), and higher photostability than mEGFP (see below). The ortholog of AausFP1 in *A*. *victoria*, AvicFP4, shares some of its unusual properties, such as narrow excitation and emission peaks (Fig 2), efficient folding at 37˚C, and a fairly high extinction coefficient, but its low quantum yield (0.10) makes it the dimmest GFP found in *A*. *victoria*.

AausFP1 was expressed at very low levels relative to other FPs in the *A*. cf. *australis* individual sequenced (see Table A in S1 Text) and would be rare or absent in most cDNA expression-cloning libraries. The transcriptomic approach used in this study is the only practical way to identify such unusual, low-abundance FPs, short of costly whole genome sequencing. Despite low expression in its native context, wild-type AausFP1 expresses and folds very efficiently in *E*. *coli* at 37˚C without any modifications. Though brightly fluorescent, AausFP1 is largely insoluble in this context, and when purified, the soluble fraction of the protein runs as a high-molecular-weight aggregate on size exclusion chromatography (Fig BB in S1 Text).

When expressed in mammalian cells, AausFP1 is excluded from the nucleus and only forms visible aggregates in the most highly expressing cells (Fig W in S1 Text), suggesting that it may form soluble but high-molecular-weight aggregates in this context as well. X-ray crystallography revealed a uniquely stabilized chromophore environment in AausFP1 that may be responsible for its unique properties (see Fig 4, Tables C–E in S1 Text, and Figs B, D, E, and G in S1 Text). Since AausFP1 crystallizes as a dimer, we speculate that it takes on this oligomeric

Table 1. Photophysical properties of fluorescent proteins described in this study derived from *A. victoria* and *A.* cf. *australis*.

Protein	λ_{abs} [a]	λ_{em} [b]	ε [c]	φ [d]	Brightness [e]	pK_a [f]	Photostability [g]
avGFP [h]	398/477 [i]	503	41 (6.5)/15 (2.2)	0.75 (0.01)	91/32	4.8 [j] (0.1)	ND
AvicFP1	481	503	64 (3.7)	0.63 (0.03)	118	4.9 (0.1)	ND
AvicFP2 [k]	480	515	59 (1.2)	0.04 (0.01)	6	ND [m]	ND
	588	— [l]	41 (3.2)	—	—		
AvicFP3 [k]	480	520	ND	<0.001 [l]	—	ND	ND
	580	—	ND	—	—		
AvicFP4	500	512	121 (2.1)	0.10 (0.01)	36	ND	ND
AausGFP [m]	398/477	503	29 (1.6)/22 (1.3)	0.73 (0.01)	62/47	4.8 [i] (0.1)	ND
AausFP1	504	510	170 (6.0)	0.97 (0.05)	485	4.4 (0.1)	129 ± 4 (N = 20) 218 ± 9 (N = 23)
AausFP2	609	—	52 (2.3)	—	—	<6.0	ND
AausFP3	587	—	59 (1.7)	—	—	<6.5	ND
AausFP4 [n]	338/477	—/510	42 (1.9)/3 (0.1)	—/<0.001	—	ND	ND
	477	513	69 (2.4)	<0.001			
mAvicFP1	480	503	65 (0.04)	0.63 (0.01)	126	4.9 (0.1)	131 ± 3 (N = 20) 121 ± 8 (N = 20)
mEGFP [o]	488	507	56	0.60	100	6.0	100 ± 4 (N = 31) 100 ± 4 (N = 13)
mNeonGreen [p]	506	517	116	0.80	274	5.7	ND

The commonly used protein mEGFP and the bright monomeric FP mNeonGreen are included for comparison. Values reported for photophysical parameters are the mean of at least 3 independent measurements on independently prepared samples; values in parentheses are standard deviation of the mean unless otherwise noted below. A dash indicates no measurable emission or negligible brightness.

[a]Peak absorbance wavelength (nm).

[b]Peak emission wavelength (nm).

[c]Molar extinction coefficient ($mM^{-1}cm^{-1}$).

[d]Fluorescence quantum yield.

[e]Brightness ($\varepsilon \times \varphi$), percent normalized to mEGFP.

[f]For FPs with a quantum yield $\geq$ 0.10, the reported pK_a is the pH at which fluorescence emission is 50% of maximal brightness; for AausFP2 and AausFP3, the pK_a was determined only approximately and represents the pH at which the long-wavelength absorbance peak is 50% of its maximal value.

[g]Mean photobleaching half-times in live cells, corrected for molecular brightness and scaled to the half-time measured for mEGFP in this study (mEGFP = 100%); half-times under widefield (upper value) or laser scanning confocal (lower value) illumination (see Methods and S1 Text) are shown ± the standard error of the mean, with the number of individual cells sampled given in parentheses.

[h]Measured in this study using the avGFP peptide sequence from the *A. victoria* individual sequenced.

[i]avGFP displays two absorbance peaks whose ratio is largely insensitive to pH changes over much of the physiological range but is somewhat sensitive to protein concentration; values separated by slashes in all columns represent those for these two distinct peaks, respectively.

[j]Fluorescence pK_a value determined by exciting the 477-nm absorbance peak.

[k]Values given are for pre- (upper value) and post-photoconverted (lower value) forms of AvicFP2 and AvicFP3.

[l]Fluorescence quantum yields less than 0.001 were not determined, even in cases with a measurable emission peak.

[m]AausGFP is the closest direct homolog to avGFP from *A.* cf. *australis* and displays a similar double-peaked absorbance; values separated by slashes represent those from each peak, respectively.

[n]AausFP4 is reversibly photoswitchable between a UV-absorbing form and a blue-absorbing form; the UV-absorbing form has a small amount of residual blue absorbance; values of photophysical parameters UV and blue absorbance peaks are separated by slashes.

[o]Values from [2] are shown in the table and were re-verified in this study.

[p]Values from [35] are shown in the table and were re-verified in this study.

avGFP, *Aequorea victoria* green fluorescent protein; FP, fluorescent protein; ND, not determined.

https://doi.org/10.1371/journal.pbio.3000936.t001

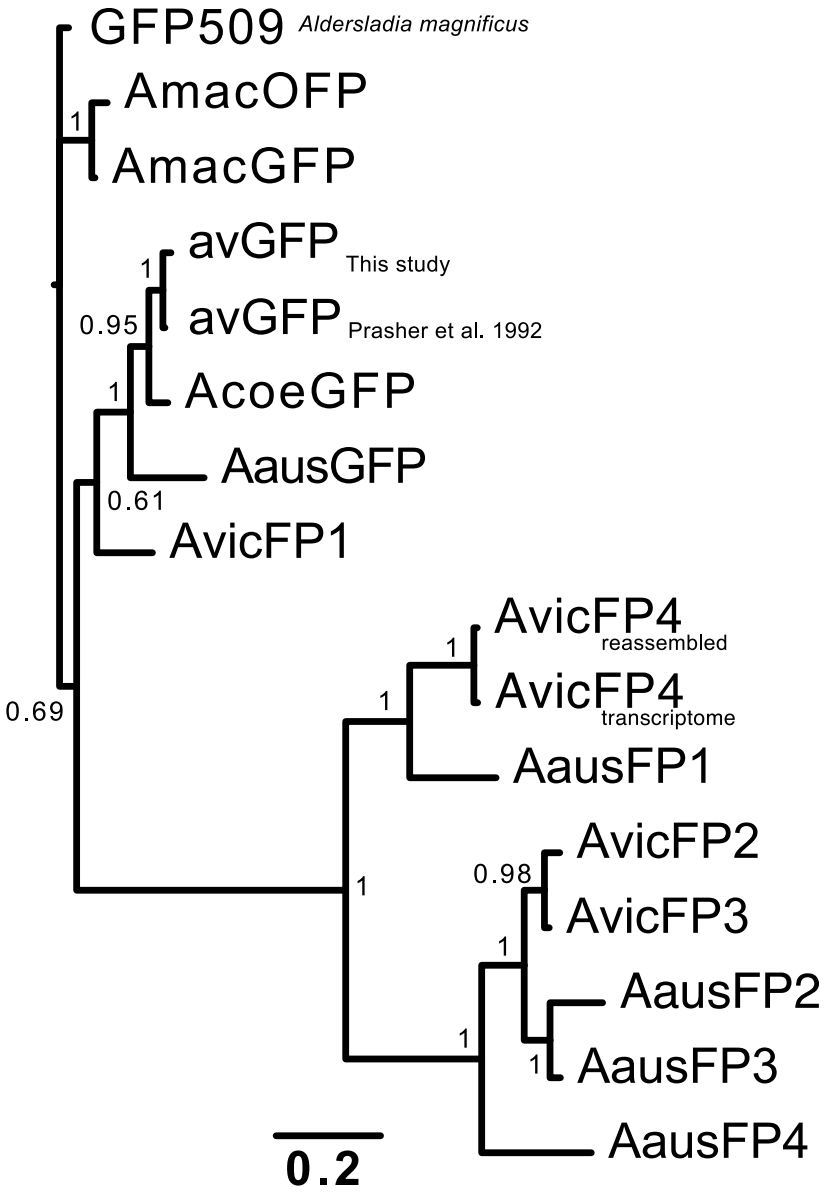

Fig 3. Phylogenetic tree for FPs cloned in this study, with *Aequorea macrodactyla* and *Aldersladia magnificus* green FPs included as outgroups. Green-emitting FPs with avGFP-like properties, including AvicFP1, fall into 1 cluster of fairly closely related sequences, while the novel fluorescent (AausFP1 and AvicFP4) and non-fluorescent homologs form 2 additional families. The data underlying this figure (nucleotide sequences of the FPs from this study) may be found in GenBank, accession numbers MN114103 through MN114112. avGFP, *Aequorea victoria* green fluorescent protein; FP, fluorescent protein.

https://doi.org/10.1371/journal.pbio.3000936.g003

state in its native context, perhaps stabilized by other interactions. It is possible that, as with other FPs such as dTomato [11], the dimer interface serves to stabilize the chromophore environment and may contribute to the unusually high brightness of AausFP1.

AausFP1 photobleaches at similar rates to mEGFP on both widefield and confocal microscopy when instrument settings are identical, but because AausFP1 emits photons at a higher rate (due to its high quantum yield and extinction coefficient), its true photostability is somewhat higher than that of mEGFP (S1 Text and Figs Z and AA in S1 Text). Because it has a

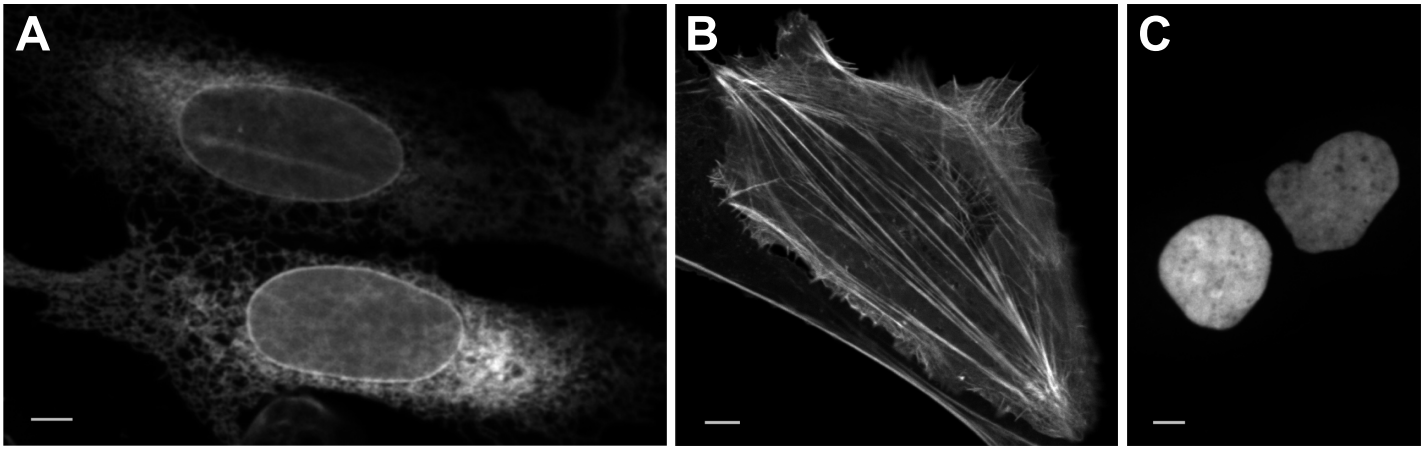

Fig 4. Expression of mAvicFP1-tagged proteins in mammalian cells. mAvicFP1 fusions to (A) CytERM, (B) LifeAct, and (C) H2B. U2-OS cells display expected localization. Scale bar is 10 mm. The data underlying this figure (raw image data) may be found at https://doi.org/10.26300/4x48-y393.

https://doi.org/10.1371/journal.pbio.3000936.g004

number of potentially useful properties, we consider AausFP1 the top candidate for future engineering among the FPs we have identified in this work.

Apart from AausFP1, an unexpected find among the newly discovered *A. victoria* FP homologs was AvicFP1, a transcript with relatively low abundance (Table A in S1 Text) but with high homology to avGFP (80% amino acid identity; see Fig A in S1 Text). At neutral pH, AvicFP1 has a single absorbance peak at 481 nm, indicating that its chromophore exists in a fully anionic state. It is curious that AvicFP1 would appear to be a superior energy transfer acceptor for the photoprotein aequorin than avGFP based on their absorbance spectra (Fig 2). However, avGFP was expressed at the sites of luminescence (bell margin), while AvicFP1 was only detected in the body of the animal (Table A in S1 Text), indicating that it is unlikely to be the natural energy acceptor for aequorin. Unfortunately, investigation of the interactions between AvicFP1 and aequorin are beyond the scope of this study.

The fluorescence pK_a of AvicFP1 (4.9) is lower than that of EGFP (6.0). We speculate that the cysteine present in the first chromophore position is partly responsible for AvicFP1's EGFP-like properties. The S65T substitution in avGFP was among the most critical early mutations introduced to generate an all-anionic chromophore, though S65C was reported at the same time [2,16]. Because of mutations derived from errors in the oligonucleotides used for synthetic gene assembly, we also identified 1 colony among the thousands of initial AvicFP1 clones that produced a much larger proportion of mature FP in *E. coli* incubated at 37˚C. This clone contained a single point mutation leading to the substitution F64L, generating a variant with optical and biochemical properties indistinguishable from those of the wild-type protein. The F64L mutation is another of those originally identified for improving the folding of avGFP to ultimately produce EGFP [2,17].

Essentially all of the side chains that participate in the weak dimer interface of avGFP are conserved in AvicFP1. We hypothesized that mutations sufficient to monomerize avGFP variants (i.e., A206K [18]) would also produce a monomeric variant of AvicFP1. Using the organized smooth endoplasmic reticulum (OSER) assay to test for oligomeric behavior in cells [19], we found that the mutant AvicFP1-F64L/A206K displays monomeric behavior equivalent to mEGFP, widely considered the "gold standard" of monomeric FPs [19] (OSER data are summarized in Table B in S1 Text). Fusions to LifeAct [20] and histone 2B (H2B) displayed the expected localization and dynamics (Fig 5, S1 Movie and S2 Movie). Additionally, cells expressing H2B-mAvicFP1 and imaged at 2-minute intervals for 72 hours at 37˚C showed no

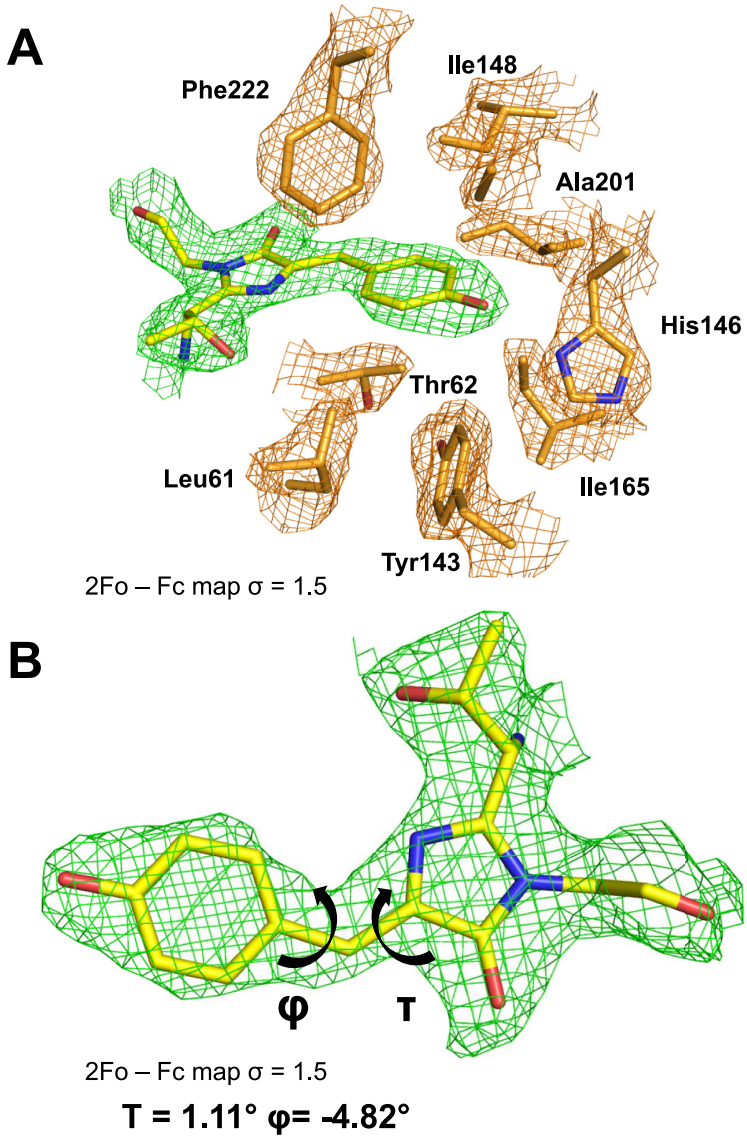

Fig 5. The AausFP1 chromophore environment. (A) $2F_{obs} - F_{calc}$ electron-density map contoured at a 1.5 σ level superimposed over the model of the chromophore and the neighboring residues in the structure of AausFP1. (B) Dihedral angle definition around the chromophore methylene bridge. The data underlying this figure may be found in PDB 6S67.

https://doi.org/10.1371/journal.pbio.3000936.g005

significant increase in doubling time (see Fig Y in S1 Text and S1 Data). We therefore decided that this variant merited an official name: mAvicFP1 (monomeric *A*. *victoria* fluorescent protein 1).

Originally, avGFP was identified as a partner to the photoprotein aequorin, and this association ultimately led to cloning the cDNA that encodes it. We speculate that other green-emitting FPs were not identified at the same time as avGFP because the brightest visible fluorescence in *A*. *victoria* is around the bell margin, while AvicFP1 appears to be expressed

exclusively in other tissues (Fig A in S1 Text). The optical properties of mAvicFP1 are superficially similar to those of mEGFP, and these FPs have similar brightness. In our hands, mAvicFP1's photostability under widefield and confocal illumination is somewhat higher than that of mEGFP (Figs Z and AA in S1 Text and S1 Data), its monomeric character is comparable, and its toxicity (as measured by the rate of cell division when expressing an H2B fusion; see S1 Text and Fig Y in S1 Text) appears to be lower that of mEGFP. However, the primary differentiating property of mAvicFP1 is its low pK_a, which may offer advantages when labeling proteins in acidic compartments. Others have also reported that mAvicFP1 spontaneously "blinks" under high illumination intensity, making it highly useful for single-molecule localization microscopy with a single excitation wavelength [21].

Unusual *Aequorea* CPs

In the context of the broad phenotypic diversity of hydrozoan FP homologs now known [22–25], the presence of green- and red-absorbing CPs in *Aequorea* species is not surprising. However, the properties of *Aequorea* CPs differ in surprising ways from those previously cloned from other organisms. Every *Aequorea* CP displays a broad absorbance spectrum (Fig 2) that lacks the well-defined sharp peak and short-wavelength shoulder typical of most FPs and CPs, suggesting that these proteins contain an unusual chromophore and/or chromophore environment. Also, none of the *Aequorea* CPs has any measurable red fluorescence emission, even on our most sensitive instruments. All CPs described here migrate as high-molecular-weight, apparently soluble aggregates or high-order oligomers on a gel filtration column when expressed in *E. coli* (see Fig BB in S1 Text).

AausFP2 has a distinctive cyan-blue pigmented appearance when expressed in *E. coli*, with a broad absorbance spectrum peaking at 610 nm. *A.* cf. *australis* expresses a second CP, AausFP3, that displays a similarly symmetrical, shoulder-less absorbance peak, but with a maximum absorbance at 590 nm. X-ray crystallography analysis of AausFP2 (Tables B and C in S1 Text) revealed a conserved dimer interface geometry containing many conserved residues between AausFP1 and AausFP2. We suspect that despite AausFP2's behavior in gel filtration experiments, the native oligomeric state of AausFP2 may be a dimer. The amino acid residues making up the dimer interface in the AausFP2 crystal structure are also largely conserved across the other *Aequorea* CPs (Fig A in S1 Text), suggesting that if this is the native oligomeric state of AausFP2, then they are all likely to be dimers.

The X-ray crystal structure of AausFP2 further revealed a chemically novel chromophore in which the side chain of a neighboring cysteine is covalently linked to the methylene bridge of a twisted GFP-like chromophore (Fig 6; Tables D, E, and G in S1 Text; Figs F and H in S1 Text). This amino acid, Cys62, is conserved in all *Aequorea* CPs. The C62S mutant of AausFP2 appears yellow and has a major absorbance peak characteristic of a GFP-type chromophore (Fig I in S1 Text), strongly suggesting that this conserved cysteine is necessary for formation of the red-shifted chromophore. The peak absorbance wavelength of alkali-denatured *Aequorea* CPs displays a 20- to 30-nm red-shift relative to that expected for a GFP-type chromophore [2], which is abolished by addition of β-mercaptoethanol (Fig CC in S1 Text), providing additional evidence for the role of this unusual bond. Quantum mechanical calculations indicate that both the presence of a sulfur atom and a twisted chromophore are required to produce long-wavelength absorbance (see S1 Text, Fig J in S1 Text, and Table F in S1 Text).

Because of its broad absorbance reaching into the far-red and near-infrared regions of the spectrum, its relatively high extinction coefficient, and its efficient folding at 37˚C, AausFP2 or its derivatives could ultimately prove very useful as photoacoustic tomography probes for deep tissue imaging.

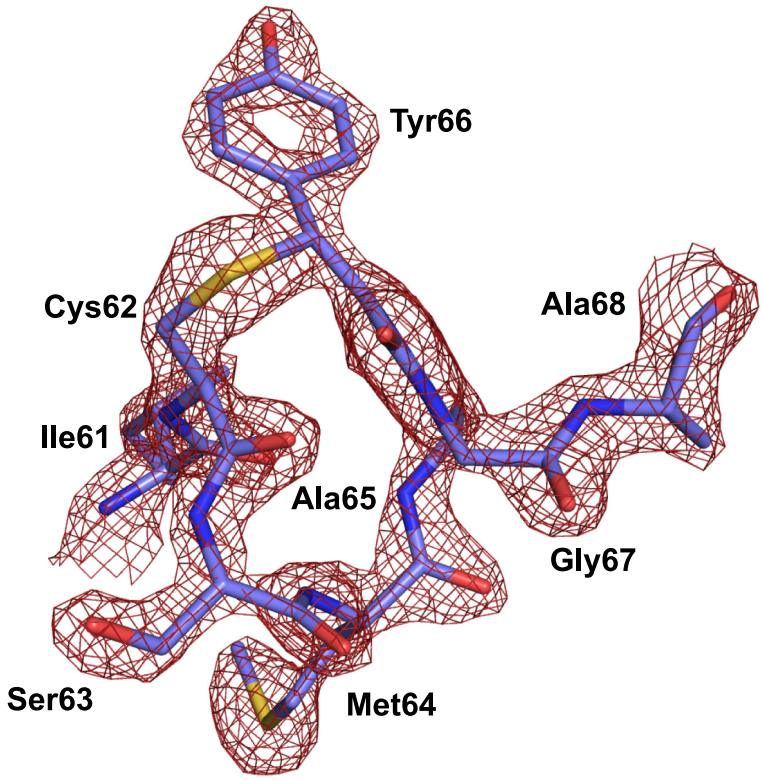

Fig 6. $2F_{obs} - F_{calc}$ electron-density map contoured at a 2.0 σ level superimposed over the model of the chromophore and the 3 covalently bonded residues in the structure of AausFP2. The data underlying this figure may be found in PDB 6S68.

https://doi.org/10.1371/journal.pbio.3000936.g006

Unlike their orthologs in *A*. cf. *australis*, which mature fully to their long-wavelength forms in the dark, the *A*. *victoria* CPs mature very slowly in the absence of blue light. When expressed in total darkness, AvicFP2 has peak absorbance in the blue region, and is weakly green fluorescent, suggesting an avGFP-type chromophore. Upon blue light exposure, AvicFP2 converts into a purple-blue CP with peak absorbance at 588 nm. In light of the quantum mechanical calculations presented (Fig J in S1 Text and Table F in S1 Text), this dramatic absorbance shift suggests that the light-induced change in AvicFP2 represents either the bonding of the Cys62 side chain to the methylene bridge of the chromophore or twisting of the chromophore from a planar to non-planar conformation.

AvicFP3 is highly homologous to AvicFP2 (96% amino acid identity; see Fig A in S1 Text), and is similarly green fluorescent when expressed and purified in the dark. Like AvicFP2, AvicFP3 converts to a green-absorbing CP when exposed to blue light, but appears to mature more efficiently than AvicFP2 in the absence of light (see pre-conversion absorbance spectrum; Fig 2). We anticipate that these proteins, if they can be engineered into monomers that mature efficiently at 37˚C, could be useful as photoactivatable Förster resonance energy transfer (FRET) quenchers.

A reversibly photochromic CP

The final FP homolog we identified in *A*. cf. *australis* is AausFP4, a very weakly fluorescent (quantum yield < 0.001) green-emitting CP with photochromic behavior strikingly similar to

that of the engineered avGFP variant Dreiklang [26]. When expressed and/or stored in the dark, AausFP4 reaches an equilibrium state with a major absorbance peak at 338 nm, indicating that the chromophore is neutral and missing at least 1 double bond relative to a mature GFP-type chromophore. With exposure to UV light, AausFP4 fully converts to an anionic GFP-like state with 477-nm peak absorbance.

This transformation is reversible by exposure to bright blue light or by storage in the dark. Together, these properties suggest a mechanism similar to that of Dreiklang, in which a structural water molecule can reversibly hydrate the imidazolinone ring of the chromophore via 2 different photochemical reactions triggered by different wavelengths of light [26]. A key difference between AausFP4 and Dreiklang is the absence of an approximately 400-nm absorbance peak in the "on" state, accompanied by off-switching mediated by blue rather than violet light. While AausFP4 is likely to be dimeric and/or aggregating like its closest relatives (AausFP2 and AausFP3), it may prove to be a useful starting material from which to engineer a new lineage of reversibly photoswitchable FPs or CPs. AausFP4 also likely represents, to our knowledge, the first naturally occurring example of Dreiklang-type photoswitching to be discovered.

Conclusion

We have identified several new *Aequorea* FPs with the potential to further diversify the landscape of fluorescent probes and biosensors. AausFP1, the brightest fluorescent protein currently known, will serve as the parent of an entirely new lineage of super-bright FP variants. As a parallel scaffold to avGFP derivatives in many ways, mAvicFP1 may be quickly adaptable to existing probes and biosensors. AausFP4 is the first natural example of Dreiklang-type photochromism and may help generate other useful variations on this mechanism. Four highly unusual *Aequorea* CPs provide truly novel engineering opportunities, including generating new far-red-emitting FPs, improved dark FRET acceptors, and photoacoustic probes, among many other potential uses.

The discovery and understanding of these new fluorescent proteins in *Aequorea* were made possible through a highly collaborative and interdisciplinary approach involving field collection work, basic molecular biology, next-generation sequencing and bioinformatics, protein engineering, microscopy, X-ray crystallography, and phylogenetics. We are optimistic that more studies with this kind of holistic approach will help elucidate many of the mysteries still hiding in the natural world. In the time that has elapsed since Osamu Shimomura's first sampling of *A*. *victoria* in Friday Harbor, it has become clear that there is an urgent need to explore and understand as much of the molecular biodiversity that exists in the world as possible before many organisms go extinct or become too rare to sample.

Materials and methods

Chemicals and other reagents

Unless otherwise noted, bacterial growth medium components were purchased from Fisher Scientific, antibiotics were purchased from Gold Biotechnology, and other chemicals were purchased from Sigma-Aldrich.

Sample collection and RNA extraction

A single specimen of *A*. cf. *australis* was collected near Heron Island (Queensland, Australia) and processed on-site at the Heron Island Research Station (University of Queensland) (see "Ethics statement" for permit and collection details). A single individual of *A*. *victoria* was obtained from the aquaculture collections of the Birch Aquarium at Scripps. The animals

being kept in the exhibit tank at this time were originally obtained from the Aquarium of the Pacific (Long Beach, CA), where they have been bred in captivity for many generations. Notably, the *A*. *victoria* jellies are fed a diet of crustaceans, and so any hydrozoan-like FP transcripts identified must come from the jelly itself rather than from contamination of the mRNA sequencing (mRNA-Seq) library with prey-derived mRNAs.

Live samples were photographed and then anaesthetized with $MgCl_2$ prior to being dissected. The bell margin, bell, and mouth were dissected separately, and total RNA was extracted using RNeasy Plus Mini Kit (Qiagen) following the manufacturer's instructions. For *A*. cf. *australis*, the purified samples were combined and dried in a GenTegra RNA tube for transport back to the US. For *A*. *victoria*, samples from the 3 body regions were kept separate.

Next-generation sequencing

Total RNA samples were used as input to generate Illumina-compatible mRNA-Seq libraries at the Scripps Research Institute Next Generation Sequencing Core facility. Total RNA underwent polyA selection prior to Illumina TruSeq library prep. Libraries were run on 1 NextSeq flowcell and generated between 25 and 35 million 150-bp paired-end reads per sample.

Transcriptomes for individual samples as well as the aggregate *A*. *victoria* transcriptome were assembled using Trinity [27,28] either on a custom workstation in the lab or using the public Galaxy bioinformatics server [29]. Read mapping was performed using Bowtie 2 alignment [30] and RSEM [31] for cross-sample comparison. Additional details on transcript verification are included in S1 Text and Figs O–W in S1 Text.

Species identification

The identity of *A*. cf. *australis* was established using phylogenetic analysis; see detailed methods and results in S1 Text, S1 Fig and S2 Fig. The identity of *A*. *victoria* was verified by the presence of an assembled transcript encoding avGFP, as well as its well-characterized morphology.

Protein tree

Protein sequences were aligned with Clustal Omega [32], and a Bayesian tree was created using a burn-in of 3,000 iterations, run length of 30,000 iterations, and 4 chains with MrBayes software [33].

Cloning and mutagenesis

Candidate FP-encoding transcripts were identified by BLAST homology searching using avGFP as the query against the assembled transcriptome databases as well as intermediate assembly files created by the Trinity workflow. Searching through intermediate assembly files allowed us to identify potential alternative transcript sequences and those that were (possibly incorrectly) collapsed into single contigs by Trinity. Putative FP-encoding transcripts were validated against raw read data and reconstructed as necessary (see below for detailed methods, results, and discussion). Sequence alignments were performed using Clustal Omega [32].

For each avGFP homolog identified, the coding region was identified and a synthetic gene was designed to produce the encoded polypeptide sequence using codons optimized for both human and *Escherichia coli* expression using an in-house BioXp 3200 instrument (SGI-DNA, La Jolla, CA) or ordered as a gBlock double-stranded gene fragment (Integrated DNA Technologies, San Diego, CA). Both PCR-amplified and synthetic cDNAs contained additional nucleotides at the 5′ end (GAAAACCTGTACTTCCAGGGT) and 3′ end

(CGTTTGATCCGGCTGC). Fragments encoding FPs were inserted using Gibson assembly [34] into the vector pNCST (modified from [35]) that had been PCR-amplified with the oligos pNCST-vec-F and pNCST-vec-R (Table H in S1 Text). The pNCST plasmid contains a synthetic promoter that drives high-level constitutive expression in most *E. coli* strains. This plasmid encodes an N-terminal 6xHis tag and linker followed by a TEV protease cleavage site just before the start codon of the inserted gene.

Site-directed mutagenesis of AvicFP1 was performed by generating 2 fragments of the FP coding sequence by standard PCR with Phusion polymerase (New England Biolabs) and primers as listed in Table H in S1 Text. Mutations were placed in the overlapping sequence between fragments to facilitate Gibson assembly of full-length mutant sequences in a 1-step insertion into the pNCST vector. Plasmids encoding AausFP1, mAvicFP1, and fusions of mAvicFP1 to H2B, LifeAct, and CytERM driven by a CMV promoter for mammalian expression were generated by Gibson assembly of the PCR-amplified FP sequence with the corresponding PCR-amplified fragment of pC1-mNeonGreen, pmNeonGreen-H2B-N-6, pmNeonGreen-LifeAct, and pmNeonGreen-CytERM [35].

Recombinant protein purification

Sequence-verified plasmids were transformed into NEB5a strain *E. coli* (New England Biolabs) (because the promoter in the pNCST vector is semi-constitutive in most strains of *E. coli*, we find it convenient to use a single strain for cloning and expression), plated on LB/agar supplemented with carbenicillin (100 μg/ml), and incubated overnight at 37˚C. For proteins that matured efficiently at 37˚C (AvicFP-F64L, mAvicFP1, AvicFP4, AausFP1, AausFP2, EGFP, mEGFP, and mNeonGreen), colonies were picked and inoculated directly into a 200-ml baffled wide-mouth flask containing 50 ml of 2xYT broth and 100 μg/ml carbenicillin, and incubated overnight at 37˚C with shaking at 250 rpm. For proteins requiring multiple days at room temperature to mature (avGFP, AvicFP1, AvicFP2, AvicFP3, AausFP3, and AausFP4), a single colony was resuspended in 10 ml of 2xYT medium, and 100 μl of this suspension was plated on 5 100-mm petri dishes containing LB/agar and 100 μg/ml carbenicillin. After overnight incubation at 37˚C to initially establish colonies, plates were then incubated at room temperature for several days in the dark.

Bacteria containing the recombinant protein were recovered by centrifuging liquid cultures in 50-ml conical tubes at 4,500*g* for 10 minutes. For proteins expressed on LB/agar plates, a razor blade was gently glided over the surface of the agar, harvesting the colonies on the blade, and then wiped into 2-ml microcentrifuge tubes and gently centrifuged to the bottom of the tube. Four milliliters of the lysis reagent B-PER (Thermo 78248) was added for every gram of *E. coli* pellet. Tubes were gently vortexed until the pellets were completely dissolved, taking care not to form bubbles from the detergent component of the B-PER. The resulting suspension was then incubated on a gentle rocker for 15 minutes and then centrifuged at >20,000*g* for 10 minutes to pellet insoluble debris. Note that we find that there is a strong correlation between true protein solubility and extraction efficiency in B-PER that is not true of other extraction methods such as sonication, which can solubilize aggregated FPs more readily.

Meanwhile, we prepared a purification column by adding 1–2 ml of Ni-NTA resin slurry (Expedeon) into a 15-ml gravity column (Bio-Rad), allowing the storage buffer to drip through. The column was equilibrated with 10 bed volumes of wash buffer (150 mM Tris [pH 7.5], 300 mM NaCl, 5 mM imidazole) and then capped at the bottom. After centrifugation, the lysate was directly added to the prepared Ni-NTA column. The column was then capped at the top and the lysate-resin slurry was tumbled end-over-end for 30 minutes at 4˚C. The top/bottom caps were removed, and the liquid was allowed to drip through by gravity flow. The

column was then washed 3 times with 3 column volumes of wash buffer. Finally, the protein was eluted from the column by gradual addition of elution buffer (50 mM Tris [pH 7.5], 150 mM NaCl, 200 mM imidazole). Clear liquid was allowed to drip through, and only the fluorescent/colorful fraction was collected.

The proteins were then concentrated further using a 3-kD MWCO column (Amicon/Millipore) until the volume of protein solution was <150 μl. Meanwhile, 2× desalting columns (Pierce) were prepared for each protein by equilibrating in 50 mM Tris (pH 8.5)/150 mM NaCl according to the manufacturer's instructions. Then 150 μl of protein solution was loaded onto the equilibrated desalting column and centrifuged at 1,500 rpm for 1 minute in a microcentrifuge. The collected protein was then passed through a second equilibrated desalting column to ensure complete buffer exchange.

Mammalian cell imaging

Experiments performed in Dr. Shaner's lab. U2-OS cells (HTB-96, ATCC) were grown in a 35-mm glass bottom dish (P35G-1.5-14-C, MatTek) or on coverslips (25CIRCLE #1.5, Fisherbrand) with DMEM (105666–016, Gibco) supplemented with 10% (v/v) FBS (10437–028, Gibco) under 5% humidified CO_2 atmosphere at 37˚C. Polyethylenimine (PEI) in double-distilled H_2O (1 mg/ml (pH 7.3), 23966, Polysciences) was used as the transfection reagent. The transfection mixture was prepared in Opti-MEM (31985047, Thermo Fisher Scientific) with 4.5 μg of PEI and 500 ng of plasmid. For static images, a coverslip was placed in an Attofluor cell chamber (A7816, Invitrogen), and FluoroBrite DMEM (A18967-01, Gibco) was added. For time series, culture medium in glass bottom dishes was replaced with FluoroBrite DMEM (A18967-01, Gibco) supplemented with GlutaMAX (35050–061, Gibco) and 10% (v/v) FBS (10437–028, Gibco).

Confocal images and time series were acquired on a Leica TCS SP8 system using a 488-nm argon laser for excitation. The sample was placed in an incubation chamber with a controlled environment at 37˚C and 5% humidified CO_2 (Okolab). For LifeAct-mAvicFP1, a 63×/1.40 oil objective (HC PL APO CS2 63×/1.40 Oil, 15506350) was used with an emission bandwidth of 500–550 nm detected with a Leica HyD. For time series, a bandwidth of 500–600 nm was used, and images were acquired at 4.6-second intervals (4× line averaging, pinhole at 510 nm, 1 A.U.). For single images of H2B-mAvicFP1, CytERM-mAvicFP1, and CytERM-mEGFP (Addgene 62237), a 20× 0.75 NA air objective (HC PL APO 20×/0.75 CS2, 15506517) was used with an emission bandwidth of 500–550 nm detected with a HyD. For time series of these constructs, a bandwidth of 500–600 nm was used, and images were acquired at 3-minute intervals (4× line averaging, pinhole at 510 nm, 1 A.U.).

Experiments performed at Harvard Medical School. U2-OS cells were grown on #1.5 35-mm glass bottom dishes (MatTek) in McCoy's 5A medium supplemented with GlutaMAX (Thermo Fisher) and 10% fetal bovine serum (Thermo Fisher) and transfected with 0.5 μg of pCytERM-mAvicFP1 and pCytERM-mEGFP plasmid DNA using fuGENE (Promega) 24 hours prior to imaging. Before imaging, the growth medium was replaced with FluoroBrite DMEM supplemented with 5% FBS (Thermo Fisher). Image acquisition was performed in a full environmental enclosure (37˚C, 5% CO_2; Okolab) on a Nikon Ti-E microscope with Perfect Focus System, a Spectral Borealis-modified spinning disc confocal (Yokogawa X1), and an Orca Flash v3 sCMOS camera (Hamamatsu). OSER data were acquired with a 40× Plan Fluor 1.3 NA objective lens, and live time-lapse imaging was acquired with a 100× Plan Apo VC 1.4 NA objective (162-nm and 65-nm pixel size, respectively). Green fluorescence was excited with a 491-nm solid state laser (Cobolt) and a Di01-T405/488/568/647 (Semrock) dichroic; emission was selected with an ET525/50m filter (Chroma). Hardware was controlled with

MetaMorph (v7.8.13). For time-lapse experiments, single-plane images were acquired every second. For OSER acquisition, a uniform grid of images was acquired covering the entire coverslip.

Images were processed with Fiji [36]. OSER assay analysis was conducted as previously described [19].

Cell division assay

U2-OS cells were grown and transfected as described above with plasmids encoding an N-terminal fusion of H2B to either mEGFP [18], AausFP1, or mAvicFP1, all with identical linker sequences. Prior to imaging, cells were stained with SiR-Hoechst (Cytoskeleton) following the manufacturer's instructions and maintaining a low concentration of stain in the medium throughout imaging. Cells were imaged on a Leica TCS SP8 system with an Okolab environmental chamber, as described above, starting 12–24 hours post-transfection. Images were collected every 2 minutes for >72 hours using 488-nm excitation with green emission to detect the H2B fusions, and with 633-nm excitation and far-red emission for the SiR-Hoechst stain to detect all DNA. For analysis, cells were selected from those expressing H2B and that underwent 1 cell division in the first half of the experiment. Control cells were selected from those neighboring the selected H2B-FP-expressing cells. The interval between cell divisions, defined as the time between visible chromosome separation, was recorded for the 2 daughter cells of each original cell.

Photostability assay

U2-OS cells were grown and transfected as described above with plasmids encoding full-length untagged mEGFP, AausFP1, or mAvicFP1. Cells were imaged on a Leica SP8 laser scanning confocal microscope with a 63×/1.40 oil objective (HC PL APO CS2 63×/1.40 Oil, 15506350) at 37˚C and humidified 5% CO_2, as described above, with 488-nm argon laser illumination and an emission bandwidth of 500–700 nm detected with a photomultiplier tube. The pinhole was set to 2 A. U. at 510 nm to illuminate a thicker optical section of the cytoplasm at high intensity, at least partly accounting for diffusion of FP molecules in and out of the focal plane.

For widefield photobleaching, cells were imaged on a Nikon Ti-E microscope with a 40×/0.95 PL APO air objective, Perfect Focus System, a Spectra X light source (Lumencor) set to 470/24-nm bandpass excitation, 495-nm dichroic, 520/35-nm emission filter, and an Orca Flash v4 camera (Hamamatsu). Cells were imaged approximately 48 hours after transfection, with focusing using white light or very low power fluorescence excitation (≥1% imaging intensity for ≥5 seconds) to prevent pre-bleaching of the FPs.

The full-power light intensity at the sample plane was measured using a power meter (model 843-R, Newport), and the illumination spectrum at the objective was measured using a mini spectrometer fitted with a fiber optic input (Hamamatsu). For confocal bleaching, the light intensity measured at the objective was 250 μW. For widefield bleaching, the intensity at the objective was 10.3 mW. Cells were imaged for 8–10 minutes with continuous illumination (widefield) or continuous laser scanning (confocal). This was sufficient time to achieve >90% loss in fluorescence in most cases.

Image stacks were processed with Fiji [36], first by registering them with the StackReg plugin [37] to eliminate any artifacts caused by drift. A region of interest (ROI) was defined in the cytoplasm of each cell as well as a background region. We then measured the mean intensity of each ROI over the image stack and interpolated to determine the time value corresponding to 50% of the initial fluorescence signal for each cell. Photobleaching half-times were then scaled by a correction factor that corresponds to the per-molecule brightness of each FP under the specific illumination condition. For confocal bleaching, the correction factor

corresponds to the molar extinction coefficient at 488 nm. For widefield bleaching, the correction factor depends on both the absorbance spectrum of the FP and the illumination spectrum at the objective:

$$\phi \int I(\lambda) \cdot A(\lambda) d\lambda$$

where ϕ is the quantum yield, $I(\lambda)$ is the illumination intensity, and $A(\lambda)$ is the absorbance of the FP. In both cases, the correction factor normalizes the photobleaching half-times to those that would be observed if the excitation were tuned to produce equal photon output per FP molecule at time 0. These experiments and the analysis of the resulting data are discussed in more detail in S1 Text.

Quantum yield and extinction coefficient determination

Purified green-emitting FPs were characterized as previously described [38] to determine quantum yield. Briefly, FPs that had been buffer-exchanged into 50 mM Tris-HCl (pH 8.5)/150 mM NaCl were diluted into the same buffer until the baselined peak absorbance was $\leq$0.05 as measured by a UV-2700 UV-Vis spectrophotometer (Shimadzu). Sample and standard (fluorescein in 0.1 M NaOH, quantum yield 0.95 [39]) absorbance were matched within 10% at 480 nm, the excitation wavelength used for fluorescence emission spectra. Immediately after measuring the absorbance spectrum, the cuvette containing the sample was transferred to a Fluorolog-3 fluorimeter (Jobin Yvon), and the emission spectrum was taken from 460 nm to 700 nm in 1-nm steps, with excitation at 480 nm and a slit width of 2 nm for both excitation and emission. Emission spectra were interpolated under the region in which scattered excitation light bleeds through into the emission path. Quantum yield was calculated by dividing the area under the sample emission curve by its absorbance at 480 nm and dividing by the same ratio for the standard, then multiplying by 0.95, the quantum yield of the standard.

Extinction coefficients for all FPs and CPs in this study were measured using the accepted standard method of measuring FP extinction coefficients, with alkali denaturation (addition of 2 M NaOH to the FP sample to a final concentration of 1 M NaOH) as previously described [38]. This method relies on the denatured chromophore absorbance and extinction coefficient to be invariant between FPs with chemically identical chromophores, and allows calculation of the extinction coefficient of the natively folded protein by comparing the peak height between native and denatured absorbance spectra. We performed this assay with the following modifications: (1) In order to avoid calculating erroneously large values of FP extinction coefficients from alkali denaturation measurements, several absorbance spectra were taken for each sample. Beginning immediately after addition of NaOH, multiple absorbance spectra were taken over several minutes to determine both the point at which the protein was fully denatured and the point at which it reached maximum absorbance at approximately 447 nm. The *maximum* measured value of the peak absorbance of fully denatured protein was used in extinction coefficient calculations. (2) For CPs containing the novel cysteine-linked chromophore, the peak absorbance of alkali-denatured protein is red-shifted 20–30 nm relative to the known 447-nm peak of GFP-type chromophores [2]. Because the extinction coefficient of this species is unknown, we also measured absorbance spectra for alkali-denatured CPs with the addition of 1 mM β-mercaptoethanol, which is expected to break the bond between the sulfur atom of the cysteine side chain and the methylene bridge of the chromophore, producing a GFP-type denatured chromophore. The maximum absorbance value of reduced, denatured chromophore was used in calculation of the extinction coefficient, which should be considered an estimate for *Aequorea* CPs pending much deeper investigation into the biochemical properties of their unique chromophore.

pK_a determination

Purified proteins were concentrated and desalted as described above into 20 mM Tris-HCl (pH 8). A solution of 50 mM Tris-HCl, 50 mM citric acid, 50 mM glycine, and 150 mM NaCl (final concentrations after pH adjustment) was prepared and split into 2 master stocks that were adjusted to pH 3 and pH 12 with HCl and NaOH, respectively. These stocks were then used to prepare buffers at pH 3, 4, 5, 6, 6.5, 7, 7.25, 7.5, 7.75, 8, 9, 10, 11, and 12 by mixing at different ratios. Each sample was then diluted (2 μl of sample + 198 μl of buffer) into each pH buffer, and its emission or absorbance was measured using an Infinite M1000 PRO (Tecan) plate reader. The pK_a was determined by interpolating the pH value at which the fluorescence or absorbance value was 50% of its maximum.

Size-exclusion chromatography and light scattering

Two milligrams of purified protein in 100 ul of running buffer was applied to a Shodex KW-802.5 column with guard column KW-G 6B (Showa Denko America, New York, NY) and run in 50 mM Na-HEPES/150 mM NaCl (pH 7.35) at a flow rate of 0.5 ml/minute using an Agilent 1100 Series HPLC system controlled by ChemStation software (Agilent Technologies, Santa Clara, CA). Protein elution was dually monitored with 280-nm absorbance and at the absorbance maxima for each fluorescent protein. In-line light scattering was performed by a Wyatt Heleos system running ASTRA software (Wyatt Technology, Goleta, CA). Clinical-grade cetuximab used as a molecular weight standard was obtained from the UCSD Moores Cancer Center pharmacy.

Protein crystallogenesis

AausFP1 and AausFP2 were first expressed and purified as aforementioned. The His-tag was cleaved off using either TEV for AausFP1 (1/100 protease/protein ratio, overnight incubation at room temperature) or proteinase K for AausFP2 (1/50 protease/protein ratio, 1-hour incubation at room temperature). The protein solution was run through an additional His-Trap column to remove cleaved tag and uncleaved protein. A final purification step consisted of a gel filtration column (Superdex 75-10/300 GL, GE Healthcare, Chicago, IL). Fractions were analyzed using 15% SDS-PAGE gels, pooled and concentrated to 40 and 51 mg/ml for AausFP1 and AausFP2, respectively, using an Amicon Ultra centrifugal filter with a molecular weight cutoff of 30 kDa (Merck, Darmstadt, Germany). Initial crystallization hits were obtained using the HTX lab platform of the EMBL Grenoble Outstation, and then manually optimized. AausFP1 was crystallized with the hanging drop method using 0.7–1.3 M trisodium citrate, 0.2 M sodium chloride in 0.1M Tris buffer (pH 6.5–8.0). AausFP2 was crystallized with the hanging drop method using 14%–24% PEG 3350 trisodium citrate and 0.2 M sodium chloride in 0.1 M HEPES buffer (pH 7.3–8.2).

Diffraction data collection

Diffraction data for AausFP1 were collected on beamline BL13-XALOC at the ALBA synchrotron in Barcelona (Spain) [40] from a crystal flash-cooled at 100 K in its mother liquor supplemented with 20% (v/v) glycerol for cryoprotection. Diffraction data for AausFP2 were collected on beamline ID30B of the European Synchrotron Radiation Facility in Grenoble (France) [41] from a crystal flash-cooled at 100 K without addition of any cryoprotectant. Diffraction data were integrated and reduced using XDS and XSCALE [42]. Data collection and reduction statistics are given in Table C in S1 Text.

Structure determination

A BLAST search (https://blast.ncbi.nlm.nih.gov/) identified the fluorescent protein phiYFPv from the jellyfish genus *Phialidium* as the closest homolog of both AausFP1 and AausFP2 (sequence identities of 61% and 50%, respectively) with a known structure (PDB entry code 4HE4 [43]). The structures of AausFP1 and AausFP2 were solved by the molecular replacement method using the 4HE4 coordinates as a search model with the program PHASER [44]. The model was progressively and interactively modified in COOT [45] and refined with REFMAC5 [46]. The asymmetrical units contain 4 molecules for AausFP1 and 1 molecule for AausFP2. Analysis of the interaction interfaces with PISA [47] strongly suggests that the AausFP1 tetramer consists of a dimer of a physiological dimer (interface areas of 1,210 $Å^2$ versus 360 $Å^2$ for the third and fourth largest areas), while the AausFP2 monomer forms a physiological dimer with a symmetry-related molecule (interface area of 1,290 $Å^2$ versus 540 $Å^2$ for the second largest area). Structure refinement statistics are given in Table C in S1 Text.

Calculation of AausFP2 absorption maxima

Eight models of the minimal part of the chromophore were constructed, modeling only the 2 conjugated cycles of the chromophore. H atoms replaced in all models the 2 alpha carbon atoms linking the chromophore to the rest of the protein. 3D coordinates for all heavy atoms of the chromophore were taken from the crystallographic structures without optimization, leading to 2 groups of models, one with the conformation of the EGFP structure and one with the conformation of the AausFP2 structure. The corresponding sets of models were labeled EGFP and AausFP2. The main difference between the 2 sets of models is the dihedral angle between the 2 cycles, i.e., −2˚ (almost planar) for EGFP and −53˚ (twisted) for AausFP2.

In each set of models, the phenol moiety was presented in its protonated form (neutral chromophore) or phenolate form (anionic chromophore). Moreover, in the AausFP2 set, the carbon between the 2 cycles of chromophore is linked to a protein's cysteine through a thioether bond, whereas this carbon is simply protonated in the case of EGFP. Therefore, in the models, this carbon was linked either to a mercapto group (–SH) or simply protonated. Structures were protonated and the position of H atoms were optimized at the B3LYP/6-31+g (d,p) level of theory with the Gaussian G09 program.

Software

The web-based data-plotting software PlotsOfData [48] was used to generate Figs Y, Z, and AA in S1 Text.

Ethics statement

All scientific collection in the field was performed under permit G17/39943.1 granted to Dr. Anya Salih, Western Sydney University, by the Great Barrier Reef Marine Park Authority. A specimen of *A*. cf. *australis* was collected within the Scientific Research Zone surrounding Heron Island (Queensland, Australia) using a hand-held net and was transported back to the lab in seawater. Live samples were kept in fresh running seawater for minimal amounts of time after collection.

Supporting information

S1 Data. Raw cell division and photobleaching data and corresponding analysis for Fig Y, Z, and AA in S1 Text.
(XLSX)

S1 Fig. 16S phylogenetic tree. The 16S tree is inconclusive as to the phylogenetic position of both the transcriptomic 16S sequences and the reference-guided assembly 16S sequence. Several species are monophyletic in this tree and *A*. *australis* is in a large polytomy. See S1 Text for additional discussion.
(PDF)

S2 Fig. COI phylogenetic tree. The COI tree shows that the reference-corrected COI sequence (sample_COI) is sister to a large *A*. *australis* clade. See S1 Text for additional discussion.
(PDF)

S1 Movie. Confocal imaging of H2B-mAvicFP1 expressed in U2-OS cells. Scale bar is 10 mm. Timestamp is in hours:minutes. Images were taken at 3-minute intervals. Video playback is at 14 frames per second (total imaging duration 3 hours 9 minutes). Raw imaging data used to generate this movie are available at https://doi.org/10.26300/4x48-y393.
(MOV)

S2 Movie. Confocal imaging of LifeAct-mAvicFP1 expressed in U2-OS cells. Scale bar is 5 mm. Timestamp is in hours:minutes. Images were taken at 4.6-second intervals. Video playback is at 100 frames per second (total imaging duration 4 hours 35 minutes). Raw imaging data used to generate this movie are available at https://doi.org/10.26300/4x48-y393.
(MOV)

S1 Text. Supporting materials and methods, results, and discussion.
(PDF)

Acknowledgments

We dedicate this manuscript to the memory of Dr. Roger Y. Tsien and Dr. Osamu Shimomura, whose studies on *A*. *victoria* and avGFP continue to inspire us and to catalyze new technologies for biological imaging. This work was also made possible by the Crystal Jelly exhibit at the Birch Aquarium at Scripps, highlighting the significance of this species in the history of biomedical research. This exhibit was the source of the *A*. *victoria* individual used in this work. The European Synchrotron Radiation Facility is acknowledged for access to beamline ID30B and facilities for molecular biology via its in-house research program. The ALBA synchrotron is acknowledged for allocation of beamtime on beamline BL13-XALOC. We thank Franck Borel, David Cobessi, and the beamline staff for help during data collection on BL13-XALOC. We also wish to thank Dr. Lauren M. Barnett for aiding in the collection of *A*. cf. *australis*, Wyatt Patry (Monterey Bay Aquarium) for helping in species identification, and Dr. Ute Hochgeschwender, Dr. Thomas Blacker, and Dr. Robert E. Campbell for helpful feedback on the manuscript.

Author Contributions

Conceptualization: Gerard G. Lambert, Isabelle Navizet, Antoine Royant, Nathan C. Shaner.

Data curation: Gerard G. Lambert, Hadrien Depernet, Guillaume Gotthard, Darrin T. Schultz, Talley Lambert, Daphne S. Bindels, Antoine Royant, Nathan C. Shaner.

Formal analysis: Gerard G. Lambert, Hadrien Depernet, Darrin T. Schultz, Isabelle Navizet, Daphne S. Bindels, Antoine Royant, Nathan C. Shaner.

Funding acquisition: Nathan C. Shaner.

Investigation: Gerard G. Lambert, Hadrien Depernet, Guillaume Gotthard, Darrin T. Schultz, Isabelle Navizet, Talley Lambert, Stephen R. Adams, Albertina Torreblanca-Zanca, Meihua Chu, Anya Salih, Antoine Royant, Nathan C. Shaner.

Methodology: Hadrien Depernet, Guillaume Gotthard, Darrin T. Schultz, Isabelle Navizet, Talley Lambert, Stephen R. Adams, Albertina Torreblanca-Zanca, Meihua Chu, Daphne S. Bindels, Antoine Royant, Nathan C. Shaner.

Project administration: Nathan C. Shaner.

Resources: Darrin T. Schultz, Talley Lambert, Vincent Levesque, Jennifer Nero Moffatt.

Software: Isabelle Navizet.

Supervision: Antoine Royant, Nathan C. Shaner.

Validation: Darrin T. Schultz.

Writing – original draft: Gerard G. Lambert, Hadrien Depernet, Guillaume Gotthard, Darrin T. Schultz, Isabelle Navizet, Talley Lambert, Daphne S. Bindels, Antoine Royant, Nathan C. Shaner.

Writing – review & editing: Gerard G. Lambert, Hadrien Depernet, Darrin T. Schultz, Talley Lambert, Jennifer Nero Moffatt, Anya Salih, Antoine Royant, Nathan C. Shaner.

References

1. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the Aequorea victoria green-fluorescent protein. Gene. 1992; 111:229–33. https://doi.org/10.1016/0378-1119(92)90691-h PMID: 1347277
2. Tsien RY. The green fluorescent protein. Annu Rev Biochem. 1998; 67:509–44. https://doi.org/10.1146/annurev.biochem.67.1.509 PMID: 9759496
3. Shaner NC, Patterson GH, Davidson MW. Advances in fluorescent protein technology. J Cell Sci. 2007; 120:4247–60. https://doi.org/10.1242/jcs.005801 PMID: 18057027
4. Rodriguez EA, Campbell RE, Lin JY, Lin MZ, Miyawaki A, Palmer AE, et al. The growing and glowing toolbox of fluorescent and photoactive proteins. Trends Biochem Sci. 2017; 42:111–29. https://doi.org/10.1016/j.tibs.2016.09.010 PMID: 27814948
5. Mishin AS, Subach FV, Yampolsky IV, King W, Lukyanov KA, Verkhusha VV. The first mutant of the Aequorea victoria green fluorescent protein that forms a red chromophore. Biochemistry. 2008; 47:4666–73. https://doi.org/10.1021/bi702130s PMID: 18366185
6. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, et al. Fluorescent proteins from nonbioluminescent Anthozoa species. Nat Biotechnol. 1999; 17:969–73. https://doi.org/10.1038/13657 PMID: 10504696
7. Salih A, Larkum A, Cox G, Kühl M, Hoegh-Guldberg O. Fluorescent pigments in corals are photoprotective. Nature. 2000; 408:850–3. https://doi.org/10.1038/35048564 PMID: 11130722
8. Karasawa S, Araki T, Nagai T, Mizuno H, Miyawaki A. Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. Biochem J. 2004; 381:307. https://doi.org/10.1042/BJ20040321 PMID: 15065984
9. Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A. A green-emitting fluorescent protein from Galaxeidae coral and its monomeric version for use in fluorescent labeling. J Biol Chem. 2003; 278:34167–71. https://doi.org/10.1074/jbc.M304063200 PMID: 12819206
10. Wiedenmann J, Schenk A, Röcker C, Girod A, Spindler K-D, Nienhaus GU. A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). Proc Natl Acad Sci U S A. 2002; 99:11646–51. https://doi.org/10.1073/pnas.182157199 PMID: 12185250
11. Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol. 2004; 22:1567–72. https://doi.org/10.1038/nbt1037 PMID: 15558047

12. Shaner NC, Lin MZ, McKeown MR, Steinbach PA, Hazelwood KL, Davidson MW, et al. Evaluating and improving the photostability of fluorescent proteins. SPIE BiOS: Biomedical Optics; 2009 Jan 23–28; San Jose, CA, US. https://doi.org/10.1117/12.814684

13. Shcherbo D, Merzlyak EM, Chepurnykh TV, Fradkov AF, Ermakova GV, Solovieva EA, et al. Bright far-red fluorescent protein for whole-body imaging. Nat Methods. 2007; 4:741–6. https://doi.org/10.1038/nmeth1083 PMID: 17721542

14. Kredel S, Oswald F, Nienhaus K, Deuschle K, Röcker C, Wolff M, et al. mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. PLoS ONE. 2009; 4:e4391. https://doi.org/10.1371/journal.pone.0004391 PMID: 19194514

15. Lam AJ, St-Pierre F, Gong Y, Marshall JD, Cranfill PJ, Baird MA, et al. Improving FRET dynamic range with bright green and red fluorescent proteins. Nat Methods. 2012; 9:1005–12. https://doi.org/10.1038/nmeth.2171 PMID: 22961245

16. Heim R, Cubitt AB, Tsien RY. Improved green fluorescence. Nature. 1995; 373:663–4. https://doi.org/10.1038/373663b0 PMID: 7854443

17. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 1996; 173:33–8. https://doi.org/10.1016/0378-1119(95)00685-0 PMID: 8707053

18. Zacharias DA. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. Science. 2002; 296:913–6. https://doi.org/10.1126/science.1068539 PMID: 11988576

19. Costantini LM, Fossati M, Francolini M, Snapp EL. Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. Traffic. 2012; 13:643–9. https://doi.org/10.1111/j.1600-0854.2012.01336.x PMID: 22289035

20. Riedl J, Crevenna AH, Kessenbrock K, Yu JH, Neukirchen D, Bista M, et al. Lifeact: a versatile marker to visualize F-actin. Nat Methods. 2008; 5:605–7. https://doi.org/10.1038/nmeth.1220 PMID: 18536722

21. Gavrikov AS, Baranov MS, Mishin AS. Live-cell nanoscopy with spontaneous blinking of conventional green fluorescent proteins. Biochem Biophys Res Commun. 2019; 522(4):852–4. https://doi.org/10.1016/j.bbrc.2019.11.163 PMID: 31801668

22. Bulina ME, Chudakov DM, Britanova OV, Yanushevich YG, Staroverov DB, Chepurnykh TV, et al. A genetically encoded photosensitizer. Nat Biotechnol. 2006; 24:95–9. https://doi.org/10.1038/nbt1175 PMID: 16369538

23. Aglyamova GV, Hunt ME, Modi CK, Matz MV. Multi-colored homologs of the green fluorescent protein from hydromedusa Obelia sp. Photochem Photobiol Sci. 2011; 10:1303–9. https://doi.org/10.1039/c1pp05068k PMID: 21614405

24. Hunt ME, Modi CK, Aglyamova GV, Ravikant DVS, Meyer E, Matz MV. Multi-domain GFP-like proteins from two species of marine hydrozoans. Photochem Photobiol Sci. 2012; 11:637–44. https://doi.org/10.1039/c1pp05238a PMID: 22251928

25. Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, et al. GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. Mol Biol Evol. 2004; 21:841–50. https://doi.org/10.1093/molbev/msh079 PMID: 14963095

26. Brakemann T, Stiel AC, Weber G, Andresen M, Testa I, Grotjohann T, et al. A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. Nat Biotechnol. 2011; 29:942–7. https://doi.org/10.1038/nbt.1952 PMID: 21909082

27. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29:644–52. https://doi.org/10.1038/nbt.1883 PMID: 21572440

28. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8:1494–512. https://doi.org/10.1038/nprot.2013.084 PMID: 23845962

29. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018; 46:W537–44. https://doi.org/10.1093/nar/gky379 PMID: 29790989

30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–9. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

31. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040

32. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019; 47:W636–41. https://doi.org/10.1093/nar/gkz268 PMID: 30976793

33. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17(8):754–5. https://doi.org/10.1093/bioinformatics/17.8.754 PMID: 11524383

34. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–5. https://doi.org/10.1038/nmeth.1318 PMID: 19363495

35. Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Methods. 2013; 10:407–9. https://doi.org/10.1038/nmeth.2413 PMID: 23524392

36. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9:676–82. https://doi.org/10.1038/nmeth.2019 PMID: 22743772

37. Thevenaz P, Ruttimann UE, Unser M. A pyramid approach to subpixel registration based on intensity. IEEE Trans Image Process. 2020; 7(1):27–41.

38. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. Nat Methods. 2005; 2:905–9. https://doi.org/10.1038/nmeth819 PMID: 16299475

39. Lakowicz JR. Principles of fluorescence spectroscopy. Berlin: Springer Science & Business Media; 2013.

40. Juanhuix J, Gil-Ortiz F, Cuní G, Colldelram C, Nicolás J, Lidón J, et al. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. J Synchrotron Radiat. 2014; 21:679–89. https://doi.org/10.1107/S160057751400825X PMID: 24971961

41. McCarthy AA, Barrett R, Beteva A, Caserotto H, Dobias F, Felisaz F, et al. ID30B—a versatile beamline for macromolecular crystallography experiments at the ESRF. J Synchrotron Radiat. 2018; 25:1249–60. https://doi.org/10.1107/S1600577518007166 PMID: 29979188

42. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–32. https://doi.org/10.1107/S0907444909047337 PMID: 20124692

43. Pletneva NV, Pletnev VZ, Souslova E, Chudakov DM, Lukyanov S, Martynov VI, et al. Yellow fluorescent protein phiYFPv (Phialidium): structure and structure-based mutagenesis. Acta Crystallogr D Biol Crystallogr. 2013; 69:1005–12. https://doi.org/10.1107/S0907444913004034 PMID: 23695245

44. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–74. https://doi.org/10.1107/S0021889807021206 PMID: 19461840

45. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. https://doi.org/10.1107/S0907444910007493 PMID: 20383002

46. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011; 67:355–67. https://doi.org/10.1107/S0907444911001314 PMID: 21460454

47. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372:774–97. https://doi.org/10.1016/j.jmb.2007.05.022 PMID: 17681537

48. Postma M, Goedhart J. PlotsOfData—a web app for visualizing data together with their summaries. PLos Biol. 2019; 17(3):e3000202. https://doi.org/10.1371/journal.pbio.3000202 PMID: 30917112

Summary

Since the cloning of the Green Fluorescent Protein (GFP) in 1992 fluorescent proteins (FPs) have increasingly become essential tools in cell imaging. GFP was first discovered in the jellyfish Aequorea victoria, then homologues were found in other types of organisms such as corals, sea anemones, lancelets and small crustaceans, forming the family of GFP-like FPs. A unique feature of GFP-like proteins is the autocatalytic formation of the chromophore, the light-absorbing and light-emitting part of the protein, from three consecutive amino acids located at the centre of the β-barrel structure of the protein. The family of GFP-like FPs have fluorescence emission maxima that cover the entire visible light spectrum from deep blue to far red. In the late 2000s, another type of FPs has been derived from phytochromes, a family of red-light photoreceptors that use bilins as chromophore. The interest of these is that their fluorescence excitation and emission spectra is in the near-infrared region of the light spectrum. This region is part of the so-called 'optical window' of living tissues, in which light absorption and scattering by haemoglobin, water and lipids is minimized and thus should be preferred for whole-body fluorescence imaging . These FPs have been called NIR FPs. In this PhD work, I have solved the crystallographic structures of three GFP-like FPs (one very bright green FP, one chromoprotein and one weakly red fluorescent FP) and three NIR FPs derived from a monomeric phytochrome. The structural information gained on the nature and environment of the chromophores has allowed me to propose explanations for their peculiar spectroscopic properties. In particular the structure of the chromoprotein revealed the existence for the first time of a chromophore which forms a covalent bond with a nearby cysteine residue, leading to a large red-shift of the UV-vis absorption maximum.

Résumé

Depuis le clonage de la protéine fluorescente verte (GFP) en 1992, les protéines fluorescentes (PF) sont progressivement devenues des outils essentiels de l'imagerie cellulaire. La GFP a été découverte chez la méduse *Aequorea victoria*, mais des homologues ont ensuite été trouvés dans d'autres types d'organismes tels que des coraux, des anémones de mer, des amphoxius (ou lancelets) et de petits crustacés, formant ainsi la famille des PF de type GFP. Une caractéristique unique des protéines de type GFP est la formation auto-catalytique du chromophore, la partie de la protéine qui absorbe la lumière visible et la réémet sous forme de fluorescence, à partir de trois acides aminés consécutifs situés au centre de la structure en tonneau β de la protéine. La famille des PF de type GFP a des maxima d'émission de fluorescence qui couvrent tout le spectre de la lumière visible, du bleu profond au rouge lointain. À la fin des années 2000, un autre type de PF a été obtenu à partir de phytochromes, une famille de photorécepteurs activés par la lumière rouge utilisant des bilines comme chromophore. L'intérêt de ces PF est que leurs maxima d'excitation et d'émission de fluorescence se situent dans la région du proche infrarouge du spectre lumineux. Cette région fait partie de la "fenêtre optique" des tissus vivants, dans laquelle l'absorption et la diffusion de la lumière par l'hémoglobine, l'eau et les lipides sont minimisées, et est donc la région du spectre à utiliser pour l'imagerie de fluorescence de corps entier. Ces protéines fluorescentes dans le proche infrarouge ont ainsi été appelées NIR-FPs. Au cours ce travail de thèse, j'ai résolu les structures cristallographiques de trois PF de type GFP (une PF verte très brillante, une chromoprotéine et une PF faiblement fluorescente dans le rouge) et de trois NIR-FPs dérivées d'un phytochrome monomérique. Les informations structurelles acquises sur la nature et l'environnement des chromophores m'ont permis d'expliquer leurs propriétés spectroscopiques particulières. Notamment, la structure de la chromoprotéine a révélé pour la première fois l'existence d'un chromophore qui forme une liaison covalente avec un résidu cystéine voisin, qui entraîne un décalage significatif vers le rouge du maximum d'absorption.