Evaluation and Development of Ligand-Dependent Fluorescent Reporters for Anaerobic Bacterial Imaging

by

Hannah E. Chia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2021

Doctoral Committee:

Associate Professor Julie S. Biteen, Co-Chair
Professor E. Neil G. Marsh, Co-Chair
Associate Professor Ajit Joglekar
Assistant Professor Kristin Koutmou
Assistant Professor Wenjing Wang

Hannah E. Chia

hechia@umich.edu

ORCID iD: 0000-0002-1132-329X

This Thesis is dedicated to my parents, to their sacrifices,

and to their dreams for us to flourish.

ACKNOWLEDGEMENTS

I am proudly presenting this dissertation with the help and support of a patient village of people by my side. As such, I would like to thank and acknowledge colleagues, friends, and family who have accompanied me on this journey.

First and foremost, I thank my advisors Dr. Julie Biteen and Dr. Neil Marsh for the scientific guidance and personal mentorship over the past five years. In their collective confidence, I have been able to grow tremendously as an independent scientist, feeling brave enough to dip my toes into new fields and techniques. I am eternally grateful for the consideration and faith both Julie and Neil extended to me as I pivoted to a multitude of disciplines that were outside of their scientific bread and butter. I appreciate the many hours both have spent with me poring over data as well as continual iterations on ideas and sentences. They have created spaces that I feel comfortable in as well as challenged to continue growing. I also thank Dr. Nicole Koropatkin for helping me learn about gut bacteria and for kindly acting as a third advisor at times. Of course, I thank my committee members, Dr. Ajit Joglekar, Dr. Kristin Koutmou, and Dr. Wenjing Wang, for their time and thoughtful feedback on my projects and career development.

I would be remiss not to mention my labmates, past and present, for their steadfast support, encouragement, and laughter. For my Marsh labmates, thank you for the hours spent together at the bench troubleshooting instruments and threatening cells to grow up and submit to our will. Thank you Soumi, Karl, Marie, Ayesha, Tim, April, Prathamesh, Srijoni, Takunda and many more for the years of fun and shared wisdom on the secrets of proteins. For my Biteen labmates, thank you for the time we've spent tinkering at the laser

tables and brainstorming new ideas to try. Thank you to Zech, Anna, Ziyuan, Chris, Saaj and more for the Slack kikis, memes, and beverage runs. I cherish the moments Stephen, Josh, Curly, and I have shared over hot sauce laden lunches and pop culture as well as the times Yilai and Ben spent fielding questions in my fledgling years. I would also like to thank Chanrith, who after starting off as my rotation mentor inadvertently became my long-term career mentor and believed in me when I dared to dream bigger.

I thank all of my friends who have helped preserve my physical and mental health and for caring for me through the highs and lows of graduate school. I am blessed to have met many life-long friends in the Midwest and away from home. Thank you Attabey, Francisco, Curly, Nhat, Zixuan, Frances, Sydney, Tom, Tiffany, Victor, and others and I will deeply miss the moments we've shared over food and nonsense. I am grateful to my friends from before graduate school who, like me, did not really know what I was getting into, but shared in my misery and triumph nonetheless. Thank you Miel, Sarah, Tianna, Michelle, Alexie, Katie, Joey, Becca, and Vivian for picking up the phone again and again.

I thank my family for the unconditional support and love throughout my life. My mother Lauren is my inspiration for finding joy and laughter in the little moments of life. Through sharing of his own journey, my father Ray continually teaches me the importance of perseverance and perspective. I thank my brother Benji for sharing photos of Olive, the queen of the Chia family. I also thank Helen and Jonathan, my new family members who have shared their home and care with me.

And finally, I thank my best friend and partner Christopher for the years of love and happiness. As he has read more miscellaneous writings than I dare to admit, this journey is one that is truly shared by both of us. Thank you for your steadfast companionship, for sharing in endless analyses of shows and movies, and for encouraging me when I took risks for our future. I look forward to our next chapters together in life.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES viii

LIST OF TABLES x

LIST OF ABBREVIATIONS xi

ABSTRACT xiv

CHAPTER

I. Introduction 1

1.1 Polymicrobial communities and the gut microbiome 2
1.2 Optical and fluorescence microscopy 4
1.3 Fluorescent reporters 8
1.4 Oxygen-independent fluorescent tools 12
1.4.1 Oxygen-independent fluorescent proteins 13
1.4.2 Bioconjugation approaches using self-labeling proteins 15
1.5 Engineering and evolution of fluorescent proteins 16
1.6 High-throughput screening of small molecules 20
1.7 Dissertation objectives 23

II. Bilin-based Fluorescent Proteins for Imaging Living Obligate Anaerobes 25

2.1 Introduction 25
2.2 Materials and methods 27
2.2.1 Bacterial strains cloning and cell cultures 27
2.2.2 Growth curves 30
2.2.3 Fluorescence microscopy 31
2.2.4 Flow cytometry 31

2.2.5 Protein expression and purification . . . 31
2.2.6 Protein characterization . . . 33
2.3 Results and discussion . . . 34
2.4 Conclusions . . . 46

III. A New Green Ligand-Dependent Fluorescent Reporter for Anaerobic Imaging . . . 48

3.1 Introduction . . . 48
3.2 Materials and methods . . . 50
3.2.1 Protein Expression . . . 50
3.2.2 HTS Assay Protocol . . . 50
3.2.3 Compound Libraries . . . 51
3.2.4 Assay Performance and Data Analyses . . . 51
3.2.5 Spectral Characterization and Titrations . . . 52
3.2.6 Cell Cultures . . . 53
3.2.7 Growth Curves . . . 54
3.2.8 Fluorescence Microscopy . . . 54
3.3 Results . . . 55
3.4 Discussion . . . 71
3.5 Conclusions . . . 74

IV. Designing a Red-Shifted Fluorescent Reporter Using Directed Evolution and Rational Design . . . 75

4.1 Introduction . . . 75
4.2 Materials and methods . . . 76
4.2.1 Cloning and Library Construction . . . 76
4.2.2 Colony-based Screening for UnaG Variants . . . 82
4.2.3 Lysate-based Screening for UnaG Variants . . . 82
4.2.4 FACS Screening for Red-Shifted UnaG Variants . . . 83
4.2.5 Protein Expression and Purification . . . 84
4.2.6 Spectral Characterization . . . 84
4.3 Results . . . 85
4.3.1 Rational design of a red-shifted bv-binding UnaG . . . 85
4.3.2 Directed evolution of a red-shifted bv-binding UnaG . . . 91
4.4 Discussion and future work . . . 98
4.5 Conclusions . . . 101

V. Conclusions . . . 102

5.1 Introduction . . . 102
5.2 BBFPs for imaging live obligate anaerobes . . . 102
5.3 HTS to identify new UnaG ligands . . . 105
5.4 Engineering ligand recognition in UnaG . . . 107

5.5 Other ligand-dependent reporters for anaerobic imaging 108
5.6 Overarching conclusions 109

APPENDICES 110

BIBLIOGRAPHY 114

LIST OF FIGURES

Figure

1.1 Optical and fluorescence microscopy . . . 5
1.2 General overview of fluorescence and epifluorescence . . . 7
1.3 GFP and DsRed FP families . . . 9
1.4 GFP and DsRed fluorescence maturation . . . 11
1.5 Summary of probes for anaerobic live-cell imaging . . . 13
1.6 Comparison of chemical genetics approaches . . . 21
2.1 Codon-optimized UnaG sequence . . . 33
2.2 GFP requires oxygen to produce fluorescence . . . 34
2.3 UnaG uses br to produce fluorescence . . . 35
2.4 Br and bv do not fluoresce . . . 35
2.5 Growth curves of *B. theta* cells expressing UnaG and IFP2.0 . . . 36
2.6 Flow cytometry results for *B. theta* expressing UnaG or GFP . . . 37
2.7 Imaging of UnaG-labeled *B. theta* cells . . . 38
2.8 IFP2.0 uses bv to produce fluorescence . . . 39
2.9 Fluorescence intensity of IFP2.0-labeled *B. theta* cells . . . 40
2.10 Outer membrane BBFP labeling of *B. theta* cells . . . 40
2.11 BBFPs can be used in mixed culture imaging and multi-color imaging . . . 42
2.12 UnaG and IFP2.0 are orthogonal fluorescent labels . . . 43
2.13 Holoprotein characterization of UnaG and IFP2.0 . . . 44
3.1 Screening methodology for identifying new, red-shifted ligand-UnaG pairs 56
3.2 Scatterplots of the entire primary HTS campaign . . . 57
3.3 Concentration-dependent curves of compounds obtained through HTS . 59
3.4 Structures of br and compounds obtained through HTS . . . 60
3.5 UV-vis absorbance and fluorescence emission spectra of compounds obtained through HTS . . . 61
3.6 UV-vis absorbance and fluorescence emission of compounds obtained through HTS incubated with UnaG . . . 62
3.7 Growth curves of *B. theta* in media supplemented with compounds obtained through HTS . . . 63
3.8 Spectral properties of UnaG-**2** and UnaG-**4** . . . 64
3.9 Fluorescence excitation and emission spectra of **2**, **4**, UnaG-**2**, and UnaG-**4** 65
3.10 Competition titrations of **2** and **4** against UnaG-br . . . 66
3.11 Titration of BSA with **2** and **4** . . . 68

3.12 Phase-contrast and fluorescence imaging of UnaG-labeled *B. theta* cells with **4** . . . 69
3.13 Phase-contrast and fluorescence imaging of *B. theta* and *E. coli* expressing UnaG with **2** . . . 70
3.14 Phase-contrast and fluorescence imaging of a mixed culture of *B. ovatus* and *B. theta* expressing different BBFPs . . . 71
3.15 Phase-contrast and fluorescence imaging of *B. ovatus* and *B. theta* expressing BBFPs . . . 72
4.1 Structures of br and bv as well as the UnaG-br binding pocket . . . 86
4.2 UV-vis absorbance spectra of UnaG point mutants bound with br or bv . 87
4.3 Fluorescence emission spectra of UnaG point mutants bound with br or bv . . . 88
4.4 Fluorescence emission spectra of Wt, M51C, and N57A UnaG bound with bv . . . 89
4.5 UV-vis absorbance and fluorescence emission of Wt, V2L, and V2L/M51C UnaG bound with br . . . 90
4.6 Summary of single UnaG variants . . . 91
4.7 Colony- and lysate-based screening for directed evolution . . . 93
4.8 Strategy for FACS-based directed evolution . . . 94
4.9 FACS analysis of bv-binding UnaG mutant library in *E. coli* . . . 95
4.10 Sequence alignment of bv-binding UnaG variants after FACS analysis . . 96
4.11 Summary of L4-Cy3 UnaG variant for br- and bv-binding abilities . . . 97
A.1 Summary of probes for anaerobic live-cell imaging . . . 112
A.2 SDS-PAGE analysis of MBP-UnaG purifications . . . 113
A.3 L4-Cy3 mtUnaG variant as a building block for protein cage design . . . 113

LIST OF TABLES

Table

2.1 Plasmids ordered and generated for BBFP evaluation 28
2.2 Oligonucleotides used for BBFP evaluation 29
3.1 Confirmed Hits from HTS Assay 52
3.2 Summary of HTS campaign 58
3.3 Properties of confirmed hit compounds and compounds bound to UnaG . 64
3.4 $K_{d(app)}$ and true K_d from fluorescence competition titrations for 2 and 4 . 65
4.1 Oligonucleotides used for bv-binding UnaG variants 77
4.2 Plasmids ordered and generated for bv-binding UnaG variants 80

LIST OF ABBREVIATIONS

BBFP Bilin-based Fluorescent Protein

BFP Blue Fluorescent Protein

br Bilirubin

bv Biliverdin

B. ovatus *Bacteroides ovatus*

B. theta *Bacteroides thetaiotaomicron*

CFP Cyan Fluorescent Protein

DOS Diversity-Oriented Synthesis

E. coli *Escherichia coli*

EMCCD Electron Multiplying Charge-Coupled Device

ep-PCR Error Prone Polymerase Chain Reaction

FACS Fluorescence Assisted Cell Sorting

FbFP Flavin Mononucleotide (FMN)-based Fluorescent Protein

FP Fluorescent Protein

FRET Fluorescence Resonance Energy Transfer

FISH Fluorescence In Situ Hybridization

FMN Flavin Mononucleotide

GI Gastrointestinal

GFP Green Fluorescent Protein

HTS High-throughput Screening

IR Infrared Radiation

LOV Light-Oxygen-Voltage

NA Numerical Aperture

OD Optical Density

PALM Photoactivated Localization Microscopy

PAmCherry Photoactivatable mCherry

PSF Point Spread Function

RFP Red Fluorescent Protein

R. bromii *Ruminococcus bromii*

SDM Site-directed Mutagenesis

SPT Single-Particle Tracking

Sus Starch Utilization System

STORM Stochastic Optical Reconstruction Microscopy

TEM Transmission Electron Microscopy

YFP Yellow Fluorescent Protein

Y-FAST Yellow Fluorescence-Activating and Absorption-Shifting Tag

ABSTRACT

Fluorescence-based investigations of living cells and biological phenomena have been largely aided by the development of fluorescent proteins (FPs) to span a wide range of different colors and specialized photophysical properties. The most commonly used families of FPs are derived from Green Fluorescent Protein (GFP) and DsRed, which require oxidative post-translational modifications to produce mature chromophores. This oxidation step precludes their use in anaerobic conditions. Nevertheless, there are countless oxygen-sensitive biological systems and mechanisms that have yet to be explored using otherwise ubiquitous techniques such as fluorescence microscopy.

This dissertation aims to bridge this gap by evaluating and developing a set of ligand-dependent reporters, Bilin-based Fluorescent Proteins (BBFPs), as oxygen-independent fluorescent probes for applications in anaerobic bacterial systems such as the gut microbiome. Chapter I provides background on the importance of the gut microbiome and how investigations of this medically relevant polymicrobial community can be aided by fluorescence-based investigations. It also provides an overview of the development of both oxygen-dependent and -independent FPs and the latters' respective advantages and disadvantages in oxygen-free imaging.

BBFPs are adapted from eel and bacterial phytochromes and bind the tetrapyrroles bilirubin (br) or biliverdin (bv) to fluoresce. These reporters are oxygen-independent due to fluorescence resulting from ligand binding, and the fluorogenic nature of these ligands enables an improved signal to noise ratio over other fluorescent dyes. Chapter II describes the first implementation of BBFPs in anaerobic gut bacteria using the blue-green UnaG

and the far-red IFP2.0. These FPs are used to label the commensal gut bacteria *Bacteroides thetaiotaomicron (B. theta)* in monoculture in the common GFP and Cy5 microscopy channels as well as used in mixed-species two-color imaging.

To broaden the utility of the UnaG FP, I set out to diversify the color palette of BBFPs in Chapters III and IV. In Chapter III, high-throughput screening (HTS) has been employed to identify new fluorogenic ligands that could bind UnaG and fluoresce in different wavelengths outside of the native UnaG-br pair's blue 488-nm channel. The new resulting UnaG-ligand pair can be used in the common green 532-nm excitation channel and again, could be used in anaerobic bacterial imaging and as an orthogonal label with IFP2.0-bv. As UnaG is a more desirable than other BBFPs in terms of size and monomeric form, Chapter IV describes efforts to engineer a bv-binding UnaG variant for a red-shifted variant that could be excited by blue or red fluorescence. These results provide insight into the br binding pocket and provide a blueprint for future engineering efforts in this BBFP.

Finally in Chapter V, future directions for oxygen-independent reporters are discussed that range from adjustments to engineering ligand recognition in UnaG to alternative ligand-dependent reporter systems that may also be implemented in anaerobic systems. This dissertation collectively validates and expands the fluorescent toolbox for probing anaerobic bacterial systems and extends fluorescence applications to previously inaccessible biological systems.

CHAPTER I

Introduction

Part of the work presented in this chapter was published in

Current Opinions in Chemical Biology:

Chia, H.E., Marsh, E.N.G., and Biteen, J.S.

Extending fluorescence microscopy into anaerobic environments.

Current Opinions in Chemical Biology, **51**, 98-104 (2019). DOI: 10.1016/j.cbpa.2019.05.008

This dissertation extends fluorescence microscopy into oxygen-sensitive polymicrobial communities, which prior to the work presented was largely inaccessible to common fluorescence and microscopy investigations. In this chapter, I comment on the importance of understanding polymicrobial communities, particularly the medically and therapeutically relevant gut microbiome. I introduce optical and fluorescence microscopy as an important method for understanding polymicrobial interactions in live-cell environments. I describe the limitations of common fluorescence labeling methods in anaerobic imaging as well as summarize past successful applications of oxygen-independent tools in fluorescence microscopy. Finally, I present an overview of the thesis in developing ligand-dependent fluorescent proteins.

1.1 Polymicrobial communities and the gut microbiome

Microbes in nature are seldom found in single species, but rather exist in complex networks of cohabitation with other species that include both symbiotic and competitive interactions. Polymicrobial communities are collections of organisms including bacteria, viruses, and fungi that live on biotic and abiotic surfaces; these diverse collections of organisms are typically suspended in self- or host-derived biofilms that are largely composed of hydrated polysaccharide matrices. Understanding the assembly and regulation of polymicrobial communities continues to have profound impacts throughout human society including in health (gut [1] and oral microbiomes [2]), in environmental ecology (soil microbiomes [3]), and in industry (fossil fuel pipelines [4]).

The gut microbiome is one of the most studied polymicrobial communities due to its importance to human health and disease: it comprises of trillions of bacteria colonizing the intestinal lumen. This microbiome provides hosts vital pathogenic protection as well as symbiotic release of nutrients by degrading host-indigestible material. Perturbations to this crucial microbiome have already been linked to inflammatory and metabolic issues as well as respiratory, digestive, cardiovascular, and even neurological illnesses [5]. Advancing understanding of the gut microbiome, both host-microbe and microbe-microbe interactions, can inform approaches to restore microbiome symbiosis and manage or prevent disease [6].

Gut microbiota are established at birth and continually shifted throughout life, largely through diet. The gut microbiome comprises multiple polymicrobial communities arranged in distinct compartments along the gastrointestinal (GI) tract [7, 8]. Although the order of first colonization of GI tract is unknown, microbial composition becomes more diverse and abundant proceeding through the lower GI tract [9]. In adult humans, the colon is largely dominated by bacteria from Bacteroidetes and Firmicutes phyla followed by those from Actinobacteria, Verrucomicrobia, and Proteobacteria phyla; these diverse bacterial species have a multitude of specialized function. For example, *Bacteroides thetaiotaomi-*

cron (B. theta) and *Bacteroides ovatus (B. ovatus)*, two extensively studied gut species, are well established as generalists capable of degrading a wide variety of poly- and oligosaccharides. Others have specialized function in degrading only specific types of carbohydrates and nutrient sources, such as *Ruminococcus bromii (R. bromii)* utilizing resistant starches [10].

In addition to the functions of individual species, communication in physical and chemical forms facilitate interactions and behaviors. Rather than build extracellular matrices scaffolds, as occurs in many biofilms, the gut microbiome is suspended in a mucus layer produced by human epithelial cells. Through a chemical communication mechanism called quorum sensing, which facilitates intra- and inter-species gene expression, gut bacteria coordinate colonization of mucosa [11]. In addition to bacterial crosstalk, quorum sensing molecules produced by bacteria can even interact with host cells [12], which may additionally explain the tight relationship between the gut microbiome and host health. A mature gut microbiome has a plethora of interactions ranging from competitive behavior for shared carbon sources [13] or spatial colonization to symbiotic behavior such as cross-feeding [14]. These interactions can be characterized for co-occurrence and co-exclusion patterns [15], which reflects the distinct communities of bacteria that coexist within the larger microbiome.

Despite the vast amount of research that has been performed on bacteria from the gut microbiome, such research is typically limited in reconstructing a full picture of polymicrobial interactions. For example, investigations into intestinal microbial diversity has largely been performed using epidemiological methods. Using 16s rRNA sequencing, advances in high-throughput sequencing enable detection of changes in intestinal microbial composition to compare healthy states with diseased states [16, 17] or different populations [18, 19]. While sequencing can identify key bacterial families or species associated with diseased states, this methodology is also limited in determining how microbial communities respond to stimuli in real time. On the other hand, in-depth biochemical and

biophysical investigations have helped parse individual species' gene expression profiles to different stimuli [20] as well as downstream protein-protein interactions [21]. However, these types of studies are typically limited to single-species contexts. Given the complexity of microbial diversity and interactions within the gut microbiome as well as the potential for medical innovation if this diversity is understood, we must push toward investigating real-time polymicrobial interactions while maintaining live-cell conditions.

1.2 Optical and fluorescence microscopy

Optical and fluorescence microscopy are ideal techniques for understanding polymicrobial interactions with spatial and temporal resolution [22]. Simple light microscopes at 100x magnification can visualize cellular morphology and some subcellular compartments in small prokaryotic bacteria (1 to 3 microns) as well as in eukaryotic organisms. Although techniques such as electron microscopy can provide much higher subcellular resolution, electron microscopy also requires fixed cells and is a destructive process [22]. In contrast, light and fluorescence microscopy are non-invasive and are thus ideal for maintaining live-cell conditions and observing samples over long periods of time [22].

In conventional light microscopy, an illuminating white light is absorbed or reflected by the sample and the remaining transmitted light creates a pattern (often known as brightfield imaging). The resulting image can distinguish the edge of the cell or specimen and the background; additional staining can increase contrast to observe large features, though staining is not compatible with live-cell imaging. Phase-contrast imaging improves upon simple illumination by using refraction and interference of light waves to allow for live-cell imaging with high contrast (Fig. 1.1 A). Phase-contrast imaging is particularly useful for analyzing monolayers of bacteria as the resulting images will have a sharp dark pattern against white backgrounds that enable analyses of cellular morphology and cell-counting algorithms [23].

The resolution of optical light microscopy hinges on the magnification and the nu-

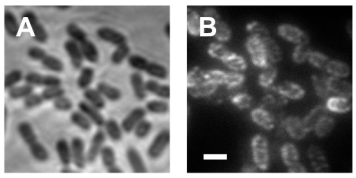


Figure 1.1: Light microscopy used to image live *B. theta*. **(A)** Phase-contrast imaging reveals cell morphology while **(B)** fluorescence imaging provides specific labeling of an outer-membrane protein. Scale bar: 1 μm.

merical aperture (NA) of an objective as well as on visible light itself. The NA describes the ability of an objective to maximally gather incoming light:

$$NA = nsin\theta \tag{1.1}$$

In equation 1.1 above, n is the refractive index of the medium between the coverslip and the objective (e.g. air, water, or oil) and θ is one-half of the objective's angular aperture. Most high-magnification objectives have an NA of 0.6 - 1.5. Due to the light source in optical light microscopy (400-700 nm), magnification is limited because of the diffraction limit of light or the Rayleigh limit:

$$d = \frac{\lambda}{2NA} \tag{1.2}$$

The illuminating light wavelength (λ, nm) and NA combined present a physical limit in distinguishing two points from one another. This limitation essentially describes the inability of a broadly illuminated sample, such as in light microscopy, to resolve fine subcellular details [24]. Hypothetically, one could image a single particle in a light microscope, but the particle would visualized, at best, as an approximately 200 nm diffraction-limited point — a poorly resolved object. This point is known as the point spread function (PSF). Now, if there are two single particles next to one another, the two PSFs could overlap if the points are close enough in physical distance and be difficult, if not impossible if they are on top of one another, to resolve. For context, actual biological samples are comprised

on countless particles that overlap with one another to create macromolecular structures and complexes that are interesting to study. Since most biological samples have low absorbance and low contrast in the 400-700 nm range of visible light, optical microscopy is mostly useful for large-scale live-cell visualizations of cell shapes or for observing spatial patterning of polymicrobial and monolayer samples.

By labeling only subsets of cellular components like organelles or proteins, fluorescence microscopy improves upon resolution over basic light microscopy to uncover finer subcellular details (Fig. 1.1 B). In short, the phenomenon of fluorescence is characterized by a fluorophore absorbing a photon (excitation) and subsequently releasing a photon at a longer wavelength (emission) (Fig. 1.2 A). Upon excitation, the fluorophore molecule is excited from ground state, S_0, to a higher energy excited state, S_1. Internal conversion, a loss of energy through vibrational relaxation, brings the fluorophore, the chemical structure of the fluorescent molecule, to the lowest energy state of the excited S_1 state. When the system returns to ground state S_0, a detectable lower energy, long wavelength photon is emitted. The difference between the maximal excitation wavelength and the maximal emission wavelength is the Stokes shift (Fig. 1.2 B). Each fluorophore has a distinct excitation and emission spectra as well as Stokes shift [25]. On the other end of imaging, mirrors inside the microscope direct filtered excitation illumination to the sample as well as filtered emitted fluorescence toward a detection source (Fig. 1.2 C).

While much of fluorescence microscopy observes bulk fluorescence signal from labeled targets, it may also omit functional or structural information on anything that is smaller than the diffraction limit of emitted light. In order to obtain information on a single-molecule level, a combination of of high-sensitivity detection, image processing of point emitters, and experimental considerations must be adjusted from standard fluorescence microscopy. Advances in fluorescence imaging methodology have yielded techniques for super-resolution on the single-molecule scale; these techniques include Photoactivated Localization Microscopy (PALM), Single-Particle Tracking (SPT), and Stochastic

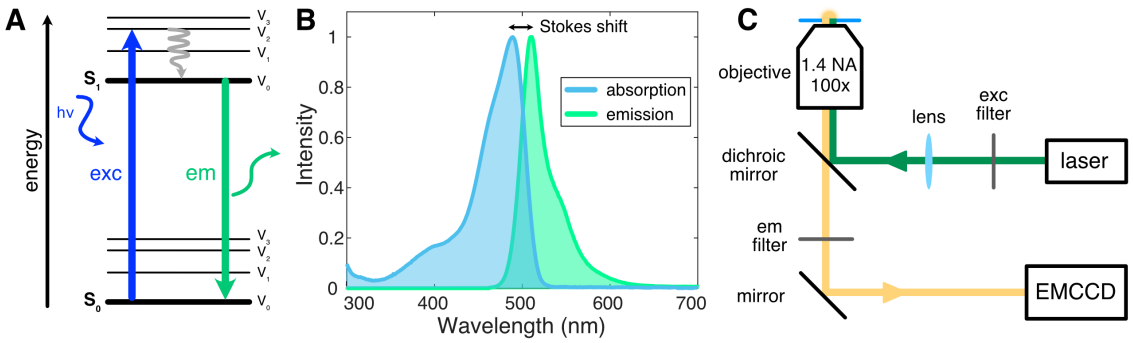


Figure 1.2: General overview of fluorescence and epifluorescence. (A) A simplified Jablonski diagram illustrates the phenomenon of fluorescence. An absorbed photon excites (exc) an electron from the singlet ground state (S_0) to a singlet excited state (S_1). Emission (em) occurs after some loss of vibrational energy and relaxation back to S_0. **(B)** EGFP absorption spectrum (blue) and fluorescence emission spectrum (green). **(C)** A simple epifluorescence microscope with excitation (exc) and emission (em) light paths as well as filters and mirrors. In our lab's experimental setup, the excitation beam is directed from the laser source to the sample and emitted light is directed to an Electron Multiplying Charge-Coupled Device (EMCCD) camera.

Optical Reconstruction Microscopy (STORM) [22, 26]. These techniques all require some form of experimental design such that only one molecule is detected at a time to avoid overlapping PSFs for precise localization fitting. Super-resolution microscopy is particularly powerful for uncovering heterogeneities that would otherwise be masked by bulk fluorescence measurements, but these techniques cannot be applied to every biological system unless compatible probes and sufficiently sensitive detection methods are available.

In investigations of the human gut microbiome, both conventional fluorescence microscopy and super-resolution techniques have been used to label species of interest [27], to visualize bacterial colonization [28], and to characterize protein dynamics [21]. As most bacteria from the gut microbiome are oxygen-sensitive, most imaging of gut bacteria are performed on fixed cell samples or on dormant, growth arrested, cells. Fluorescence In Situ Hybridization (FISH) is a popular strategy for targeted labeling of bacteria through probes specific to 16s rRNA, though the technique is incompatible with live-cell imaging. Using FISH, researchers have quantified spatial patterning of gut bacteria in colonized

mice to investigate differences in colonization density in different parts of the intestines as well as in response to changed diet [29, 30]. These studies have uncovered distinct colonization patterns in the colon and polymicrobial organization on a species level. Since a key function of most intestinal bacteria is sugar degradation, fluorescent labeling of carbohydrates is another emerging technology to label intestinal bacteria. One approach can provide taxonomic specific labeling of Gram-negative and Gram-positive bacteria through click chemistry directed at different lipopolysaccharides presented on the cell surface [31]. Another approach provides single cell resolution of cells during uptake of fluorescently labeled sugars to uncover heterogeneous metabolic phenotypes [32]. Fluorescently labeled polysaccharides can even be used to track a species colonization in a mouse intestine in real time [33]. Our lab has also extensively used fluorescence microscopy for live-cell characterization of protein complexes in *B. theta*. Using HaloTag technology [34] as well as the Photoactivatable mCherry (PAmCherry) fluorescent reporter, proteins involved in starch utilization were tracked by SPT to uncover heterogeneities in starch binding dynamics and coordinated protein complex assembly [21, 27].

1.3 Fluorescent reporters

While applications of fluorescence microscopy to investigate biological phenomena are countless, fluorescent proteins (FPs) are the gold standard for labeling species and proteins. Through the extensive development of FPs with different colors ranging from blue to near-infrared radiation (IR) as well as interesting photophysical properties like photoactivation [35, 36] and photoswitching [37, 38], FPs have enabled expansive applications of fluorescence microscopy from multi-color imaging to super-resolution microscopy. FPs are particularly advantageous over other labeling strategies because FPs can be genetically encoded, easily inserted into genetic material for expression or appended to another protein's termini.

The green fluorescent protein (GFP) is the most significant and commonly utilized FP

and was the basis of the 2008 Nobel Prize in Chemistry. The protein was first isolated from jellyfish *Aequorea victoria* (avGFP) and through a single point mutation (S65T) yielded the blue-green GFP variant that is commonly utilized in research laboratories [39]. In addition to its use in basic science research such as fluorescence labeling for microscopy or as gene expression reporters, GFP and other FPs have been developed in biotechnology and medicine for use as assay markers and biosensors. Variants of GFP with different colors and distinct excitation and emission maxima were engineered including blue fluorescent protein (BFP) [40], cyan fluorescent protein (CFP) [41], and yellow fluorescent protein (YFP) [42] (Fig. 1.3). These variants cover approximately 80 nm of the visible light spectrum with emission ranging from BFP (445 nm) to YFP (527 nm). Now, the BFP-YFP pair remains a popular pair for Fluorescence Resonance Energy Transfer (FRET) experiments. These variants retain the same β-barrel fold as the wild-type GFP, but the internal chromophore is altered by lengthening π-conjugation through amino acid substitution, changing chromophore protonation state, increasing π-π stacking, or shifting the electrostatics of the local environment. Other mutations have improved GFP by enhancing brightness (EGFP) [43] and promoting monomeric forms through dimerization disruption.

DsRed is the other FP that has profoundly impacted biology and dramatically ex-

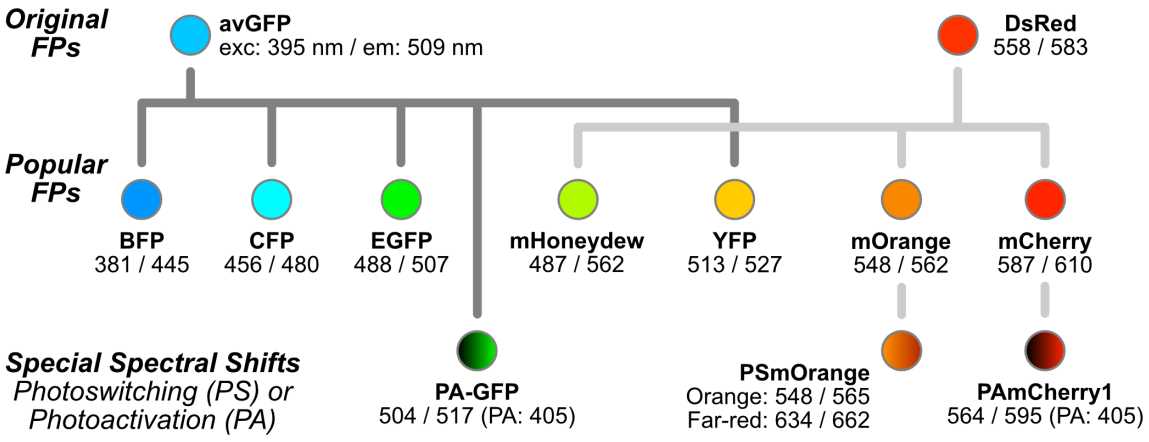


Figure 1.3: GFP and DsRed FP families. Popular fluorescent proteins used in research and sensor applications are typically variants derived from GFP and DsRed. Figure adapted from FPbase.org [44].

panded the color palette of FPs, and is a red fluorescent protein (RFP) isolated from *Discosoma* coral [45]. The discovery and development of RFPs enables additional multi-color fluorescence labeling and overall greater imaging capacity. RFPs are particularly advantageous over blue-green FPs, including GFP, because red fluorescence is far away from the blue intrinsic cellular background. RFPs have also been coupled with other FPs for multicolor imaging as well as additional spectral pairs for FRET. The original DsRed was improved for live-cell imaging through protein engineering to change the natively tetrameric protein into robust monomeric forms [46]. Monomeric DsRed was further evolved into a wide range of colors named after various fruits such as mHoneydew, mOrange, and mCherry [47] (Fig. 1.3). These engineering efforts generated variants that cover a different range of the visible light spectrum than GFP variants with emission ranging from green mHoneydew (562 nm) to yellow mTangerine (585 nm) and red mCherry (610 nm). The mFruits similarly retain β-barrel structure as monomeric DsRed, which is homologous to evolved GFP variants. Again, the mFruits have amino acid substitutions that extend the π-conjugation of the chromophore as well as stabilize fluorescence lifetimes to improve brightness. The mFruit variants also improve upon the wild-type DsRed in terms of faster time to full chromophore maturation, reduced pH sensitivity, and tolerance for N-terminal fusions [47].

In spite of the advances in making a rainbow of FP colors and their vast applications to investigating biology, GFP and DsRed variants are not compatible with live-cell investigations in oxygen-free environments. GFP and DsRed proteins alike internally house a chromophore that must undergo oxidative post-translational modifications before fluorescing (a process also known as fluorescence maturation) (Fig. 1.4), which typically takes place on a time scale of 15 to 25 minutes. For GFP, the chromophore is formed out of three residues (Thr65, Tyr66, and Gly67) and three intramolecular reactions: (1) the amide of Gly67 performs a nucleophilic attack on the carbonyl of Thr65 to form an imidazoline, (2) a dehydration following cyclization, and (3) an oxidation to the Cα-Cβ Tyr66 backbone

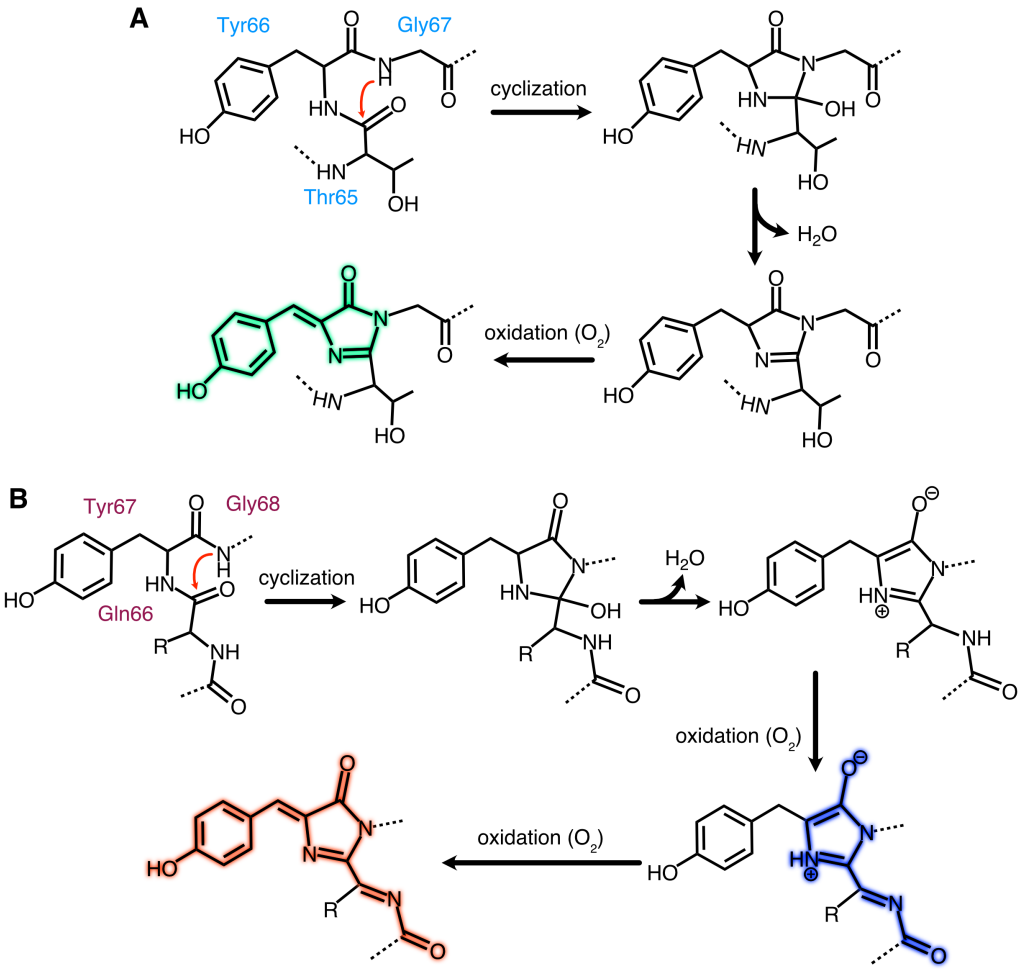


Figure 1.4: GFP and DsRed fluorescence maturation. Fluorescent protein variants derived from **(A)** GFP and **(B)** DsRed undergo internal intramolecular reactions and require oxidative post-translational modifications for chromophore maturation. Figure adapted from [48] and [49].

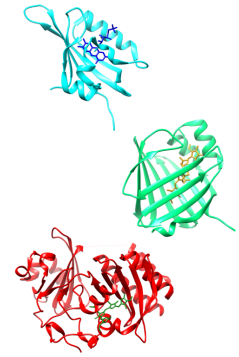
to complete conjugation between the formed imidazoline and the Tyr sidechain [48] (Fig. 1.4 A). DsRed and its variants similarly rely on key amino acid residues and intramolecular reactions to form the fluorescent chromophore; while multiple explanations for how DsRed reaches full maturation exist, past literature best supports a mechanistic scheme involving a BFP-like intermediate [49]. Cyclization occurs through nucleophilic attack of the Gly68 peptide bond on the amide sidechain of Gln66. Subsequent dehydration and aerial oxidation reactions form a choromophore similar to GFP. However, an additional dehydrogenation of the Cα-N bond of Gln66 extends the π-conjugation to generate a

strongly red-shifted chromophore [45, 49] (Fig. 1.4 B).

The oxidation steps precludes the use of popular GFP and DsRed variants to study obligate anaerobes. To date, microscopy applications of these FPs in gut microbiome research are performed on fixed (dead cell) samples that have been exposed to air. In our lab, applications of the DsRed variant PAmCherry were performed on dormant cells that, again, do not necessarily reflect biological function in live cells. To take advantage of fluorescence enabled methods to investigate anaerobic live-cell populations, we must turn toward other oxygen-independent fluorescent tools.

1.4 Oxygen-independent fluorescent tools

Developing robust fluorescent probes suited for anaerobic imaging would allow biological exploration of anaerobic systems and extend live-cell fluorescence imaging to medically important organisms and microbial communities. While three distinct alternative strategies have been, or could be, employed to label oxygen-sensitive bacterial cells, I will focus on two techniques that have been demonstrated extensively in the literature: (1) oxygen-independent fluorescent proteins and (2) bioconjugation using self-labeling tags (Fig. 1.5). In this section, I will also address advantages and disadvantages of each labeling strategy for live-cell imaging as well as future efforts that could make an approach more robust and easier to employ. An ideal probe should balance brightness (for high contrast), minimal biological disturbance (to not affect the interrogated system), and ease of implementation (such that the highest possible proportion of target molecules are labeled). A more comprehensive overview of tools that can potentially be used for anaerobic imaging can be found in my previously published review [50] and in the appendix (full table in Fig. A.1.



Probe	Color Brightness Size*	Advantages	Disadvantages
Oxygen-independent Fluorescent Protein Approaches			
Flavin-mononucleotide-based fluorescent proteins			
FbFPs (BsFbFP, PpFbFP) iLOV	Cyan Dim Small	• Genetically encoded • Ligand is available in most systems • Demonstrated in live anaerobic bacterial cells	• Low contrast: FbFPs are dim and similar in color to cellular autofluorescence
Bilin-binding fluorescent proteins			
UnaG	Green Same Small	• Genetically encoded • Ligand is cell permeable in some systems • Ligands are fluorogenic • Demonstrated in live anaerobic bacterial cells	• Ligand is not water-soluble • Low contrast: IFPs are dim • IFPs dimerize at high concentrations
IFP1.4 IFP2.0 mIFP	Red Dim Same		
Bioconjugation Approaches			
Self-labeling proteins			
SNAP CLIP HaloTag TMP	Varies Varies Same	• Genetically encoded • Tags can be multiplexed • Demonstrated in live anaerobic bacterial cells (HaloTag)	• Low contrast: Most ligands are not fluorogenic • Most ligands are not cell-permeable

Figure 1.5: Summary of probes for anaerobic live-cell imaging. *Brightness and size relative to GFP. Figure adapted from [50].

1.4.1 Oxygen-independent fluorescent proteins

Flavin Mononucleotide (FMN)-based Fluorescent Proteins (FbFPs) were the first described FP approach for anaerobic labeling of anaerobic bacteria. This class of FPs contains the photoactive Light-Oxygen-Voltage (LOV) domain to produce blue fluorescence. Native LOV proteins covalently bind the Flavin Mononucleotide (FMN) cofactor and are found in bacterial and plant photosensors. While native LOV proteins are typically non-fluorescent, FbFPs have been engineered to fluoresce [51]. FbFPs have already been used in many biological conditions including as fluorescent markers for labeling anaerobic gut bacteria [52] and hypoxically cultured mammalian cells [53] as well as gene expression reporters for anaerobically cultured bacteria [54, 55]. The development of FbFPs has been summarized recently in the literature [56, 57].

FbFPs provide distinct advantages over GFP variants and other oxygen-independent

FPs. FbFPs are smaller than GFP-like FPs (10–15 kDa versus 27 kDa), and may therefore be less disruptive to cellular signalling or protein-protein interactions. Whereas GFP variants may oligomerize, FbFPs like iLOV are monomeric [56,58], which further decreases the risk that these tags will introduce artifacts into the biological system under investigation. An advantage of FbFPs over other oxygen-independent FPs is that FMN and its precursor molecule riboflavin are essential molecules in metabolism and thus do not need to be externally supplied in live-cell imaging.

Despite these advantages, the weak fluorescence signal of FbFPs has prevented wide adoption of these tools. While engineering efforts have made FbFPs brighter and more photostable [59], the relatively weak fluorescence signal is difficult to distinguish from the intrinsic cellular auto-fluorescence background. Further directed evolution and protein engineering targeted at the FMN-binding pocket may increase FbFP brightness and might red-shift the excitation and emission peaks away from blue intrinsic fluorescence signal.

Bilin-based Fluorescent Proteins (BBFPs) are another class of FPs that can enable oxygen-independent fluorescence labeling. Generally, BBFPs belong to a subset of proteins that reversibly bind and transport fatty acids between intra- and extra-cellular membranes. BBFPs, in particular, bind the poryphin-derived chromophores bilirubin (br) and biliverdin (bv), which are breakdown products of heme metabolism, to fluoresce. Since these proteins bind exogenously added br and bv, there is no oxygen requirement for fluorescence production.

One natively fluorescent BBFP, UnaG, is isolated from Japanese unagi eels and only becomes fluorescent when reconstituted with br [60,61]. UnaG binds br with high affinity and specificity, and has similar excitation and emission wavelengths to GFP. UnaG has previously been used to label mammalian HeLa cells [60] as well as developed into a protein-protein interaction sensor [62] and a calmodulin sensor [63]. Similar to FbFPs, UnaG is smaller in size (15 kDa) than oxygen-dependent FPs.

Infrared and far-red BBFPs have also been developed and are comparable in spectral

characteristics to the far-red mKate FP family. The first BBFP, IFP1.4, was developed from a bacterial phytochrome [64] and binds bv [65], though it suffers from dim fluorescence due to low quantum yield. The cellular brightness of IFP1.4 was improved by directed evolution protein engineering to generate IFP2.0, which has been successfully utilized for imaging neurons in *Drosophila* [66]. IFP2.0 was further improved upon for labeling purposes by protein engineering to create a truly monomeric form in mIFP [67].

I demonstrated the first application of BBFPs in anaerobic bacterial systems and will discuss their implementation in Chapter II of this dissertation. While I have demonstrated a successful application of UnaG and IFP2.0, other factors may preclude implementation of BBFPs at large. BBFPs are by nature restricted to the two bilin cofactors for usable fluorescent colors, which may restrict multi-color imaging beyond two targets. Likewise, bacterial cell walls are largely impermeable to many extracellular material and may limit BBFP labeling to outer membrane targets if br or bv cannot diffuse or be imported into the cell.

1.4.2 Bioconjugation approaches using self-labeling proteins

Alternatives to genetically encoded FPs are genetically encoded "self-labeling" proteins. Similar to FPs, a self-labeling protein tag can be appended to a target of interest. However, self-labeling proteins are not intrinsically fluorescent and instead react with ligands containing a fluorophore.

Self-labeling proteins combine the specificity and ease of a genetically encoded tag with the functional diversity of synthetic chemistry, as swapping the reactive ligand changes the functionality of the tag. Large libraries of functionalized ligand substrates encompassing a wide range of spectral properties for fluorescence microscopy, including enhanced brightness and photo-activation, have been generated [68]. The orthogonal HaloTag, SNAP, and CLIP tags can be used in conjunction for simultaneous multi-color fluorescence microscopy [69]. The use of fluorescent ligands is also advantageous as synthetic dyes are

typically brighter and more photostable than FPs.

These self-labeling proteins do not require oxidation for ligand conjugation and fluorescence. The HaloTag, derived from the bacterial halogenase, covalently binds molecules containing a chloroalkane moiety [34]; HaloTag technology has previously been used in anaerobic single-molecule tracking experiments in the gut microbe *B. theta* [21, 27]. Similarly, fusing a target protein to dihydrofolate reductase (eDHFR) allows for labeling with trimethoprim fluorophores [70]. The SNAP [71] and CLIP [72] tags, derived from the human DNA repair protein O^6-alkylguanine-DNA alkyltransferase, are respectively reactive toward O^6-alkylguanine (AG) and O^2-benzylcytosine (BC) substrates; multiple AG and BC substrates with attached fluorescein (green), rhodamine (red), and Cy5 (far-red) fluorescent probes have been generated.

As with all fluorescent probes, self-labeling proteins require some degree of optimization with regard to ligand choice and tag placement. Ligands can be easily introduced to a biological sample by supplementing the growth medium, but non-specific ligand interactions may occur if excess ligand is not removed since the unbound ligands themselves may also be dimly fluorescent and result in increased background noise. The ligands are not typically cell-permeable in bacterial systems and are limited to outer membrane labeling. Like traditional FPs, the self-labeling protein tags may be disruptive and require strategic placement to minimize artificial interactions in *in vitro* and *in vivo* experiments.

1.5 Engineering and evolution of fluorescent proteins

From imaging to biosensor applications, FPs of all origins have been extensively engineered to improve brightness, to enhance protein stability, and to generate spectrally unique variants [73]. So many variants have been collectively created that resources such as the site FPbase.org have been created to track the lineage and spectral properties of 200+ proteins [44]. Perhaps more so than other enzymes and proteins, FPs are particularly attractive candidates for modification as changes from introduced mutations can be

readily and directly observed by detectable fluorescence.

The two preeminent approaches to engineering FPs are through directed evolution and rational protein design. Directed evolution continues on a surge of interest and investment from academic and industrial communities, especially due the recent 2018 Chemistry Nobel Prize [74]. In short, this approach utilizes the natural process of evolution, in which beneficial mutations are accumulated slowly overtime through generations, and speeds up this process in laboratory settings. This process involves iterative cycles of gene mutagenesis, expression, and screening. Mutagenesis can be applied to an entire protein as well as confined to specific regions known to affect enzymatic activity or substrate binding. The latter confinement is more feasible when structural information is available and allow for the selection of a set of residues that in the active site and/or are in direct contact with a substrate. The power of directed evolution approaches comes from the total, and often simultaneous, randomization of residues that may impart enhanced properties that may not otherwise be predicted by mutating individual residues. However, this randomized approach also necessitates the creation of thousands, if not tens of thousands to millions, of variants that have no functional improvement, ablated activity, or even structurally unsound and unfolded mutants. As such, stringent screening and selection are imperative to a successful directed evolution campaign.

On the other hand, rational design is a more deliberate process than directed evolution in that it that relies heavily on structural information, sequence conservation, and known mechanistic details for highly specific engineering opportunities [75]. This approach is limited in the number of variants that can be generated relative to an entire protein, but can be utilized for fine tuning of desired properties and activity. Given the deep biochemical knowledge that is required for a rational design approach, machine learning [76], molecular dynamics [77], and other computational approaches are particularly well suited for overcoming these barriers. The greatest challenge of rational design is that proteins are complex macromolecules where putatively beneficial mutations may not be actually

realized when inserted into the protein.

The color palette and spectral properties of FPs have been diversified largely through directed evolution efforts. As mentioned briefly in Section 1.3, the earliest diversification of FPs occurred using the wild-type avGFP as the parent protein. The parent gene was subjected to random mutagenesis and in an simple and elegant screening method, variants grown in bacterial colonies on agar could be screened visually for different emission colors [40, 41], resulting in BFP, CFP, and YFP to name a few (Fig. 1.3). A similar directed evolution and visual screening approach was employed to rescue red fluorescence in the DsRed-derived variants that eventually became mRFP1 [46], and again to evolve mRFP1 into the mFruit FPs ranging from yellow to red-orange emission [47]. Moreover, the popular PAmCherry was generated by applying random mutagenesis to regions of mCherry that were spatially close to the chromophore and presumed determinants of the probe's color [35]; due to selection for photoactivation, a more specialized spectral property, fluorescence assisted cell sorting (FACS) was employed to select for the brightest clones instead of a classic visual screening method.

Rational design on FPs has been most impactful in overcoming oligomerization. The avGFP tends to dimerize natively and this oligomerization poses problems for labeling proteins of interest for *in vitro* and *in vivo* studies as the fluorescent tag may interfere with the biological phenomena that is being probed. To overcome this issue in GFP, key hydrophobic surface residues are swapped to basic residues (Arg or Lys) to reduce or even abolish the dimerization surface [78]. Likewise, the aforementioned mRFP1 variant was generated by creating surface mutations to break apart the natively tetrameric DsRed parent protein, which led to better understanding and control of the dimerization interface across the tetramer, tandem-dimer, and monomer forms of DsRed variants [46]. Collectively, these engineering efforts to make robust FPs have resulted in not only their common utilization throughout scientific research and applications, but also their ability to function as control or signal tags for FACS and other fluorescence-based directed

evolution endeavors.

Ligand-dependent reporters have also been subjected to protein engineering efforts to improve on oligomerization and photophysical properties. Rational design was employed to create the versatile HaloTag probe [34]: utilizing structural and mechanistic information about the parent dehalogenase, the HaloTag is a variant in which the catalytic His272 residue is altered to Phe to help trap the intermediate and allow for covalent linkage to ligands with corresponding chloroalkane linkers. Additional semi-rational mutagenesis was used to improve the binding kinetics in the final HaloTag variant [34]. Similar directed evolution approaches were taken to generate a faster reacting SNAP-tag [79] and to create and distinguish the CLIP-tag from the parental SNAP-tag [72].

As mentioned previously in Section 1.4, bv-binding FPs are derived from DrBphP, a parental bacteriophytochrome protein from *Deinococcus radiodurans* [65]. Using structural information about DrBphP, the protein was truncated to only its chromophore-binding domain, which was a weakly fluorescent and dimeric variant named IFP1.0. IFP1.0 was subjected to mutagenesis at residues near the bv D pyrrole ring and visually screening for variants that emitted fluorescence in the far-red to create IFP1.2 and IFP1.4 variants that were brighter and red-shifted [65]. Further directed evolution was applied to improve the brightness of IFP1.4 and generate a monomeric form in IFP2.0 [66], though we and the original authors have since discovered that IFP2.0 still maintains some propensity for dimerization and aggregation at high concentrations. This dimerization issue has since been solved by utilizing sequence comparisons with other BphP proteins to identify residues that may promote monomeric forms in Nature; the identified monomer was subsequently evolved to restore bv-fluorescence into the mIFP variant [67].

The BBFP UnaG has not been evolved nearly as extensively as oxygen-dependent GFP- and Ds-Red-like reporters nor its bv-binding counterparts. To our knowledge, the only attempt of protein engineering on UnaG was in developing a br-inducible fluorescence sensor of transporter-coupled activity; the work involved a random mutagenesis approach

with a FACS screen to identify the brighter Val2Leu variant, eUnaG [80]. In Chapter IV, we discuss preliminary work utilizing rational design and directed evolution to attempt to swap UnaG recognition of br to bv to create a red-shifted variant.

1.6 High-throughput screening of small molecules

In Chapter III, I have employed a high-throughput screening approach (HTS) to identify new binding partners for UnaG and to create a new UnaG-ligand pair for fluorescence microscopy. This section provides an overview of chemical genomics and HTS. Canonically, the field of chemical genomics describes the use of HTS of small molecules to study biological systems, rather than through direct genetic perturbations [81]. This field enables the elucidation of proteins that are directly responsible for specific cellular morphologies and phenotypes. As such, chemical genomics has enabled countless drug discovery endeavours including the development of novel drug scaffolds, deeper understandings of drug-drug interactions, and identification of new druggable protein targets [82].

Approaches in chemical genomics can be divided into forward and reverse chemical genetics studies (Fig. 1.6). Forward chemical genetics provides a phenotype-to-protein perspective, in which small molecule screens are used to select for a different cellular phenotype and the protein responsible for the desired phenotype is subsequently identified. The reverse approach provides a protein-to-phenotype perspective in that a known protein of interest is used as the bait in small molecule screens to discover protein-molecule binding partners before further characterizing cellular level phenotypic changes.

When selecting chemical libraries to explore, one may choose to search vast chemical structural space or choose a more focused search using molecules with similar structural features. Libraries of compounds have been curated by academics and by pharmaceutical companies collectively interested in drug discovery applications. These libraries can be broadly classified into three categories: commercial, natural products, and diversity-oriented synthesized (DOS) [81]. Commercial libraries are ones that can be purchased

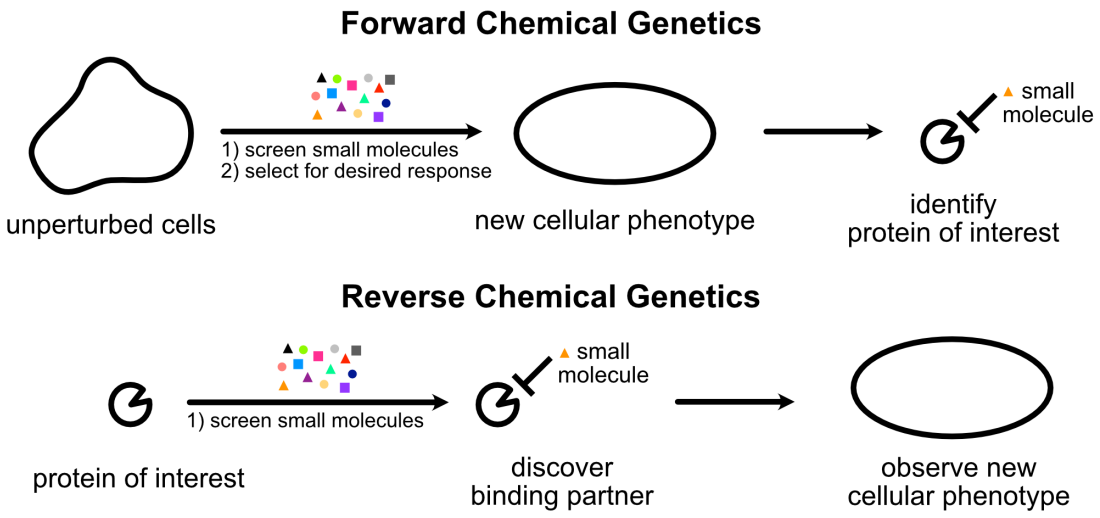


Figure 1.6: Comparison of chemical genetics approaches. Figure adapted from [81].

directly from pharmaceutical companies or other companies that specialize in curating compound collections; one may specify the size and structural diversity of a collection but may be limited by the high cost of commercial libraries. Natural product libraries are more structurally diverse and complex than commercial libraries, which make natural product libraries attractive collections for finding biologically active molecules. However, natural products are not always extracted as single isolated compounds and subsequent screening in heterogeneous mixtures can make it difficult to identify a bioactive compound. DOS approaches bridge the gap between the other two approaches to generate compounds covering chemical space that is not available in commercial compound libraries or in natural products [83]. Nevertheless, it remains difficult to predict which types of structural changes to compound scaffolds will result in a desired phenotypic effect. Thus, the selection of an appropriate chemical library is crucial for any successful assay.

As compounds are typically tested only once in most HTS primary screens, assay optimization and robust statistical analysis are required to confidently identify active compounds or "hits." For both biochemical and cell-based assays, a positive and a negative control must be established for the entire assay and be supplied on each assay plate. The signal obtained from both controls are utilized to establish the dynamic range that signal

from tested compounds can be statistically evaluated. This distribution can be defined by the Z-factor:

$$Z = 1 - \frac{3\sigma_s + 3\sigma_c}{|\mu_s - \mu_c|} \tag{1.3}$$

where σ_s is the standard deviation of sample signal, σ_c is the standard deviation of control signal, μ_s is the mean of sample signal, and μ_c is the mean of control signal. The Z-factor describes the quality of an overall assay for a specific HTS workflow where an ideal assay is one with a large dynamic range coupled with small data variability [84]. As such, the Z-factor is a superior measure of assay quality than more general the signal-to-background ratio (S/B) that does not account for sample and background variability.

Additionally, a Z'-factor is used for evaluating overall assay quality. The Z'-factor describes an assay using only the signal from controls:

$$Z' = 1 - \frac{3\sigma_{c+} + 3\sigma_{c-}}{|\mu_{c+} - \mu_{c-}|} \tag{1.4}$$

Since the Z'-factor only utilizes the signal from positive ($c+$) and negative controls ($c-$), it can be used to evaluate an entire assay during the optimization process [84]. A Z' value that is negative or close to zero indicates a small dynamic range between the positive and negative control signal, resulting in low confidence for "hits" with signal about the typical $3\sigma_{c-}$ threshold. For most HTS-drug discovery applications, assays are optimized for a Z' value of 0.5 or greater.

Even so, such careful statistical consideration of HTS assays may overlook real hits or obscure artifacts introduced during assay optimization [85]. For example, a robust assay with a Z' of 0.8 or higher may have low sensitivity, especially if the positive control used produces an arbitrary signal unrelated to the biochemical or cellular assay. Likewise, an assay with a Z' below 0.5 may deliver high sensitivity that merely needs to be evaluated with additional replicates, concentration-dependent, or counter screen formats. As with

all screening campaigns, subsequent hits must be further characterized with more in-depth analyses, ideally with a readout different than the one used for the HTS assay. In the work discussed in Chapter 3, we devise a HTS campaign to identify new ligands for the BBFP with a low Z'-factor that is complemented with replicates and counter-screening measures to identify real hits.

1.7 Dissertation objectives

Understanding how different species from the gut microbiome organize in relation to carbohydrate accessibility or the order of colonization on a mucosal surface will not only further knowledge on the symbiosis of this complex microbial community, but also may lead to insights into developing treatments for treating dysbiosis in diseased states. Given growing interest in studying live-cell behavior with spatial and temporal resolution, a suite of robust oxygen-independent fluorescent probes are required. As such, the aims of this dissertation are to evaluate, implement, and further develop BBFPs for labeling anaerobic gut bacteria.

In Chapter II, I describe the implementation of BBFPs in anaerobic gut bacteria. The green UnaG and the red IFP2.0 are used to label *B. theta* in monocultures in anaerobic fluorescence imaging as well as to distinguish *B. theta* from *R. bromii* in mixed-species imaging and from *B. ovatus* in multicolor imaging. These studies are the first demonstration of BBFP applications for labeling obligate anaerobic bacteria in oxygen-free imaging conditions, highlighting the high potential for utilizing BBFPs for studying the gut microbiome.

In order to label multiple species in polymicrobial studies, a diverse palette of FP colors that are compatible with oxygen-free environments need to be developed; as such, BBFPs are ideal candidates for diversification. In particular, UnaG is preferable over IFP2.0 for further spectral shifting due to UnaG's small monomeric form and because the native ligand br is non-covalently bound. In Chapter III, I detail the use of high-throughput screening of

small molecule libraries to identify new ligand-UnaG pairs useful for fluorescence outside of the blue 488-nm excitation used by the native br-UnaG pair. After assessing hit compounds based on fluorescence intensity and triaging results across different fluorescence channels, I have identified a new ligand-UnaG pair that can be used in the green 532-nm excitation channel. Like the native br-UnaG pair, this novel UnaG-benzothiazole ligand pair can also be used to label Bacteroides strains in oxygen-free environments.

In Chapter IV, I elaborate on efforts to introduce bv-binding in UnaG through directed evolution and protein engineering. While the selective mutations made to UnaG do not result in selective binding of bv over the native br ligand, I have gained insight into the binding pocket and importance of key residues. I also discuss efforts in creating robust screens to select for red-shifted bv-binding UnaG variants. FACS-based screening was determined best for screening variants with red fluorescence from violet 405-nm excitation or red 640-nm excitation. After four rounds of selection, we identify a V100D mutation that is highly selected for as well as an UnaG mutant with improved solubility. This work provides the groundwork for creating a toolbox of distinct UnaG-ligand pairs and colors that can be used for anaerobic fluorescence investigations.

Finally, in Chapter V, I summarize the applications and development of BBFPs for live-cell anaerobic fluorescence microscopy. I provide recommendations for types of biological problems in polymicrobial studies that can be accessed by BBFP implementation. The work presented in this thesis broadly addresses the necessity for oxygen-independent probes and how ligand-dependent FPs can fulfill this gap.

CHAPTER II

Bilin-based Fluorescent Proteins for Imaging Living Obligate Anaerobes

The work presented in this chapter was previously published in

Current Research in Microbial Sciences.

Chia, H.E., Zuo, T., Koropatkin, N.M., Marsh, E.N.G., and Biteen, J.S.

Imaging living obligate anaerobic bacteria with bilin-binding fluorescent proteins.

Current Research in Microbial Sciences, **1**, 1-6 (2020). DOI: 10.1016/j.crmicr.2020.04.001

2.1 Introduction

While Green Fluorescent Protein (GFP) and other genetically encodable fluorescent proteins (FPs) are powerful and ubiquitous tools in biology, these GFP-like FPs are not fluorescent in the absence of oxygen [48]. This oxygen dependence precludes their ability to provide additional insight into anaerobic environments. Many anaerobic systems, like the gut microbiome [86, 87] and soil microbiomes [3, 88], are medically and ecologically important to study, yet remain underexplored because of a lack of appropriate fluorescent probes. Thus, the development of oxygen-independent fluorescent reporters is essential to discover and understand biological and biophysical processes in anaerobic systems.

Previously, fluorescence microscopy studies of anaerobic systems have circumvented the challenges of oxygen-free environments by investigating fixed cells. For example, much of our understanding of the spatial organization of obligate anaerobes comes from imaging of fixed, antibody-stained cells [20, 29] or cells that express FPs that are exposed to oxygen during fixation [28]. Flavin Mononucleotide (FMN)-based Fluorescent Proteins (FbFPs) are among the few tools that have been demonstrated to label live anaerobic bacteria [51]; however, FbFPs are confined to the blue spectral range, which overlaps with intrinsic cellular fluorescence. Despite engineering [59, 89], FbFP variants also remain dim compared to EGFP. We have previously labeled and studied live anaerobic bacteria with several other probes, but each labeling approach comes with significant limitations [50]. In one approach, enzymes in the Starch Utilization System (Sus) [90] of the prevalent gut bacterium *Bacteroides thetaiotaomicron (B. theta)* were tagged with Photoactivatable mCherry (PAmCherry) [21]. PAmCherry is suitable for single-molecule super-resolution and tracking experiments but this probe, like GFP, requires oxygen exposure for fluorescence maturation, so measurements of Sus-PAmCherry fusions were done in dormant *B. theta*. On the other hand, we achieved fully anaerobic live-cell imaging by using the HaloTag system to label Sus proteins in single-particle tracking [27]. The bright fluorescent dye ligands used in HaloTag systems provide high signal, but significant washing is required to avoid increased background. These ligands are also largely impermeable to bacterial cells and are restricted to outer membrane labeling without the aid of electroporation or osmotic shock [69]. These invasive washing and labeling processes prevent in situ HaloTag labeling.

To achieve non-pertubative labeling for investigations of microbial communities, we therefore seek other reporters [50]. Bilin-based Fluorescent Proteins (BBFPs) are particularly attractive options for anaerobic live-cell imaging. BBFPs, such as the green UnaG [60] and the far-red IFP2.0 [66], bind ligands to produce a fluorescent holoprotein rather than relying on an oxidative reaction for chromophore maturation. The ligands bilirubin (br)

and biliverdin (bv), bound by UnaG and IFP2.0, respectively, are fluorogenic molecules that produce no background fluorescence even when supplied in excess [60, 66]. These ligands are cell-permeable in bacteria [91], thus allowing labeling of both cytoplasmic and extracellular targets. Both UnaG and IFP2.0 can be used in genetically tractable organisms and in most conventional microscopes. UnaG excitation and emission are within the GFP range and UnaG was previously characterized to be approximately as bright as GFP [60]. Although IFP2.0 is dimmer than UnaG or GFP, IFP2.0 emits in the far-red range of visible light [66], at wavelengths that suffer from less intrinsic background fluorescence in biological samples.

Here, we implement these two BBFPs, UnaG and IFP2.0, to fluorescently label living *B. theta* cells in the cytoplasm and on the outer membrane. We demonstrate that unlike GFP-labeled *B. theta* cells which only fluoresce in aerobic conditions, UnaG- and IFP2.0-labeled cells fluoresce in live-cell, anaerobic imaging conditions. UnaG-labeled cells can also be used to differentiate *B. theta* from another prevalent gut bacterium, *Ruminococcus bromii (R. bromii)*. Furthermore, UnaG-labeled *Bacteroides ovatus (B. ovatus)* can be distinguished from IFP2.0-labeled *B. theta* cells, indicating that BBFPs can also be used in two-color imaging. Overall, we demonstrate the utility of BBFPs for non-invasive investigations of anaerobic microbial ecologies.

2.2 Materials and methods

2.2.1 Bacterial strains cloning and cell cultures

Oligonucleotides and plasmids used in cloning are described in Table 2.1 and 2.2, respectively. DNA sequencing was used to confirm assembled genes in plasmids.

B. theta and *B. ovatus* were generated by counter-selectable allelic exchange in a thymidine kinase deletion (Δ tdk) mutant and grown as previously described [86]. In summary, *B. theta* and *B. ovatus* were first cultured in rich media containing tryptone-yeast extract-

Table 2.1: Plasmids ordered and generated for BBFP evaluation.

Plasmid	Description	Resistance	Reference
pcDNA3_UnaG-Flag-Sec61B	For general cloning; contains original (not bacterial codon-optimized) gene	Amp	Addgene #83413
IFP2.0-N1	For general cloning; contains IFP2.0 gene	Amp	Addgene #54785
pExchange-tdk (pExch-tdk)	For general cloning; insertion into *B. theta* (containing thymidine kinase knockout)	Amp	
pExch-tdk_SusG-HaloTag	pExchange-tdk vector containing SusG-HaloTAg	Amp	Tuson et al. (2018)
pExch-tdk_SusG-UnaG	pExchange-tdk vector containing codon-optimized UnaG gene; generated by Gibson Assembly between XbaI and SalI cloning sites	Amp	
pExch-tdk_SusE-IFP2.0	pExchange-tdk vector containing SusE-IFP2.0 gene; generated by Gibson Assembly between XbaI and SalI cloning sites	Amp	
pNBU2_P_{BfP1E6}_GFP	For general cloning and insertion into *B. theta*; results in high constitutive expression of GFP	Amp	Whitaker et al. (2017)
pNBU2_P_{BfP1E6}_UnaG	Insertion into *B. theta*; results in high constitutive expression of UnaG; generated by Gibson Assembly between XhoI and SapI cloning sites	Amp	
pNBU2_P_{BfP1E6}_IFP2.0	Insertion into *B. theta*; results in high constitutive expression of IFP2.0; generated by Gibson Assembly between XhoI and SapI cloning sites	Amp	
pMAL-c5x_UnaG	For protein expression; results in the production of codon-optimized UnaG with a maltose-binding-protein (MBP) tag on the N-terminus and a 6x-His tag on the C-terminus	Amp	Genscript
pBAD_EGFP	For protein expression; results in the production of EGFP with a 6x-His tag on the N-terminus	Amp	Addgene #54762
pBAD_IFP2.0	For protein expression; results in the production of IFP2.0 with a 6x-His tag on the N-terminus; generated by Gibson Assembly	Amp	

Table 2.2: Oligonucleotides used for BBFP evaluation.

Primer name	Sequence (5' → 3')	Use
pExchXbaI_f	TAGTTCTAGAGCGGCCGCC	Linearize pExchange-tdk vector
pExchSalI_r	TACCGTCGACTCGAATGTTATCTTCTTTC	Linearize pExchange-tdk vector
PBfP1E6_f	CCTAGCTGATTAGCTCGAGAAGG	Linearize pNBU2_P_{BfP1E6} vector
PBfP1E6_r	CAGCTCTTCGCCTTTACGCATTGAAAATAAAT	Linearize pNBU2_P_{BfP1E6} vector
coUnaG_f	TGCGTAAAGGCGAAGAGCTGATGGTGGAAAAGTTCGTTGG	UnaG into pNBU2_P_{BfP1E6}
coUnaG_r	TCTCGAGCTAATCAGCTAGGTTCGGTCGCACGACGATAG	UnaG into pNBU2_P_{BfP1E6}
N-SusG_f	TAACATTCGAGTCGACGGTAATGAATAAACATCTCCACTTTTTATCTCTTTTGTG	SusG-UnaG into pExchange-tdk
N-SusG-coUnaG_r	TTTCCACCATTGCTGCTGCATTGCCTGAG	SusG-UnaG into pExchange-tdk
N-SusG-coUnaG_f	TGCAGCAGCAATGGTGGAAAAGTTCGTTGG	SusG-UnaG into pExchange-tdk
coUnaG-SusG-C_r	CCCACTTGAATTCGGTCGCACGACGATAG	SusG-UnaG into pExchange-tdk
coUnaG-SusG-C_f	TGCGACCGAATTCAAGTGGGAGCGCGTG	SusG-UnaG into pExchange-tdk
SusG-C_r	TGGCGGCCGCTCTAGAACTATTAGTTGCCCAACTTGAATACTACAG	SusG-UnaG into pExchange-tdk
IFP2.0_f	TGCGTAAAGGCGAAGAGCTGATGGCTCGGGACCCTCAAC	IFP2.0 into pNBU2_P_{BfP1E6}
IFP2.0_r	TCTCGAGCTAATCAGCTAGGTCAGGCTTCTTTCCTCTGCAC	IFP2.0 into pNBU2_P_{BfP1E6}
IFP2.0_f	GGCAGCCGCGATGGCTCGGGACCCTCAAC	SusE-IFP2.0 into pExchange-tdk
IFP2.0_r	TGGCGGCCGCTCTAGAACTATCAGGCTTCTTTCCTCTGCAC	SusE-IFP2.0 into pExchange-tdk

glucose and incubated anaerobically at 37°C in a Coy chamber. Cultures were subsequently back-diluted into minimal media with a carbohydrate source (0.05% w/v glucose or maltose). Bilirubin (br, Sigma Aldrich) and biliverdin (bv, Sigma Aldrich) stocks were created in DMSO and added to media as required; final concentrations in media were 25 μM and 2.5 μM for br and bv, respectively.

R. bromii was grown in specialized Ruminococcus (Rum) media (per 50 mL of 2× media): yeast extract (0.25g), $NaHCO_3$ (0.4 g), L-cysteine (0.1 g), $(NH_4)_2SO_4$ (0.09 g), K_2HPO_4 (0.045 g), KH_2PO_4 (0.045 g), NaCl (0.09 g), $MgSO_4$ (0.004 g), $CaCl_2$ (0.009 g), biotin (20 μg), cobalamine (20 μg), p-aminobenzoic acid (60 μg), folic acid (100 μg), pyridoxamine (300 μg), thiamine (100 μg), riboflavin (100 μg), D-pantothenoic acid hemicalcium salt (100 μg), nicotinamide (100 μg), and resazurin (50 μg). Rum media also contained (concentrations per 50 mL of 2× media): hematin (30 μM), L-histidine (3 mM), acetic acid (63.7 mM), propinoic acid (17.8 mM), isobutyric acid (5.75 mM), isovaleric acid (1.95 mM), and valeric acid (1.95 mM). Rum media was diluted with an equal volume of a carbohydrate source (0.05% w/v maltose or fructose and glycogen) to culture cells. Co-cultures were made by growing *R. bromii* anaerobically at 37°C overnight and adding *B. theta* the following day for continued cell growth.

2.2.2 Growth curves

B. theta and *B. ovatus* cells were cultured in minimal media with an appropriate carbohydrate source and back-diluted 1:200 into 96-well clear bottom plates with media. Each growth experiment condition were performed in triplicate. Plates were loaded into a Biostack automated plate-handling device (Biotek Instruments). Absorbance at 600 nm (Optical Density $(OD)_{600}$) was measured in each well every 20 min by a Powerwave HT absorbance reader (Biotek Instruments). Data were recorded using Gen5 software (BioTek Instruments) and processed using Prism (GraphPad).

2.2.3 Fluorescence microscopy

Cells were grown to early to mid-log phase and imaged. All imaging was performed at room temperature and anaerobically on cells sealed between coverslips with epoxy as previously described [92] unless otherwise noted for "+O_2" samples in the figure captions. For immunostaining, *B. theta* cells were incubated with custom rabbit polyclonal SusF antibody (1:100 dilution; Cocalico Biologicals, Inc.) [21] for 30 min and washed with phosphate-buffered saline (PBS) prior to incubation with AlexaFluor 594 secondary anti-rabbit goat antibody (4 μg in 100 μL; ThermoFisher Scientific). Imaging was done in an Olympus IX71 inverted epifluorescence microscope with a 100 $\times$ 1.4 Numerical Aperture (NA) wide-field oil-immersion objective. Samples were illuminated by a 488-nm laser (Coherent Sapphire 488-50; 8-18 W/cm^2), 561-nm laser (Coherent Sapphire 561-50; 120 W/cm^2), or 640-nm laser (Coherent CUBE 640-40C; 80 W/cm^2). Fluorescence emission was filtered with appropriate filter sets and imaged on a 512 $\times$ 512 pixel Photometrics Evolve Electron Multiplying Charge-Coupled Device (EMCCD) camera at 25 frames/s or on a 512 $\times$ 512 pixel Andor iXon EMCCD camera at 100 frames/s. For consistency and noise reduction, final fluorescence images were created by summing frames for a total of 400 ms total integration time. Recorded images and movies were analyzed using ImageJ; all images presented are presented on the same color scale.

2.2.4 Flow cytometry

Flow cytometry was performed on unlabeled and labeled *B. theta* cells expressing GFP or UnaG with varying concentrations of bilirubin. Cells were analyzed on an Attune NxT Flow Cytometer in the BL1 channel. Data was processed using Attune NxT software.

2.2.5 Protein expression and purification

Like others, we had difficulty expressing the original *UnaG* gene in high enough quantities for successful purification [80]. To improve solubility and protein expression, we or-

dered *UnaG* that was codon-optimized for *Escherichia coli (E. coli)* expression in the vector pMAL-c5x (GenScript) (Fig. 2.1). The transcribed protein was UnaG with maltose-binding-protein (MBP) on the N-terminus and 6x-His on the C-terminus; the protein herein is referred to as UnaG. UnaG was expressed in *E. coli* (NEB-Express) with IPTG induction (0.4 mM final concentration). Cell pellets were lysed by sonication and after centrifugation, the supernatant was collected and loaded onto affinity columns.

UnaG was purified by two rounds of affinity chromatography using fast protein liquid chromatography (Akta Systems): (1) MBP-trap (5 mL, GE Healthcare) and (2) His-trap (5 mL, GE Healthcare) (Fig. A.2). In the first amylose-based purification, the protein loaded onto the MBP-trap column was washed with 10 column equivalents of buffer A-M (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA) and eluted using buffer B-M (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM maltose) until protein was no longer detected by UV signal. These fractions were pooled and mixed with an equivalent volume of buffer A-I (20 mM Tris-HCl, 200 mM NaCl, 20 mM imidazole) and loaded onto a His-trap column equilibrated with the same buffer A-I. The column was washed with an 2 column volumes of buffer A-I and eluted using a gradient with buffer B-I (20 mM Tris-HCl, 200 mM NaCl, 500 mM imidazole). The final elution fractions were dialyzed in buffer A-M to remove imidazole (ThermoFisher Slide-A-Lyzer Dialysis Cassette G2, 10 kDa cutoff) and concentrated by centrifugation (Sartorius Vivaspin 500, 30 kDa cutoff). For the holoprotein, UnaG was incubated with 2-fold excess ligand (br or bv) and isolated by PD-10 desalting column (5 mL, GE Healthcare).

EGFP was obtained from Addgene (plasmid #54762) and expressed from the pBAD vector in *E. coli* (BL21 DE3) with arabinose induction (0.005% final concentration). Cell pellets were lysed by sonication and purified by nickel column chromatography. Elution fractions were dialyzed to remove imidazole and concentrated.

IFP2.0 was obtained from Addgene (plasmid #54785) and inserted into the pBAD vector using Gibson Assembly cloning. IFP2.0 was expressed in *E. coli* (BL21 DE3) with arabinose

Original sequence:
ATGGTTGAAAAATTTGTTGGTACGTGGAAAATAGCGGATTCTCATAATTTTGGCGAATATCTGA
AGGCTATCGGGGCGCCTAAAGAGCTGAGTGATGGAGGCGATGCCACGACTCCAACTCTGTATAT
TTCACAGAAGGACGGTGATAAAATGACCGTTAAGATCGAGAATGGCCCGCCCACCTTCCTGGAT
ACACAGGTAAAGTTTAAACTTGGTGAAGAATTTGATGAGTTCCCGAGCGACAGACGCAAAGGTG
TTAAATCCGTAGTCAACTTGGTCGGAGAAAAACTCGTATATGTCCAAAAATGGGATGGCAAAGA
GACAACCTACGTGCGTGAAATTAAAGACGGGAAACTGGTTGTGACGCTGACAATGGGCGACGTG
GTGGCAGTGCGCTCGTACCGGCGTGCAACCGAA

Optimized sequence:
ATGGTGGAAAAGTTCGTTGGTACCTGGAAAATCGCTGATAGCCATAATTTTGGTGAATATCTGA
AGGCGATTGGCGCTCCGAAAGAACTGAGCGATGGTGGCGACGCTACCACCCCGACCCTGTACAT
CAGCCAGAAAGATGGTGACAAGATGACCGTGAAAATTGAAAATGGTCCGCCGACCTTCCTGGAT
ACCCAAGTTAAATTTAAGCTGGGTGAAGAATTCGATGAATTTCCGAGCGACCGTCGTAAAGGTG
TGAAGAGCGTGGTTAACCTGGTTGGCGAAAAGCTGGTGTATGTTCAGAAGTGGGACGGCAAAGA
AACCACCTACGTGCGTGAAATCAAAGATGGTAAACTGGTGGTGACCCTGACCATGGGCGACGTG
GTTGCAGTTCGTAGCTATCGTCGTGCGACCGAACATCACCATCACCATCACTAA

Protein sequence:
MVEKFVGTWKIADSHNFGEYLKAIGAPKELSDGGDATTPTLYISQKDGDKMTVKIENGPPTFLD
TQVKFKLGEEFDEFPSDRRKGVKSVVNLVGEKLVYVQKWDGKETTYVREIKDGKLVVTLTMGDV
VAVRSYRRATE

Figure 2.1: Codon-optimized UnaG sequence: For protein expression and other fusion proteins, we used a codon-optimized UnaG gene that was ordered from Genscript.

induction (0.005% final concentration). Cell pellets were lysed by sonication and purified by nickel column chromatography. IFP2.0 was incubated with 2-fold excess ligand (br or bv) and excess ligand was removed by dialysis for 3 h.

2.2.6 Protein characterization

All proteins were prepared for measurements in the same buffer: Tris-HCl (20 mM), NaCl (200 mM), and EDTA (1 mM). UV-Vis absorbance measurements were performed in 96-well clear bottom plates (Molecular Devices SpectraMax iD3 microplate reader). Fluorescence measurements were carried out on an Agilent Varian Cary Eclipse Fluorescence Spectrophotometer. Plotted fluorescence excitation and emission spectra are averaged plots from three technical replicates. Data was processed using MATLAB.

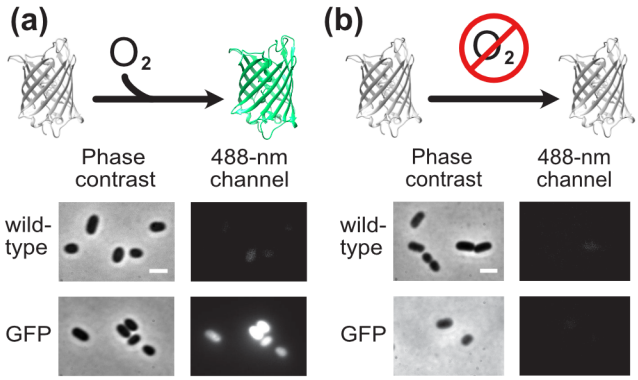


Figure 2.2: GFP requires oxygen to produce fluorescence. *B. theta* cells expressing GFP (a) fluoresce in aerobic conditions after 20 min of air exposure and (b) do not fluoresce in anaerobic conditions. Scale bars: 2 μm.

2.3 Results and discussion

The use of GFP as a fluorescent tag is restricted to aerobic environments. Though dormant *B. theta* cells labeled with GFP are fluorescent after exposure to oxygen in air (Fig. 2.2a), maintaining the oxygen-free environments for continued cell growth of these anaerobic aerobic cells results in no fluorescence from GFP-labeled *B. theta* cells (Fig. 2.2b).

On the contrary, BBFPs are oxygen-independent fluorescence reporters that bind to a ligand to produce fluorescence [50, 60, 66]. UnaG fluoresces only after addition of br (Fig. 2.3a). The purified holoprotein of UnaG with bound br has maximal excitation and emission that is only slightly red-shifted compared to GFP excitation and emission (Fig. 2.3b); UnaG can therefore be imaged with these common microscopy filters. The br ligand is not fluorescent in solution (Fig. 2.3c and d, Fig. 2.4); this fluorogenic ligand is therefore suitable for supplementing in high concentrations in live-cell imaging without creating background or requiring washing steps. *B. theta* cells expressing codon-optimized UnaG in the cytoplasm and grown with br were fluorescent in both aerobic and anaerobic environments (Fig. 2.3c and d). The addition of excess br ligand does not affect the growth kinetics of UnaG-labeled cells (Fig. 2.5), nor does 25 μM br affect *B. theta* cell morphology (Fig. 2.6), thus we selected a br concentration of 25 μM to optimize fluorescent labeling efficiency while minimizing toxicity (Fig. 2.7).

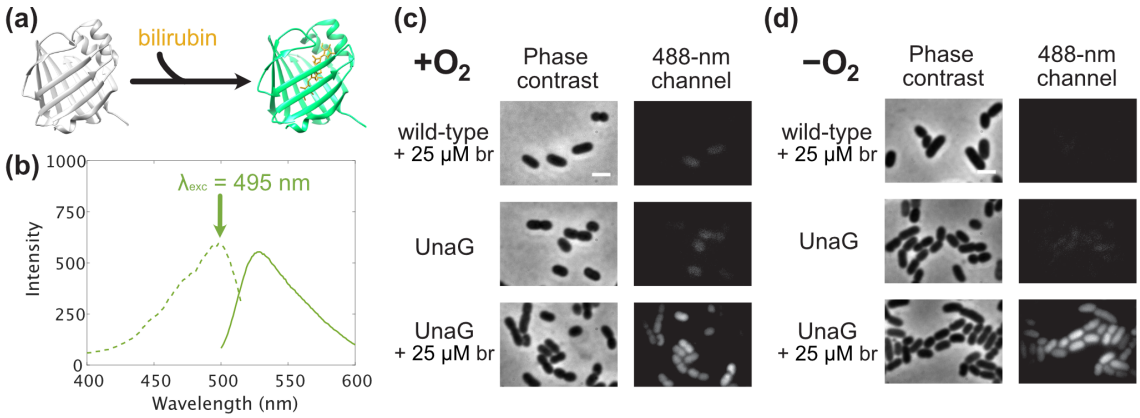


Figure 2.3: UnaG uses bilirubin to produce fluorescence. (a) The protein UnaG binds br to become fluorescent. (b) Purified protein with bound br is maximally excited at 495 nm and has maximal emission at 525 nm. *B. theta* cells expressing UnaG and grown with br fluoresce upon 488-nm excitation in (c) aerobic and (d) anaerobic conditions. Scale bars: 2 μm. PDB ID: 4I3B.

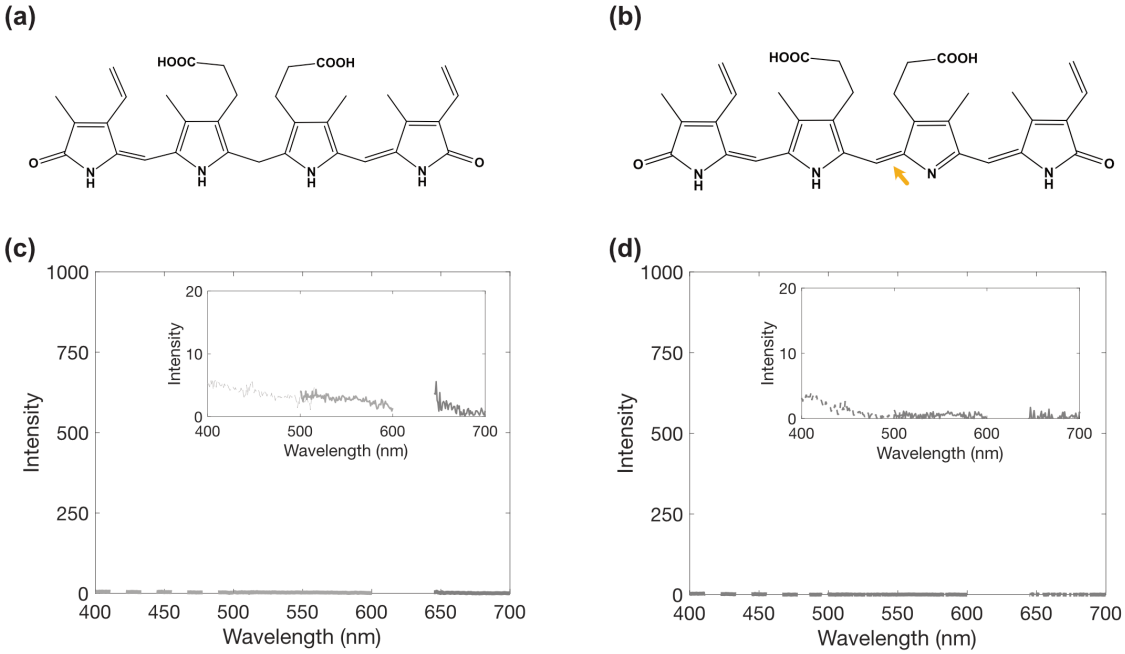


Figure 2.4: (a,b) Chemical structures and (c,d) fluorescence excitation (light grey) and fluorescence emission (dark grey) spectra for br and bv, respectively. Insets: fluorescence spectra on a zoomed-in scale. Yellow arrow in (b) indicates the additional double bond that extends bv conjugation relative to br.

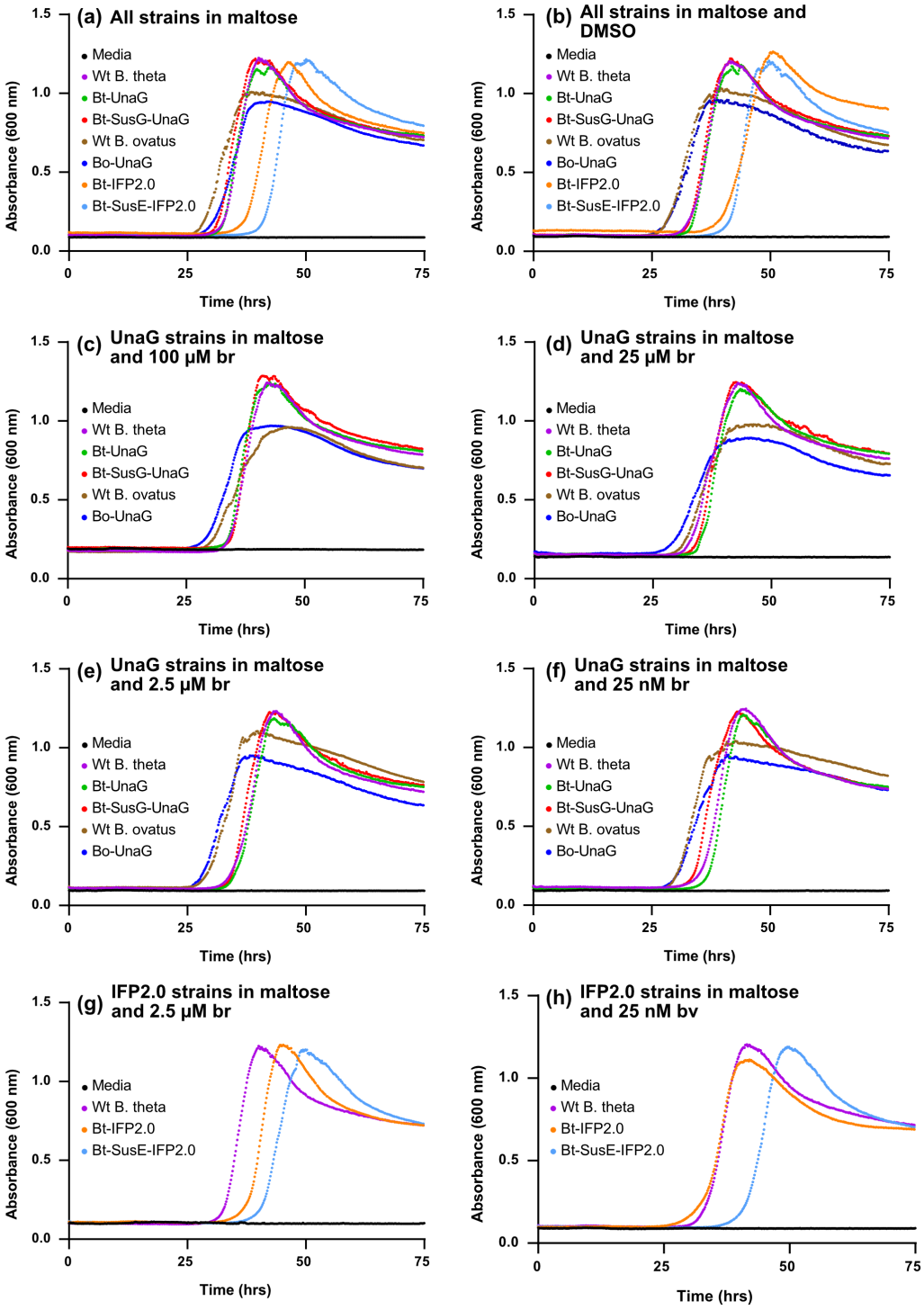


Figure 2.5: Growth measured by absorbance (600 nm) for cells grown on minimal media containing (a) 5 mg/mL maltose, (b) an additional 0.001% DMSO, (c–f) different concentrations of br, or (g–h) different concentrations of bv. *B. ovatus (Bo)* and *B. theta (Bt)* growth is not affected by supplementing br or bv in media. None of the strains show defects when grown in media supplemented with br or bv, though a small growth lag is detected for *B. theta* expressing IFP2.0 and SusE-IFP2.0.

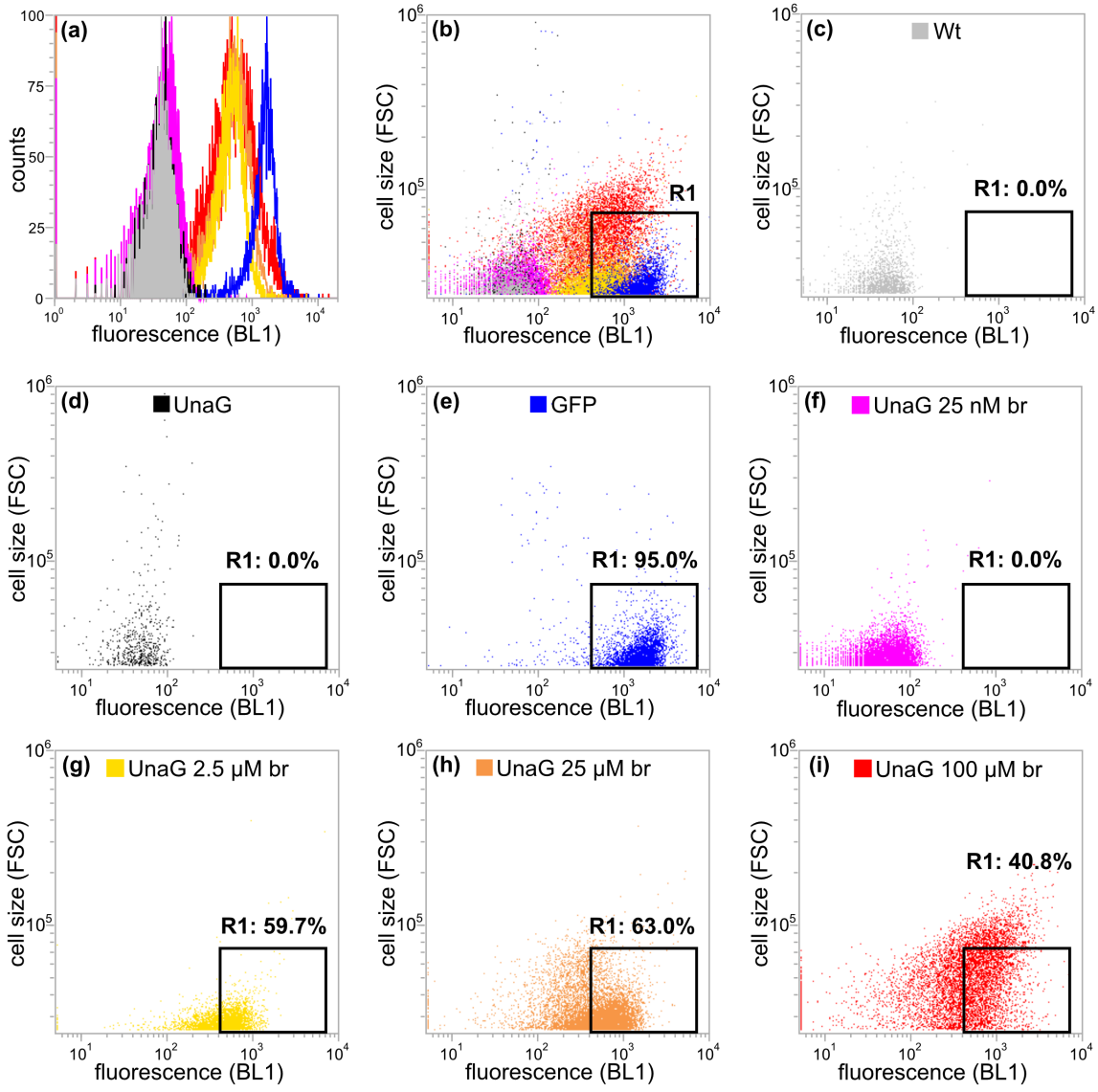


Figure 2.6: The (a) histogram and (b) scatter plot compiled data for all cell types: (c) unlabeled wild-type *B. theta* (wt), (d) UnaG-labeled, (e) GFP-labeled, and UnaG-labeled with increasing br concentrations at (f) 2.5 nM, (g) 2.5 μM, (h) 25 μM, and (i) 100 μM. The fluorescence intensity of *B. theta* expressing UnaG increases with br concentration. High br concentration (100 μM) also led to significant changes in cell size; main text experiments are performed using 25 μM br.

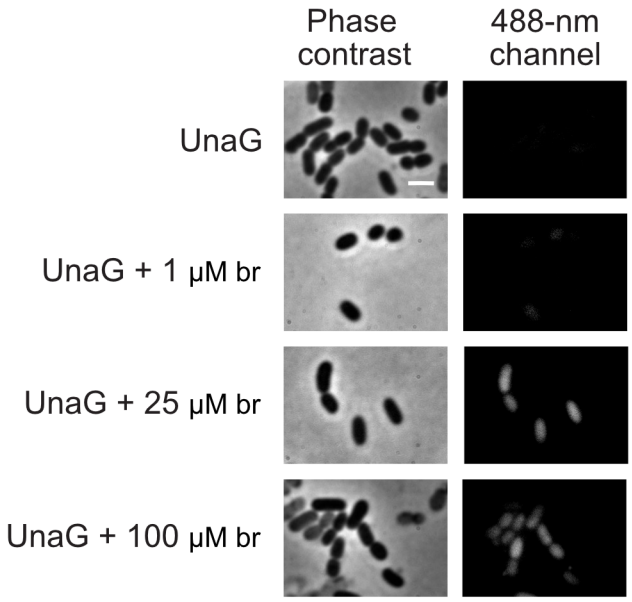


Figure 2.7: Phase-contrast and fluorescence images of UnaG-labeled *B. theta* cells grown in a range of br concentrations and excited by 488-nm illumination. The fluorescence intensity depends on br concentration. Scale bar: 2 μm.

We implemented IFP2.0 as a second oxygen-independent fluorescence reporter in *B. theta*. Like UnaG, IFP2.0 fluoresces only upon addition of its ligand, bv (Fig. 2.8a), which differs from br in the extent of their conjugated π-systems. The purified holoprotein is maximally excited and emissive in the near-IR range of light (Fig. 2.8b, red line), which is the laser wavelength used to excite conventional red dyes like Cy5 and AlexaFluor 633. Like br, the bv ligand has the benefit of being fluorogenic (Fig. 2.8c and d, Fig. 2.4). *B. theta* cells expressing IFP2.0 grown with bv were fluorescent in both aerobic and anaerobic environments (Fig. 2.8c and d), and in fact, we observed that IFP2.0-labeled cells photobleached quickly in aerobic conditions (seconds timescale), whereas IFP2.0-labeled cells did not bleach appreciably on the timescale of anaerobically sealed samples (Fig. 2.8c and d, Fig. 2.9). The bleaching of IFP2.0-labeled *B. theta* cells in oxygen-exposed conditions may be related to the low photostability of far-red BBFPs [67].

The labeled *B. theta* strains in Fig. 2.3 and Fig. 2.8 constitutively express UnaG or IFP2.0 in the cytoplasm. We also investigated BBFP labeling in fusions to Sus outer-membrane

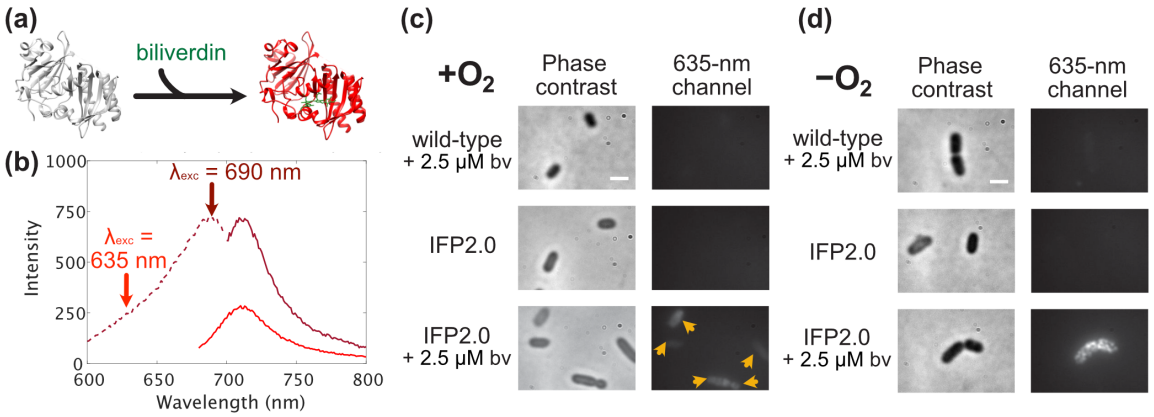


Figure 2.8: IFP2.0 uses biliverdin to produce fluorescence. (a) The protein IFP2.0 binds bv to become fluorescent. (b) Purified protein with bound bv is maximally excited at 690 nm and has maximal emission at 710 nm (burgundy curve. Purified IFP2.0 incubated with bv has reduced emission when excited at 635 nm (red curve). *B. theta* cells expressing IFP2.0 and grown with bv fluoresce upon 635-nm excitation in (c) aerobic and (d) anaerobic conditions. Scale bars: 2 μm. PDB ID: 4CQH.

proteins on the *B. theta* surface. We created two *B. theta* strains expressing SusG-UnaG and SusE-IFP2.0, respectively, at the native promoter; we have previously demonstrated that fusions of these proteins to HaloTag and PAmCherry do not disrupt protein function or outer-membrane localization [21]. The expression level of the outer membrane fusion proteins was more than 40-fold lower than UnaG or IFP2.0 in cytoplasm expressed FP strains [28]. Fluorescence signal was only weakly detected in the SusG-UnaG strain, possibly because the low copy number of the fusion protein was not enough to overcome signal from the intrinsic background fluorescence at this wavelength (Fig. 2.10). On the contrary, though IFP2.0 is dimmer than UnaG [66], the intrinsic background fluorescence is reduced at longer wavelengths and SusE-IFP2.0 can still be visualized despite the relatively low copy number (Fig. 2.10).

To demonstrate the utility of UnaG and IFP2.0 for microscopy of live, anaerobic polymicrobial communities, we imaged these BBFPs in mixed culture. Like *B. theta*, *R. bromii* is another prevalent bacterial species in the human gut microbiome; cross-feeding between these two species is important in establishing a healthy microbial community [10].

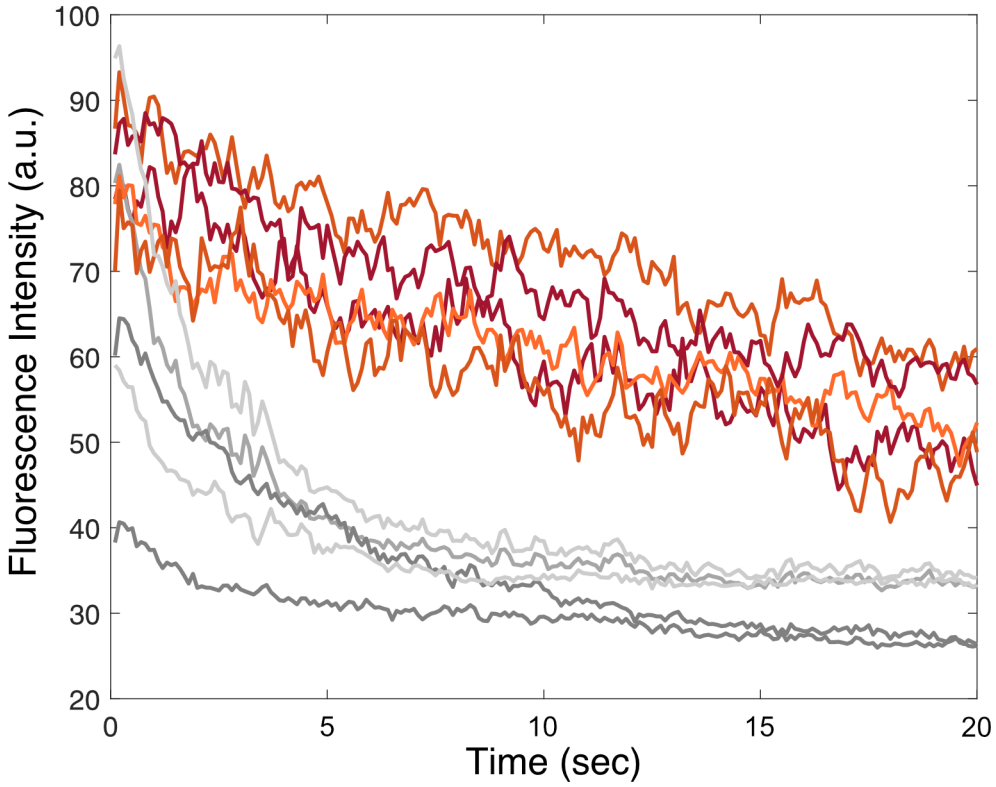


Figure 2.9: Fluorescence intensity of IFP2.0-labeled *B. theta* cells in bv under constant illumination by a 80 mW/cm^2 635-nm laser. Each curve plots the fluorescence intensity of a single cell normalized to cell area for anaerobic conditions (red curves) and aerobic conditions (grey curves). Five representative cells are plotted per condition; the anaerobic conditions led to slightly brighter and more photostable fluorescence.

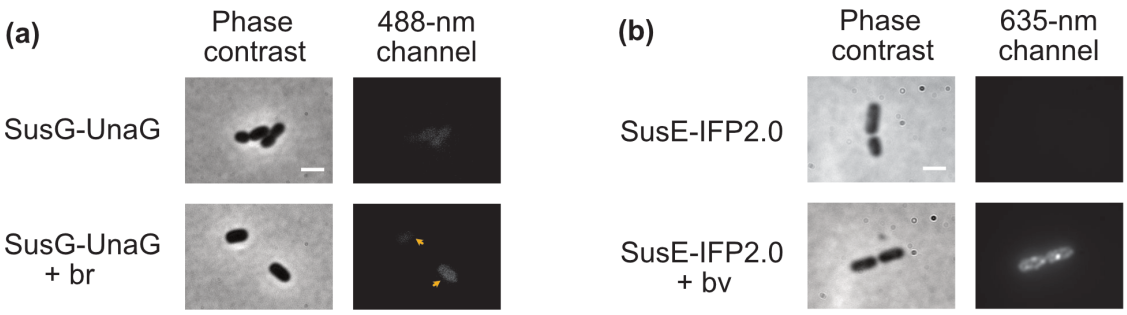


Figure 2.10: To investigate BBFP extracellular labeling, IFP2.0 and UnaG were fused to the outer membrane proteins SusE and SusG, respectively. (a) *B. theta* cells expressing SusG-UnaG grown with br (yellow arrows) were dim. (b) *B. theta* cells expressing SusE-IFP2.0 fluoresce upon 635-nm excitation when grown with bv. Scale bars: 2 μm.

R. bromii is not genetically tractable and cannot be fluorescently labeled with genetically encoded tools such as FPs. In phase-contrast mixed cultures of *B. theta* and *R. bromii*, dividing *B. theta* cells are difficult to distinguish from the relatively small and circular *R. bromii* (Fig. 2.11a). However, when mixed cultures of *B. theta* expressing UnaG and *R. bromii* were grown with br, the fluorescent *B. theta* cells (green) were distinguishable from the non-fluorescent *R. bromii* cells (Fig. 2.11a).

Furthermore, two-color imaging can be attained using UnaG and IFP2.0 as a labeling pair. We determined that *B. theta* cells expressing UnaG are fluorescent in the 488-nm channel but not the 635-nm channel whether grown in br or bv. Conversely, *B. theta* cells expressing IFP2.0 are fluorescent in the 635-nm channel but not the 488-nm channel whether grown in br or bv (Fig. 2.12). This protein specificity was further verified in vitro for purified UnaG and IFP2.0 protein incubated with excess br or bv (Fig. 2.13). Again, the protein spectrum is independent of the bound ligand. Purified UnaG had a similar fluorescence spectrum when incubated with excess br or bv, albeit with a two-fold reduction in emission efficiency in bv. Surprisingly, though IFP2.0-labeled cells grown in br were fluorescent (Fig. 2.12), we did not measure any fluorescence for purified IFP2.0 incubated with excess br excited at 635 nm; we attributed these differences to the fact that the single-molecule microscope used in these imaging experiments is more sensitive than a fluorescence spectrophotometer [22].

Based on this ability to distinguish UnaG from IFP2.0 regardless of the ligand identity, we imaged a mixed culture of UnaG-labeled *B. ovatus* and IFP2.0 labeled *B. theta* cells grown in both br and bv, and differentiated between the two cell strains based on imaging in separate color channels (Fig. 2.11b). Immunofluorescence staining of *B. theta* using polyclonal antibodies against SusF, a *B. theta*-specific protein, was imaged in a third channel to independently identify the *B. theta* cells.

To our knowledge, this work demonstrates the first example of BBFPs to live-cell imaging of obligate anaerobic bacteria. The blue-green UnaG and the red IFP2.0 can fluoresce in

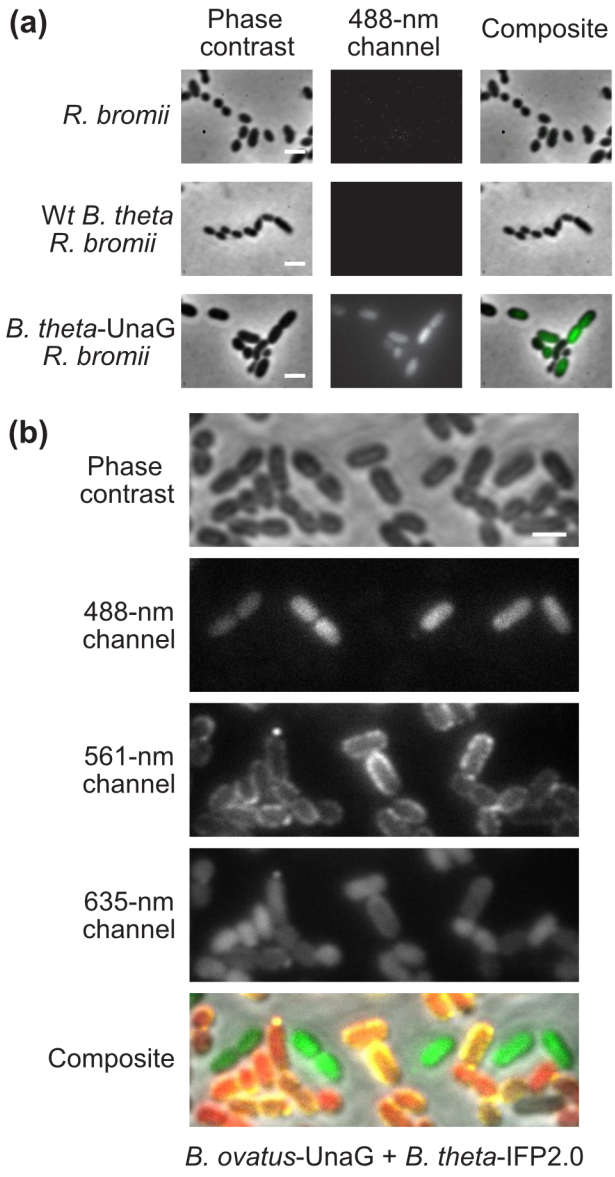


Figure 2.11: BBFPs can be used in mixed culture imaging and multi-color imaging. (a) *B. theta* expressing UnaG (green) is distinguished from unlabeled *R. bromii* when grown in mixed culture with br. Scale bar: 2 μm. (b) When grown in media containing br and bv, *B. ovatus* expressing UnaG is differentiated from *B. theta* expressing IFP2.0 in separate color channels using 488-nm (green) and 635-nm (red) excitation, respectively. Immunofluorescence staining imaged at 561 nm independently identified *B. theta* cells. Scale bar: 2 μm.

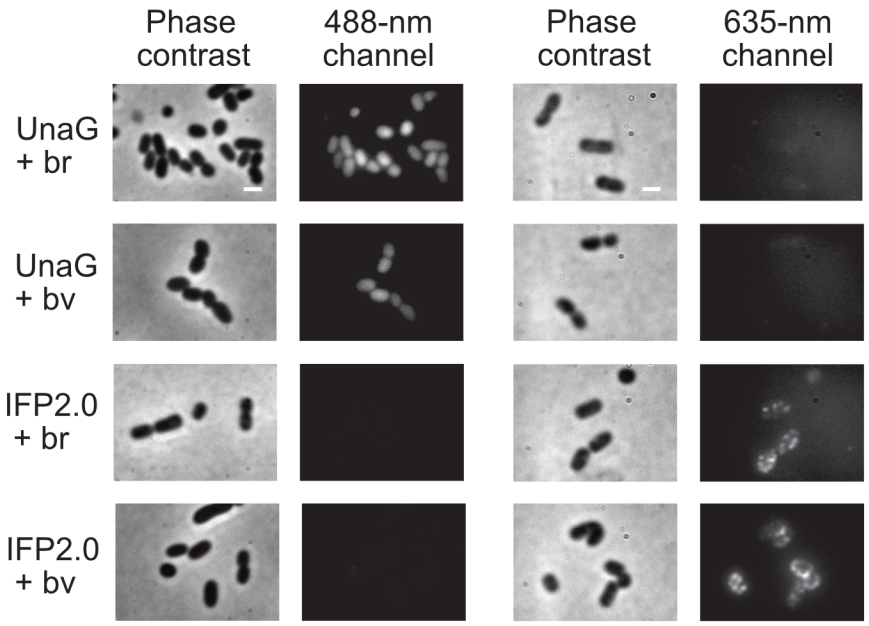


Figure 2.12: To evaluate ligand and excitation specificity, *B. theta* strains expressing UnaG or IFP2.0 were grown in br or bv and excited by 488-nm or 635-nm illumination. *B. theta* cells expressing UnaG are fluorescent in the 488-nm channel but not the 635-nm channel whether grown in br or bv and conversely, *B. theta* cells expressing IFP2.0 are fluorescent in the 635-nm channel but not the 488-nm channel whether grown in br or bv. Scale bars: 2 μm.

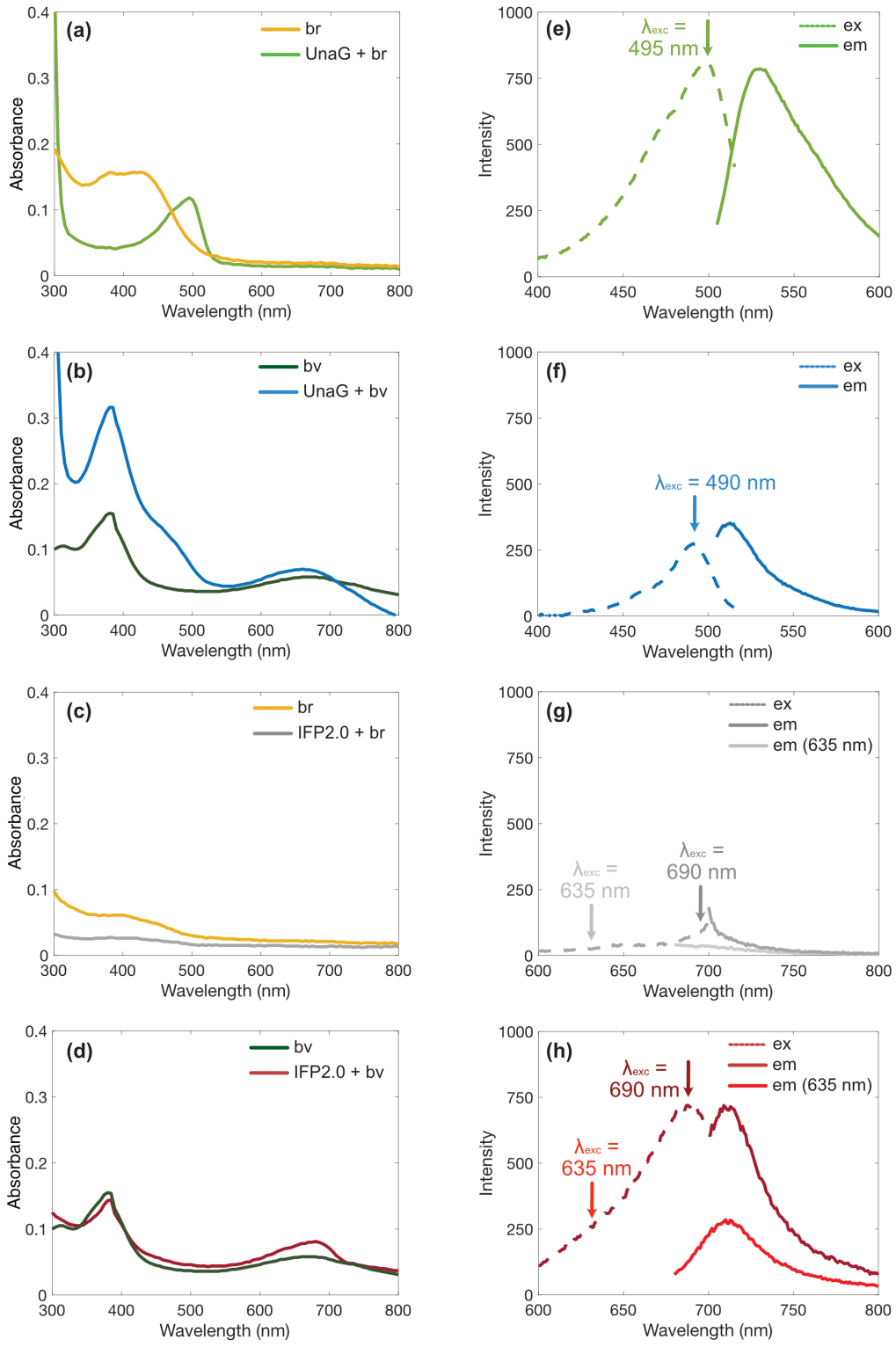


Figure 2.13: (a–d) UV-Vis absorbance spectra and (e–h) fluorescence emission spectra of UnaG and IFP2.0 incubated with excess br or bv. The fluorescence excitation wavelength is indicated with an arrow. (a,e) UnaG with br, (b,f) UnaG with bv, (c,g) IFP2.0 with br, and (d,h) IFP2.0 with bv.

both aerobic and anaerobic conditions and the wavelength used for fluorescence depends on the protein rather than the ligand bound. Both UnaG and IFP2.0 can fluorescently label *B. theta* in monocultures. In co-cultures, *B. theta* labeled with UnaG can be distinguished from unlabeled *R. bromii*. Finally, UnaG and IFP2.0 labeled cells could be distinguished from one another in two-color imaging.

Among the options for oxygen-independent fluorescent reporters [50], BBFPs are an attractive option for further optimization. Flavin Mononucleotide (FMN)-based FPs work in live-cell imaging because the FMN cofactor is readily accessible, but these reporters are confined to the blue spectral region and are weakly fluorescent [57]. HaloTag [34] and other self-labeling tags like SNAP [71] and CLIP [72] specifically bind bright fluorescent dyes but require a wash step to eliminate excess ligand; this treatment may not be compatible with continuous live-cell imaging and may disrupt microbial communities. Recent developments have yielded fluorogenic dyes for HaloTag [93] and SNAP-tag [94–96] that circumvent the need for washing and increase applications in single-molecule imaging. However, these ligands may still suffer from the low cell permeability in bacteria that limits their utility in labeling cytoplasmic targets. In our past experiments using HaloTag technology, we have not successfully labeled bacterial proteins in anaerobes other than targets expressed on the outer membrane [27]. It may also be feasible to implement other ligand-dependent reporter systems such as Y-FAST [97], antibody-based fluorogen anctivating proteins [98], and engineered fluorogen-dependent proteins [99–101] in anaerobic bacteria using cell-permeable fluorogenic ligands to provide the same advantages of oxygen-independent labeling as BBFPs. Previously, BBFPs have been demonstrated in live-cell mammalian systems [66,67]. In these applications, as in the current paper, br and bv may be endogenously released by heme degradation and interconverted in cells by endogenous oxidoreductases.

IFP2.0 has already been engineered to be a bright and monomeric FP variant from phytochrome protein family [66, 67]. Here, we have shown that IFP2.0 is an appropriate red

label for live-cell imaging in anaerobic bacteria. Future developments in the IFP family of FPs will further improve this application. For instance, IFP2.0 dimerizes at high concentrations and was recently further modified to produce mIFP, a monomeric red BBFP [67]; IFP2.0 aggregation may explain the distinct puncta we observed in the *B. theta* strains with highly expressed IFP2.0 (Fig. 2.8d). Overall, these red-shifted FPs are highly desirable for in vivo imaging with applications in deep tissue imaging and thick biofilms.

While UnaG has been recently demonstrated as a useful tool for single-molecule imaging as a dark-to-green photoswitchable FP [102], this FP has yet to be as extensively optimized as IFP2.0, and it is possible that UnaG can be diversified to a palette of different fluorescent colors. Others engineered UnaG to create a brighter and more stable variant containing a single V2L mutation [80]. While no one has attempted to do so, it is possible that UnaG can be diversified to a palette of different fluorescent colors. Since UnaG binds br non-covalently using a hydrophobic pocket, we initially postulated that the protein could accommodate different fluorogenic small molecules as a ligand. We hypothesized that UnaG could bind bv as an alternative ligand and that extended π-conjugation of bv would red-shift UnaG holoprotein excitation and emission. Others have noted that UnaG does not fluoresce using bv [60]. Similarly, we did not observe any red fluorescence when UnaG was bound to bv instead of br. However, we observed some fluorescence in the original 495/525 nm range, albeit with a diminished intensity (Fig. 2.12 and 2.13). These results suggest that there may be some spontaneous or endogenous interconversion between br and bv. Likewise, both the BBFP binding pocket and the ligand itself are important factors in determining fluorescence intensity and color.

2.4 Conclusions

We have demonstrated that BBFPs UnaG and IFP2.0 are well-suited probes for anaerobic live-cell imaging. Unlike GFP, both UnaG and IFP2.0 can fluorescently label obligate anaerobic bacteria regardless of oxygen exposure. Furthermore, fluorescence was

detectable without perturbing the sample in a washing step. These proteins can be used to label strains in mixed bacterial cultures and can be used for two-color imaging as complementary probes, using common GFP and Cy5 filter cubes. We therefore foresee the use of BBFPs as labels for imaging more complex, polymicrobial communities and in general for studying living bacterial systems that require oxygen-free environments.

CHAPTER III

A New Green Ligand-Dependent Fluorescent Reporter for Anaerobic Imaging

The work presented has been submitted for publication as:

Chia, H.E., Koebke, K.J., Koropatkin, N.M., Marsh, E.N.G., and Biteen, J.S.
A new green ligand-dependent fluorescent reporter for anaerobic imaging.

3.1 Introduction

Fluorescence imaging has become the gold standard for visualizing biological phenomena in living and fixed cells. Advances in fluorescence microscopy have yielded techniques such as multicolor imaging, single-molecule tracking, and super-resolution microscopy [22, 26]; all of these advances were aided by the development of diverse small-molecule dyes and fluorescent proteins (FPs) [103]. In particular, FPs are among the most commonly used tools to label proteins and cells of interest and have been engineered to fluoresce over a wide range of wavelengths for use in imaging [73,104]. However, popular FPs derived from GFP and DsRed require an oxidative post-translational modification to fluoresce [48, 49]. This requirement precludes the use of common FPs to probe oxygen-sensitive samples such as the medically relevant polymicrobial communities in the gut

microbiome [50].

On the other hand, ligand-dependent FPs are promising tools that have already been demonstrated in obligate bacterial anaerobes [27,105]. Unlike GFP or DsRed-derived FPs, ligand-dependent FPs confer oxygen-independent labeling because fluorescence depends only on the molecular structure of the fluorophore and, if the ligand is fluorogenic [106], upon the ligand binding to the protein.

Several approaches have addressed the limitations of GFP-like FPs for labeling anaerobic bacteria. These approaches include the Flavin Mononucleotide (FMN)-based Fluorescent Proteins (FbFPs) or flavin-binding Light-Oxygen-Voltage (LOV) FPs [51, 56, 107], which are constrained to blue emission by the molecular structure of FMN [108], and covalent self-labeling systems like HaloTag [34], which rely on organic dye ligands [68,109] that are not always cell-permeable in prokaryotes and which require multiple washing steps that may not be compatible with continuous imaging of live-cell samples. Both FbFPs [55, 89] and HaloTag [21, 27, 92] have been used to label anaerobic bacteria with varying degrees of implementation ease due to the aforementioned limitations. Another ligand-dependent reporter, Y-FAST [97] and its derivatives [110], utilize synthetically tailored fluorogenic ligands for bright fluorescence without the need for washing before imaging, but have not yet been demonstrated in anaerobic bacterial systems.

Bilin-based Fluorescent Proteins (BBFPs) reconcile the advantages of FbFPs and self-labeling tags for oxygen-independent fluorescent labeling. BBFPs bind bilirubin (br) and biliverdin (bv), which are fluorogenic ligands derived form the heme degradation pathway, and can be used to label intra- and extra-cellular targets [105]. BBFPs include the green UnaG [60] and the far-red IFP2.0 [66], and we recently demonstrated these probes in *Bacteroides thetaiotaomicron (B. theta)* and *Bacteroides ovatus (B. ovatus)*, including in anaerobic live-cell mixed-species imaging [105]. The advantages of UnaG include its small size (15.6 kDa compared to GFP-like FPs at 27 kDa), a monomeric structure that is not prone to oligomerization, and much higher brightness compared to FbFPs. UnaG-br was

also recently demonstrated as a dark-to-green photo-switchable FP for super-resolution imaging [102, 111], which further extend its versitility as an imaging tool. Despite these advantages, the attainable BBFPs colors are limited by their ligands: UnaG with br is green and IFP2.0/mIFP/smURFP with bv is red [67, 112].

Utilizing the plasticity of the UnaG binding pocket, we identified new fluorogenic ligands via high-throughput screening (HTS) of a library of small molecules. This screen ultimately yielded a novel UnaG-ligand pair. Although the ligand is nominally fluorescent under 532-nm excitation, binding UnaG results in a 10 nm shift, resulting in an emission peak at 581 nm, and a 2.5-fold increase in fluorescence intensity. We demonstrated the novel pair's implementation in anaerobic imaging in the common green 532-nm fluorescence channel using UnaG-labeled *B. theta* and *E. coli*. this pair is also useful for two-color mixed-culture imaging, which we demonstrated in a mixture of UnaG-labeled *B. ovatus* and IFP2.0-labeled *B. theta*. Overall, this work expands the toolbox of ligand-dependent FP imaging agents that can be used for non-invasive fluorescence investigations of anaerobic microbial systems.

3.2 Materials and methods

3.2.1 Protein Expression

MBP-UnaG was prepared as previously reported in Section 2.2 or [105] and stored at 4°C in Buffer A, consisting of filtered 20 mM Tris-HCl (pH 7.2), 200 mM NaCl, and 1 mM EDTA. MBP-UnaG is referred to as UnaG throughout the *in vitro* HTS and spectral characterization experiments.

3.2.2 HTS Assay Protocol

For primary assays, black, low-volume, non-coated 384-well plates (Greiner Bio-One, Ref. 784900) were prepared by dispensing 10 μL of Buffer A using a Multidrop Combi

Reagent Dispenser (Thermo Scientific). Compounds weree added into sample wells using a Sciclone liquid handler with a pin tool (50 nL of 2 mM stock in DMSO, 5 μM final concentration); an equivalent volume of DMSO was added to control wells. UnaG protein was subsequently dispensed into plates (10 μL of 1 mM stock, 500 nM final concentration). Negative controls were buffer-only wells and protein-only wells. Bilirubin (br) was added to positive control wells (2 μL, 5 μM final). Plates were covered with aluminum foil to prevent light exposure and incubated by shaking on the Multidrop dispenser (10 min, 500 rpm). Plates were centrifuged (1 min, 1000 rpm) before measurements.

For confirmation assays, compounds were dispensed with a Mosquito picker (50 nL of 2 mM stock in DMSO, 5 μM final; TTP Labtech); plates were prepared with Buffer A and UnaG protein as was done for the primary assays above. For secondary assays, plates were prepared similar to confirmation assays with the exception of adding 10 μL Buffer A instead of UnaG protein.

Fluorescence intensity was detected using a PHERAstar plate reader (BMG Labtech) using three optics modules centered at excitation/emissiion wavelengths of 485/520 nm, 540/570 nm, and 580/610 nm, respectively.

3.2.3 Compound Libraries

All compounds screened were stored at the University of Michigan Center for Chemical Genomics. For the primary screen, 7,680 compounds from the ChemDiv 100K library were used.

3.2.4 Assay Performance and Data Analyses

Data collected in the 485/520 nm and the 540/570 nm channels were analyzed using MScreen [113]; data collected in the 580/610 channel was analyzed manually using Microsoft Excel.

In primary assays, compounds incubated with protein that displayed fluorescence in-

tensity greater than ≥ 3 standard deviations (SD) away from the mean of the negative controls were considered as initial hits. Compounds that fluoresced in the 540/570 nm optics module but not in the 485/520 nm module were considered primary hits; the same analysis was used to identify preliminary hits that fluoresced in the 580/610 nm module but not the 485/520.

In confirmation and secondary assays, preliminary hits were evaluated in triplicate. Compounds incubated with protein that exhibited an increase in fluorescence intensity greater than ≥ 3 SD from the mean of the negative controls were considered hits in the confirmation assays. However, these compounds were excluded from the final hit list if the compound-only plates in secondary screening also produced fluorescence intensity greater than ≥ 3 SD from the mean of the negative controls. Compounds that were hits in primary and confirmation assays but not in secondary assays were considered confirmed hits and ordered from MolPort (Table 3.1). These confirmed hits were analyzed in concentration-dependent curves in duplicate.

3.2.5 Spectral Characterization and Titrations

UnaG and compounds were all prepared for measurements in Buffer A. UV-visible absorbance measurements were performed in 96-well clear bottom plates using the Molec-

Table 3.1: Confirmed Hits from HTS Assay.

ID in this study	Hit in HTS channel	MolPort ID	Catalog No.	Supplier
2	540/570	MolPort-035-896-678	4090-1986	ChemDiv, Inc.
3	580/610	MolPort-001-836-754	1959-0257	ChemDiv, Inc.
4	580/610	MolPort-047-118-081	0898-0008	ChemDiv, Inc.
5	540/570	MolPort-000-564-975	AG-690/12510375	Specs
6	580/610	MolPort-000-445-083	AG-690/12890124	Specs
7	580/610	MolPort-001-931-262	AG-690/10379022	Specs
8	540/570	MolPort-000-717-433	STK874239	Vitas-M Laboratory, Ltd.
9	540/570	MolPort-001-965-798	STK831400	Vitas-M Laboratory, Ltd.
10	580/610	MolPort-001-848-862	STK094419	Vitas-M Laboratory, Ltd.
11	580/610	MolPort-001-004-902	STK039750	Vitas-M Laboratory, Ltd.
12	540/570	MolPort-000-225-490	STK396289	Vitas-M Laboratory, Ltd.
13	580/610	MolPort-001-931-334	STK372609	Vitas-M Laboratory, Ltd.
14	540/570	MolPort-001-935-437	STK084537	Vitas-M Laboratory, Ltd.
15	540/570	MolPort-002-116-650	STK094803	Vitas-M Laboratory, Ltd.
16	580/610	MolPort-001-951-131	STK893803	Vitas-M Laboratory, Ltd.

ular Devices SpectraMax iD3 microplate reader in a black-walled quartz cuvette (Hellma, 1 cm pathlength) using a Hewlett Packard 8453 UV-vis spectrophotometer. Molecular extinction coefficient (ε) was calculated using the Beer-Lambert law. Plotted UV-vis spectra are background corrected using buffer blanks and processed using MATLAB.

Fluorescence measurements were carried out on an Agilent Varian Cary Eclipse Fluorescence Spectrophotometer. Plotted fluorescence excitation and emission spectra are averaged plots from three technical replicates.

Fluorescence titrations were performed by taking fluorescence emission spectra at 495-nm excitation of UnaG bound with br, followed by parallel additions of competing ligand to a br-only sample and competing ligand to the UnaG-br complex. After each addition of competing ligand, the sample was allowed to equilibrate for 10 min before fluorescence measurements [61]. The compound-only spectra were used for background subtraction for competing ligand-UnaG-br readings. Assuming a 1:1 displacement of br by the competing ligand, the concentration of free competing ligand was calculated to use for fitting. Data was fit using Prism (GraphPad) using a single-site binding model with a linear baseline correction:

$$Y = C + S[L] + \frac{B_{max}[L]}{K_{d(app)} + [L]} \tag{3.1}$$

Where Y is the signal of the UnaG-br complex, and $[L]$ is the concentration of the competing ligand, C and S are the intercept and slope of the baseline, B_{max} is the fluorescence in the absence of L, and $K_{d(app)}$ is the apparent dissociation constant of L. $K_{d(app)}$ was corrected to the true K_d for L using the known K_d for br.

3.2.6 Cell Cultures

The *B. ovatus* and *B. theta* strains used in this study have been previously described [105]. *B. ovatus* and *B. theta* strains were intially started in rich media containing tryptone-yeast extract-glucose and incubated anaerobically at 37°C in a Coy chamber. Cultures

were subsequently back-diluted into minimal media with 0.05% w/v maltose as a carbohydrate source.

E. coli (NEBExpress, New England Biolabs) were transformed with puc19 or pMAL_c5x_UnaG plasmids using standard high-efficiency transformation protocols and plated on LB/Agar plates with ampicillin (amp, final 100 μg/mL). Colonies were picked into liquid LB cultures with amp to grow overnight at 37°C on a shaker. cultures weree subsequently back-diluted 1:200 into LB with amp to grow on a shaker at 37°C.

3.2.7 Growth Curves

B. theta cells were cultured with minimal media with 0.05% w/v maltose and back-diluted 1:200 into 96-well clear bottom plates with media and respective compounds. Each growth experiment condition was performed in triplicate. Plates were loaded into a Biostak automated plate-handling device (BioTek Instruments). Absorbance at 600 nm (OD_{600}) was measured in each well every 20 minutes by a Powerwave HT absorbance reader (BioTek Instruments). Data was recorded using Gen5 software (BioTek Instruments) and processed using Prism (GraphPad).

3.2.8 Fluorescence Microscopy

E. coli were grown to OD_{600} 0.4 and induced with 0.4 mM IPTG to grow for 3 hours before imaging. *B. ovatus* and *B. theta* cells were grown to early to mid-long phase and imaged. For labeling with **2** and **4**, cells were incubated with a final ligand concentration of 2.5 μM for 2.5-3 hours before washing or directly used for imaging. For IFP2.0-labeled cells, bv was directly supplemented into the overnight culture (2.5 μM final concentration).

All imaging was performed at room temperature and anaerobically on cells sealed between coverslips with epoxy as previously described by our lab [27]. Imaging was performed in an Olympus IX71 inverted epifluorescence microscope with a 100× 1.4 N.A. wide-field oil-immersion objective. Samples were illuminated by 532-nm laser (Crysta-

Laser CL-532-025-O; 2 mM/cm^2) or with a 640-nm laser (Coherent CUBE 640-40C; 80 W/cm^2). Fluorescence emission was filtered with appropriate filter sets and imaged on a 512 × 512 Andor iXon EMCCD camera at 100 frames/s. For consistency and noise reduction, all phase-contrast and fluorescence images were created by summing frames for a total of 400 ms integration time. Recorded images were analyzed using ImageJ; all images presented on the same color scale.

3.3 Results

Using a high-throughput and small molecule approach, we devised a screening method to identify new fluorogenic ligands that could bind UnaG and to create new UnaG-ligand pairs that were red-shifted relative to the original UnaG-br complex (GFP channels, 488 nm excitation). The HTS strategy is depicted in Figure 3.1. In the primary screening step, 7680 compounds were loaded into individual wells in 384-well plates containing UnaG protein and were illuminated at three excitation wavelengths (485 nm, 540 nm, and 580 nm) for fluorescence endpoint reading at 520 nm, 570 nm, and 610 nm, respectively (Fig. 3.1a). For each fluorescence channel , compounds were considered hits if their fluorescence signal was greater than 3 SD above the negative control included on each plate (Fig. 3.1b and Fig. 3.2).

As this screening methodology was a non-standard HTS protocol, Z' scores [84] could not be calculated for the entire primary assay given that a positive control could only be established for the 485/520 nm optics module using UnaG-br. While UnaG-br signal could be read in the 540/570 nm module, the dynamic range established made it difficult to calculate robust Z' scores; furthermore, no positive control could be established on each plate for 580/610 nm, which again made it unfeasible to calculate Z' scores. The Z' score for the primary assay 485/520 nm channel was low (0.27), reflecting significant variability in the signal magnitude in positive control wells. However, very little variability was observed between negative controls in each channel (Fig. 3.2), which allowed us to

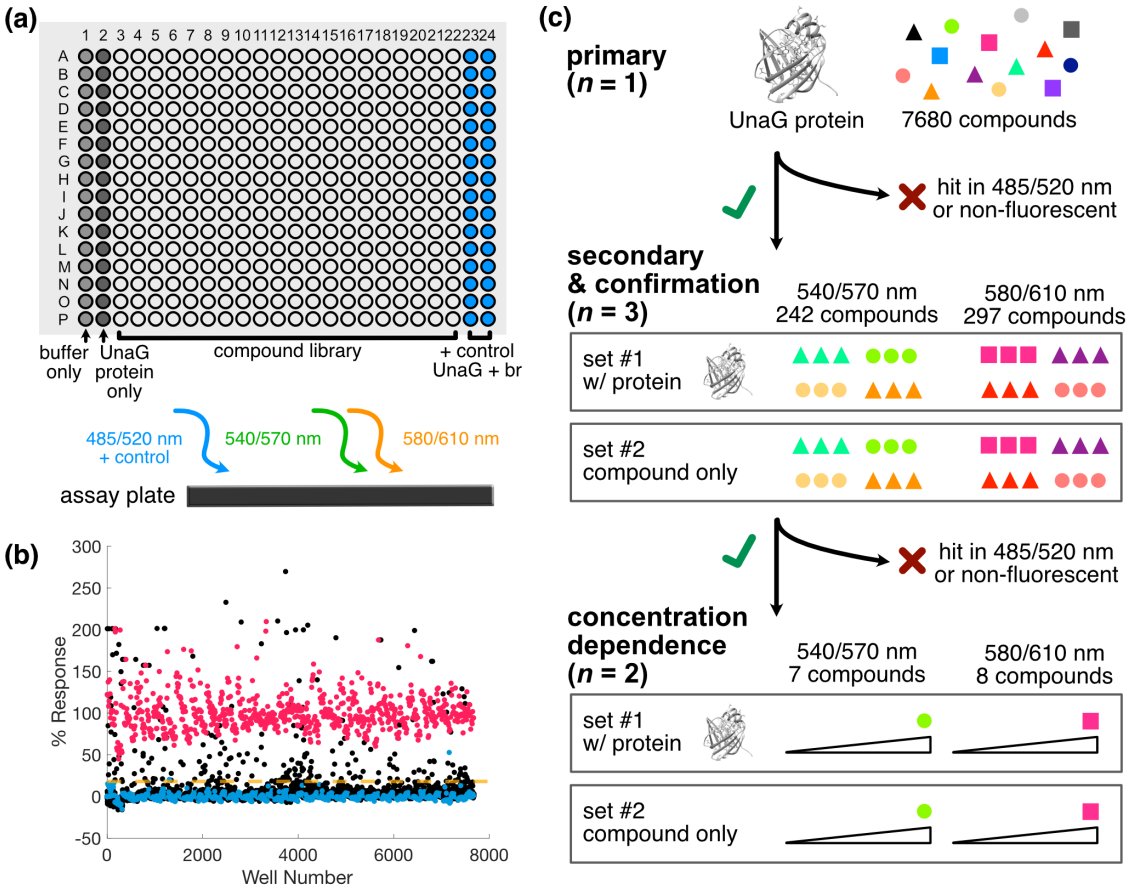


Figure 3.1: Screening methodology for identifying new, red-shifted ligand-UnaG pairs. (a) Schematic of primary HTS assay. Compounds were incubated with UnaG protein in a 384-well black-walled plates before fluorescence intensity endpoint reads using three optics modules; columns 1 and 2 were reserved for negative controls and columns 23 and 24 for positive controls of UnaG and br for the 485/520 nm module. (b) 485/520 nm primary screen. Each dot represents the activity result of a well containing a text compound (black), negative control of buffer or protein only (blue), and positive control of UnaG and br (red). In each module, wells that returned signal above the 3 × SD threshold cutoff (yellow dashed line) were considered hits. (c) Workflow logic used to cull hit lists from three optics modules.

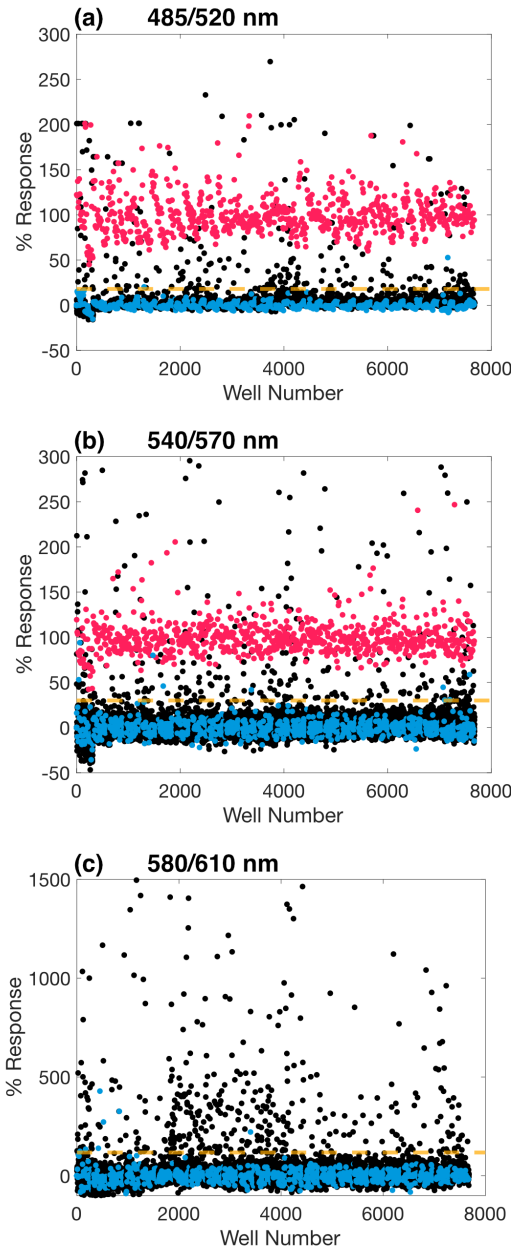


Figure 3.2: Scatterplots of the entire primary HTS. Fluorescence endpoint readings were taken at (a) 485-nm excitation, 520-nm emission (485/520 nm), (b) 540-nm excitation, 570-nm emission (540/570 nm), and (c) 580-nm excitation, 610-nm emission (580/610 nm). Each dot plotted represents the percent response of a well containing a test compound (black), negative control of buffer or protein only (blue), or positive control of UnaG with bilirubin (red); the positive control exhibited no signal in the 580/610 nm module and is thus omitted in the plot in (c). Preliminary hits were selected based on signal above the 3 × SD threshold cutoff (yellow dashed line).

draw a low activity cut off for identifying primary hits.

To identify compounds that bind UnaG and fluoresce at longer wavelengths, the workflow required counter-selection at each screening step (Fig. 3.1c). In the primary screen, UnaG-ligand pairs that showed fluorescence in either the 540/570 nm or 580/610 modules but not in the 485/520 nm module were considered preliminary hits (Table 3.2, 539 compounds with 7.01% hit rate). These preliminary hits were evaluated in triplicate in confirmation screens with UnaG. Secondary screens with compounds alone were plated in triplicate and measured in parallel; this counter-screening was necessary to select for compounds with fluorescence that is enhanced by UnaG rather than compounds with high intrinsic fluorescence. A concentration-dependent screen confirmed that the detected fluorescence signal was due to ligands binding UnaG rather than non-specific fluorescence emission (Fig. 3.3). In total, we identified fifteen promising compounds for imaging: seven of which are fluorescent in the 540/570 nm channel and eight of which are fluorescent in the 580/610 nm channel. Interestingly, the structures of the selected compounds, while highly π-conjugated, are not structurally similar to popular commercial organic dyes such as the AlexaFluor dyes (Fig. 3.4).

We subsequently tested the 15 candidates (referred to by boldface numbers **1 – 15**

Table 3.2: Summary of HTS campaign.

	Screen Type	**Readout**	**# Compounds Tested (Wavelength)**	**# Selected Compounds**	**% Hit Rate**
1	Primary	Fluorescence endpoint	7680	539	7.01
2a 2b	Confirmation Secondary	Fluorescence endpoint	242 (540/570) 297 (580/610)	15	2.78
3	Concentration dependence	Fluorescence endpoint	7 (540/570) 8 (580/610)	10	66.67
4	Validation	Fluorescence spectrum; UV-vis spectrum; K_d	10	2	20.00

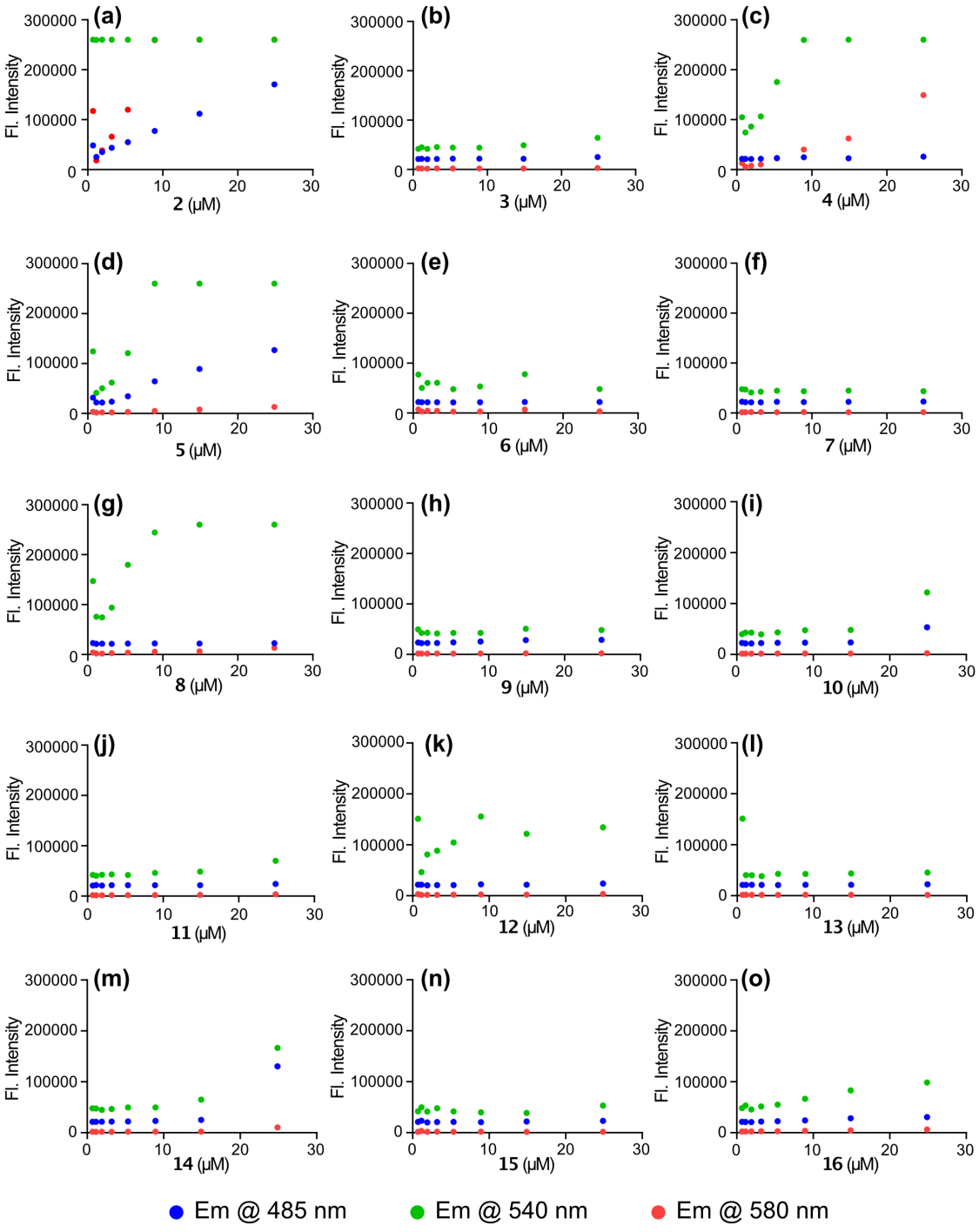


Figure 3.3: Concentration-dependent curves of compounds obtained through HTS. (a-o) **2-16** in increasing concentration were added to wells with a fixed UnaG concentration (500 nM).

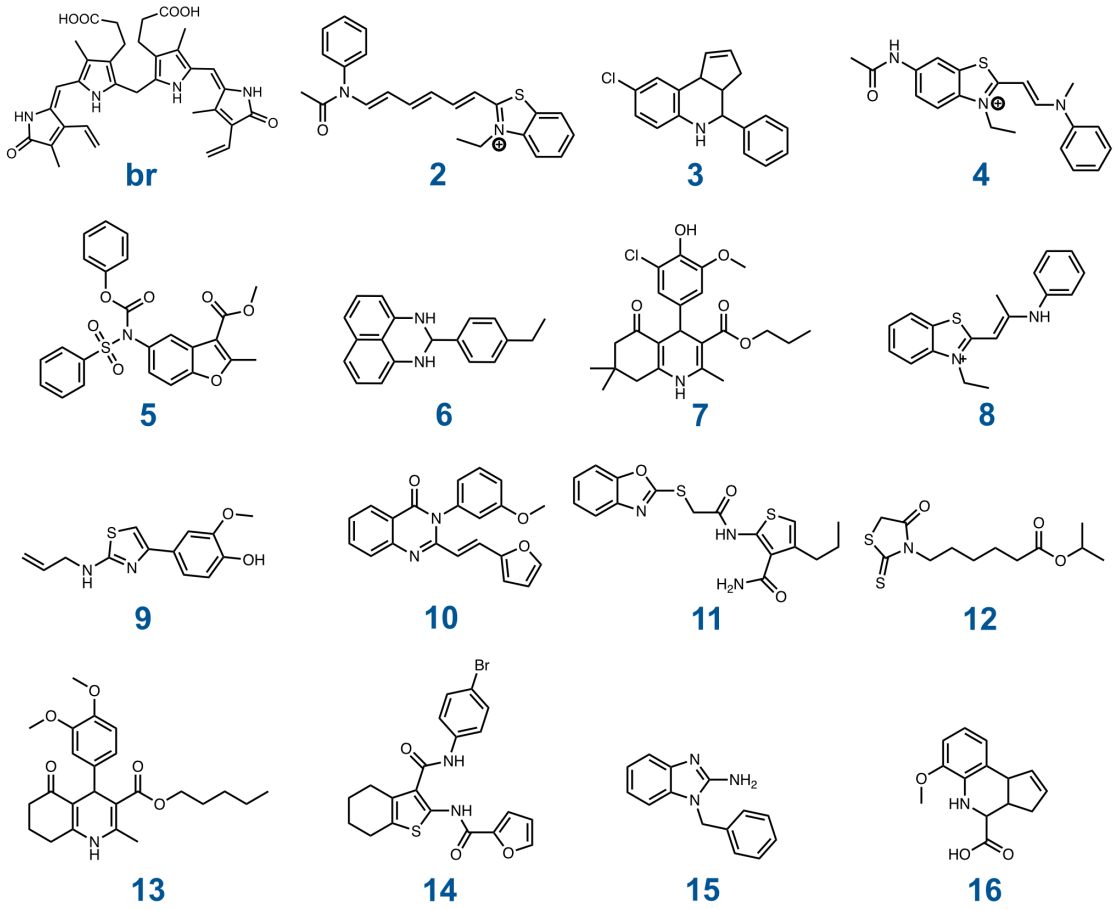


Figure 3.4: Structures of the native UnaG ligand br and the compounds obtained through HTS (**2**-**16**).

by analysis of the compounds alone as well as when incubated in excess with UnaG using UV-vis absorbance and fluorescence emission (Fig. 3.5 and 3.6). Furthermore, adding compounds in the media (2.5 μM) did not affect the growth kinetics of *B. theta* (Fig. 3.7), indicating that these compounds were non-toxic and could be supplemented to media to label growing cells.

Of the 15 confirmed hit compounds, **2** and **4** were most promising for further investigation due to observed red-shifted emission and fluorescence enhancement upon binding UnaG (Fig. 3.8). Spectral properties of the UnaG-**2** and UnaG-**4** pairs are reported in Table 3.3. Both compounds are benzothiazoles with moieties that extend the π-conjugation of the molecules (Fig. 3.8d and g) and are similar in size and planarity to the original br

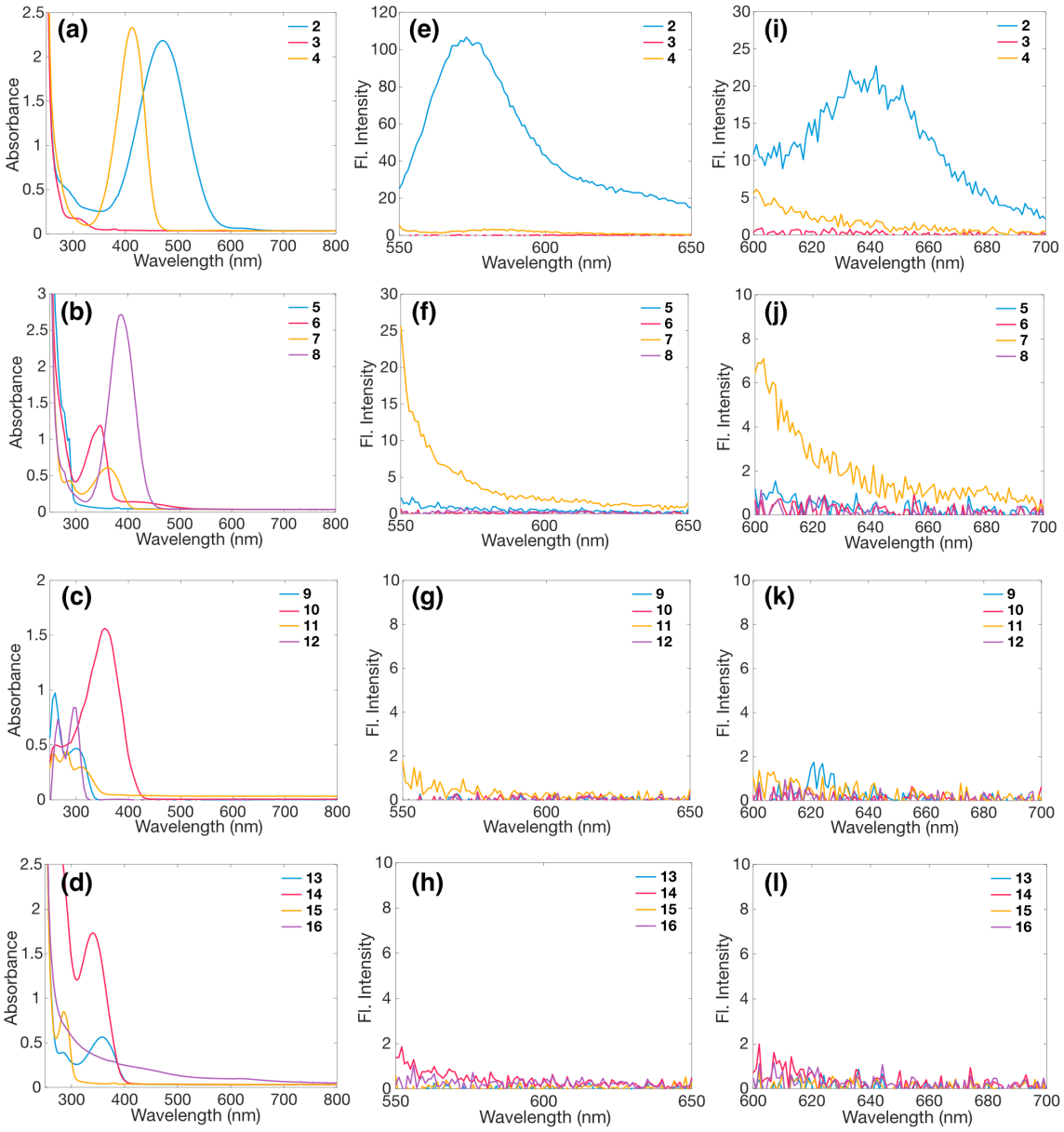


Figure 3.5: (a-d) UV-vis absorbance spectra and (e-l) fluorescence emission spectra of compounds obtained through HTS. UV-vis spectra were collected in DMSO at concentration of 100 μM. Fluorescence emission of compounds in buffer were recorded from (e-h) 540-nm excitation and (i-l) 580-nm excitation at concentration of 1 μM.

ligand (Fig. 3.8a). Surprisingly, binding of **2** and **4** to UnaG does not significantly shift the UV-vis absorbance maximum, whereas binding of br to UnaG shifts the absorbance maximum from 437 nm to 495 nm (Fig. 3.8b, e, h). Incubating **2** with UnaG red shifts the maximal emission by 10 nm (from 571 to 581 nm) under excitation at 532 nm and gives

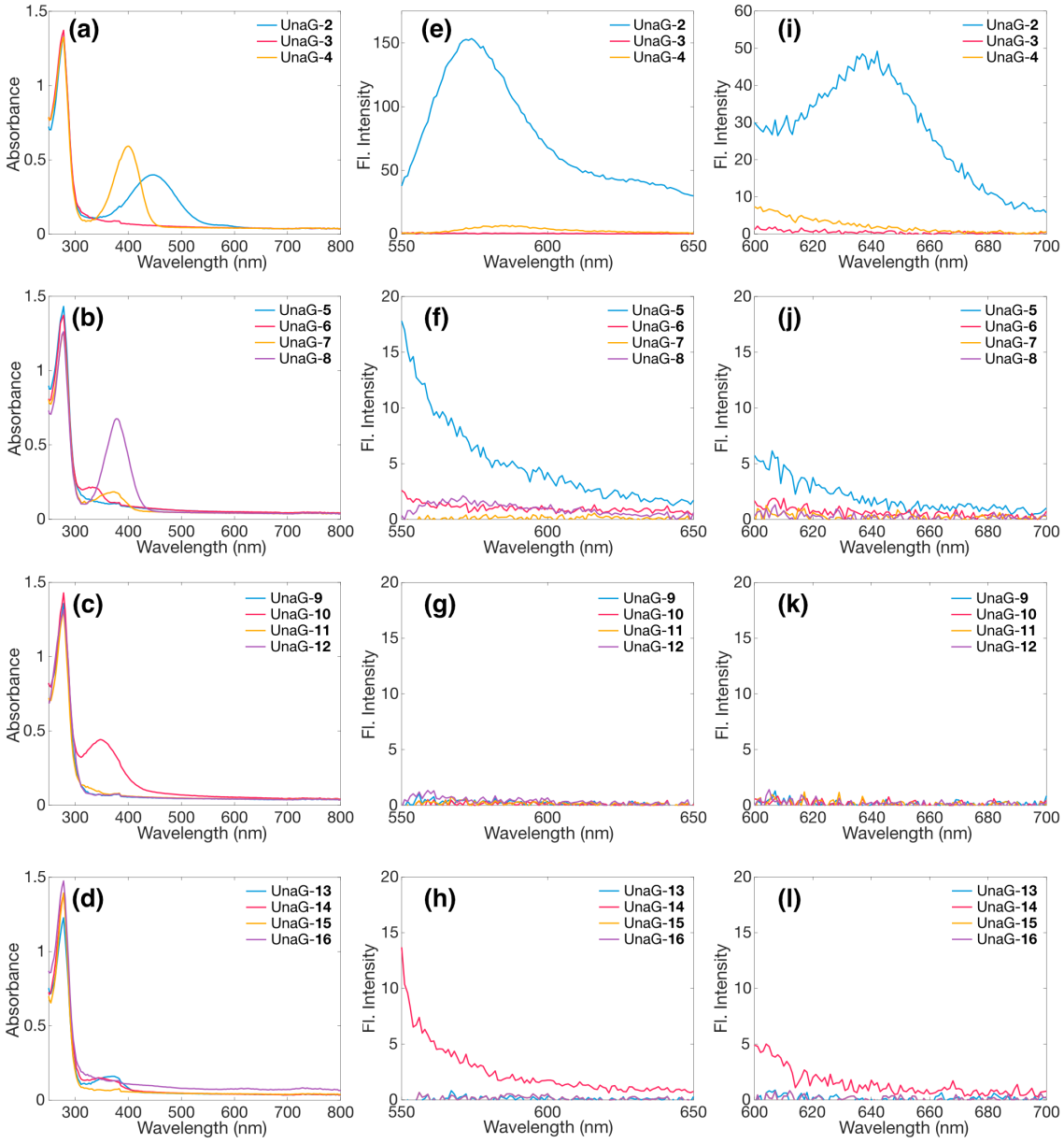


Figure 3.6: (a-d) UV-vis absorbance spectra and (e-l) fluorescence emission spectra of compounds obtained through HTS incubated with UnaG. UV-vis spectra were collected in buffer at concentration of 20 μM. Fluorescence emission of UnaG-ligands in buffer were recorded from (e-h) 540-nm excitation and (i-l) 580-nm excitation at concentration of 2 μM for ligands and 100 nM for UnaG.

rise to a 2.5-fold fluorescence intensity enhancement (Fig. 3.8f). Likewise, incubation of **4** with UnaG red shifts the emission maximum by 13 nm (from 582 to 595 nm) under excitation at 532 nm and produces a 3-fold fluorescence intensity enhancement (Fig. 3.8i). Through 532 nm is not the peak fluorescence excitation wavelength for either UnaG-**2** or

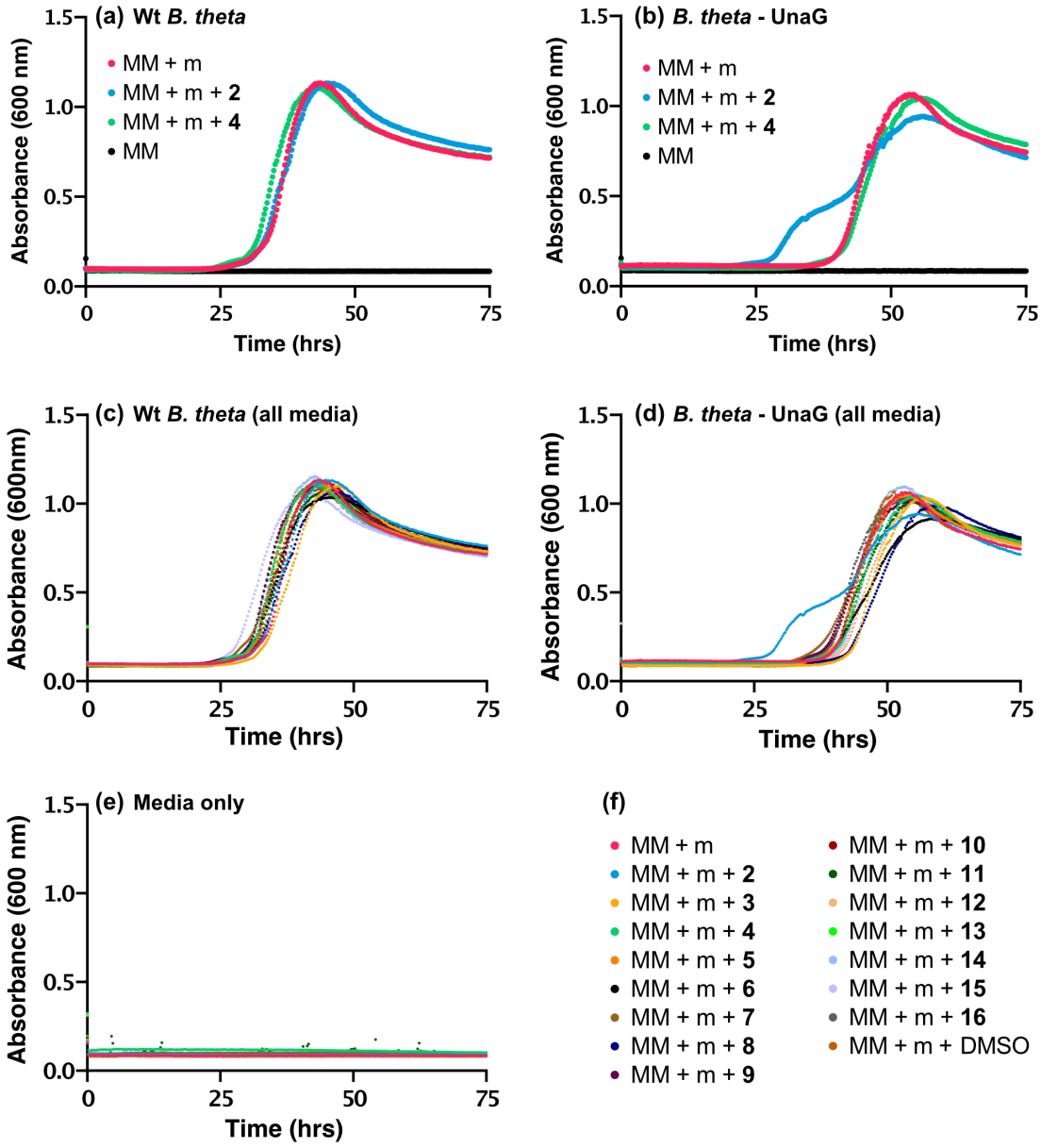


Figure 3.7: Growth measured by absorbance (600 nm) for cells grown in minimal media (MM) containing 5 mg/mL maltose (m) and supplemented with 2.5 μM compounds selected from the ligand-screening process. (a) Wild-type *B. theta* and (b) *B. theta* cells expressing UnaG are unaffected by the presence of **2** and **4**, with the exception of *B. theta* - UnaG, which exhibited a slight rapid growth phase in the presence of **2**. The addition of compounds **2 - 16** as well as 0.005% DMSO does not affect the growth of (c) Wt. *B. theta*, (d) *B. theta* - UnaG, or (e) media-only conditions. (f) Strain legend for panels c-e.

UnaG-**4**, this excitation wavelength produces detectable fluorescence emission (Fig. 3.9).

We evaluated the affinity and specificity of **2** and **4** binding to UnaG with respect to

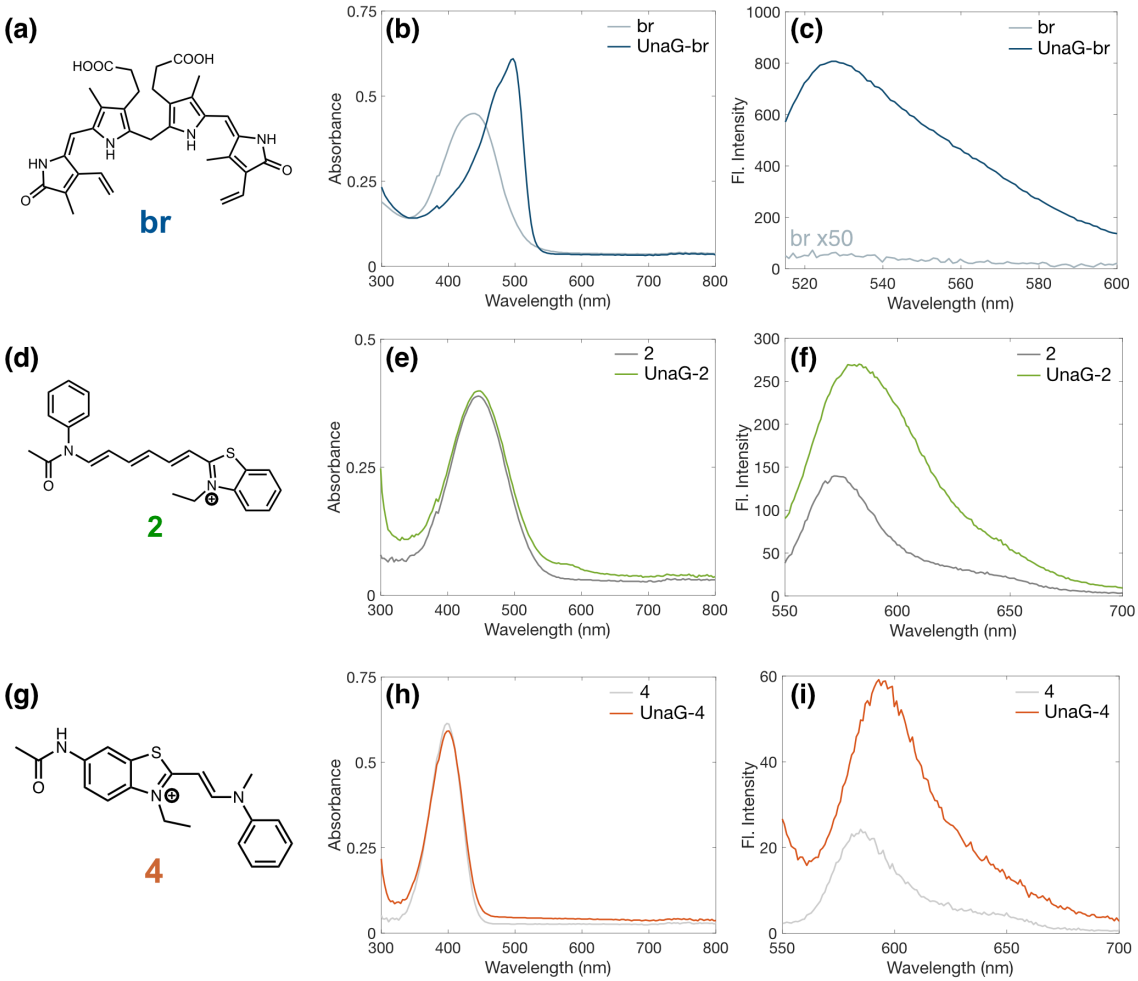


Figure 3.8: Spectral properties of UnaG-2 and UnaG-4. (a,d,g) Compound structures, (b,e,h) UV-vis absorbance spectra, and (c,f,i) fluorescence emission spectra of compounds only and compounds bound to UnaG. (a-c) UnaG and its original binding partner br, (d-f) UnaG and compound **2**, and (g-i) UnaG and compound **4**. Fluorescence emission spectra were taken at (c) 495-nm excitation and (f,i) 535-nm excitation.

Table 3.3: Properties of confirmed hit compounds and compounds bound to UnaG.

	Hit in HTS channel	λ_{exc} **(nm)***	λ_{em} **(nm)**	ε **(M^{-1} cm^{-1})****
2 alone	540/570	455	571	45606
UnaG-**2**		463	581	49119
4 alone	580/610	409	582	50918
UnaG-**4**		580	595	63808

*Maximal excitation wavelength was determined through fluorescence excitation scan at 535 nm.

**Extinction coefficient measured at peak absorbance wavelength in Tris-HCl buffer (pH 7.4).

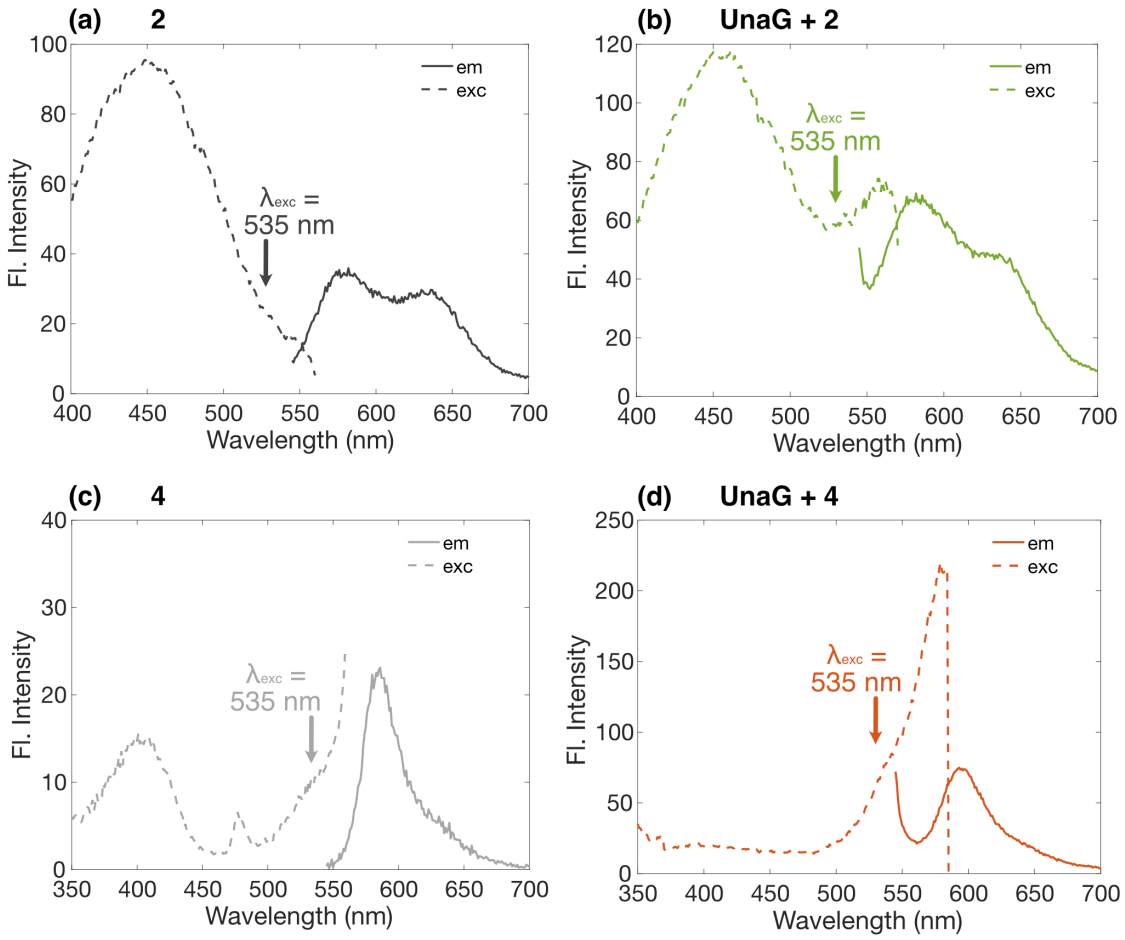


Figure 3.9: Fluorescence excitation (exc) and emission (em) spectra of (a) **2**, (b) UnaG incubated with **2**, (c) **4**, and (d) UnaG incubated with **4**. The fluorescence excitation wavelength, λ_{exc}, is indicated with an arrow. (a) **2** and (b) UnaG-**2** were measured at 5 μM in buffer; (c) **4** and (d) UnaG-**4** were measured at 15 μM in buffer.

Table 3.4: $K_{d(app)}$ and true K_d from fluorescence competition titrations for **2** and **4**.

Compound	$K_{d(app)}$ **nM**	K_d **(nM)**
2	13 ± 4	3
4	38 ± 8	10

the native br ligand in competitive fluorescence titrations (Fig. 3.10 and Table 3.4). The K_ds for compounds **2** and **4** were determined by following the decrease in fluorescence as the compounds were titrated against a fixed concentration of UnaG and br (50 nM of protein and ligand for compound **2** and 25 nM of protein and ligand for compound **4**). The titration curves (Fig. 3.10) were fitted to a simple binding isotherm, which provided

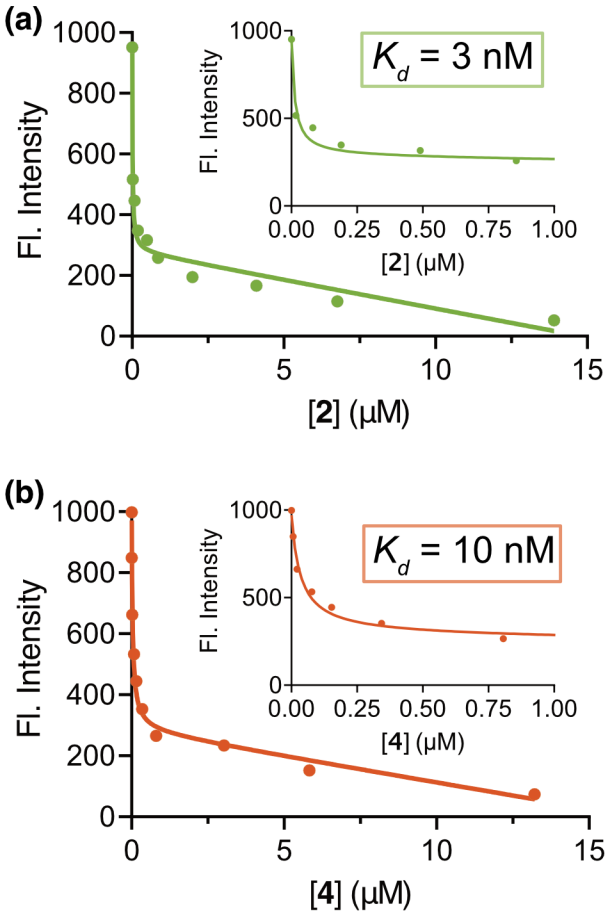


Figure 3.10: Competition titrations of (a) 2 and (b) 4 against UnaG-br. Increasing amounts of **2** and **4** were added to UnaG-br (50 nM and 25 nM for each titration, respectively) and the decreasing fluorescence emission intensity of the UnaG-br complex was recorded (495-nm excitation). A parallel set of titrations was performed using br and compounds **2** and **4** only for background subtractions. Insets: fits on a zoomed-in scale. Data shown are the mean of three technical replicates.

an apparent K_d ($K_{d(app)}$) for **2** and **4** in the presence of a fixed concentration of br. For a competitive binding model, the relationship between $K_{d,c}$ and $K_{d(app)}$ is given by equation 3.2:

$$K_{d(app} = K_{d,c}(1 + \frac{L_{br,free}}{K_{d,br}}) \tag{3.2}$$

Where $K_{d(app)}$ is the experimentally determined dissociation constant and $K_{d,c}$ is the true disassociation constant for the titrated competitor ligand; $K_{d,br}$ is the disassociation con-

stant for br binding to UnaG, which was measured as 3 nM and $L_{br,free}$ is the *free* concentration of br in solution.

To determine the free concentration of br ($L_{br,free}$), we calculated the concentration of the UnaG-br complex (EL_{br}) from the known, total UnaG concentration (E_t) and total br concentration (L_t) using the following standard relationships:

$$\mathrm{E} + \mathrm{L_{br}} \rightleftharpoons \mathrm{EL_{br}} \tag{3.3}$$

$$K_{d,br} = \frac{(E_{free})(L_{br,free})}{EL_{br}} \tag{3.4}$$

$$K_{d,br} = \frac{(E_t - EL_{br})(L_t - EL_{br})}{EL_{br}} \tag{3.5}$$

$$EL_{br} = \frac{E_t + L_t + K_{d,br} \pm \sqrt{E_t^2 - 2E_tL_t + 2E_tK_{d,br} + L_t^2 + K_{d,br}^2 + 2L_tK_{d,br}}}{2} \tag{3.6}$$

Equation 3.6 was solved to find the equilibrium concentration of the UnaG-br complex and hence the free concentration of br. This value was used in equation 3.2 to correct the apparent K_d to calculate the true K_d. We determined true K_d values of 3 nM and 10 nM for UnaG-**2** and UnaG-**4**, respectively, indicating suitably high-affinity binding of the new ligands to the protein in the same binding pocket utilized by br (Fig. 3.10). An additional direct fluorescence titration of **2** and **4** in the presence of BSA did not show any enhanced fluorescence signal (Fig. 3.11). These results demonstrate that the red-shifted and brighter fluorescence emission of UnaG-**2** and UnaG-**4** are directly due to the ligand binding UnaG rather than to any other non-specific protein intercalation.

We implemented these new ligand pairs for anaerobic imaging of commensal gut bacteria. While the UnaG-**4** fluorescence was not bright enough to be detected in cellular imaging (Fig. 3.12), UnaG-**2** fluorescence was detected in *B. theta*, both for cytosolic

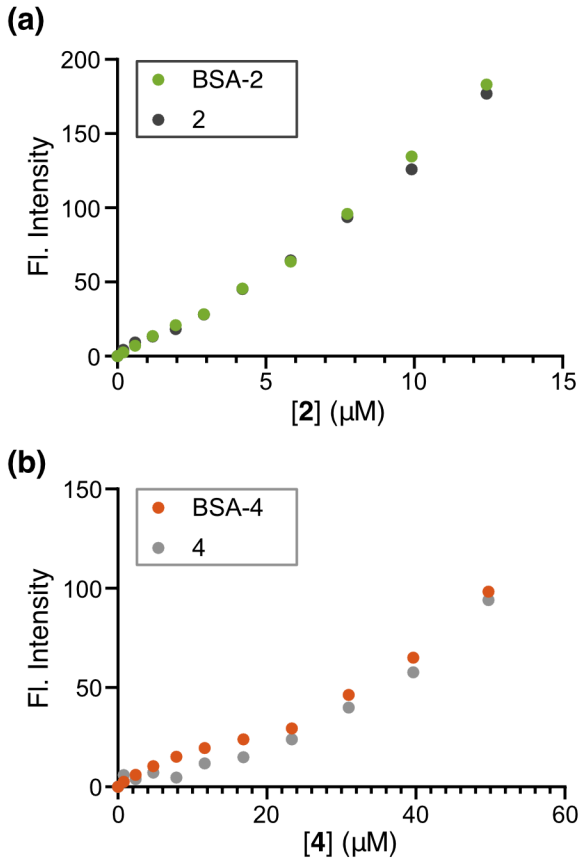


Figure 3.11: Titration of BSA with (a) **2** and (b) **4**; fluorescence emission spectra were taken at 535-nm excitation and the peak emission intensities (580 nm and 596 nm) were plotted. BSA concentrations were (a) 5 μM and (b) 15 μM.

expression and for membrane-expression of UnaG (UnaG+**2** and SusG-UnaG+**2**, respectively, in Fig. 3.13a). This fluorescence was observed under 532-nm excitation, a common green channel in microscope systems. In the absence of UnaG, **2** did not appreciably label wild-type (wt) *B. theta* cells, indicating that it can be supplemented into live-cell imaging conditions without additional washing steps. In the SusG-UnaG *B. theta* strain, the additional accessibility of an outer membrane bound UnaG increased the fluorescence signal of labeled cells.

We also demonstrated UnaG-**2** labeling of *E. coli* in anaerobic imaging conditions. UnaG-**2** labeled *E. coli* is significantly brighter than the puc19 control that did not express UnaG and washing UnaG-**2** labeled *E. coli* after labeling does not impact fluorescence sig-

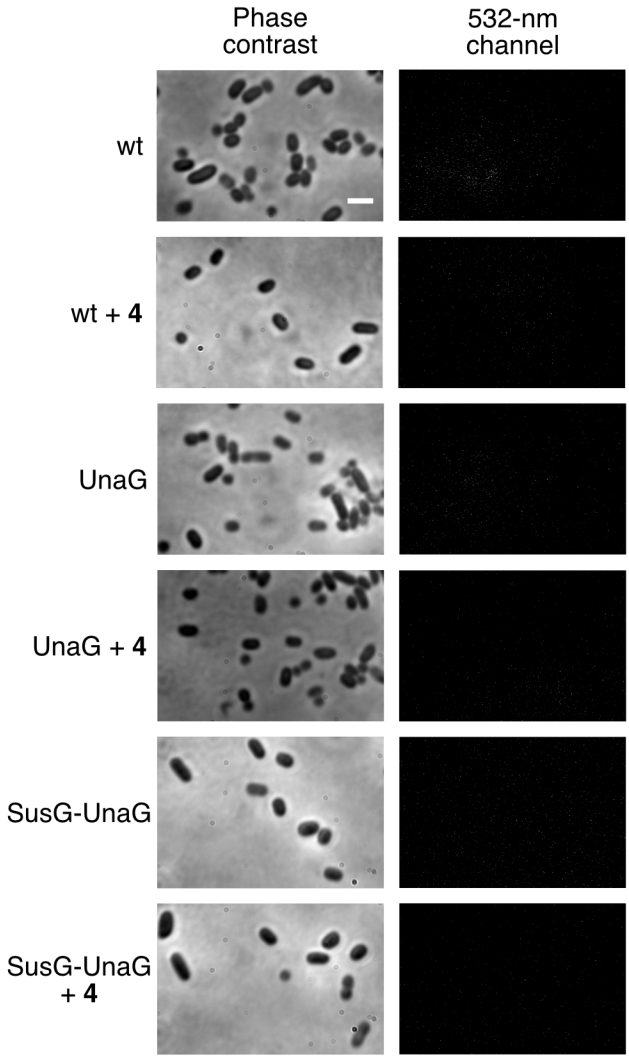


Figure 3.12: (Left) phase-contrast and (right) fluorescence images of wild-type (wt) *B. theta* and *B. theta* cells expressing UnaG or SusG-UnaG and grown to mid-log phase before incubation with **4**. Cells were illuminated with 532-nm excitation. The addition of **4** did not yield any fluorescence signal from UnaG-labeled cells. Scale bar: 2 μm.

nal (Fig. 3.13b). These imaging conditions indicate that while **2** itself is subtly fluorescent, the fluorescence intensity enhancement upon binding UnaG is sufficient such that **2** can be used similarly to a fluorogenic ligand such as br. This effective fluorogenicity enables imaging of bacterial systems without requiring additional washing steps that are required of other labeling schemes such as immunostaining or HaloTag approaches.

Just as the bluer UnaG-br and the redder BBFP IFP2.0-bv can be used as two-color imaging FP pairs [105], UnaG-**2** and IFP2.0-bv can also label different species in mixed

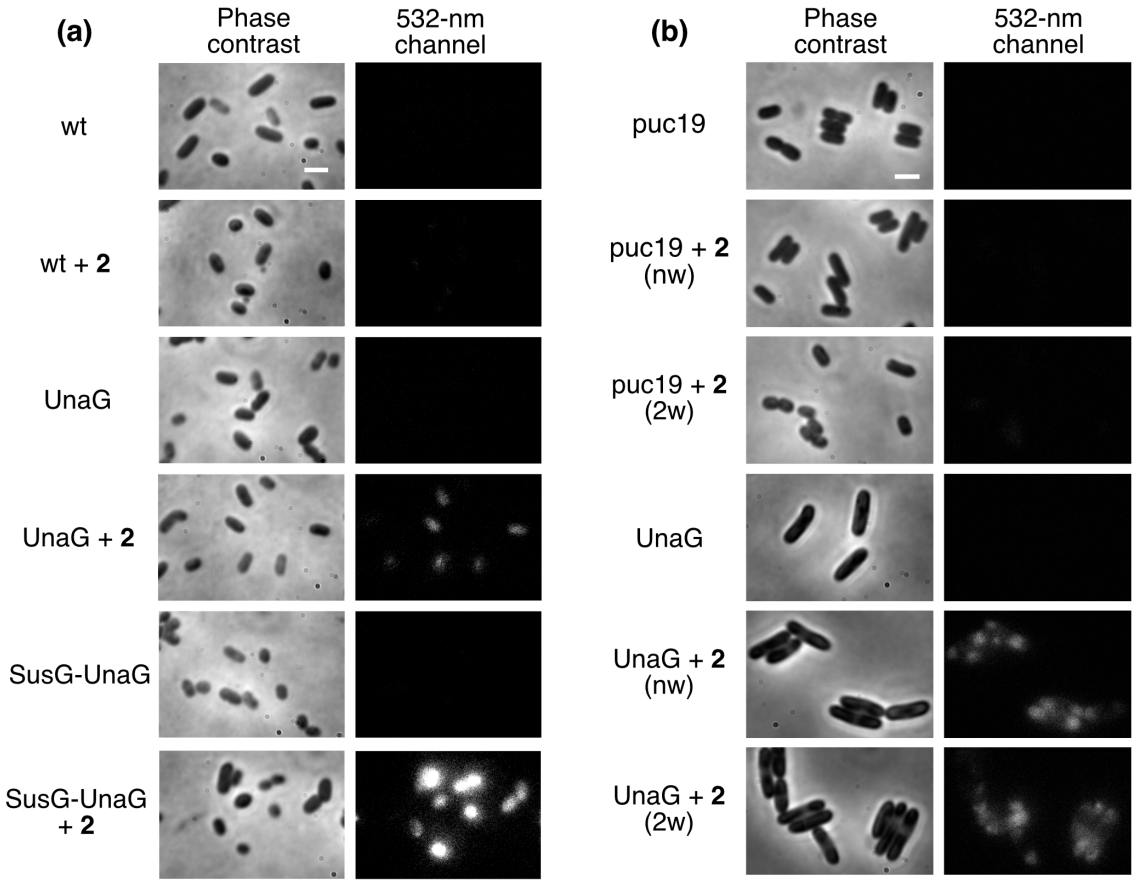


Figure 3.13: Phase-contrast and fluorescence imaging of (a) *B. theta* and (b) *E. coli* expressing UnaG. (a) *B. theta* expressing UnaG in the cytosol (UnaG) or on the outer membrane (SusG-UnaG) and incubated with **2** was compared to wild-type (wt) *B. theta* with **2**. (b) *E. coli* cells expressing UnaG and incubated with **2** were imaged without washing (nw) or after 2 washes (2w), and they were compared to the puc19 control, which did not express UnaG. Cells were imaged with 532-nm illumination. Scale bars: 2 μm; all fluorescence images are on the same brightness scale.

anaerobic bacterial cultures. We imaged a mixed culture of UnaG-labeled *B. ovatus* and IFP2.0-labeled *B. theta* cells grown in bv and incubated with **2**; the different species can be differentiated as they fluoresce in different color channels (532-nm excitation and 635-nm excitation, respectively) (Fig. 3.14). Separately, we found that including **2** or bv in the labeling media provided exclusive labeling of UnaG and IFP2.0, respectively (Fig. 3.15), demonstrating the specificity of each ligand-dependent FP for each color channel.

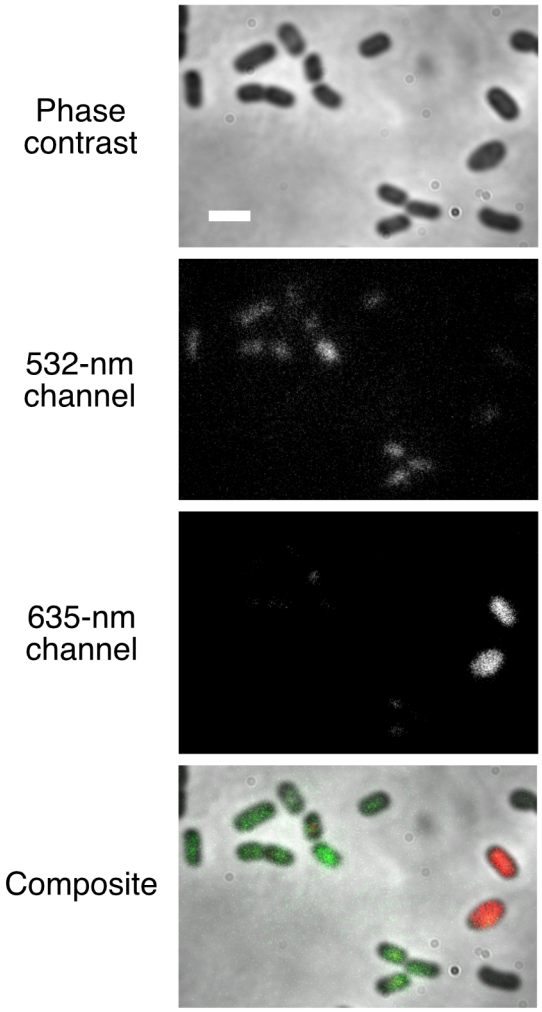


Figure 3.14: Phase-contrast and fluorescence imaging of a mixed culture of *B. ovatus* and *B. theta* expressing different BBFPs. *B. ovatus* expressing UnaG and labeled with **2** (green, 532-nm excitation) is distinguished from *B. theta* expressing IFP2.0 and labeled with bv (red, 635-nm excitation) in separate color channels. Scale bar: 2 μm.

3.4 Discussion

Here, we developed a new UnaG-ligand pair that is suitable for anaerobic fluorescence microscopy and to probe oxygen-sensitive bacterial systems. Using HTS, we identified 15 new ligands that UnaG can utilize as fluorescent binding partners (Fig. 3.1) and found one ligand (**2**) that was particularly useful for fluorescence imaging. UnaG binds **2** with high affinity (K_d 3 nM, Fig. 3.10) resulting in a 10-nm red shift of the peak fluorescence emission wavelength (to 581 nm) as well as a 2.5-fold fluorescence intensity enhancement over the

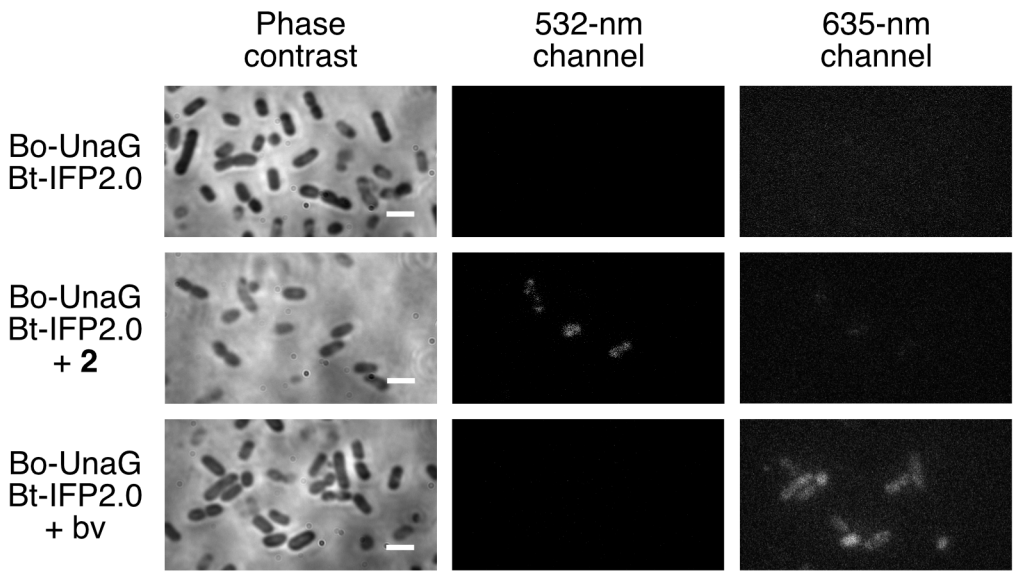


Figure 3.15: Phase-contrast and fluorescence images of *B. ovatus* (*Bo*) expressing UnaG and *B. theta* (*Bt*) expressing IFP2.0. When **2** is supplemented into media for labeling or when cells are grown in bv, UnaG-labeled *B. ovatus* can be distinguished from IFP2.0-labeled *B. theta* in separate color channels using 532-nm and 635-nm excitation, respectively. Scale bar: 2 μm.

signal produced by **2** alone (Fig. 3.8).

Due to the nature of the compound library chosen for HTS, the confirmed hits found through our screening workflow are structurally diverse (Fig. 3.4) yet dissimilar to classic families of organic fluorescent dyes like the xanthenes (fluorescein, rhodamine) and coumarins. Unsurprisingly, almost all the hit compounds were extensively conjugated heterocyclic molecules, although most did not fluoresce significantly to carry on in spectral characterization during validation steps. We hypothesize that the high concentrations of compounds utilized in HTS contributed to the discrepancy: fluorescence was detected in HTS but not detected during the validation experiments at the lower concentrations that were closer to experimental conditions suitable for imaging. Of note, the two best hits **2** and **4**, are derivatives of benzothiazole, which is a common heterocycle in medicinal chemistry and in studies of bioactive molecules [114], but not a common structural feature in popular commercial fluorescent dyes. To our knowledge, this work describes the first application of a benzothiazole-based ligand in bacterial imaging and in oxygen-independent

microscopy.

Like the previously characterized BBFP pairs [105], UnaG-br and IFP2.0-bv, UnaG-**2** is a ligand-dependent FP reporter for imaging obligate anaerobic bacteria (Fig. 3.13). Importantly, UnaG-**2** diversifies the colors of the available UnaG-ligand pairs by adding a probe that fluoresces in the green 532-nm excitation channel, commonly used for Cy3 and Red Fluorescent Protein (RFP) imaging. In imaging *B. theta*, we observed UnaG-**2** fluorescence for both cytoplasmic and outer-membrane labeling conditions, though the latter positioning increases the accessibility to yield brighter cells. Relative to UnaG-br, the red-shifted UnaG-**2** fluorescence will enable better signal-to-background measurements as intrinsic cellular background decreases with increasing excitation wavelength. Anaerobic bacterial systems with high cellular background in the blue region are not suitable for labeling by the UnaG-br pair but could potentially be probed by the red-shifted UnaG-**2** labeling system. Likewise, UnaG-**2** can be coupled with IFP2.0-bv as complementary labeling pairs in two-color fluorescence microscopy using the 532-nm and 635-nm excitation channels, respectively (Fig. 3.14). As we have also demonstrated demonstrated the utility of UnaG-**2** in *E. coli* in anaerobic conditions (Fig. 3.13b), we envision that studies of oxygen-sensitive mechanisms in this more common laboratory microbe will be further accessible via fluorescence-based investigations.

Since the UnaG-**2** labeling system utilizes the wild-type UnaG protein, this system cannot be coupled with the native UnaG-br pair (blue, 488-nm excitation) for two-color imaging. However, our HTS methodology and hits reveal that the UnaG binding pocket is not exclusive to the original br ligand and can indeed bind other potentially fluorogenic molecules. Additional HTS screening may discover further molecules that bind UnaG and fluoresce in other microscopy channels, further diversifying UnaG into a wide palette of fluorescent colors. For more extensive multi-color setups, protein engineering will also be required to create exclusive pairs of UnaG variants-ligands such that each UnaG variant can recognize only one ligand.

3.5 Conclusions

We discovered and characterized a new ligand that the ligand-dependent FP UnaG can utilize for anaerobic live-cell fluorescence microscopy. Like the native blue-green UnaG-br pair, the fluorescence of the new green UnaG-**2** pair produces red-shifted and intensity enhanced fluorescence that is detectable without a wash step. The UnaG-**2** pair was used in conjunction with the red IFP2.0-bv pair as complementary probes to label strains in mixed bacterial cultures with two-color imaging. This work adds another color to the ligand-dependent fluorescent reporter toolbox for imaging and other fluorescence-based analyses that can be used to probe oxygen-sensitive biological systems.

CHAPTER IV

Designing a Red-Shifted Fluorescent Reporter Using Directed Evolution and Rational Design

4.1 Introduction

As previously discussed, Bilin-based Fluorescent Proteins (BBFPs) are ideal tools for probing anaerobic microbial systems. In addition to their oxygen-independent fluorescence, BBFPs bind fluorogenic ligands that are non-toxic to bacteria and do not increase background signal during imaging. Of the BBFPs that have been reported, the bilirubin (br)-binding fluorescent protein (FP) UnaG [60] is particularly promising to further develop as a probe.

As demonstrated in Chapter III, the binding pocket of UnaG permits the binding of a wide range of ligands and extends the possibility of utilizing new fluorogenic molecules for imaging in more color channels. UnaG is also a more attractive probe than biliverdin (bv)-binding FPs such as IFP2.0 [66], mIFP [67], Sandercyanin [115], and smURFP [112]; UnaG is small and natively monomeric [61], unlike most bv-FPs that are di- or tetrameric and/or prone to oligomerization.

However, bv-binding BBFPs hold advantages over UnaG due to their far-red to near-Infrared Radiation (IR) fluorescence. Far-red emission is desirable over the blue-green emission exhibited by the original UnaG-br pair because red excitation and emission is

spectrally far away from intrinsic cellular background fluorescence. Molecularly, bv and br differ only by a double bond, but this double bond confers elongated π-conjugation for red-shifted fluorescence and a more planar structure. This additional double bond gives rise to two absorbance peaks that can be utilized by the far-red BBFPs that can absorb and emit red light or by the Sandercyanin FP that can absorb blue light and emits red light. Ideally, a red BBFP would have both bv recognition as well as the small monomeric form of UnaG to enable maximal flexibility and implementation in bacterial systems with minimal biological perturbation.

In this Chapter, I discuss efforts to engineer a red-shifted UnaG by altering the UnaG binding pocket to switch ligand recognition from br to bv. Based on the crystal structure of the protein, I chose to mutate residues in the UnaG binding pocket that appeared to interact with the hydrogen bonding network or with br itself. Given the challenge of altering ligand recognition through a small number of selected rational design mutations, I also employed a randomized directed evolution approach to generate red-shifted UnaG variants with bv recognition that could fluoresce upon blue or red excitation. From our two-pronged protein engineering approach, the preliminary work discussed in this Chapter provides additional insight into the UnaG binding pocket.

4.2 Materials and methods

4.2.1 Cloning and Library Construction

Oligonucleotides and plasmids used in cloning are described in Table 4.1 and 4.2, respectively. PCR products and plasmids were purified using commercially available kits (Qiagen and Zymo Research). Plasmids were assembled using commercial HiFi Gibson gene assembly kits (New England Biolabs). DNA sequencing (Eurofins) was used to confirm assembled genes in plasmids. Sequences were aligned using CLC Sequence Viewer (Qiagen).

Table 4.1: Oligonucleotides used in this study.

Name	**Sequence**	**Use**
UnaG-SDM-V2L_f	GGGCATTTCACATATGATGTTGGAAAAGTTCGTTGGTAC	Install V2L into UnaG
UnaG-SDM-V2L_r	GTACCAACGAACTTTTCCAACATCATATGTGAAATGCCC	Install V2L into UnaG
UnaG-SDM-M51C_f	CAGCCAGAAAGATGGTGACAAGTGCACCGTGAAAATTGAAAATGGTC	Install M51C into UnaG
UnaG-SDM-M51C_r	GACCATTTTCAATTTTCACGGTGCACTTGTCACCATCTTTCTGGCTG	Install M51C into UnaG
UnaG-SDM-M51K_f	CCAGAAAGATGGTGACAAGAAGACCGTGAAAATTGAAAATG	Install M51K into UnaG
UnaG-SDM-M51K_r	CATTTTCAATTTTCACGGTCTTCTTGTCACCATCTTTCTGG	Install M51K into UnaG
UnaG-SDM-N57A_f	TGACAAGATGACCGTGAAAATTGAAGCTGGTCCGCCGACC	Install N57A into UnaG
UnaG-SDM-N57A_r	GGTCGGCGGACCAGCTTCAATTTTCACGGTCATCTTGTCA	Install N57A into UnaG
UnaG-SDM-N57H_f	CAAGATGACCGTGAAAATTGAACATGGTCCGCCGA	Install N57H into UnaG
UnaG-SDM-N57H_r	TCGGCGGACCATGTTCAATTTTCACGGTCATCTTG	Install N57H into UnaG
UnaG-SDM-D81H_f	CGATGAATTTCCGAGCCACCGTCGTAAAGGTGT	Install D81H into UnaG
UnaG-SDM-D81H_r	ACACCTTTACGACGGTGGCTCGGAAATTCATCG	Install D81H into UnaG
UnaG-SDM-R112K_f	CGGCAAAGAAACCACCTACGTGAAGGAAATCAAAGATGGTAAACTGG	Install R112K into UnaG

Continuation of Table 4.1

UnaG-SDM-R112K_r	CCAGTTTACCATCTTTGATTTCCTTCACGTAGGTGGTTTCTTTGCCG	Install R112K into UnaG
UnaG-SDM-R112W_f	GGCAAAGAAACCACCTACGTGTGGGAAATCAAAGATGGTAAACTG	Install R112W into UnaG
UnaG-SDM-R112W_r	CAGTTTACCATCTTTGATTTCCCACACGTAGGTGGTTTCTTTGCC	Install R112W into UnaG
UnaG-SDM-R132K_f	TGGGCGACGTGGTTGCAGTTAAGAGCTATCGTCGTGCG	Install R132K into UnaG
UnaG-SDM-R132K_r	CGCACGACGATAGCTCTTAACTGCAACCACGTCGCCCA	Install R132K into UnaG
UnaG-SDM-R132W_f	GGCGACGTGGTTGCAGTTTGGAGCTATCGTCGTG	Install R132W into UnaG
UnaG-SDM-R132W_r	CACGACGATAGCTCCAAACTGCAACCACGTCGCC	Install R132W into UnaG
pMAL-c5x-seq_f	GGTCGTCAGACTGTCGATGAAG	For sequencing pMAL-c5x
pMAL-c5x-seq_r	TGTCCTACTCAGGAGAGCGTTCAC	For sequencing pMAL-c5x
pRSFDuet_UnaG_f	ACCATCATCACCACAGCCAGATGGTGGAAAAGTTCGTTGG	Install UnaG into pRSFDuet-1 (MCS #1)
pRSFDuet_UnaG_r	TTAAGCATTATGCGGCCGCACTATTCGGTCGCACGACG	Install UnaG into pRSFDuet-1 (MCS #1)
pRSFDuet_EGFP_f	GTATAAGAAGGAGATATACAATGGTGAGCAAGGGCGAG	Install EGFP into pRSFDuet-1 (MCS #2)
pRSFDuet_EGFP_r	CAGCGGTTTCTTTACCAGACTTACTTGTACAGCTCGTCCATG	Install EGFP into pRSFDuet-1 (MCS #2)

Continuation of Table 4.1

pRSFDuet-mcs1_f	TGCGGCCGCATAATGCTTAAG	For amplifying pRSFDuet-1 from MCS #1 for Gibson Assembly
pRSFDuet-mcs1_r	CTGGCTGTGGTGATGATGGT	For amplifying pRSFDuet-1 from MCS #1 for Gibson Assembly
pRSFDuet_mtUnaG_f	ACCATCATCACCACAGCCAGATG	Used in error-prone PCR to generate UnaG variant library; UnaG into pRSFDuet-1 (MCS #1)
pRSFDuet_mtUnaG_f	TTAAGCATTATGCGGCCGCACTA	Used in error-prone PCR to generate UnaG variant library; UnaG into pRSFDuet-1 (MCS #1)
ACYCDuet Up1_f	GGATCTCGACGCTCTCCCT	For sequencing pRSFDuet-1 MCS #1
DuetDOWN1_r	GATTATGCGGCCGTGTACAA	For sequencing pRSFDuet-1 MCS #1
DuetUP2_f	TTGTACACGGCCGCATAATC	For sequencing pRSFDuet-1 MCS #2
DuetDOWN2_r	GCTAGTTATTGCTCAGCGG	For sequencing pRSFDuet-1 MCS #2

Table 4.2: Plasmids ordered and generated for this study.

Plasmid	Description	Resistance	Reference
pMAL-c5x_UnaG	For general cloning and protein expression; results in the production of codon-optimized UnaG with a maltose-binding-protein (MBP) tag on the N-terminus and a 6x-His tag on the C-terminus	Amp	Genscript [105]
pMAL-c5x-TEV-UnaG	For general protein expression; installs a TEV protease cleavage site between the Factor Xa and UnaG protein	Amp	
pMAL-c5x-TEV-UnaG_V2L	Produces MBP-UnaG V2L variant	Amp	
pMAL-c5x-TEV-UnaG_V2L/M51C	Produces MBP-UnaG V2L/M51C variant	Amp	
pMAL-c5x-TEV-UnaG_M51C	Produces MBP-UnaG M51C variant	Amp	
pMAL-c5x-TEV-UnaG_M51K	Produces MBP-UnaG M51K variant	Amp	
pMAL-c5x-TEV-UnaG_N57A	Produces MBP-UnaG N57A variant	Amp	
pMAL-c5x-TEV-UnaG_N57H	Produces MBP-UnaG N57H variant	Amp	
pMAL-c5x-TEV-UnaG_D81H	Produces MBP-UnaG D81H variant	Amp	
pMAL-c5x-TEV-UnaG_R112K	Produces MBP-UnaG R112K variant	Amp	
pMAL-c5x-TEV-UnaG_R112W	Produces MBP-UnaG R112W variant	Amp	
pMAL-c5x-TEV-UnaG_R132K	Produces MBP-UnaG R132K variant	Amp	
pMAL-c5x-TEV-UnaG_R132W	Produces MBP-UnaG R132W variant	Amp	
pBAD_EGFP	For general cloning; resulting EGFP protein used for FACS control	Amp	Addgene #54762
pRSFDuet-1	For general cloning and FACS control	Kan	Novagen / EMD Millipore
pRSFDuet-1 _UnaG_EGFP	For directed evolution using FACS; results in expression of UnaG with a 6x-His tag on the N-terminus and EGFP with an S-tag on the C-terminus; generated by Gibson Assembly (UnaG inserted between BamHII and HindIII sites in MCS #1 and EGFP inserted between NdeI and XhoI sites in MCS #2)	Kan	

Single and double mutants were generated through site-directed mutagenesis (SDM) using the commercial QuikChange SDM kit (Agilent). All single mutation variants were built into a pMAL-c5x-TEV_UnaG plasmid modified from one I have previously described [105], which contains an *E. coli* codon-optimized form of the *UnaG* gene. A TEV protease cleavage site was inserted in-frame between the Factor Xa protease site and the *UnaG* gene to cleave the maltose-binding-protein (MBP) tag from UnaG if desired. Clones were transformed into chemically competent XL10-gold *E. coli* cells (Agilent) for plasmid storage.

All constructs used for directed evolution were built onto a pRSFDuet-1 vector (Novagen, EMD Millipore). First, EGFP was inserted into the second multiple cloning site (MCS) in pRSFDuet-1 and confirmed by sequencing. Error Prone Polymerase Chain Reaction (ep-PCR) was then used to introduce mutations to the parent UnaG gene. The following protocol was adapted from previous literature [116] and added to the reaction mixture sequentially (per 100 μL for 1× concentration): Tris-HCl (pH 8.3, 10 mM), KCl (50 mM), $MgCl_2$ (7 mM), dCTP (1 mM), dTTP (1mM), dATP (0.2 mM), dGTP (0.2 mM), forward primer (2 μM), reverse primer (2 μM), DNA template (20 pg), $MnCl_2$ (0.5 mM), and Taq polymerase (0.05 Units, New England Biolabs). The PCR conditions were 95°C 3 min, 12 cycles of 95°C 30 sec / 60°C 30 sec / 72°C 3 min, and finished with 72°C extension for 5 min. The resulting insert was purified before insertion into the first MCS in the sequence confirmed pRSFDuet_EGFP plasmid. The assembly reaction was transformed into electrocompetent MegaX DH10B *E. coli* cells (Invitrogen). A small volume of transformed cells was plated onto LB/Agar plates (50 μg/mL kanamycin) to calculate transformation efficiency; the rest of the transformed cells were used to inoculate fresh LB media (50 μg/mL kanamycin) and grown overnight to harvest the plasmid library.

4.2.2 Colony-based Screening for UnaG Variants

pMAL-c5x_UnaG plasmid was transformed into chemically competent BL21 DE3 *E. coli* cells (New England Biolabs). Standard LB/Agar plates were prepared with ampicillin (100 μg/mL) and IPTG (100 mM) and varying amounts of br were spread over the plate surface (10-40 μL of 100 μM stock). M9/Agar plates were prepared by autoclaving a 95 mL base of M9 salts (20× 1 liter stock: 33.9 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 5 g NH_4Cl) with agar (0.75 g for soft plates or 0.3 g for extra-soft plates), followed by addition of 0.2 mL 1 M $MgSO_4$, 20 μL 1 M $CaCl_2$, 0.8 mL 20% glucose, 50 μL 1% thiamine, and IPTG (final concentration 100-400 mM). Once dried, br was spread over the plate surface (10-40 μL of 100 μM stock). Cells were streaked over the plates in the classic three streak dilution method to isolate single colonies. Following overnight growth at 37°C, colonies on plates were visualized in a BioRad ChemiDoc Imager using AlexaFluor488 excitation and filters.

4.2.3 Lysate-based Screening for UnaG Variants

pMAL-c5x-TEV_UnaG plasmid was transformed into chemically competent NEB-Express *E. coli* cells (New England Biolabs) using standard transformation protocols. For minimal media conditions, a colony was picked into standard M9 media (2% glucose as carbon source) and grown in a shaker at 20°C for 24 hours. For LB media conditions, a colony was picked into LB media to grow overnight at 37°C and subsequently backdiluted 1:200 into fresh LB media; these cells were induced when OD_{600} reached 0.4 with IPTG (0.4 mM final concentration) to grow at 37°C for another 3 hours before lysing. To normalize cell quantity, cultures from both M9 and LB conditions were diluted to the same OD_{600} before lysing.

Lysozyme (50 mg/mL, Sigma Aldrich) and DNase (200 mg/mL, Sigma Aldrich) were prepared in PBS and added to a buffer consisting of 20 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Tween-20 (Sigma Aldrich) and TritonX-100 (Sigma Aldrich) were added to lysis buffer or to B-PER bacterial extraction reagent (ThermoFisher) as indicated. Cultures were

mixed in equal volume with lysis buffer and incubated at room temperature for 15 min. Half of this suspension was visualized directly in a UV-light box and the other half was subjected to an additional incubation at 50°C for 5 min before visualization. Br (20 μM final concentration) was added to tubes before visualization.

4.2.4 FACS Screening for Red-Shifted UnaG Variants

Plasmid libraries were transformed into chemically competent NEB-Express *E. coli* cells (New England Biolabs). Following recovery, the transformed cells were used to inoculate fresh LB media (50 μg/mL kanamycin) and grown overnight. The overnight culture was used to inoculate 20 mL of rich autoinduction media (per liter: 6 g Na_2HPO_4, 3 g KH_2PO_4, 20 g tryptone, 5 g yeast extract, 5 g NaCl, 10 mL 60% v/v glycerol, 5 mL 10% w/v glucose, 8% w/v lactose; 50 μg/mL kanamycin) and grown at 25°C for 48 hrs.

Fluorescence Assisted Cell Sorting (FACS) was performed on live cells expressing mutant UnaG and EGFP to collect red-shifted UnaG variants. Cells were diluted and washed in 1× PBS to 8×10^6 – 2.4×10^7 cells/mL, incubated with bv (1 μM) for 30 min on ice, followed by DAPI stain (10 μg/mL) for another 30 min on ice. Cells were washed twice with ice-cold PBS to remove excess ligand and stain. Cells with no bv incubation as well as cells transformed with empty pRSFDuet-1 vector were prepared with the same protocol as negative controls. Only cells excited with red-shifted fluorescence in the Brilliant Violet 605 (405-nm excitation, filter 617/30 nm) or the APC-Cy5.5 (640-nm excitation, filter 720/60 nm) channels were collected. Collected cells were grown up in fresh LB media (50 μg/mL kanamycin) to repeat more rounds of mutagenesis and cell sorting.

As the collected cells from FACS sorting was heterogeneous, variants were isolated as single colonies by plating dilutions of the overnight sorted culture onto LB/Agar plates (50 μg/mL kanamycin). These colonies were picked into LB media (50 μg/mL kanamycin) in 24-well plate format to use for saving glycerol freezer stocks and for sequencing.

4.2.5 Protein Expression and Purification

All UnaG variants, MBP-UnaG mutants and individual variants from FACS-based sorting, were expressed in NEB-Express *E. coli* cells (New England Biolabs) with IPTG induction (0.4 mM final concentration) upon OD_{600} 0.4-0.6 and harvested after 3.5 hrs outgrowth at 37°C. Cell pellets were lysed by sonication or microfludizer and after centrifugation, the collected supernatant was collected and loaded onto Ni-NTA resin (New England Biolabs). Protein was purified by gravity flow using a step gradient of imidazole (base buffer: 20 mM Tris-HCl, 200 mM NaCl): 10 mM imidazole (also used for equilibrating supernatant and resin), 25 mM, 100 mM, 200 mM, and 500 mM. Fractions were analyzed for size and purity by SDS-PAGE (4-20% acrylamide, BioRad). Final elution fractions were passively concentrated (10 kDa cutoff, Sartorius Vivapore) or by centrifugation (10 kDa cutoff, Millipore-Sigma Amicon). Concentrated protein was buffer exchanged into imidazole free buffer and the final concentration was assessed by A_{280} measurements.

4.2.6 Spectral Characterization

Br (100 mM, Sigma Aldrich) and bv (1 mM, Sigma Aldrich) stocks were dissolved in DMSO. All compounds and proteins were prepared for measurement in the same buffer: 20 mM Tris-HCl (pH 7.2), 200 mM NaCl, and 1 mM EDTA.

UV-vis absorbance measurements were performed in 96-well clear bottom plates using the Molecular Devices SpectraMax iD3 microplate reader or in a black-walled quartz cuvette (Hellma, 1 cm pathlength) using a Hewlett Packard 8453 UV-Visible spectrophotometer. Plotted UV-vis spectra are background corrected using buffer blanks and processed using MATLAB.

Fluorescence measurements were carried out on an Agilent Varian Cary Eclipse Fluorescence Spectrophotometer. Plotted fluorescence excitation and emission spectra are averaged plots from three technical replicates and processed using MATLAB.

4.3 Results

4.3.1 Rational design of a red-shifted bv-binding UnaG

As bv only differs from br by a single double bond (Fig. 4.1a), I hypothesized that rational design of a bv-binding UnaG variant would rely upon altering the hydrogen bonding network within the original binding pocket (Fig. 4.1b). Utilizing structural information, I chose to selectively mutate residues that are involved in direct hydrogen bonding with br or involved in coordinating water molecules that help rigidify the fluorogenic ligand within the hydrophobic protein core. For example, Asn57 is positioned at the midpoint of the br molecule between the two planar dipyrrole sections and has extensive hydrogen bond interactions with the *endoo*-vinyl pyrrinone as well as the C-ring propionate; the N57A point mutation abolishes all of these hydrogen bond interactions [60] and opens up the binding pocket.

The mutations I chose to introduce were ones that would alter the sterics of the binding pocket to accommodate the entirely planar bv ligand (N57A, M51C, R112K, R132K), alter the electrostatics of the binding pocket (M51K, N57H, D81H), or affect potential π-stacking between residues and the ligand (R112W, R132W). I also chose to include the previously published single V2L mutation from eUnaG [80], a variant that was evolved for increased br-induced fluorescence, to investigate the effects of this mutation on br and bv binding; although Val2 does not appear to directly interact with br in the crystal structure, the V2L mutation may be stabilizing other interactions within the binding pocket.

I expressed, purified, and characterized all single mutants bound with br and bv by UV-vis absorbance (Fig. 4.2) and fluorescence emission (Fig. 4.3). The M51K and R132K single mutants did not express well and could not be purified, indicating that these residue changes introduced structural instability to the protein. Most variants did not enhance the primary br absorbance peak at 495 nm over Wt-br with the exception of the M51C variant (Fig. 4.2b-c). In addition to enhanced absorbance, the M51C-br pair was also slightly more

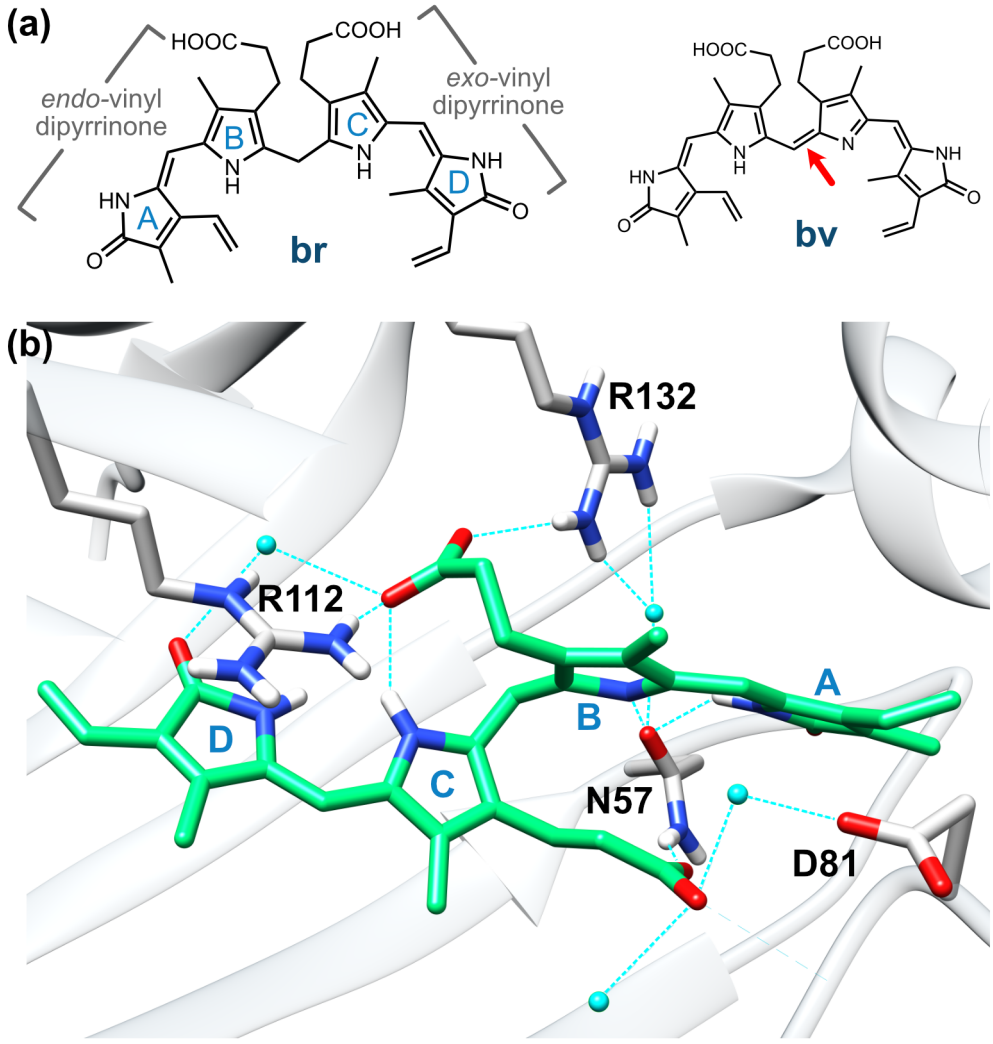


Figure 4.1: (a) Structures of br and bv. Red arrow indicates the additional double bond in bv. (b) Detailed view of the UnaG binding pocket with residues critical for br binding and interaction. Water (cyan molecules) and hydrogen bonds (dashed cyan lines) are critical for coordinating and stabilizing the br ligand (green). PDB: 4I3B.

fluorescent upon excitation at 495 nm than Wt-br (Fig. 4.3a-b). Surprisingly, D81H-br was also slightly more fluorescent than Wt-br with 495-nm excitation (Fig. 4.3a and e) despite having lower absorbance at the corresponding peak (Fig. 4.2b and d). The rest of the introduced mutations lowered the fluorescence emission intensity when bound with br and excited at 495 nm.

Similarly, most variants did not enhance the two primary bv absorbance peaks at 345 nm (blue) and 650 nm (red) over Wt-bv, with the exceptions of M51C and N57A (Fig.

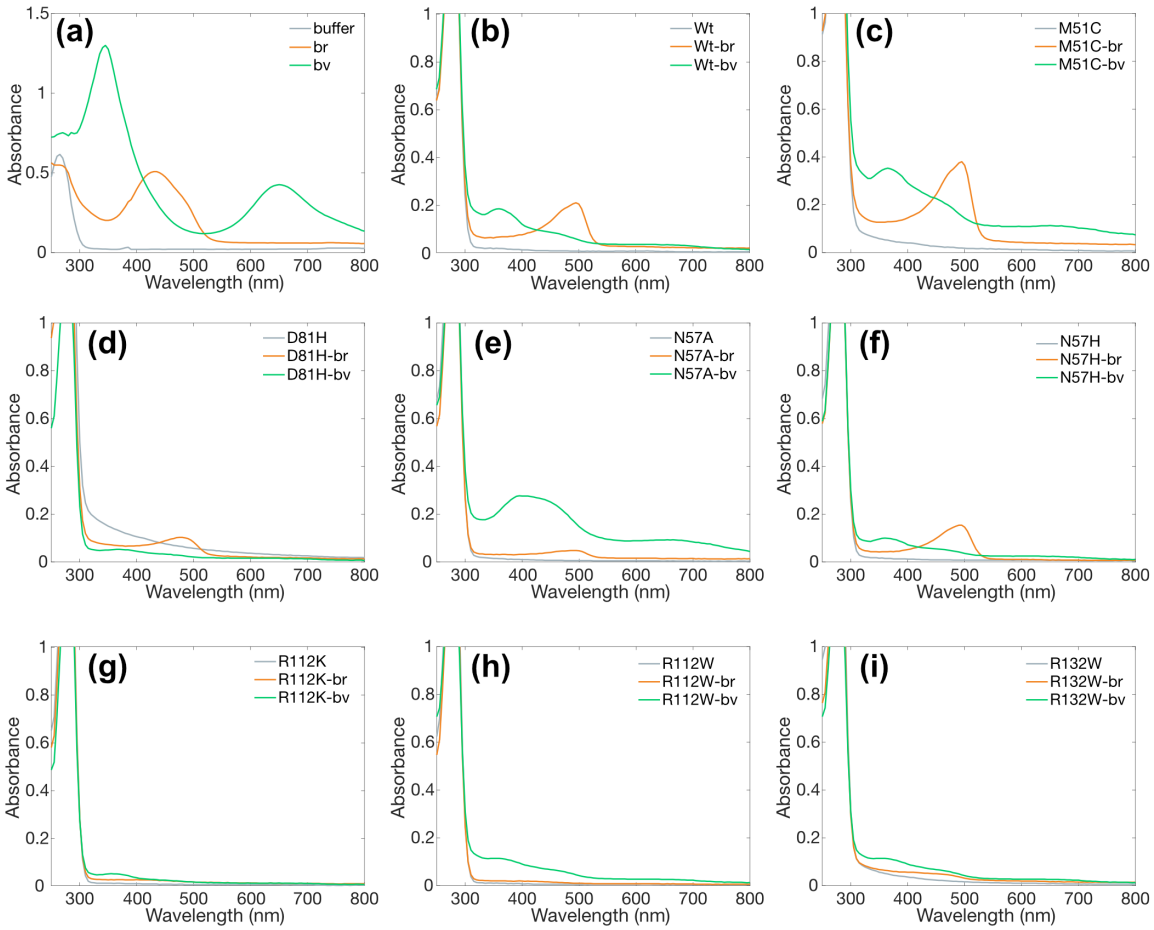


Figure 4.2: UV-vis absorbance spectra of (a) br (orange) and bv (green) ligands (50 μM), (b) Wt UnaG with br and bv, and (c-i) single mutant variants with br and bv. (b-i) All protein and ligand concentrations were measured at 30 μM.

4.2b, c, and e). For M51C-bv and N57A-bv, both the dominant blue absorbance peak and secondary red absorbance peak were broader relative to the bv ligand alone (Fig. 4.2a). When compared to other mutants that had no discernible absorbance when incubated with bv, the broader peaks observed with M51C-bv and N57A-bv indicates these two mutations confer some bv-binding capacity. However, neither mutation was able to confer bv-induced fluorescence signal at any excitation wavelength (Fig. 4.4).

As the M51C variant complexed with br appeared to impart the greatest absorbance enhancement and fluorescence emission increase, I installed the M51C mutation onto the eUnaG variant containing V2L. I observed that the absorbance of the V2L/M51C double

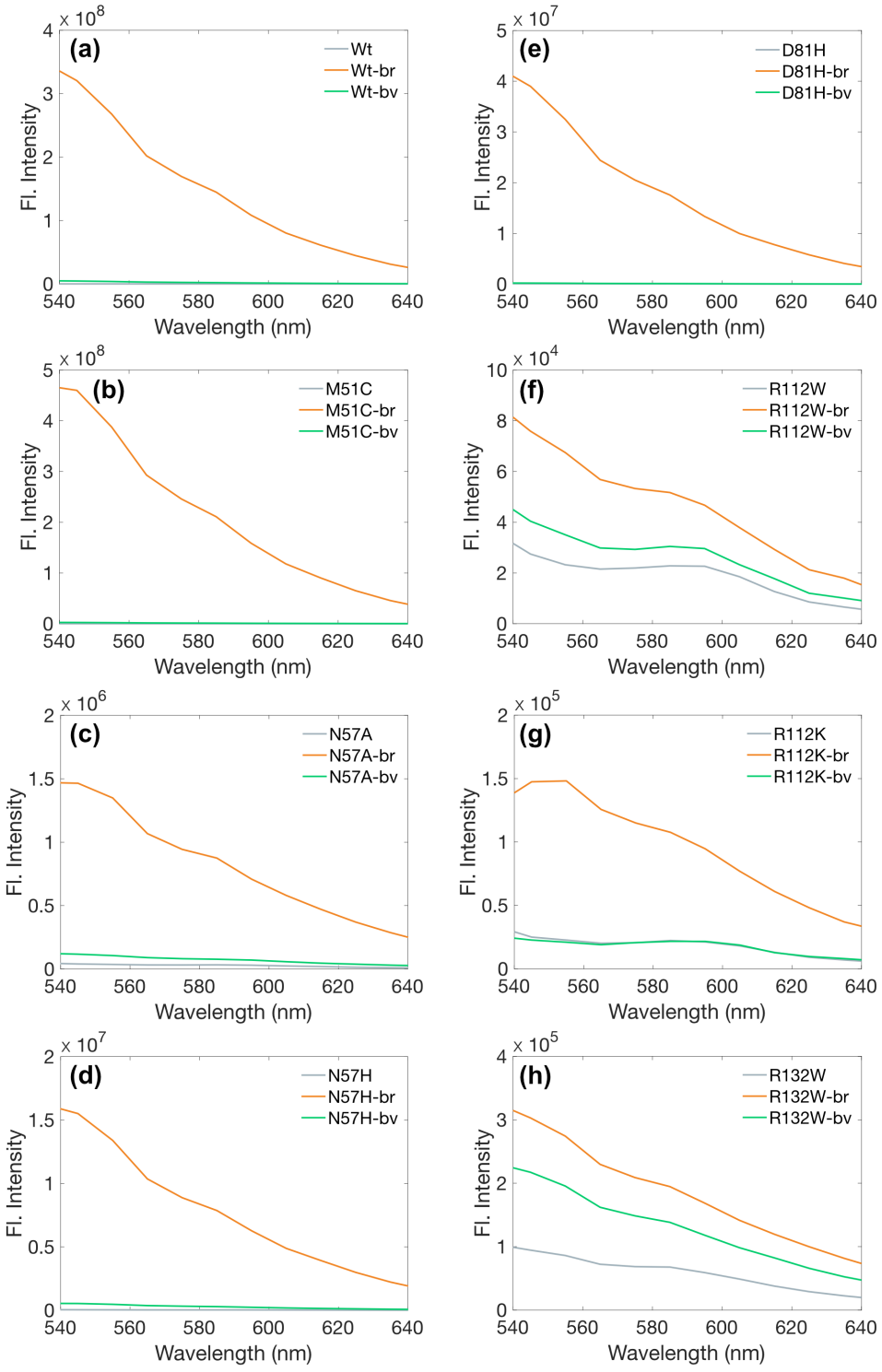


Figure 4.3: Fluorescence emission spectra of (a) Wt and (b-h) single mutant variants bound with br (orange) or bv (green) and excited at 495 nm. All protein and ligand concentrations were measured at 30 μM.

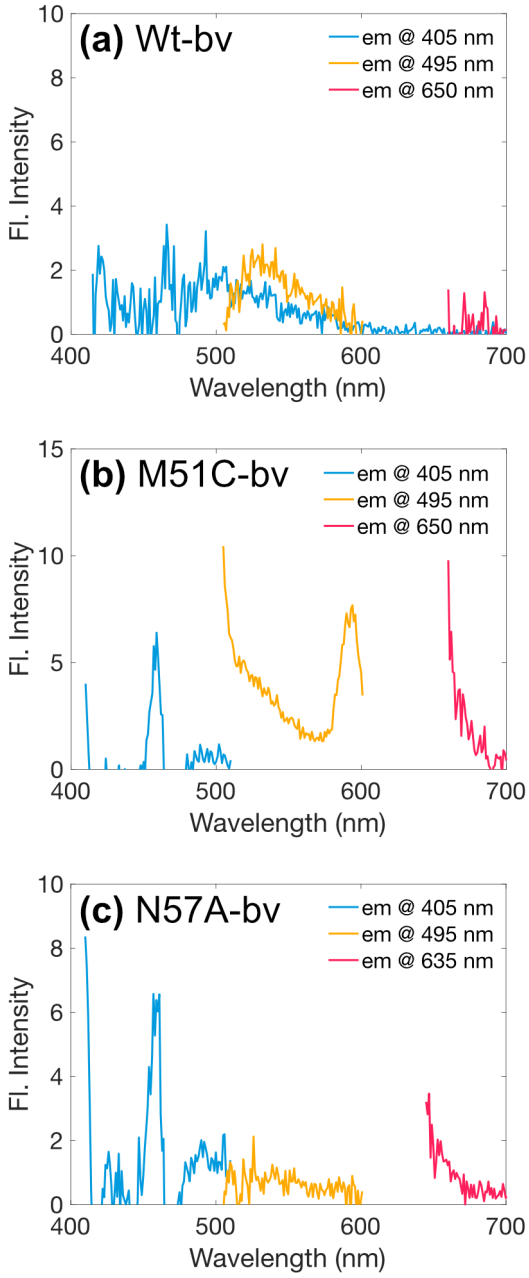


Figure 4.4: Fluorescence emission spectra of (a) Wt UnaG, (b) M51C, and (c) N57A bound with bv at excitation at 405 nm (blue), 495 (yellow) and 635 or 650 nm (red).

mutant bound with br had higher absorbance at 495 nm than Wt-br but lower than V2L-br (Fig. 4.5a). Contrary to published literature, I observed that the V2L-br pair had the same fluorescence emission intensity as the native Wt-br pair at 495-nm excitation (Fig. 4.5b-c). Furthermore, the addition of M51C to V2L decreased the fluorescence intensity

from the V2L single mutant (Fig. 4.5d). A summary of designed rational mutants' spectral characterization with respect to each residue's structural placement within the protein is detailed in Figure 4.6.

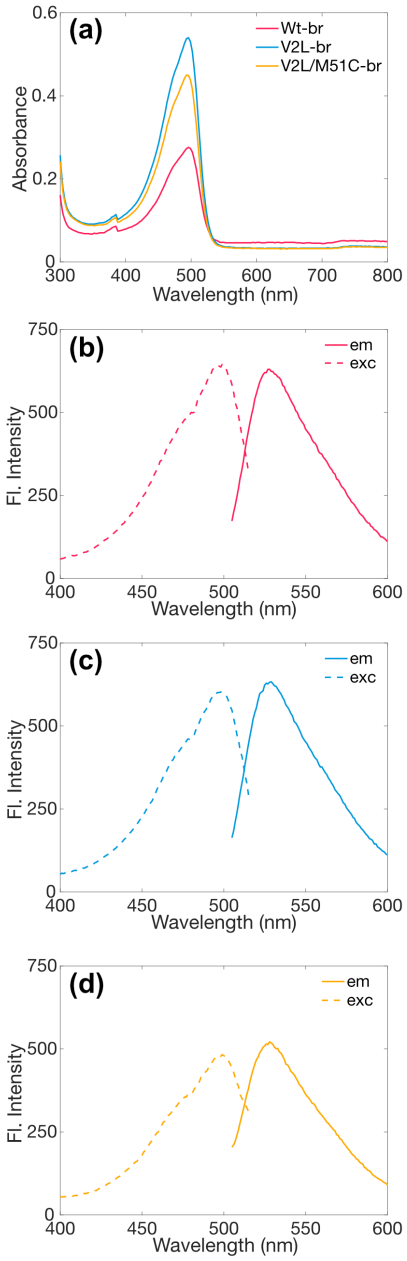


Figure 4.5: (a) UV-vis absorbance and fluorescence emission spectra of (b) Wt UnaG, (c) V2L, and (d) V2L/M51C bound with br. Protein and ligand were incubated in equimolar amounts at (a) 30 μM and (b-d) 9 nM.

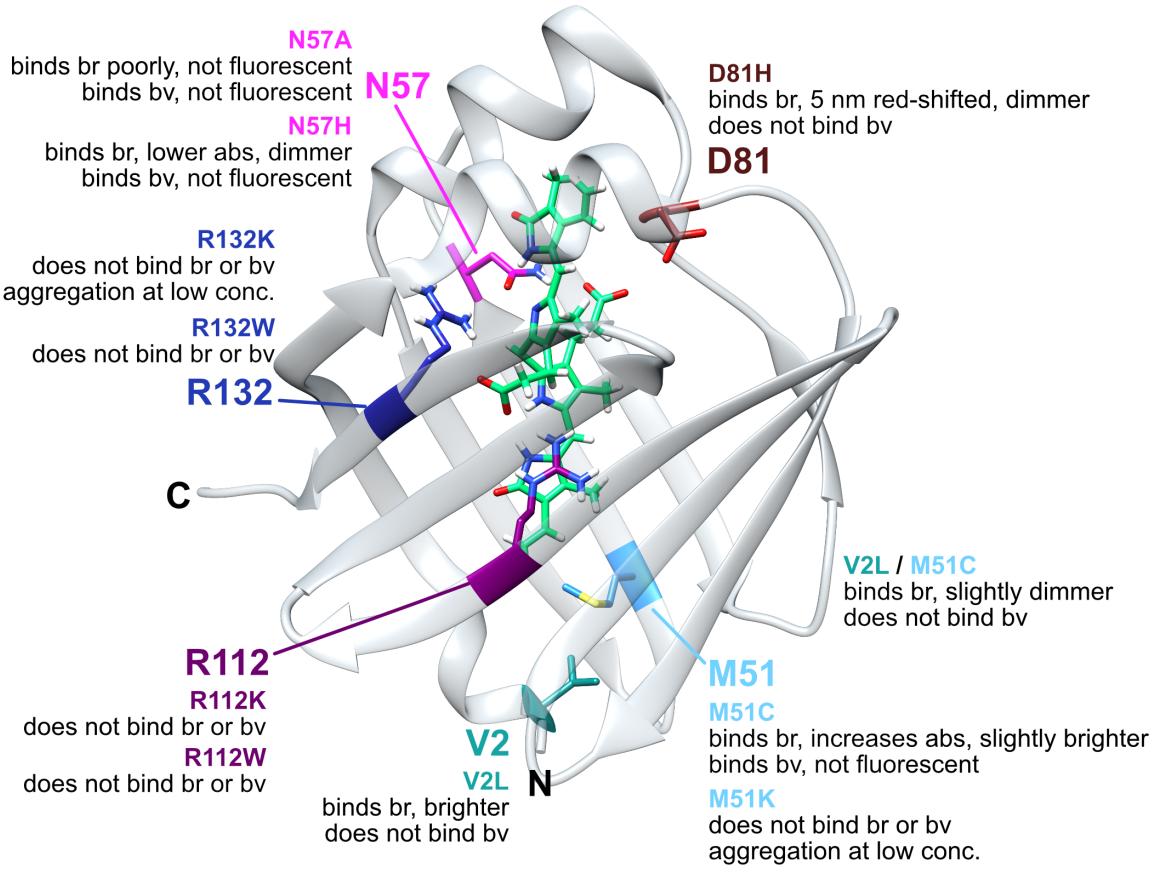


Figure 4.6: Summary of single UnaG variants and the effects of each mutation on br or bv binding. All comparisons are relative to Wt UnaG. PDB: 4I3B.

4.3.2 Directed evolution of a red-shifted bv-binding UnaG

Given the difficulty of predicting beneficial mutations through a rational design approach, I turned toward a random mutagenesis and directed evolution approach to create a red-shifted bv-binding UnaG variant. Similar to eUnaG [80], mIFP [67], and other FP evolution approaches described in literature [41, 42, 47], I sought to use fluorescence to evaluate mutants in a high-throughput manner. Using Wt UnaG and br, I tried to establish a screening format using bacterial colonies (Fig. 4.7a) as it would be simple to visualize thousands of colonies on LB/Agar plates. Despite altering the concentrations of ligand added to plates, varying agar thickness, and switching to agar made of minimal media, I were not able to detect blue-green fluorescence from colonies expressing UnaG. I hypoth-

esize that this lack of fluorescence is due to the inaccessibility of the br ligand to cells in the plate format. Since a bacterial colony is a biofilm, the br ligand incorporated into the agar may have only been accessible to the bottom layer of cells in direct contact with the agar and the signal from this thin layer of cells would be masked by other layers of cells that did not have any br. Another possibility is that the blue-green signal expected from cells with UnaG-br may be too blue or dim to be detectable in an colony-agar format.

I also attempted to establish a lysate-based screening method. Using *E. coli* cells that expressed UnaG, I lysed cells grown from liquid culture in a variety of conditions with detergents (Tween-20, TritonX-100), with enzymes to facilitate cellular lysis (DNase, lysozyme), and in commercial lysis buffers before incubating the lysates with br (Fig. 4.7b). While I could not discern any blue-green fluorescence from lysed cells grown in nutrient rich LB media, I could detect fluorescence in lysed cells grown in M9 minimal media. I also tried to take advantage of UnaG's high thermal stability (T_m = 60°C [80]) by incubating lysates in a heat block to further precipitate other cellular debris before visualization. Unfortunately, this additional step did not improve fluorescence read-out. I ultimately decided not to pursue this lysate-based approach due to its low-throughput and resource intensive nature.

I turned toward fluorescence assisted cell sorting (FACS) to accomplish sensitive fluorescence detection with high-throughput screening of our mutant UnaG libraries. I devised a construct with dual expression of UnaG variants and EGFP, with the latter serving as an internal control for cells that expressed protein (Fig. 4.8a). Each UnaG variant library was constructed using ep-PCR to introduce mutations to the parent gene and the dual expression construct was transformed into *E. coli* cells for FACS sorting (Fig. 4.8b).

I performed four successive rounds of selection for UnaG variants that appeared to bind bv and exhibit red fluorescence upon blue 405-nm excitation in the Brilliant Violet 605 (BV605) channel or red 640-nm excitation in the APC-Cy5.5 channel (Fig. 4.9). Each successive round of selection was more stringent in gating for red-shifted fluorescence; I

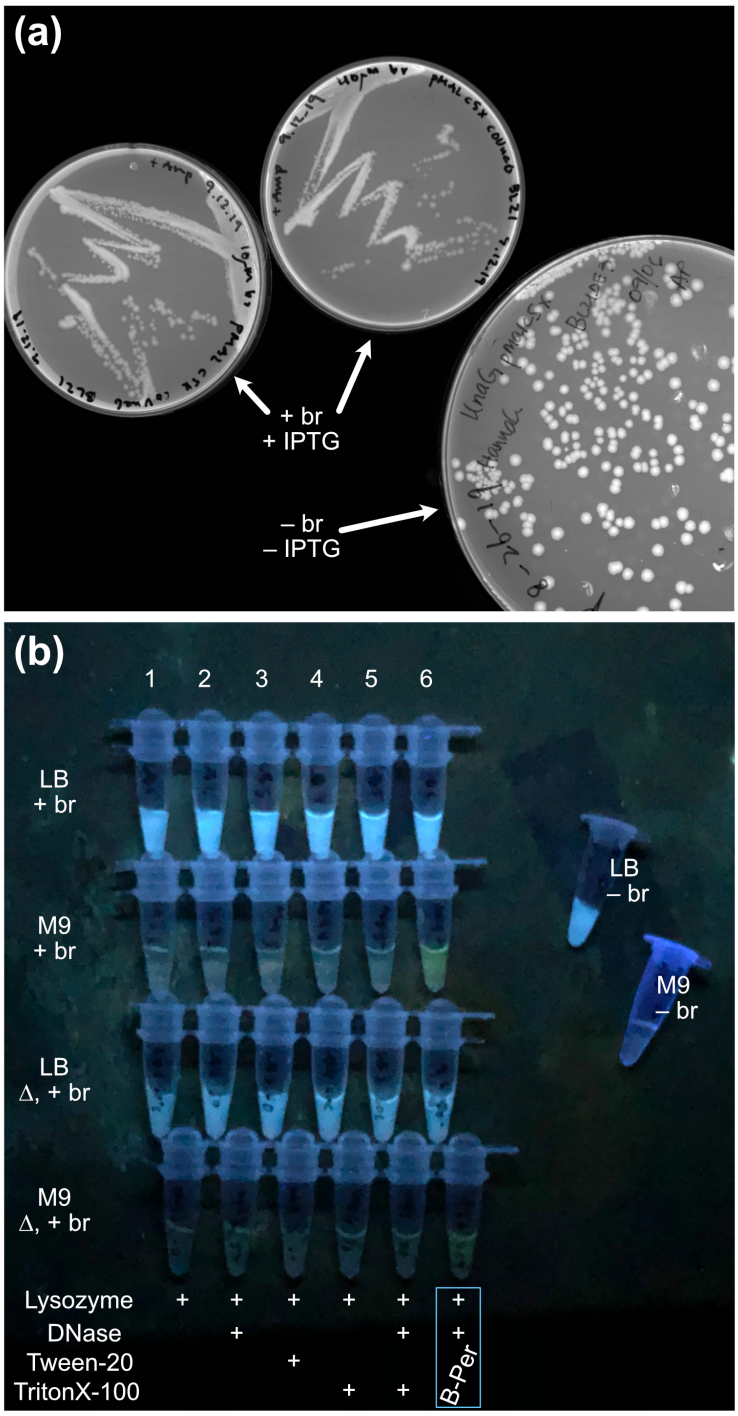


Figure 4.7: (a) Colony-agar and (b) lysate-based screening to detect fluorescence in *E. coli* cells expressing UnaG protein and incubated with br ligand. Cells were visualized under (a) 488-nm excitation using Alexa488 filters or (b) under UV light.

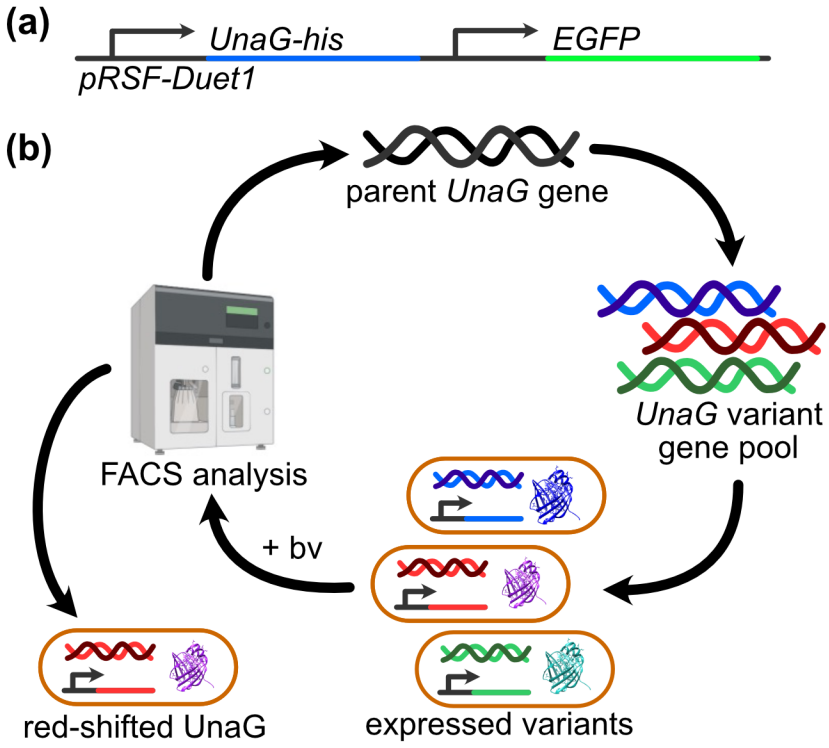


Figure 4.8: (a) Construct used for FACS-based evolution for red-shifted bv-binding UnaG. (b) Schematic of FACS-based evolution strategy.

observed somewhat enhanced fluorescence in each round per channel, though this "tail" of greater fluorescence signal tapered off in round 4. Due to our constructed dual expression system, it is possible that I inadvertently selected for mutants that could be more easily expressed by cells throughout all four libraries as EGFP functioned as a control for cells that generally expressed protein rather than as a direct readout for amount of mutant UnaG expressed. Likewise, the dropoff in signal in the fourth library may have been due to incomplete expression of protein by cells and may need repeating.

Sequencing after four rounds of sorting indicated that the most enriched mutation was V100D (Fig. 4.10). This mutation appeared in a handful of round 2 mutants, was present in the majority of round 3 mutants, and was found in nearly all round 4 mutants. Clones from library 4 that did not possess the V100D mutation instead posessed V100D, the only other possible mutation to an acidic residue. While mutations were scattered throughout nearly all of the protein sequence, the residues identified as critical for br binding in Wt

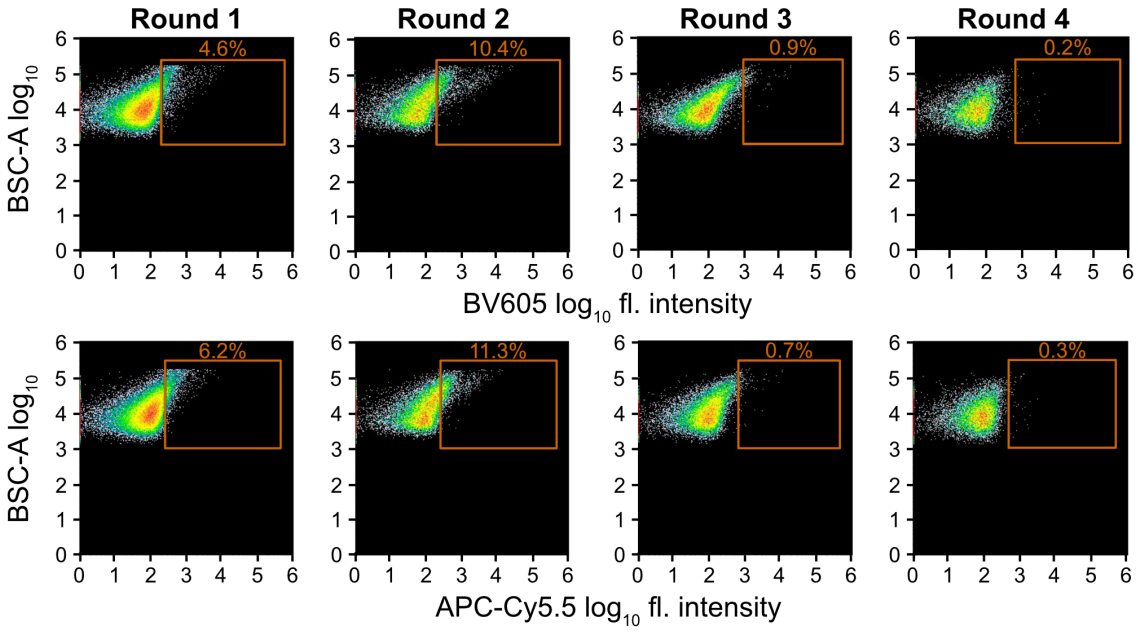


Figure 4.9: FACS analysis of bv-binding UnaG mutant library in *E. coli* sorted by red fluorescence emission using 405-nm excitation (BV605) or 640-nm excitation (APC-Cy5.5). Orange gate was used to collect cells with the most intense, red-shifted fluorescence.

UnaG (Asn57, Thr61, Glu77, Ser80, Asp81, Arg112) [60] remained nearly untouched when selecting for bv binding, indicating that these residues may remain as conserved positions for binding bilin ligands.

I attempted to characterize a handful randomly selected mutants from round 4. Most of these proteins were difficult to purify in despite troubleshooting to improve protein expression, to more thoroughly lyse cell pellets, and to maximize binding of protein to Ni-NTA resin. One mutant, L4-Cy3 (named after library 4 - APC-Cy5.5 mutant #3), stood out as a particularly easy variant to work with as the protein could be purified in large and pure quantities (yield 23 mg protein/L) and was extremely soluble in high concentrations. Like other clones in library 4, the L4-Cy3 variant possessed the V100D mutation as well as a handful of other mutations (Fig. 4.11a). Upon spectral characterization, I found that L4-Cy3 did not appear to bind not fluoresce with bv (Fig. 4.11c and e), but still bound br with the characteristic 48 nm absorbance shift that is also observed in Wt-br (Fig. 4.11b). However, the mutations in L4-Cy3 appear to destabilize br-induced fluorescence as suc-

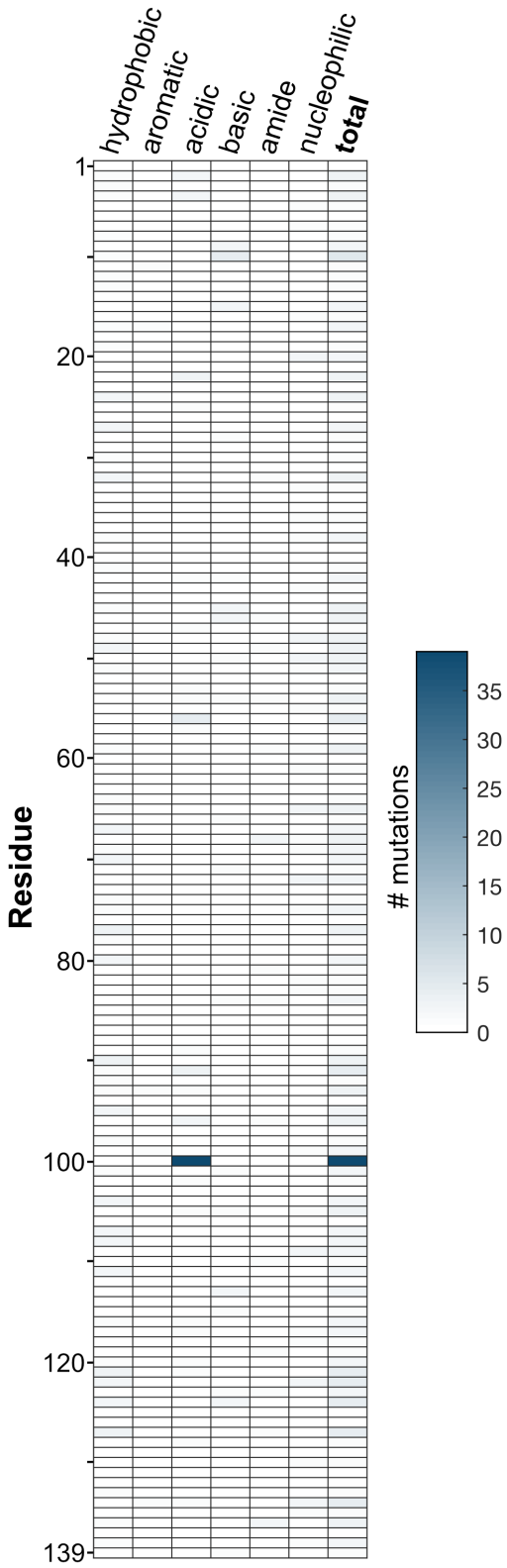


Figure 4.10: Sequence alignment of mutations in bv-binding UnaG variants after four rounds of library generation and selection.

cessive scans using 495-nm excitation resulted in decrease in fluorescence intensity (Fig. 4.11d).

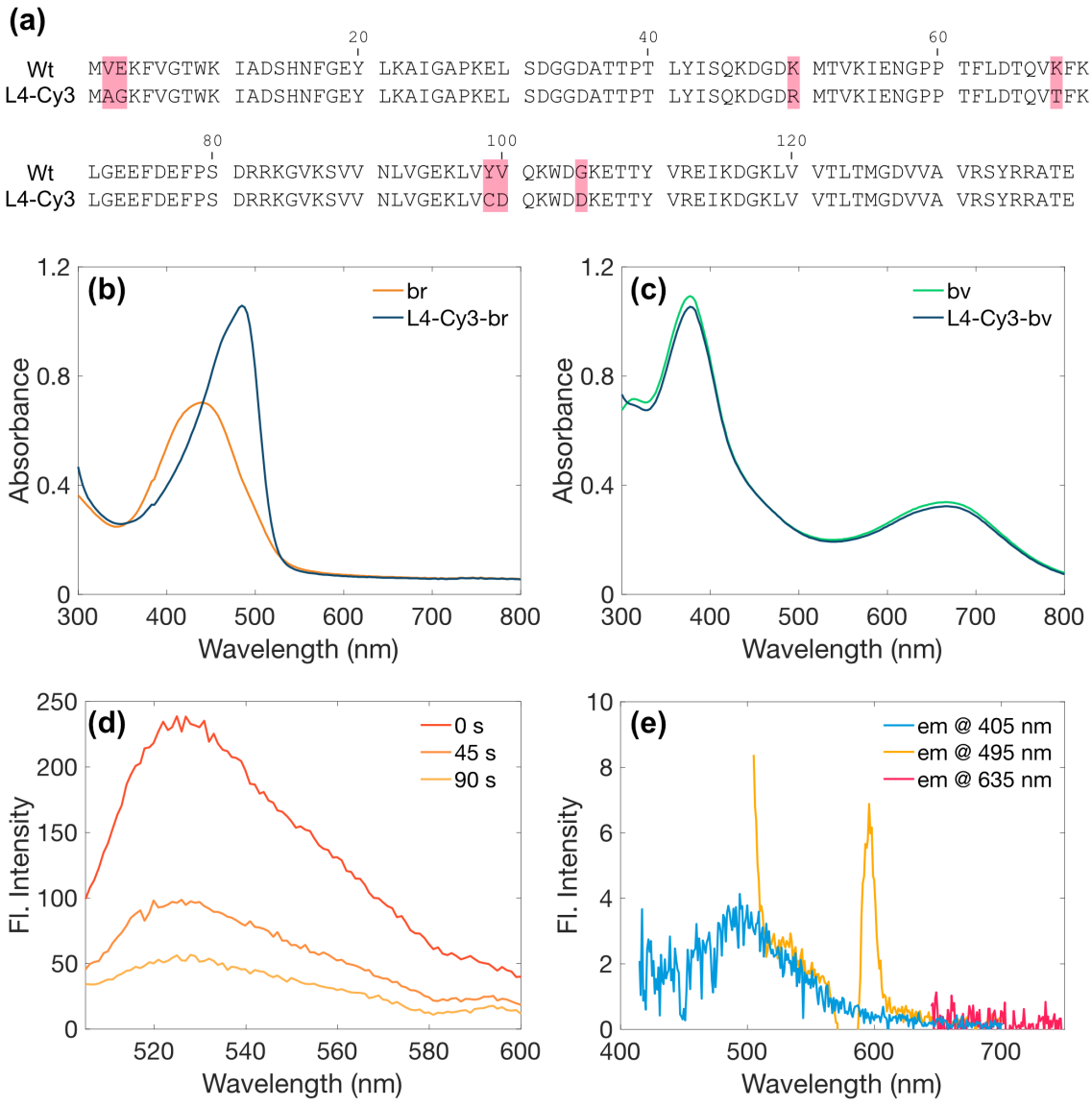


Figure 4.11: (a) Sequence alignment of Wt UnaG and L4-Cy3 variant. Mutated residues are highlighted in pink. (b-c) UV-vis absorbance and (d-e) fluorescence emission spectra of L4-Cy3 variant bound with (b,d) br or (c,e) bv. (d) Fluorescence emission was taken after incubating L4-Cy3 protein and br for 5 min and recording spectra at 45-sec intervals.

4.4 Discussion and future work

Designing a bv-binding UnaG variant would combine two advantages of BBFPs over other ligand-dependent FPs: the small monomeric UnaG protein and bv-induced far-red fluorescence. The wild-type UnaG exists as a monomer in both apo-and holo-protein forms [60,61], which allows for introduction of the tag into systems without fear of biological perturbation. On the other hand, the bv ligand provides a far-red fluorescence signal that is spectrally separated from the blue intrinsic fluorescence region. This red fluorescence not only allows for multi-color imaging when using BBFPs [105], but also opens up the possibility of probing multi-layer cellular samples like biofilms or tissues [66, 67]. The existence of red-emitting bv-binding FPs that can be excited by blue light (Sandercyanin [115]) or red light (IFP2.0 [66], mIFP [67], smURFP [112]) provides additional spectral flexibility in choosing tags for fluorescence-based applications.

UnaG has proven to be a versatile probe for biosensors and for advanced super-resolution imaging techniques. Applications of UnaG as a biosensor include uPPI, a split protein approach utilizing UnaG as a protein-protein interaction reporter [62], BReleaCa, an UnaG-calmodulin chimeric protein that detects Ca^{2+} [63], and most recently GOD-POD-UnaG, a multi-protein construct that measures unconjugated br in clinical serum samples [117]. Beyond typical bulk fluorescence imaging uses, UnaG-br and the eUnaG variant-br have been established as green-to-dark photoswitching FPs for super-resolution imaging [102, 111].

Most of these approaches rely on the non-covalent UnaG-br interaction, allowing for binding or displacement of the br ligand to function as detectable fluorescence signal. However, bv-binding FPs, with the exception of Sandercyanin, covalently bind one of the bv pyrrole rings via a cysteine residue in the binding pocket. This covalent linkage limits these FPs from biosensor applications using similar ligand displacement techniques to UnaG-br. Furthermore, the oligomeric forms of many bv-binding FPs introduce additional complexity in utilizing bv-induced fluorescence for stricter stoichiometric measure-

ments in biosensor applications. A bv-binding FP engineered from cyanobacteriochromes with a similar size and monomeric form to UnaG has been recently described, but this miRFP670nano protein also requires covalent binding of bv [118], again rendering it unhelpful for any application reliant on ligand association or displacement. Based on available crystal structures of bv-binding FPs, it does not appear that a covalent thioether bond is required for inducible fluorescence, but rather to anchor the ligand in the binding pocket.

While I was unsuccessful in engineering a fluorescent bv-binding UnaG variant through our rational design and directed evolution approaches, I have gained insight into the br-binding pocket. Only two rationally designed mutants, M51C and N57A, appeared to bind and broaden bv absorbance peaks (Fig. 4.2c and e), though neither variant is capable of bv-induced fluorescence (Fig. 4.4). N57A was a previously described mutant that has no br-induced fluorescence because the Asn57 residue is critical in stabilizing the planar conformation of the *endo*-vinyl dipyrrinone (rings A and B). I hypothesize that the N57A mutation may open up the ligand binding pocket to accomodate the entirely planar bv molecule, though structural investigations are required to confirm how the N57A mutation interacts with the bv ligand.

Previous molecular dynamics simulations performed on the eUnaG variant indicates that the V2L mtuation stabilizes a loop near the cavity entrance and pushes Met51 toward br [80]. Based on these simulations, I hypothesize that the M51C mutation alone may have promoted favorable contacts with the ligand to further stabilize br-induced fluorescence, similar to the single V2L mutation. However, the effects of these mutations were not additive as the V2L/M51C double mutant did not have enhanced green fluorescence when bound to br (Fig. 4.5). The M51K mutation that introduced a sterically similar but charged residue instead resulted in poor expression and presumably structural instability, suggesting that the Met51 residue is more important than previously thought in positioning ligands and stabilizing fluorescence. Given that the M51C mutant was among the best

of the designed variants at binding bv and enhancing absorbance, yet did not fluoresce when binding bv, suggests that this residue may act as a gatekeeper for accommodating the planar bv liigand and additional mutations are thus required to stabilize bv-induced fluorescence.

The preeminent mutation that arose after four rounds of FACS-based evolution was V100D. However, the benefit of this mutation is not obvious based on the crystal structure of Wt UnaG, as the original Val100 residue points outward and away from the binding pocket. It is possible that V100D may alter the positioning of the neighboring Tyr99 and Gln101 residues that interact with the second solvent shell, and these alterations would need to be probed with deeper structural characterization. Nevertheless, this mutation would not have persisted through four rounds of screening without conferring some benefit to the mutant UnaG-bv pair selection process. Additional characterization of a point mutation variant is required to confirm its benefit for bv binding.

As even the high-throughput and high-sensitivity method of FACS did not yield a red-emitting and bv-binding UnaG variant, future work would likely require a redesign of the construct used for FACS-based evolution. Since our construct resulted in dual expression of mutant UnaG and EGFP, the EGFP expression functioned as a control for cells expressing protein overall, but not for normalized protein expression of UnaG. As a result, it is possible that our screening inadvertently selected for variants that could be expressed more readily by *E. coli* or was more generally permissible for cells that were able to take in more bv to fluoresce. This construct design may also explain why I did not see the main density of cells sorted gradually shift to be more fluorescent in each analyzed round (Fig. 4.9). Likewise, I may have not performed enough rounds of selection to notably shift UnaG recognition to exclusively recognize bv.

Even so, the most promising L4-Cy3 variant that possesses enhanced solubility may prove to be useful as a building block in a symmetry-based protein cage assembly project being conducted by another member of the Marsh lab (Fig. A.3). A future iteration of this

FACS-based evolution approach should instead construct the mutant UnaG and EGFP proteins in tandem such that EGFP becomes a direct readout for protein expressed per cell. Overall, an established directed evolution approach for engineering ligand recognition into UnaG could be generalizable and be utilized to make UnaG variants that can exclusively recognize new ligand, such as the ones detailed in Chapter III.

4.5 Conclusions

In our attempts to engineer a red-shifted bv-binding UnaG variant, I have identified residues that may confer the most benefit for switching ligand recognition from br to bv: M51C, N57A, and V100D. These residues appear to help the protein accommodate the more planar bv molecule and have enhanced bv binding over the Wt UnaG protein, though additional mutations are required to restore bv-induced fluorescence. This work lays a foundation for engineering non-native ligand recognition into UnaG and be further used to expand the colorful toolbox of oxygen-independent fluorescent tools.

CHAPTER V

Conclusions

5.1 Introduction

Fluorescent proteins (FPs) remain the gold standard tools for probing biological systems and there exists the need for new tools for probing oxygen-sensitive or anaerobic environments. As I have discussed throughout this dissertation, the ligand-dependent family of Bilin-based Fluorescent Proteins (BBFPs) is an attractive family of reporters to utilize for anaerobic fluorescence microscopy. In this final chapter, I will review the conclusions from the preceding chapters as well as present future directions for these oxygen-independent reporter proteins.

5.2 BBFPs for imaging live obligate anaerobes

In Chapter II, I presented the first application of BBFPs in live-cell imaging of anaerobic bacteria. I demonstrated how the blue-green UnaG and the far-red IFP2.0 BBFPs can label the commensal gut bacteria *Bacteroides thetaiotaomicron (B. theta)* in oxygen-free imaging where the commonly used green fluorescent protein (GFP) reporter fails. These reporters are readily compatible with most home-built and commercial microscopy setups. UnaG-labeled cells can be visualized using GFP excitation and filter sets as can IFP2.0-labeled cells using Cy5 excitation and filter sets. These BBFPs' utility also extends beyond mono-

culture applications to multi-species and two-color imaging. In a mixed culture, I distinguished UnaG-labeled *B. theta* cells from unlabeled *Ruminococcus bromii (R. bromii)*, the latter a species that is not genetically tractable. Likewise, a mixed culture of *Bacteroides ovatus (B. ovatus)* and *B. theta* labeled with UnaG and IFP2.0, respectively, was exclusively differentiated in separate color channels in multicolor imaging.

This work collectively validates not only the use of BBFPs for anaerobic fluorescence applications, but also highlights the potential for utilizing this family of reporters in investigations of polymicrobial systems. The genetic encodability of FPs is desirable for directly labeling different species in mixed-species conditions, but most of the common FPs in biology are ones that require an oxidative post-translational modification to fluoresce. As such, experiments that utilize GFP- or DsRed-like FPs are conducted on dormant or fixed cells that are exposed to oxygen to mature FP chromophores, but result in static snapshots of bacterial communities or biological phenomena.

Ligand-dependent reporters like BBFPs, HaloTag, SNAP-, and CLIP-tags avoid the necessity for any oxygen exposure [50]. While the Biteen lab has previously used the HaloTag system to track protein complex assembly in *B. theta*, these experiments were limited to tracking targets on the outer membrane as HaloTag ligands were unable to reach cytoplasmic targets. Moreover, most HaloTag ligands are fluorescent dyes that not only provide a bright signal to labeled targets, but also result in a bright background signal that must be removed by washing away excess ligand that may non-specifically stick to cells. Although new HaloTag ligands have been developed with special spectral properties, most notably the caged Janelia Fluors, which are capable of photoswitching and photoactivation [68, 119, 120], these ligands still require additional washing steps before imaging.

BBFPs solve this problem by utilizing the truly fluorogenic ligands bilirubin (br) and biliverdin (bv). Fluorogenic ligands, which are non-fluorescent or weakly emissive before binding a partner protein or target, are versatile molecules for live-cell and multicolor

imaging. There is great interest within the field in developing fluorogenic probes from existing rhodamine fluorescent dyes [121] and other fluorogen-activated FP systems [97, 110]. Wash-free labeling is the most attractive property imparted by fluorogenicity, as this quality allows continual addition of ligand for long term imaging without increased background signal. The ligands br and bv are unique within this class of molecules as they are among the few that can be found in Nature and do not require additional synthetic manipulation for stable fluorescence upon binding protein partners.

One of the long term goals of the Biteen lab is to continuously track the growth of mixed-species cultures of gut bacteria and observe spatial pattern distribution of different species as they share nutrients and carbon sources. BBFPs enable these types of polymicrobial studies by providing oxygen-independent fluorescent labeling as well as the flexibility to label targets that are intra- or extracellular. For experiments requiring multiple days of monitoring cells in monolayer or even multilayers like biofilms, the br and bv ligands will allow for robust and non-toxic labeling of BBFPs-tagged cells.

Given that UnaG has also been recently demonstrated as a reporter suitable for super-resolution experiments [102, 111], an outsider would naturally ask if the Biteen lab could also utilize UnaG as a tag for super-resolution microbial experiments. However, the Biteen lab specializes in Single-Particle Tracking (SPT) and localization, whereas studies of UnaG in super-resolution experiments rely on Stochastic Optical Reconstruction Microscopy (STORM)-like experimental parameters. In the latter, the non-covalent association between UnaG and br allows for fresh ligand to bind after br is photo-oxidized [102], resulting in a series of imaging cycles with "blinking" fluorophores that can be individually resolved for super-resolution localization [122]. SPT differs through the tracking of individual molecules for periods of time, longer relative to STORM emitters, to uncover dynamics of heterogeneous protein populations with temporal and spatial resolution. To accomplish SPT, the fluorophores need to be bright and photostable to stitch together consecutive imaging frames of the same molecule and fit "tracks".

While the photon count of the br ligand is improved in oxygen-depleted conditions close to purely anaerobic imaging [102], the ligand may require further improvement to the photostability before it is suitable for SPTs of molecules with long tracks. Furthermore, an UnaG-br system lacks the convenience of photoactivation like with Photoactivatable mCherry (PAmCherry) or HaloTag with Janelia Fluor ligands, so additional experimental considerations are required to control the number of molecules that are labeled at a time for single-molecule detection. I imagine that an UnaG-br system for SPT experiments would be in a microfluidic device that would allow for continual flow of low concentrations of br over time such that photobleached holoUnaG can be reconstituted with fresh ligand for continuous imaging.

5.3 HTS to identify new UnaG ligands

To begin to build out a toolbox of BBFPs colors, I set out to find new ligands that bound UnaG as discussed in Chapter III. Utilizing high-throughput screening (HTS), I identified new compounds that bound UnaG and fluoresced in channels that were red-shifted with respect to the native blue-green UnaG-br pair. The best hit was a benzothiazole-based ligand that binds UnaG with high affinity (K_d = 3 nM) and could be used in the common green 532-nm microscopy channel. While this new ligand was nominally fluorescent on its own, the fluorescence is enhanced and red-shifted upon binding UnaG. Again, I demonstrated new UnaG-ligand pair could be used as a reporter in anaerobic live-cell imaging of *B. theta* in monoculture and in mixed-species two-color imaging with IFP2.0-labeled *B. ovatus*. To our knowledge, this work demonstrates the first application of benzothiazole-based ligands for live-cell microscopy and anaerobic imaging.

When UnaG was initially characterized, the tight coordination of the br ligand was a key feature of the FP and distinguished it from other homologous fatty-acid-binding proteins with broader ligand binding promiscuity. However, the work in this chapter demonstrates that the UnaG binding pocket has some degree of plasticity that can accommodate

and stabilize new fluorogenic ligands. This finding opens the door to creating new UnaG-ligand pairs of more colors, similar to how the toolbox of GFP and DsRed-like FPs were diversified into a rainbow.

There are far more colors that UnaG can be diversified into for multicolor fluorescence applications. In this Chapter, I have only described one new UnaG-ligand pair that fluoresces in the commonly found 532-nm green channel, and we were unsuccessful in finding a ligand that could fluoresce in our other HTS optics channel of red 580/610 nm. Further HTS may be used to identify molecules that bind UnaG and fluoresce in this unused 580/610 nm channel. Likewise, I was restricted by the prefabricated optics modules that were available at the University of Michigan's Center for Chemical Genomics core; however, the assay I have developed is generalizable to other common microscopy channels such as the yellow-green 561 nm channel or the often underutilized teal 515 nm channel. While we have successfully used IFP2.0-bv in the red 635-nm channel, the pair is actually maximally excited in the far-red and should instead be used with 670-nm, 685-nm, or 690-nm excitation sources and filters. I expect that a combination of additional HTS and synthetic chemistry can discover and refine new color palette of fluorogenic ligands suitable for anaerobic and advanced imaging applications.

Although it was unsurprising that the molecules we found in our HTS assay were highly π-conjugated compounds, the structural diversity of heterocycles in our top hits was unexpected. Furthermore, our two best hits were benzothiazole-based ligands but this structural feature is not common in commercial fluorescent dyes. Future work could involve virtual screening of chemical libraries to find other molecules with structural similarity to the top hits that can also fluoresce in microscopy channels outside of the blue 488 nm channel of the native UnaG-br pair.

My top hits from HTS indeed contained the properties I selected for, which was red-shifted fluorescence relative to the UnaG-br pair and fluorogenicity or enhanced fluorescence upon binding UnaG; even so, these ligands are dimmer than commercial dyes and

required high concentrations in the micromolar range for *in vitro* characterization for detectable fluorescence signal. I also noticed degradation of the ligand over time (weeks-long timescale) as detectable by UV-vis absorbance and fluorescence emission intensity reduction, though I was unable to detect any change caused by oxidation or hydrolysis by mass spectrometry analysis. Future work using these benzothiazole-based ligands should focus on improving the stability of the ligands both for practical storage and for improved quantum yield and brighter signal. It would also be interesting if benzothiazole proved to be a useful scaffold for designing fluorogenic molecules with tunable emission or other special spectral properties, as this possible work could open up a new family of useful fluorescent ligands or dyes.

5.4 Engineering ligand recognition in UnaG

In Chapter IV, I described two approaches toward designing a bv-binding UnaG variant, with the goal of creating a monomeric red-emitting BBFPs. In the first, rational design approach, I selectively mutated residues in the binding pocket that interacted with the br ligand based on the published holoUnaG-br crystal structure. In the second, directed evolution approach, I employed random mutagenesis across the entire protein sequence and selected for bv-binding and red-emitting UnaG variants that could be excited by blue or red light. While I was ultimately unsuccessful in creating a red-shifted UnaG variant that could bind bv, the two approaches I have taken have collectively provided greater insight into the surprisingly malleable UnaG binding pocket.

A handful of key mutations will require deeper structural characterization to understand their effects on ligand binding. From the rational design approach, I found that two separate mutations, M51C and N57A, enabled the protein to bind bv but did not result in bv-induced fluorescence. At a cursory glance, these mutations appear to help accommodate the planar bv ligand by opening up the binding pocket. On the other hand, the purpose of V100D, the mutation most enriched from our directed evolution approach, is

unclear because this residue does not directly interact with the ligand binding pocket. Either computational or experimental structural characterization is needed to understand how bv-binding is enabled by any of these mutations. These studies could also be complemented with computational modeling to understand the difference in energetics and photophysical properties between br and bv, such that these understandings can be used to guide UnaG binding pocket design to enhance holoprotein brightness.

While I only used Fluorescence Assisted Cell Sorting (FACS)-based workflow for evolving bv recognition into UnaG, I envision that the workflow can be generalizable to evolving other ligand recognition into UnaG, such as the benzothiazole-based ligands described in Chapter III. For future work, the dual expression construct of EGFP and UnaG variants should be switched to a chimeric EGFP-UnaG system such that EGFP becomes a direct readout for protein expression. Likewise, I only utilized two fluorescence channels in FACS screening but the cell sorters in the University of Michigan's Flow Cytometry core are equipped with more excitation sources and filter sets. As I discussed in the previous section, these additional excitation wavelengths will enable selection of variant UnaGs with new ligands that fluoresce in other microscopy channels. The work presented in Chapter IV lays the groundwork for robust selection of unique UnaG-ligand variants.

5.5 Other ligand-dependent reporters for anaerobic imaging

The focus of my dissertation is on the evaluation of BBFPs and development of the br-binding FP UnaG into new colors, but other ligand-dependent FP systems may prove to be as fruitful as BBFPs for anaerobic fluorescence microscopy of microbes. The Yellow Fluorescence-Activating and Absorption-Shifting Tag (Y-FAST) is the closest analog to UnaG in that Y-FAST is similarly small and monomeric as well as also utilizes a fluorogenic ligand [97]. Recently, the Y-FAST system has been diversified into orthogonal greenFAST and redFAST probes with specific protein recognition of unique fluorogens for multicolor fluorescence microscopy [110]. Like UnaG-br, these FAST tags have also been used for

single-molecule localization experiments [123]. Another similar system to UnaG-br are the hCRBPII variants developed from retinol binding proteins that bind synthetically enhanced fluorogens with large stokes shifts and far-red emission [101]. Theoretically, the Y-FAST and hCRBPII systems are compatible for live-cell anaerobic microscopy and implementable like BBFPs. Future work could validate the use of these two systems to verify that their fluorogenic ligands are non-toxic as well as cell permeable for targeted labeling of cells and proteins of interest. A parallel investigation of BBFPs, Y-FAST, and hCRBPII systems should be carried out in live-cell conditions to compare respective no-wash procedures and determine which system is most compatible and easily implementable for long-term imaging of polymicrobial communities.

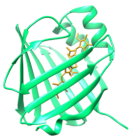
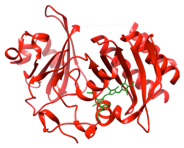
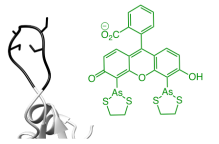
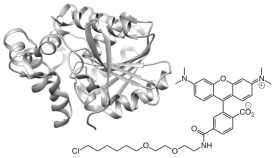
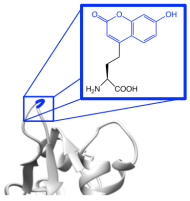
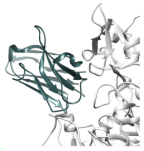
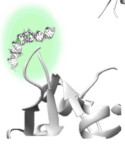
5.6 Overarching conclusions

In this thesis, I have demonstrated the importance of evaluating and implementing BBFPs for studying important oxygen-sensitive microbial systems. FPs remain the most important tools for probing biological phenomena and the expansion of oxygen-independent reporters opens new frontiers in understanding microbial communities and biochemical mechanisms. While the applications in this thesis are largely centered on fluorescence microscopy, these BBFPs may also be used for other biochemical and biophysical techniques that require fluorescence readouts. These ligand-dependent fluorescent reporters will provide the greatest impact at the intersections of microbiology and biochemistry that require strict oxygen-free investigations to answer outstanding biological questions.

APPENDICES

APPENDIX A

Supporting Information for Chapters I - IV

	Probe	Color Brightness Size*	Advantages	Disadvantages
Oxygen-independent Fluorescent Protein Approaches				
	*Flavin-mononucleotide-based fluorescent proteins*			
	FbFPs (BsFbFP, PpFbFP) iLOV	Cyan Dim Small	▪ Genetically encoded • Ligand is available in most systems • Demonstrated in live anaerobic bacterial cells	• Low contrast: FbFPs are dim and similar in color to cellular autofluorescence
	*Fatty-acid-binding fluorescent proteins*			
	UnaG	Green Same Small	▪ Genetically encoded ▪ Demonstrated in live anaerobic cells	• Ligand is cell-impermeable • Ligand is not water-soluble
	IFP1.4 IFP2.0	Red Dim Same	▪ Genetically encoded ▪ Demonstrated in live anaerobic cells	• Ligand is cell-impermeable • Ligand is not water-soluble • Low contrast: IFPs are dim
Bioconjugation Approaches				
	*Biarsenical-tetracystine tag*			
	FlAsH ReAsH	Green/Red Bright Very Small	▪ Genetically encoded ▪ Limited demonstration live anaerobic bacterial cells	• Low contrast: Ligand is not fluorogenic • Ligands may be cytotoxic
	*Self-labeling proteins*			
	SNAP! CLIP HaloTag TMP	Varies Varies Same	• Genetically encoded • Demonstrated in live anaerobic bacterial cells (HaloTag) • Tags can be multiplexed	• Low contrast: Most ligands are not fluorogenic
	*Unnatural amino acids (UAAs)*			
	Fluorescent UAA	Blue Dim Very Small	▪ No ligand required	• Technically difficult and restricted in implementation to genetically tractable organisms • Not yet demonstrated in live anaerobic bacterial cells • Low contrast: Fluorescent UAAs are dim
	Bioconjugating UAA	Varies Bright Very Small	▪ Ligand is bright dye	
Target-based Approaches				
	Nanobodies & Chromobodies	Varies Varies Small – Same	▪ Can be modified to be cell-permeable ▪ Can be conjugated to bright organic dyes ▪ Can be attached to an FP	• Not yet demonstrated in live anaerobic bacterial cells
	Aptamers (Spinach, Malachite Green (MG), RNA-mango)	Varies Varies Very Small	▪ Ligand is fluorogenic (Spinach) ▪ Ligand is bright (MG and Mango) ▪ Ligand is cell-permeable (Spinach, modified MG and Mango)	• Not yet demonstrated in live anaerobic bacterial cells • Limited stability *in vivo* • Ligand is cytotoxic (MG)

*Brightness and Size relative to GFP.

Figure A.1: Summary of probes for anaerobic live-cell imaging. Figure adapted from [50].

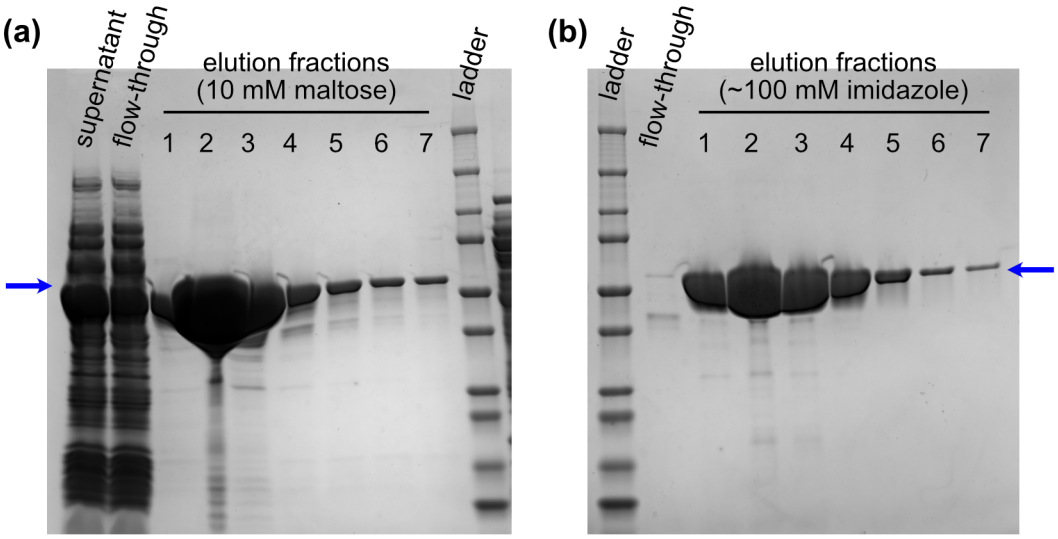


Figure A.**2:** SDS-PAGE analysis of MBP-UnaG from (a) MBP-trap followed by (b) His-trap purification. Blue arrows indicate the correct protein size of 59 kDa. Protein yield per prep was 8.2 mg/L.

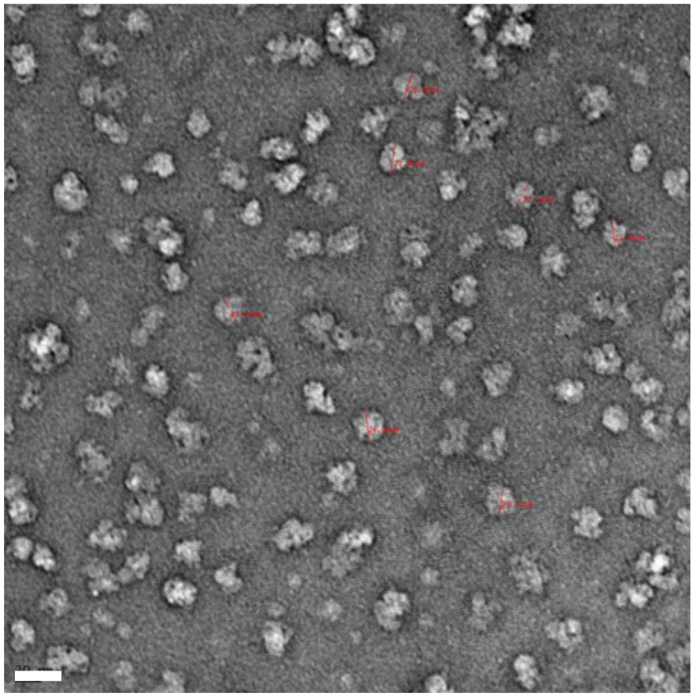


Figure A.**3:** Transmission Electron Microscopy (TEM) images of protein cages assembled using the L4-Cy3 mtUnaG variant as a building block. Scale bar: 20 μm. Figure provided courtesy of Dr. Karl J. Koebke.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] P. D. Cani, "Human gut microbiome: hopes, threats and promises", *Gut* **67**, 1716 (2018). DOI: 10.1136/gutjnl-2018-316723. Publisher: BMJ Publishing Group Section: Recent advances in basic science.

[2] D. Verma, P. K. Garg, A. K. Dubey, "Insights into the human oral microbiome", *Archives of Microbiology* **200**, 525 (2018). DOI: 10.1007/s00203-018-1505-3.

[3] N. Fierer, "Embracing the unknown: disentangling the complexities of the soil microbiome", *Nature Reviews Microbiology* **15**, 579 (2017). DOI: 10.1038/nrmicro.2017.87.

[4] A. Vigneron, *et al.*, "Succession in the petroleum reservoir microbiome through an oil field production lifecycle", *The ISME Journal* **11**, 2141 (2017). DOI: 10.1038/ismej.2017.78.

[5] B. M. Peters, M. A. Jabra-Rizk, G. A. O'May, J. W. Costerton, M. E. Shirtliff, "Polymicrobial Interactions: Impact on Pathogenesis and Human Disease", *Clinical Microbiology Reviews* **25**, 193 (2012). DOI: 10.1128/CMR.00013-11.

[6] J. Durack, S. V. Lynch, "The gut microbiome: Relationships with disease and opportunities for therapy", *Journal of Experimental Medicine* **216**, 20 (2019). DOI: 10.1084/jem.20180448. Publisher: The Rockefeller University Press.

[7] C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg, "The Gut Microbiome: Connecting Spatial Organization to Function", *Cell Host & Microbe* **21**, 433 (2017). DOI: 10.1016/j.chom.2017.03.010.

[8] J. S. Lichtman, *et al.*, "The effect of microbial colonization on the host proteome varies by gastrointestinal location", *The ISME Journal* **10**, 1170 (2016). DOI: 10.1038/ismej.2015.187.

[9] G. P. Donaldson, S. M. Lee, S. K. Mazmanian, "Gut biogeography of the bacterial microbiota", *Nature Reviews Microbiology* **14**, 20 (2016). DOI: 10.1038/nrmicro3552. Number: 1 Publisher: Nature Publishing Group.

[10] X. Ze, S. H. Duncan, P. Louis, H. J. Flint, "Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon", *The ISME Journal* **6**, 1535 (2012). DOI: 10.1038/ismej.2012.4.

[11] S. Macfarlane, E. J. Woodmansey, G. T. Macfarlane, "Colonization of Mucin by Human Intestinal Bacteria and Establishment of Biofilm Communities in a Two-Stage Continuous Culture System", *Applied and Environmental Microbiology* **71**, 7483 (2005). DOI: 10.1128/AEM.71.11.7483-7492.2005. Publisher: American Society for Microbiology Section: PHYSIOLOGY AND BIOTECHNOLOGY.

[12] E. Wynendaele, *et al.*, "Crosstalk between the microbiome and cancer cells by quorum sensing peptides", *Peptides* **64**, 40 (2015). DOI: 10.1016/j.peptides.2014.12.009.

[13] Y. E. Tuncil, *et al.*, "Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence", *mBio* **8** (2017). DOI: 10.1128/mBio.01068-17. Publisher: American Society for Microbiology Section: Research Article.

[14] E. H. Crost, *et al.*, "Mechanistic Insights Into the Cross-Feeding of Ruminococcus gnavus and Ruminococcus bromii on Host and Dietary Carbohydrates", *Frontiers in Microbiology* **9** (2018). DOI: 10.3389/fmicb.2018.02558. Publisher: Frontiers.

[15] K. Faust, *et al.*, "Microbial Co-occurrence Relationships in the Human Microbiome", *PLoS Computational Biology* **8** (2012). DOI: 10.1371/journal.pcbi.1002606.

[16] E. G. Zoetendal, M. Rajilić-Stojanović, W. M. d. Vos, "High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota", *Gut* **57**, 1605 (2008). DOI: 10.1136/gut.2007.133603.

[17] J. Qin, *et al.*, "A metagenome-wide association study of gut microbiota in type 2 diabetes", *Nature* **490**, 55 (2012). DOI: 10.1038/nature11450.

[18] G. Falony, *et al.*, "Population-level analysis of gut microbiome variation", *Science* **352**, 560 (2016). DOI: 10.1126/science.aad3503. Publisher: American Association for the Advancement of Science Section: Research Article.

[19] P. Vangay, *et al.*, "US Immigration Westernizes the Human Gut Microbiome", *Cell* **175**, 962 (2018). DOI: 10.1016/j.cell.2018.10.029.

[20] T. E. Rogers, *et al.*, "Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures", *Molecular Microbiology* **88**, 876 (2013). DOI: 10.1111/mmi.12228.

[21] H. H. Tuson, M. H. Foley, N. M. Koropatkin, J. S. Biteen, "The Starch Utilization System Assembles around Stationary Starch-Binding Proteins", *Biophysical Journal* **115**, 242 (2018). DOI: 10.1016/j.bpj.2017.12.015.

[22] H. H. Tuson, J. S. Biteen, "Unveiling the Inner Workings of Live Bacteria Using Super-Resolution Microscopy", *Analytical Chemistry* **87**, 42 (2015). DOI: 10.1021/ac5041346.

[23] E. T. Arena, *et al.*, "Quantitating the cell: turning images into numbers with ImageJ", *WIREs Developmental Biology* **6**, e260 (2017). DOI: 10.1002/wdev.260. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wdev.260.

[24] R. Heintzmann, G. Ficz, *Methods in Cell Biology* (Academic Press, 2007), vol. 81 of *Digital Microscopy, 3rd Edition*, pp. 561–580.

[25] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Springer US, 2006), third edn.

[26] A. M. Sydor, K. J. Czymmek, E. M. Puchner, V. Mennella, "Super-Resolution Microscopy: From Single Molecules to Supramolecular Assemblies", *Trends in Cell Biology* **25**, 730 (2015). DOI: 10.1016/j.tcb.2015.10.004.

[27] K. S. Karunatilaka, E. A. Cameron, E. C. Martens, N. M. Koropatkin, J. S. Biteen, "Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts", *mBio* **5**, e02172 (2014). DOI: 10.1128/mBio.02172-14.

[28] W. R. Whitaker, E. S. Shepherd, J. L. Sonnenburg, "Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome", *Cell* **169**, 538 (2017). DOI: 10.1016/j.cell.2017.03.041.

[29] K. Earle, *et al.*, "Quantitative Imaging of Gut Microbiota Spatial Organization", *Cell Host & Microbe* **18**, 478 (2015). DOI: 10.1016/j.chom.2015.09.002.

[30] J. L. M. Welch, Y. Hasegawa, N. P. McNulty, J. I. Gordon, G. G. Borisy, "Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice", *Proceedings of the National Academy of Sciences* p. 201711596 (2017). DOI: 10.1073/pnas.1711596114.

[31] W. Wang, Y. Zhu, X. Chen, "Selective Imaging of Gram-Negative and Gram-Positive Microbiotas in the Mouse Gut", *Biochemistry* **56**, 3889 (2017). DOI: 10.1021/acs.biochem.7b00539.

[32] J.-H. Hehemann, *et al.*, "Single cell fluorescence imaging of glycan uptake by intestinal bacteria", *The ISME Journal* **13**, 1883 (2019). DOI: 10.1038/s41396-019-0406-z. Number: 7 Publisher: Nature Publishing Group.

[33] N. Geva-Zatorsky, *et al.*, "In vivo imaging and tracking of host–microbiota interactions via metabolic labeling of gut anaerobic bacteria", *Nature Medicine* **21**, 1091 (2015). DOI: 10.1038/nm.3929.

[34] G. V. Los, *et al.*, "HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis", *ACS Chemical Biology* **3**, 373 (2008). DOI: 10.1021/cb800025k.

[35] F. V. Subach, *et al.*, "Photoactivatable mCherry for high-resolution two-color fluorescence microscopy", *Nature Methods* **6**, 153 (2009). DOI: 10.1038/nmeth.1298.

[36] G. H. Patterson, J. Lippincott-Schwartz, "A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells", *Science* **297**, 1873 (2002). DOI: 10.1126/science.1074952.

[37] S. Habuchi, *et al.*, "Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa", *Proceedings of the National Academy of Sciences* **102**, 9511 (2005). DOI: 10.1073/pnas.0500489102. Publisher: National Academy of Sciences Section: Biological Sciences.

[38] O. M. Subach, *et al.*, "A photoswitchable orange-to-far-red fluorescent protein, PSmOrange", *Nature Methods* **8**, 771 (2011). DOI: 10.1038/nmeth.1664. Number: 9 Publisher: Nature Publishing Group.

[39] R. Heim, A. B. Cubitt, R. Y. Tsien, "Improved green fluorescence", *Nature* **373**, 663 (1995). DOI: 10.1038/373663b0. Number: 6516 Publisher: Nature Publishing Group.

[40] R. Heim, R. Y. Tsien, "Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer", *Current Biology* **6**, 178 (1996). DOI: 10.1016/S0960-9822(02)00450-5.

[41] R. Heim, D. C. Prasher, R. Y. Tsien, "Wavelength mutations and posttranslational autoxidation of green fluorescent protein", *Proceedings of the National Academy of Sciences* **91**, 12501 (1994). DOI: 10.1073/pnas.91.26.12501.

[42] M. Ormö, *et al.*, "Crystal Structure of the Aequorea victoria Green Fluorescent Protein", *Science* **273**, 1392 (1996). DOI: 10.1126/science.273.5280.1392.

[43] B. P. Cormack, R. H. Valdivia, S. Falkow, "FACS-optimized mutants of the green fluorescent protein (GFP)", *Gene* **173**, 33 (1996). DOI: 10.1016/0378-1119(95)00685-0.

[44] T. J. Lambert, "FPbase: a community-editable fluorescent protein database", *Nature Methods* **16**, 277 (2019). DOI: 10.1038/s41592-019-0352-8. Number: 4 Publisher: Nature Publishing Group.

[45] L. A. Gross, G. S. Baird, R. C. Hoffman, K. K. Baldridge, R. Y. Tsien, "The structure of the chromophore within DsRed, a red fluorescent protein from coral", *Proceedings of the National Academy of Sciences* **97**, 11990 (2000). DOI: 10.1073/pnas.97.22.11990. Publisher: National Academy of Sciences Section: Biological Sciences.

[46] R. E. Campbell, *et al.*, "A monomeric red fluorescent protein", *Proceedings of the National Academy of Sciences* **99**, 7877 (2002). DOI: 10.1073/pnas.082243699. Publisher: National Academy of Sciences Section: Biological Sciences.

[47] N. C. Shaner, *et al.*, "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein", *Nature Biotechnology* **22**, 1567 (2004). DOI: 10.1038/nbt1037.

[48] R. Y. Tsien, "The Green Fluorescent Protein", *Annual Review of Biochemistry* **67**, 509 (1998). DOI: 10.1146/annurev.biochem.67.1.509.

[49] F. V. Subach, V. V. Verkhusha, "Chromophore Transformations in Red Fluorescent Proteins", *Chemical Reviews* **112**, 4308 (2012). DOI: 10.1021/cr2001965.

[50] H. E. Chia, E. N. G. Marsh, J. S. Biteen, "Extending fluorescence microscopy into anaerobic environments", *Current Opinion in Chemical Biology* **51**, 98 (2019). DOI: 10.1016/j.cbpa.2019.05.008.

[51] T. Drepper, *et al.*, "Reporter proteins for in vivo fluorescence without oxygen", *Nature Biotechnology* **25**, 443 (2007). DOI: 10.1038/nbt1293.

[52] J. M. Landete, *et al.*, "Anaerobic green fluorescent protein as a marker of Bifidobacterium strains", *International Journal of Food Microbiology* **175**, 6 (2014). DOI: 10.1016/j.ijfoodmicro.2014.01.008.

[53] J. Walter, *et al.*, "Flavin Mononucleotide-Based Fluorescent Proteins Function in Mammalian Cells without Oxygen Requirement", *PLOS ONE* **7**, e43921 (2012). DOI: 10.1371/journal.pone.0043921.

[54] T. Drepper, *et al.*, "Flavin Mononucleotide-Based Fluorescent Reporter Proteins Outperform Green Fluorescent Protein-Like Proteins as Quantitative In Vivo Real-Time Reporters", *Applied and Environmental Microbiology* **76**, 5990 (2010). DOI: 10.1128/AEM.00701-10.

[55] L. A. Lobo, C. J. Smith, E. R. Rocha, "Flavin mononucleotide (FMN)-based fluorescent protein (FbFP) as reporter for gene expression in the anaerobe Bacteroides fragilis", *FEMS Microbiology Letters* **317**, 67 (2011). DOI: 10.1111/j.1574-6968.2011.02212.x.

[56] A. Mukherjee, C. M. Schroeder, "Flavin-based fluorescent proteins: emerging paradigms in biological imaging", *Current Opinion in Biotechnology* **31**, 16 (2015). DOI: 10.1016/j.copbio.2014.07.010.

[57] M. Wingen, *et al.*, "The photophysics of LOV-based fluorescent proteins – new tools for cell biology", *Photochemical & Photobiological Sciences* **13**, 875 (2014). DOI: 10.1039/C3PP50414J.

[58] S. Chapman, *et al.*, "The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection", *Proceedings of the National Academy of Sciences* **105**, 20038 (2008). DOI: 10.1073/pnas.0807551105.

[59] A. Mukherjee, *et al.*, "Engineering and Characterization of New LOV-Based Fluorescent Proteins from Chlamydomonas reinhardtii and Vaucheria frigida", *ACS Synthetic Biology* **4**, 371 (2015). DOI: 10.1021/sb500237x.

[60] A. Kumagai, *et al.*, "A Bilirubin-Inducible Fluorescent Protein from Eel Muscle", *Cell* **153**, 1602 (2013). DOI: 10.1016/j.cell.2013.05.038.

[61] Y. Shitashima, T. Shimozawa, A. Kumagai, A. Miyawaki, T. Asahi, "Two Distinct Fluorescence States of the Ligand-Induced Green Fluorescent Protein UnaG", *Biophysical Journal* **113**, 2805 (2017). DOI: 10.1016/j.bpj.2017.10.022.

[62] T.-L. To, Q. Zhang, X. Shu, "Structure-guided design of a reversible fluorogenic reporter of protein-protein interactions", *Protein Science* **25**, 748 (2015). DOI: 10.1002/pro.2866.

[63] Y. Shitashima, T. Shimozawa, T. Asahi, A. Miyawaki, "A dual-ligand-modulable fluorescent protein based on UnaG and calmodulin", *Biochemical and Biophysical Research Communications* **496**, 872 (2018). DOI: 10.1016/j.bbrc.2018.01.134.

[64] J. R. Wagner, J. S. Brunzelle, K. T. Forest, R. D. Vierstra, "A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome", *Nature* **438**, 325 (2005). DOI: 10.1038/nature04118.

[65] X. Shu, *et al.*, "Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome", *Science (New York, N.Y.)* **324**, 804 (2009). DOI: 10.1126/science.1168683.

[66] D. Yu, *et al.*, "An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging", *Nature Communications* **5**, 3626 (2014). DOI: 10.1038/ncomms4626.

[67] D. Yu, *et al.*, "A naturally monomeric infrared fluorescent protein for protein labeling in vivo", *Nature Methods* **12**, 763 (2015). DOI: 10.1038/nmeth.3447.

[68] J. B. Grimm, *et al.*, "A general method to fine-tune fluorophores for live-cell and in vivo imaging", *Nature Methods* **14**, 987 (2017). DOI: 10.1038/nmeth.4403.

[69] F. Stagge, G. Y. Mitronova, V. N. Belov, C. A. Wurm, S. Jakobs, "Snap-, CLIP- and Halo-Tag Labelling of Budding Yeast Cells", *PLOS ONE* **8**, e78745 (2013). DOI: 10.1371/journal.pone.0078745.

[70] S. S. Gallagher, J. E. Sable, M. P. Sheetz, V. W. Cornish, "An In Vivo Covalent TMP-Tag Based on Proximity-Induced Reactivity", *ACS Chemical Biology* **4**, 547 (2009). DOI: 10.1021/cb900062k.

[71] A. Keppler, *et al.*, "A general method for the covalent labeling of fusion proteins with small molecules in vivo", *Nature Biotechnology* **21**, 86 (2003). DOI: 10.1038/nbt765.

[72] A. Gautier, *et al.*, "An Engineered Protein Tag for Multiprotein Labeling in Living Cells", *Chemistry & Biology* **15**, 128 (2008). DOI: 10.1016/j.chembiol.2008.01.007.

[73] E. A. Rodriguez, *et al.*, "The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins", *Trends in Biochemical Sciences* **42**, 111 (2017). DOI: 10.1016/j.tibs.2016.09.010.

[74] F. H. Arnold, "Directed Evolution: Bringing New Chemistry to Life", *Angewandte Chemie International Edition* **57**, 4143 (2018). DOI: https://doi.org/10.1002/anie.201708408. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201708408.

[75] S. Lutz, "Beyond directed evolution—semi-rational protein engineering and design", *Current Opinion in Biotechnology* **21**, 734 (2010). DOI: 10.1016/j.copbio.2010.08.011.

[76] E. C. Alley, G. Khimulya, S. Biswas, M. AlQuraishi, G. M. Church, "Unified rational protein engineering with sequence-based deep representation learning", *Nature Methods* **16**, 1315 (2019). DOI: 10.1038/s41592-019-0598-1. Number: 12 Publisher: Nature Publishing Group.

[77] M. Carter Childers, V. Daggett, "Insights from molecular dynamics simulations for computational protein design", *Molecular Systems Design & Engineering* **2**, 9 (2017). DOI: 10.1039/C6ME00083E. Publisher: Royal Society of Chemistry.

[78] D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, "Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells", *Science* **296**, 913 (2002). DOI: 10.1126/science.1068539. Publisher: American Association for the Advancement of Science Section: Report.

[79] T. Gronemeyer, C. Chidley, A. Juillerat, C. Heinis, K. Johnsson, "Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling", *Protein Engineering, Design and Selection* **19**, 309 (2006). DOI: 10.1093/protein/gzl014.

[80] J. T.-H. Yeh, K. Nam, J. T.-H. Yeh, N. Perrimon, "eUnaG: a new ligand-inducible fluorescent reporter to detect drug transporter activity in live cells", *Scientific Reports* **7**, 41619 (2017). DOI: 10.1038/srep41619.

[81] D. R. Spring, "Chemical genetics to chemical genomics: small molecules offer big insights", *Chemical Society Reviews* **34**, 472 (2005). DOI: 10.1039/B312875J. Publisher: Royal Society of Chemistry.

[82] E. Cacace, G. Kritikos, A. Typas, "Chemical genetics in drug discovery", *Current Opinion in Systems Biology* **4**, 35 (2017). DOI: 10.1016/j.coisb.2017.05.020.

[83] I. Pavlinov, E. M. Gerlach, L. N. Aldrich, "Next generation diversity-oriented synthesis: a paradigm shift from chemical diversity to biological diversity", *Organic & Biomolecular Chemistry* **17**, 1608 (2019). DOI: 10.1039/C8OB02327A. Publisher: The Royal Society of Chemistry.

[84] J.-H. Zhang, T. D. Y. Chung, K. R. Oldenburg, "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays", *Journal of Biomolecular Screening* **4**, 67 (1999). DOI: 10.1177/108705719900400206. Publisher: SAGE Publications Inc STM.

[85] J. Inglese, *et al.*, "High-throughput screening assays for the identification of chemical probes", *Nature Chemical Biology* **3**, 466 (2007). DOI: 10.1038/nchembio.2007.17. Number: 8 Publisher: Nature Publishing Group.

[86] N. M. Koropatkin, E. C. Martens, J. I. Gordon, T. J. Smith, "Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices", *Structure* **16**, 1105 (2008). DOI: 10.1016/j.str.2008.03.017.

[87] J. R. Marchesi, *et al.*, "The gut microbiota and host health: a new clinical frontier", *Gut* **65**, 330 (2016). DOI: 10.1136/gutjnl-2015-309990.

[88] J. M. Chaparro, A. M. Sheflin, D. K. Manter, J. M. Vivanco, "Manipulating the soil microbiome to increase soil health and plant fertility", *Biology and Fertility of Soils* **48**, 489 (2012). DOI: 10.1007/s00374-012-0691-4.

[89] A. M. Buckley, *et al.*, "Lighting Up Clostridium Difficile: Reporting Gene Expression Using Fluorescent Lov Domains", *Scientific Reports* **6**, 23463 (2016). DOI: 10.1038/srep23463.

[90] E. C. Martens, N. M. Koropatkin, T. J. Smith, J. I. Gordon, "Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm", *Journal of Biological Chemistry* **284**, 24673 (2009). DOI: 10.1074/jbc.R109.022848.

[91] D. Chen, J. D. Brown, Y. Kawasaki, J. Bommer, J. Y. Takemoto, "Scalable production of biliverdin IXα by Escherichia coli", *BMC Biotechnology* **12**, 89 (2012). DOI: 10.1186/1472-6750-12-89.

[92] K. S. Karunatilaka, *et al.* (2013), vol. 8590, pp. 85900K–85900K–7.

[93] Y. Liu, *et al.*, "The Cationπ Interaction Enables a Halo-Tag Fluorogenic Probe for Fast No-Wash Live Cell Imaging and Gel-Free Protein Quantification", *Biochemistry* **56**, 1585 (2017). DOI: 10.1021/acs.biochem.7b00056. Publisher: American Chemical Society.

[94] T. Komatsu, *et al.*, "Real-Time Measurements of Protein Dynamics Using Fluorescence Activation-Coupled Protein Labeling Method", *Journal of the American Chemical Society* **133**, 6745 (2011). DOI: 10.1021/ja200225m. Publisher: American Chemical Society.

[95] X. Sun, *et al.*, "Development of SNAP-Tag Fluorogenic Probes for Wash-Free Fluorescence Imaging", *ChemBioChem* **12**, 2217 (2011). DOI: 10.1002/cbic.201100173. Publisher: John Wiley & Sons, Ltd.

[96] G. Lukinavičius, *et al.*, "Fluorogenic Probes for Multicolor Imaging in Living Cells", *Journal of the American Chemical Society* **138**, 9365 (2016). DOI: 10.1021/jacs.6b04782. Publisher: American Chemical Society.

[97] M.-A. Plamont, *et al.*, "Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo", *Proceedings of the National Academy of Sciences* **113**, 497 (2016). DOI: 10.1073/pnas.1513094113. Publisher: National Academy of Sciences Section: Physical Sciences.

[98] C. Szent-Gyorgyi, *et al.*, "Fluorogen-activating single-chain antibodies for imaging cell surface proteins", *Nature Biotechnology* **26**, 235 (2008). DOI: 10.1038/nbt1368. Number: 2 Publisher: Nature Publishing Group.

[99] N. G. Bozhanova, *et al.*, "Protein labeling for live cell fluorescence microscopy with a highly photostable renewable signal", *Chemical Science* **8**, 7138 (2017). DOI: 10.1039/C7SC01628J. Publisher: Royal Society of Chemistry.

[100] N. V. Povarova, *et al.*, "Red-Shifted Substrates for FAST Fluorogen-Activating Protein Based on the GFP-Like Chromophores", *Chemistry – A European Journal* **25**, 9592 (2019). DOI: 10.1002/chem.201901151. Publisher: John Wiley & Sons, Ltd.

[101] W. Sheng, *et al.*, "A Near-Infrared Photoswitchable Protein–Fluorophore Tag for No-Wash Live Cell Imaging", *Angewandte Chemie International Edition* **57**, 16083 (2018). DOI: 10.1002/anie.201810065. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201810065.

[102] J. Kwon, *et al.*, "Bright ligand-activatable fluorescent protein for high-quality multicolor live-cell super-resolution microscopy", *Nature Communications* **11**, 1 (2020). DOI: 10.1038/s41467-019-14067-4.

[103] D. M. Shcherbakova, P. Sengupta, J. Lippincott-Schwartz, V. V. Verkhusha, "Photocontrollable Fluorescent Proteins for Superresolution Imaging", *Annual review of biophysics* **43**, 303 (2014). DOI: 10.1146/annurev-biophys-051013-022836.

[104] R. N. Day, M. W. Davidson, "The fluorescent protein palette: tools for cellular imaging", *Chemical Society Reviews* **38**, 2887 (2009). DOI: 10.1039/B901966A. Publisher: Royal Society of Chemistry.

[105] H. E. Chia, T. Zuo, N. M. Koropatkin, E. N. G. Marsh, J. S. Biteen, "Imaging living obligate anaerobic bacteria with bilin-binding fluorescent proteins", *Current Research in Microbial Sciences* **1**, 1 (2020). DOI: 10.1016/j.crmicr.2020.04.001.

[106] T. Péresse, A. Gautier, "Next-Generation Fluorogen-Based Reporters and Biosensors for Advanced Bioimaging", *International Journal of Molecular Sciences* **20**, 6142 (2019). DOI: 10.3390/ijms20246142. Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

[107] A. M. Buckley, J. Petersen, A. J. Roe, G. R. Douce, J. M. Christie, "LOV-based reporters for fluorescence imaging", *Current Opinion in Chemical Biology* **27**, 39 (2015). DOI: 10.1016/j.cbpa.2015.05.011.

[108] A. Losi, K. H. Gardner, A. Möglich, "Blue-Light Receptors for Optogenetics", *Chemical Reviews* **118**, 10659 (2018). DOI: 10.1021/acs.chemrev.8b00163. Publisher: American Chemical Society.

[109] L. D. Lavis, "Teaching Old Dyes New Tricks: Biological Probes Built from Fluoresceins and Rhodamines", *Annual Review of Biochemistry* **86**, 825 (2017). DOI: 10.1146/annurev-biochem-061516-044839.

[110] A. G. Tebo, *et al.*, "Orthogonal fluorescent chemogenetic reporters for multicolor imaging", *Nature Chemical Biology* **17**, 30 (2021). DOI: 10.1038/s41589-020-0611-0. Number: 1 Publisher: Nature Publishing Group.

[111] S. Ko, J. Kwon, S.-H. Shim, "Enhanced UnaG With Minimal Labeling Artifact for Single-Molecule Localization Microscopy", *Frontiers in Molecular Biosciences* **8** (2021). DOI: 10.3389/fmolb.2021.647590. Publisher: Frontiers.

[112] E. A. Rodriguez, *et al.*, "A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein", *Nature Methods* **13**, 763 (2016). DOI: 10.1038/nmeth.3935.

[113] R. T. Jacob, *et al.*, "MScreen: An Integrated Compound Management and High-Throughput Screening Data Storage and Analysis System", *Journal of Biomolecular Screening* **17**, 1080 (2012). DOI: 10.1177/1087057112450186. Publisher: SAGE Publications Inc STM.

[114] A. Rouf, C. Tanyeli, "Bioactive thiazole and benzothiazole derivatives", *European Journal of Medicinal Chemistry* **97**, 911 (2015). DOI: 10.1016/j.ejmech.2014.10.058.

[115] S. Ghosh, *et al.*, "Blue protein with red fluorescence", *Proceedings of the National Academy of Sciences* **113**, 11513 (2016). DOI: 10.1073/pnas.1525622113.

[116] D. S. Wilson, A. D. Keefe, "Random Mutagenesis by PCR", *Current Protocols in Molecular Biology* **51**, 8.3.1 (2000). DOI: 10.1002/0471142727.mb0803s51.

[117] S. Iwatani, *et al.*, "A Novel Method for Measuring Serum Unbound Bilirubin Levels Using Glucose Oxidase–Peroxidase and Bilirubin-Inducible Fluorescent Protein (UnaG): No Influence of Direct Bilirubin", *International Journal of Molecular Sciences* **21**, 6778 (2020). DOI: 10.3390/ijms21186778. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

[118] O. S. Oliinyk, A. A. Shemetov, S. Pletnev, D. M. Shcherbakova, V. V. Verkhusha, "Smallest near-infrared fluorescent protein evolved from cyanobacteriochrome as versatile tag for spectral multiplexing", *Nature Communications* **10**, 279 (2019). DOI: 10.1038/s41467-018-08050-8. Number: 1 Publisher: Nature Publishing Group.

[119] Q. Zheng, *et al.*, "Rational Design of Fluorogenic and Spontaneously Blinking Labels for Super-Resolution Imaging", *ACS Central Science* **5**, 1602 (2019). DOI: 10.1021/acscentsci.9b00676. Publisher: American Chemical Society.

[120] J. B. Grimm, *et al.*, "A general method to optimize and functionalize red-shifted rhodamine dyes", *Nature Methods* **17**, 815 (2020). DOI: 10.1038/s41592-020-0909-6. Number: 8 Publisher: Nature Publishing Group.

[121] L. Wang, *et al.*, "A general strategy to develop cell permeable and fluorogenic probes for multicolour nanoscopy", *Nature Chemistry* **12**, 165 (2020). DOI: 10.1038/s41557-019-0371-1. Number: 2 Publisher: Nature Publishing Group.

[122] M. J. Rust, M. Bates, X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)", *Nature Methods* **3**, 793 (2006). DOI: 10.1038/nmeth929. Number: 10 Publisher: Nature Publishing Group.

[123] E. M. Smith, A. Gautier, E. M. Puchner, "Single-Molecule Localization Microscopy with the Fluorescence-Activating and Absorption-Shifting Tag (FAST) System", *ACS Chemical Biology* **14**, 1115 (2019). DOI: 10.1021/acschembio.9b00149. Publisher: American Chemical Society.