UNIVERSITY OF CALIFORNIA, MERCED

Engineering Protein Scaffolds for Ultrafast, Broad Range, Single-Molecule Analog pH Biosensors

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Bioengineering

by

Ameed Haikal Hashmi

Committee in charge:

Professor Wei-Chun Chin, Chair

Professor Andy LiWang

Associate Professor Eva De Alba

Professor Victor Muñoz (Advisor)

The Dissertation of Ameed Haikal Hashmi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Professor Andy LiWang

Associate Professor Eva De Alba

Professor Victor Muñoz, Faculty Advisor

Professor Wei-Chun Chin, Chair

University of California, Merced

2021

Dedication

Dedicated to Abbi

Acknowledgement

I would like to thank my advisor Dr. Victor Muñoz for his support and guidance through this research.

I would also like to thank my committee members, Dr. Andy LiWang, Dr. Eva De Alba, and Dr. Wei-Chun Chin for their mentorship and invaluable feedback.

To my wonderful family and friends, thank you deeply for your love and support through this rewarding journey.

This research was supported by the KECK Foundation and the NSF-CREST Center for Cellular and Biomolecular Machines (National Science Foundation-HRD-1547848.

Table of Contents

List of Figures

List of Tables

Vita

AMEED HASHMI

Education

- **Ph.D. (2021), Bioengineering,** University of California, Merced, California, USA.
- **B.Tech. (2012), Biotechnology,** ICFAI University, Dehradun, Uttarakhand, India.

Field of study

- **Protein Engineering,** UC Merced, California, Aug 2014 – Jul 2021.
    - Designing protein based ultra-fast biosensors having micro-second response time and optical readout for real-time applications.
    - Engineering proteins to behave as conformational rheostats having analog readouts rather than switches with binary readouts.
    - Demonstrating application of biosensors at the single-molecule level for sensing pH and ionic strength.

- **Nanoparticle Synthesis,** Indian Institute of Chemical Technology (IICT), Hyderabad, India, Jan 2012 – Apr 2013.
    - Used soft chemical techniques to control the particulate sizes of the nanomaterials, it resulted in six publications.
    - Worked on optimizing wet chemistry approaches to scale production of metal oxide nanoparticles.

Abstract

The research described in this thesis aims at exploiting recent important theoretical and experimental developments in protein folding towards developing advanced biosensors with improved/novel properties. Transforming this basic protein folding knowledge into engineering strategies for designing novel biosensors is an enormous challenge. Here we demonstrate the very first steps in that direction.

Biosensors based on proteins that naturally toggle between two states (unfolded/folded) upon specific binding to a target molecule have been successfully used for real-time sensing. These devices behave as conformational switch sensors. Principles on how to engineer this type of conformational transducer onto any protein of interest have also been laid out. Our work takes this state of the art forward to design high-performance conformational rheostat sensors. The rationale is to develop sensors with expanded dynamic range and faster response time by using as conformational transducer the coupling of binding to the analyte and the gradual folding process of fast folding protein modules, and fluorescence as optical readout. As proof of concept, we investigate the pH sensing capabilities of engineered proteins based on two scaffolds: i) an anti-parallel coiled-coil, and ii) a tandem array of the small downhill folding domain BBL. Our results reveal that such pH sensors exhibit a linear response over at least 4 orders of magnitude in proton concentration. We also demonstrate that a pH sensor based on a conformational rheostat transducer can produce analog pH readouts at the single-molecule level together with ultrafast response (< 1ms). These results lay the ground for the development of fluorescence biosensors for the analog monitoring of pH in real time and with nanoscale spatial resolution. Finally, we introduce a platform for the plug and play implementation of fully genetically encoded fluorescence conformational biosensors.

1. BIOSENSORS

Biosensors are integrated, self-contained, analytical systems that use biomolecular recognition to quantify and process signals of interest. Biosensors provide a basis to monitor and understand processes in and around us. Non-biologists will benefit from portable semi-quantitative or qualitative devices to make informed decisions about themselves and their relationships with the environment. In addition, quantitative biologists require technological improvements to probe -omics, such as genomics, proteomics and metabolomics *in vivo*.

To develop innovative biosensors that could address the sophisticated high throughput needs of biologists requires integration of multifaceted approaches. Classical titer methods that combine spectroscopic determination with immobilized aptamers, affibodies, peptides and molecule imprinted polymers are inadequate at this level. Integrated approaches provide a better way-forward to developing specific and sensitive biosensors capable of real-time, broadband, analog detection of analytes and biomarkers with nanoscale spatial resolution, ideally inside living systems.

Single-molecule biosensors can address the requirement for detection and quantification of biomarkers with ultra-high spatial resolution. Measuring an analog signal generated by an individual molecule upon its interaction with biological partners can enable creation of superior biosensors that can monitor processes inside living cells and their organelles at nanoscale resolutions. These biosensors can also benefit from the recent technical advances in single-molecule fluorescence detection of biomolecular interactions, which have attracted numerous applications to probe long-standing questions in biomolecular interactions (Holzmeister, Acuna et al. 2014), real-time conformational changes (Hohng, Lee et al. 2014), heterogenous catalysis (Chen, P., Zhou et al. 2014), enzymatic systems (Janssen, De Cremer et al. 2014) and complex biological systems (Larson, Rodgers et al. 2014). Detection of cells in complex mixtures or single-proteins can enable diagnosis of disease at its onset and has great potential for future point-of-care devices.

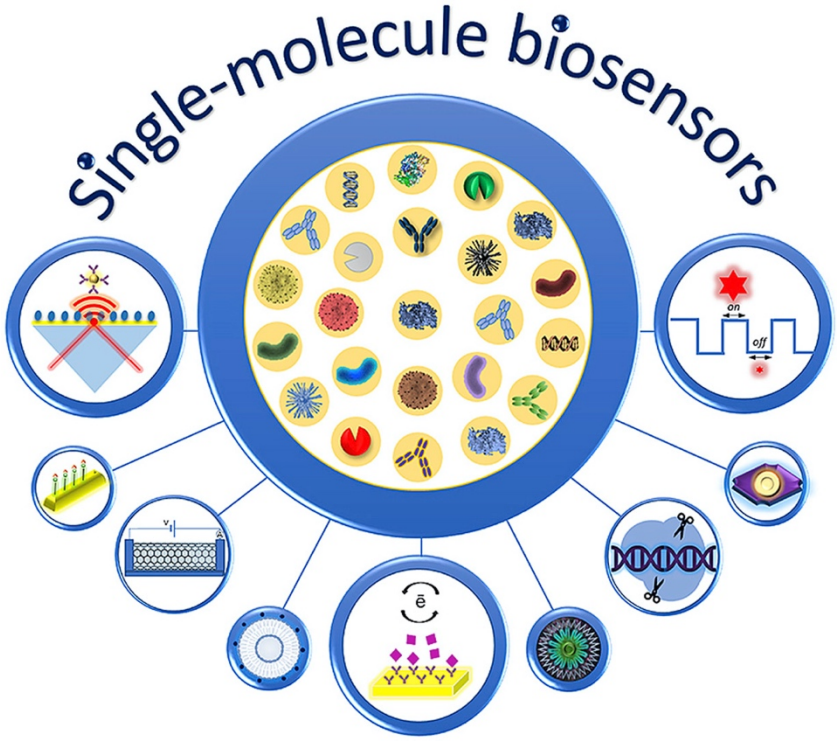

Figure 1.1 Approaches in single-molecule biosensors for in vitro and in vivo detections (Akkilic, Geschwindner et al. 2020). Figure reproduced with permission.

Major breakthroughs in single-molecule biosensors have made it possible to probe biologically important molecules in vitro and in vivo. Although the field is in its nascent stage, recent advances in electrochemical, plasmonic and SERS- based biosensors, as shown in Figure 1.1, aim to fulfill the need for real time detection of individual biological molecules and are helping uncover unique properties of individual molecules which are hidden when measured using ensemble averaging methods (Akkilic, Geschwindner et al. 2020). However, single-molecule methods are extremely challenging for biosensor development. Electrochemical biosensors require a homogenous surface electrode where the molecules need to be attached with a well-defined orientation and typically requires thousands of electrons flow to obtain a reasonable signal (Fan, Anderson et al. 2018, Gu, Q., Nanney et al. 2018). Plasmonic biosensors have great potential for measuring the size and concentration of the molecules but can only provide the dynamic process of the molecular translocation (Wang, Yuqin, Wang et al. 2019, Spitzberg, Zrehen et al. 2019). SERS reporter molecules can be heavily influenced by interactions with the metal and SERS based redox biosensors have not been applied yet to biologically relevant molecules (Zrimsek, Chiang et al. 2017). Detection is mostly limited by

the instrument's sensitivity, electric field drop, protein folding and nanomaterial incompatibility for these approaches. Fortunately, optical techniques and experimental methods have advanced further over the years and are becoming a promising tool for single molecule biosensing. Most widely, confocal microscopy and total internal reflection fluorescence microscopy (Balzarotti, Eilers et al. 2017, Gunnarsson, Snijder et al. 2015, Gu, L., Li et al. 2019, Day, Tao et al. 2016).

1.1. Fluorescence biosensors: strengths and limitations

Optical biosensors use the interaction of light with molecules to determine the concentration of analytes. These interactions can be measured through absorbance, scattering and/or fluorescence (Chen, C., Wang 2020). Optical sensors have widespread applications from healthcare industry to personal use, ranging from traditional sensors such as pulse oximeters to modern photo-plethysmograms for heart rate monitoring in wearable devices like smartwatches (Inui, Kohno et al. 2020). Today, optical biosensors continue to evolve in numerous applications, including cancer diagnostics (Balaji, Zhang 2017), DNA sensing (Lan, Yao et al. 2019), environmental monitoring (A. Halilović, E. Merdan et al. 2019, Liu, X., Yao et al. 2019) and food safety (Scognamiglio, Arduini et al. 2014).

Fluorescence biosensors offer numerous advantages that continue to spur research into its applications. Foremost, fluorescence biosensors have demonstrated high sensitivity by measuring analytes in femtomolar concentrations (Muneer, Ayoko et al. 2020). Such sensitivity makes optical biosensors ideal for measuring concentrations of very dilute samples in conditions that do not require stringent experimental conditions. However, it is also important to note that the performance of biosensors may be compromised when deployed outside of laboratory conditions. But optical biosensors can easily be integrated with microfluidic structures, providing a stable and closed biosensing environment. Microfluidics chips confer additional advantages, such as portability, reduced sample volumes and increased efficiency through multi-analyte detection (Liao, Zhang et al. 2019). Simultaneous measurement of multiple analytes, also known as multiplexing, is possible by measuring signals at different wavelength that are unique to each individual analyte of interest. Apart from improving efficiency of optical biosensors, multiplexing can be used to quantify an internal standard for self-calibration. Hence, improving the reliability and long-term accuracy of biosensors. These features complement the recent development and application of smartphone-based spectrometers and fluorimeters which are comparable in performance with research laboratory equipment (Kong, Gan et al. 2020, Rao, Liu et al. 2020).

As with all emerging technology, numerous regulatory and technical challenges impede adoption of novel optical biosensors for clinical applications. For

widespread acceptance in clinical settings optical biosensors must produce accurate and reproducible data, be sufficiently robust and ideally operate non-invasively. However, at present, the reliability, repeatability and stability of optical biosensors are seldom evaluated and reported in research articles (Ong, Pollard et al. 2021). Stability is a particularly important aspect of label-based biosensors. Protein-based biosensors having enzymes and antibodies may denature and lose their conformational shape and functionality. Even the performance of aptamers is influenced by pH, ionic strength and temperature, which could result in discrepancies between tests (Charbgoo, Soltani et al. 2016). Amongst the few studies that do evaluate the stability of their sensors, testing protocols are not standardized. Optical biosensors also do not have fast response times compared to their electrochemical counterparts, current optical biosensors suffer from need for an amplification reaction, that takes time and requires skilled technicians (Roda, Arduini et al. 2020).

1.2. Proteins as basis for biomolecular recognition

Biomolecular recognition is the process by which biomolecules recognize and bind to their molecular targets. Biomolecular recognition is typically driven by many weak (non-covalent) interactions working in concert. The most important of these interactions include: electrostatic interactions, hydrogen bonds and the hydrophobic effect. Because the aqueous environment significantly reduces the impact of electrostatic and induction interactions, the hydrophobic effect is often the dominant force stabilizing the formation of biomolecule-target complexes (Bruno da Silva, Contessoto et al. 2018, Munshi, Subramanian et al. 2019). Molecules in solution collide billions of times a second and in most cases the complexes formed by these collisions are weak, short-lived, and nonspecific (Baker 2000, Dobson 2003). But when the surface features of biomolecules are complementary to those of its target; stronger, long-lived, and specific interactions can be established. These specific complexes have proven to have significant biological and technological significance (Caro, Harpole et al. 2017, Ganguly, Otieno et al. 2012, Rossmann, J. Greive et al. 2017, Luzarowski, Skirycz 2019). Consequently, it has been exploited for a wide range of diagnostic and synthetic technologies (Azad, Tashakor et al. 2014, Zhang, X., Cui et al. 2016).

Proteins are molecules involved in key processes inside and outside of cells and are an ideal scaffold for biomolecular recognition. Although they are formed with a limited number of building blocks (twenty types of amino acids), proteins have complex three-dimensional (3D) structures from which their functions arise. Proteins vary in size from a few amino acids (peptides) to thousands of amino acids with the largest characterized being titin, having about 34,000 residues (Zacharchenko, von Castelmur et al. 2015). Proteins help in the survival of organisms by regulating and carrying out cellular and extra-cellular activities.

Evolution led to a wide variety of proteins that are responsible for structural and metabolic functions inside a cell. Consequently, some proteins are involved in molecular recognition and have ability to specifically bind to organic and inorganic molecules. Apart from being specific by acquiring selective active sites, proteins have structural motifs that interact with DNA or RNA (Ren, Robert et al. 2000). Peptides, as in the case of glutathione (Glu-Cys-Gly) can also be secreted by cells to influence the extracellular redox state (Castillo, Gáspár et al. 2004) or bind and promote absorption of nutrients (Dimise, Widboom et al. 2008). The siderophore like methanobactin peptide (Cys-Gly-Ser-Cys-Tyr-Pro-Cys-Ser-Cys-Met) secreted by bacteria has 6-7 $\times$ 10^{20} M^{-1} affinity for ionic copper Cu(I), although being just ten amino acid long (El Ghazouani, Baslé et al. 2011). Features like these make proteins ideal scaffolds for biosensing.

Furthermore, proteins are naturally evolved to work in complex crowded environments such as the cytosol that contains 2-4 million protein molecules per cubic micrometer (Milo 2013). To be able to recognize their target and form functional complexes, proteins evolved to be highly specific. Moreover, their structure and function can be engineered to a certain extent to achieve specific and desired physiochemical properties. Protein engineering has enabled the introduction of binding to novel molecules, such as converting the maltose binding protein into a zinc binding protein (Marvin, Hellinga 2001), enhancing of affinity for its epitopes of an antibody (Ashworth, Havranek et al. 2006), and altering the substrate specificity of enzymes (Harris, Craik 1998). Enzymes can also be engineered to develop protease biosensors, such as the peptide (Asp-Glu-Val-Asp), which has been inserted into enzyme luciferase and works as an apoptosis biosensor. When a cell enters apoptosis, caspase is expressed which cleaves the forementioned peptide resulting in the acquisition of luciferase activity and the production of luminescent signal when the substrate luciferin is present (Galbán, Jeon et al. 2013). Finally, proteins can be recombinantly produced even at industrial scales at affordable costs (Faccio 2019).

1.3. Protein folding coupled to binding: conformational switch

Nature solved the problem of real-time molecular sensing in complex environments long ago. Biomolecules respond to their targets by undergoing specific, binding induced changes in conformation or oligomerization (Gerstein, Krebs 1998). Lessons learnt from these natural, protein-based sensing can therefore assist in the development of improved sensor technologies. Inspired by the sensitivity, specificity and versatility of these naturally occurring sensors, ongoing advances in the field of protein engineering (e.g., computational design, directed evolution, selection strategies and labeling chemistries) are resulting in synthetic biosensors (Shi, S., Ang et al. 2018, Wang, Zhiqing, Doshi et al. 2020, Calzini, Malico et al. 2021, Koch, Pandi et al. 2019, d'Oelsnitz, Ellington 2018).

There are numerous advantages in the use of proteins as conformational switches for the design of biosensors. Foremost, is the ability of proteins to undergo binding-specific conformational changes, which offers a robust means of molecular transducing a binding event into output signal. This conformational switching is induced by the formation of numerous intramolecular and intermolecular weak interactions (e.g., hydrogen bonding, hydrophobic effect and van der Waals forces) and is generally specific to a given ligand-protein interface and thus largely insensitive to the presence of other molecules in complex environments. Additionally, signal transduction through conformational changes is rapid and reversible, potentially supporting real-time detection even inside living cells. Such transduction can also be implemented with a number of optical reporters with sufficient sensitivity to eventually function even at the single molecule level. Finally, the conformational equilibria are related to both the underlying thermodynamics of the protein as well as the ligand concentration, which makes it possible to rationally optimize the dynamic range of sensing without altering its binding specificity and getting quantitative results.

$$Non\ Binding\ State \underset{K_S}{\overset{switch}{\rightleftharpoons}} Binding\ Competent\ State \underset{K_A^L}{\overset{ligand}{\rightleftharpoons}} Bound\ State \qquad (1.1)$$

Conformational switching in proteins typically occurs via a population shift mechanism between a non-binding state and a state formed upon binding of the ligand, as shown in scheme 1.1. The thermodynamics of binding-induced biomolecular switches can be described via a 3-state thermodynamic model, in which ligand binding shifts a preexisting equilibrium between the binding-competent (native) and non-binding states (unfolded). The switching equilibrium constant (K_S) is a key factor defining the performance of a biosensor. Specifically, K_S must favor the non-binding state in order to obtain a large population change when it eventually folds upon binding to the ligand (L) and forms the bound/folded state (Vallée-Bélisle, Ricci et al. 2009). Folding and binding become effectively coupled when the binding free energy is sufficient to overcome the unfavorable free energy of folding. Conversely, coupling between binding and folding makes the overall affinity for the ligand to also depend on K_S. As K_S becomes smaller, binding must overcome the increasing cost of shifting the folding equilibrium and thus the overall affinity for the ligand is reduced. For these reasons, it is important to tune the K_S and binding affinity to optimize the detection level required for the specific application (Vallée-Bélisle, Plaxco 2010).

This recently developed strategy for engineering biosensors based on macromolecular conformational switches relies on manipulating the macromolecule's stability so that the binding of a ligand triggers folding. It has been

implemented on nucleic acids (Tyagi, Kramer 1996) and in some two-state folding proteins (Kohn, Plaxco 2005). However, there are intrinsic limitations for using conformational switches as biosensors. A critical one is their somewhat narrow dynamic range: a switch sensor is only sensitive to ~20-fold differences in concentration above and below the L_{50} (i.e., the concentration of ligand that results in 50% binding saturation). Additionally, the slow kinetics of two-state folding coupled to binding hamper their application as fast sensors. Finally, binary readouts make it impractical to employ conformational switches as single-molecule sensors (Cerminara, Desai et al. 2012).

To investigate the time response of sensors based on the conformational switch concept a simple kinetic model of folding coupled to binding is used. In this model, the protein is assumed to fold in a two-state manner (i.e., binary switch) and engineered to have an intrinsically unstable native state, so that it only populates the unfolded state when it is unbound to the ligand. The ligand is assumed to only bind to the native state. Therefore, the binding free energy compensates the intrinsic cost in free energy to folding, inducing the refolding of the protein, and thus eliciting the readout signal. This simple model can be represented with the following kinetic scheme:

$$U + L \underset{k_U}{\overset{k_F}{\rightleftharpoons}} N + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} FL \tag{1.2}$$

where k_F and k_U are the folding and unfolding rate constants for the protein sensor (in s^{-1}), k_{on} is the rate constant for the formation of the complex (in $M^{-1}s^{-1}$ units), and k_{off} is the complex dissociation rate constant (in s^{-1}). The microscopic dissociation equilibrium constant is defined as $K_d = k_{off} / k_{on}$, whereas the apparent dissociation equilibrium constant is determined by $L_{50} = k_U\ k_{off} / k_F k_{on}$. Therefore, a switching behavior is achieved when $k_U \gg k_F$ and the concentration of ligand compensates this penalty such as $k_F k_{on}[L]/k_U k_{off} > 1$. For the kinetic treatment of this system of second-order differential equations, the Eigen linearization method can be used, which assumes that, in conditions in which [L]>>[protein], the squared concentration displacement terms can be neglected. Using the linearization method, the differential equations can be solved by finding the eigenvalues and eigenvectors of the rate matrix K

$$K = \begin{bmatrix} -k_F & k_U & 0 \\ k_F & -(k_U + k_{on}[L]) & k_{off} \\ 0 & k_{on}[L] & -k_{off} \end{bmatrix} \quad (1.3)$$

This system has three eigenvalues: zero (corresponds to the equilibrium conditions) and two kinetic phases. From the eigenvectors of the zero eigenvalue one can obtain the equilibrium populations of the three species at different ligand concentrations. The eigenvectors of the two non-zero eigenvalues provide the changes in populations for the two kinetic phases of the system. For any condition in which k_U >> k_F, the population of N is always negligible, and the fast phase (largest non-zero eigenvalue) corresponds to a fast pre-equilibration between U and N. Under these conditions >99% of the kinetic flux is typically contained in the slow kinetic phase (smallest non-zero eigenvalue), which represents the relaxation kinetics between FL and the pre-equilibrated U and N species. Therefore, the slow phase is the one that produces the conformational change and the one that is relevant for the response time of a switch-based sensor (Cerminara, Desai et al. 2012).

1.4. Downhill folding protein: conformational rheostat

Protein folding transitions have been classically associated with a two-state mechanism, in which the protein toggles between the native (folded) state and a disordered (unfolded) ensemble, in analogy to an on-off switch. However, from a theoretical viewpoint a two-state mechanism is not a requirement for efficient folding (Naganathan, Muñoz 2010) or for function (Shoemaker, Portman et al. 2000). For example, the existence of partially disordered proteins such as insulin and protease inhibitors, whose folding is coupled to binding highlights that folding can be a regulatory mechanism (Sosnick, Krantz et al. 2006). Regulatory pathways commonly employ mechanisms controlled by allosteric transitions which function as quasi on-off switches depending on structural conformation (Passam, Chiu 2019, Mott, Owen 2018). The most common examples are the ATP-coupled chaperonin activity of the groEL/groES complex and the R- to T-state transitions of the oxygen transporter protein hemoglobin (Ishii 2017, Tekpinar, Zheng 2013). In contrast, conformational rheostats explore a continuum of partially disordered conformations, and can be thought of as analogs to electronic potentiometers (Nagpal, Luong et al. 2020). An example of such a system is the active site CXXC motif in thiol: disulfide oxireductases which have been shown to differentially control the flux of electrons involved in redox catalysis reactions. There are numerous examples of structural moderators (Li, Q., Nance et al. 2007), transcription factors (Rizzino 2008), and sensors (Moskvin, Kaplan et al. 2007) which are thought to commonly involve several chemical and physical interactions.

Signal transduction pathways are similarly controlled by concentration gradients of inducers and repressors (Wang, Da-Zhi, Jin et al. 2017, Bonnot, Nussaume et al. 2018).

Experimental and theoretical protein folding work over the last two decades has shown that folding free energy barriers are small (Akmal, Muñoz 2004), in some cases the barriers are marginal ($\leq$ 3RT) or disappear leading to downhill folding (Muñoz 2007, Gruebele 2008). With no folding barrier, the protein populates a single conformational ensemble (one-state) that becomes gradually unstructured as denaturational stress increases. The discovery of downhill folding challenged the simple assumption that specific conformational transitions in proteins operate as binary switches. Accordingly, two-state folders may act as binary switches owing to their binary conversion over a free energy barrier, but this leaves the door open to explain the function of downhill folders with an equally simple rationalization: that they behave as conformational rheostats (Garcia-Mira, Sadqi et al. 2002). It has been speculated that evolution could select among switching and rheostatic behaviors depending on the functional needs and the characteristics of the biological process (Muñoz, Campos et al. 2016, Rizzino 2008, Cerminara, Desai et al. 2012, Sharma, Sancho et al. 2017). It is likely that the importance of downhill folding is entrenched in the continuous structural landscape imparted by the single free energy well. Analogous to a concentration gradient, a signal could be tightly regulated by the degree of structure formation present in the folding ensemble as shown in Figure 1.2 (Zhu, Chen 2014, Cerminara, Campos et al. 2013, Cerminara, Desai et al. 2012). Additionally, a coordination event between a multi-enzyme or multi-domain complex could be controlled by the oscillatory fluctuations in the structure within such complex (Wainio-Theberge, Wolff et al. 2021, Kembro, Cortassa et al. 2014).

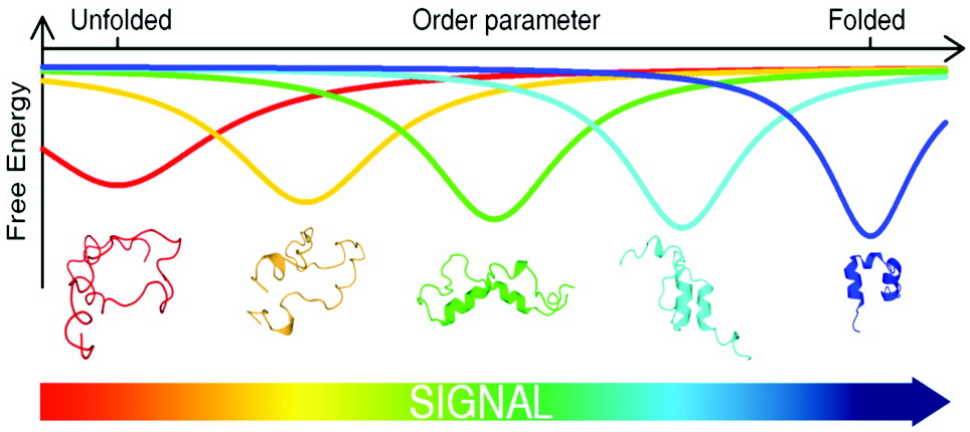

Figure 1.2 Molecular rheostat based on the coupling of a signal (e.g., proton binding) to the folding ensemble of a one-state downhill folding protein module (Cerminara, Desai et al. 2012). Figure reproduced with permission.

As a strategy to circumvent the limitations of conformational switches for biosensing applications, our study proposes to design and engineer proteins to operate as conformational rheostats and demonstrate their application in the development of high-performance fluorescence pH biosensors. Our premise was that the application of the "binding coupled to folding (or unfolding)" sensor-design principle to downhill folding (or fast folding) proteins could lead to conformational rheostat-based biosensors. These sensors are expected to display extended dynamic range, faster response (the rate limiting folding step is much faster) and most importantly analog readouts at the single-molecule level.

1.5. pH in living cells

For the implementation of rheostatic fluorescence biosensors we focused on measuring pH as analyte of reference. Thus, we use pH sensing as paradigm to develop proofs of concept for conformational rheostatic biosensors. However, being able to accurately measure pH is important to rationalize myriad biological phenomena. Virtually every known biological process is pH dependent reflecting the importance of the local pH. The concentration of hydrogen ions effects protein stability and activity (Talley, Alexov 2010). Protein-protein (Hom, Vora et al. 2007), protein-ligand (Re, Sesana et al. 2008) and protein-membrane (Re, Sesana et al. 2008) associations are greatly affected by the pH as well. Having the capabilities to monitor real-time pH fluctuations inside living cells would greatly improve our

understanding of these biological processes involving macromolecules and their assemblages (Yu, Hou et al. 2015).

In cells, unique functions of organelles determine the ambient intracellular pH distributions (pH 4-8) (Casey, Grinstein et al. 2010). For example, the lumen of mitochondria is basic (Ernster, Schatz 1981, Abad, Di Benedetto et al. 2004, Hajnóczky, Csordás et al. 2006, Rizzuto, De Stefani et al. 2012), nucleus and cytochylema are neutral (Roos, W 1981, Grinstein, Furuya et al. 1986), whereas endosomes and lysosomes are acidic (Turk, Turk 2009). Changes in pH can have detrimental effects in the cell such as generation of toxic free radicals, membrane contractility, cellular necrosis and apoptosis, etc. (Piwon, Günther et al. 2000, Edinger, Thompson 2004, Hara-Chikuma, Wang et al. 2005, Vaughan-Jones, Spitzer et al. 2009, Johansson, Appelqvist et al. 2010). It is for this reason, the proton gradient across intracellular compartment membranes and pH homeostasis is significant for cellular survival, function and proliferation (Trombetta, Ebersold et al. 2003, Rivinoja, Kokkonen et al. 2006, Urano, Asanuma et al. 2009).

2. METHODS FOR CHARACTERIZATION

2.1. Bulk Techniques

2.1.1. Circular Dichroism

Biomolecules of interest, such as proteins and DNA are chiral and hence are optically active. Chiral molecules are asymmetric, and such molecules are said to be optically active. When light of a certain wavelength passes through an absorbing sample, the intensity of the transmitted light is lower. In an optically active sample, due to molecular asymmetry, the refractive index (propagation velocity in the medium) for the two forms of circularly polarized light are different, giving rise to a change in the angle of the linearly polarized light. This is known as circular birefringence, or optical rotatory dispersion. In addition, there is also a difference in the extent of absorption of the two circularly polarized components of the linearly polarized light wave, by the optically active sample. This property is called circular dichroism (CD), with the measured quantity being the difference in absorption between right and left circularly polarized light by the asymmetric sample.

Proteins are highly asymmetric because of the asymmetric α-carbon centers and also their secondary and tertiary structure. Due to this super-asymmetry, optical activity arises because of the interaction between transition dipoles of absorbing chromophores. In the far-UV wavelength region (190-250 nm), an isolated amide group has an electronic absorption peak at 222 nm due to a weak $n\pi^*$ electronic transition, and another electronic absorption band with a peak at 195 nm due to a $\pi\pi^*$ transition.

In proteins, owing to the interaction between the several amide transition dipoles in an ordered α-helix, the $\pi\pi^*$ transition is predicted by the Exciton theory to have two components. One transition occurs at a lower energy with the transition dipole parallel to the helix axis, and the other at a higher energy, with the transition dipole perpendicular to the helix axis. Hence, the absorption due to $\pi\pi^*$ transitions is split into a negative at 208 nm due to the parallel component of the transition and positive at 190 nm due to the perpendicular component. A random coil on the other hand shows a negative peak at 195 nm, similar to an isolated amide bond, due to the $\pi\pi^*$ transition and another positive, but very weak, band above 210 nm due to the $n\pi^*$ transition. β-strands have a negative absorption band at about 215 nm and a positive band at about 200 nm. Proteins with a combination of α-helices, β-sheets, and β-turns show CD spectra with a linear combination from the different sources of absorbance.

An instrument measuring CD is called a spectropolarimeter and its output is typically in millidegrees. The differential absorption of the two circularly polarized components by an optically active substance yields a transmitted light beam with unequal amplitudes of the electric field vector of the two components. These combine to give an electric field vector which traces an elliptical path. The tangent of this ellipticity is related to the different amplitudes of the two circularly polarized components.

$$tan\theta = \frac{E_r - E_l}{E_r + E_l} = \frac{I_r^{\frac{1}{2}} - I_l^{\frac{1}{2}}}{I_r^{\frac{1}{2}} + I_l^{\frac{1}{2}}} \tag{2.1}$$

where, E represents magnitude of the electric field vectors for the right and left circularly polarized light, and I, their respective intensities. Upon approximating the tangent of the ellipticity to be numerically equal to the ellipticity itself (due to the small magnitude of the ellipticity), and replacing transmitted light intensities by the absorbance of the sample from Beer's Law, one can arrive at the expression for ellipticity in degrees:

$$\theta = \frac{\Delta A \cdot (ln10) \cdot 180}{4 \cdot \pi} = 32.99 \cdot \Delta A \tag{2.2}$$

where, ΔA is the difference in absorbance of the left and right circularly polarized light. Since the typical instrumental readout is in millidegrees, this read can be converted to give molar ellipticity $[\theta]$ in deg.cm^2.dmol^{-1}, which accounts for the path length l in units of cm traversed by light in the optically active medium, and the concentration C of the macromolecule in units of M.

$$[\theta] = \frac{\theta}{10 \cdot C \cdot l} \tag{2.3}$$

An additional normalization, included for proteins because of the repeating chromophoric amide bond, yields the mean residual ellipticity $[\theta]_{MRE}$ where, N is the number of amide bonds in the protein.

$$[\theta]_{MRE} = \frac{\theta_{readout}}{10 \cdot C \cdot l \cdot N} \tag{2.4}$$

Analysis of CD-spectra can be done using singular value decomposition (SVD). Singular value decomposition refers to the theory and procedure which states that any matrix A can be decomposed into a product of three matrices:

$$A = USV \tag{2.5}$$

where U is a unitary and orthogonal matrix, whose columns are an orthonormal basis set for the set of vectors in A. S is a diagonal matrix with the diagonal values representing singular values for the basis vectors in U in decreasing order of magnitude along the diagonal. Finally V is another unitary and orthogonal matrix containing the least-squared coefficients that regenerate the original matrix A upon multiplication with US.

This procedure finds immediate application in the field of spectroscopy, as any series of spectra that are a linear combination of a set of basis representative spectra can be split into their individual, linearly independent contributing components; i.e. a procedure similar to but more robust than traditional principal component analysis. These components are found in the columns of the first matrix U, with the first vector in U being the average representative vector or spectrum of the dataset in A. The singular values in S are numerical estimates of the contribution from each of the orthonormal basis vectors or spectra in U, to the overall series of spectra in A. Finally, the matrix V charts the changes in amplitude for each basis vector in U.

This procedure is advantageous for various reasons relative to principal component analysis. It offers the possibility of eliminating or reducing spectral noise, as the noise components often turns out as distinct vectors in the matrix U. Moreover, in spectroscopy, changes at particular wavelengths are indicative of changes in a specific structural property or spectral transition being observed. The vectors in the matrix V however are calculated after taking into account the entire spectrum, and contain the varying amplitude for each component vector in U.

For example, in CD experiments this would mean looking at the changes in the amplitude at 222 nm for the α-helix from the corresponding vector in V for the basis set in U that most resembles the average helical spectrum, usually the first column

for a helical protein, instead of simply the signal at 222 nm in the raw spectra. The second column in U can be representative of other structural signatures such as random coil, β-sheet, or β-turn or could be a complex mixture of signals. In general, red or blue shifts of a spectra, and relative correlations or anti-correlations between different structural signatures present in each spectral band measured by any procedure such as absorbance or fluorescence can also be judged from the orthonormal basis set of vectors produced in the matrix U by SVD. Finally, SVD allows for the possibility of reducing the effective rank of a rank-degenerate large data-set, thus reducing the noise without introducing artificial digital filters.

2.1.2. Fluorescence: Resonance Energy Transfer

Electronic absorption of light refers to the interaction of a quantum of light having a specific energy with an electron of a molecule resulting in its excitation from a ground state (S_0) to a higher quantized electronic energy state. The excitation from the ground state can lead electrons to jump to its first singlet state (S_1). Excited electrons can undergo several processes: traversing several vibrational levels or rotational relaxations in their excited singlet state or crossing over to the excited triplet state, by non-radiative processes before returning to its ground state by spontaneously emitting a photon (Kasha's rule).

This phenomenon involving spontaneous emission of photons by molecules from its excited singlet state to its ground state is called fluorescence. Fluorescing molecules typically emit light at wavelengths longer than the excitation wavelength due to what is known as Stoke's shift. This shift is seen because of rapid interconversion between vibrational or rotational states and manifests as heat loss to its surrounding thereby reducing the amount of energy carried by the emitted photon. In practice, the probability of spontaneous emission of a photon by an excited molecule, the intensity of fluorescence and the final wavelength of the emitted photon depend on the properties of the molecule, properties of the surrounding solvent and specific interactions that the fluorophore may experience when in the excited state.

It is for these reasons that fluorescence is widely used in the study of protein folding, since it allows the possibility of tracking changes in the local environment of specific fluorescing groups. Tryptophan, tyrosine and phenylalanine comprise the intrinsic fluorophores in proteins. It is common to chemically attach extrinsic fluorophores to proteins to monitor conformational changes, since fluorophore absorption, emission, rotational correlation time and anisotropy are sensitive to the environment. These several possibilities are summarized in the Jablonski diagram shown in Figure 2.1.

Resonance energy transfer (RET) describes the non-radiative transfer of energy from an excited donor to an acceptor in its vicinity. The class of RET that is described here is heterotransfer, wherein the energy is transferred between two different kinds of chromophores. This energy transfer process occurs through interaction between the transition dipoles of the donor and acceptor chromophores. Theodor Förster connected the rate of this energy transfer phenomenon to the rates of fluorescence decay and non-radiative energy dissipation in the donor by the Förster rate equation and the distance of separation between the donor and acceptor.

$$k_T = (k_D + k_{Di})\left(\frac{R_0}{r}\right)^6 = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \tag{2.6}$$

where k_T, k_D and k_{Di} are the rates of energy transfer from the donor to the acceptor, the rate of decay of fluorescence of the donor, and the rates of non-radiative dissipation of the excited electron in the donor, respectively. τ_D is the intrinsic fluorescence lifetime of the donor, r is the inter-dye distance and R_0 is the characteristic Förster distance or Förster radius. The energy transfer efficiency E describes the fraction of energy transferred from the donor to the acceptor in comparison with the total energy absorbed by the donor.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} = 1 - \frac{\tau_D}{\tau_{DA}} = 1 - \frac{Q_D}{Q_{DA}} \tag{2.7}$$

where τ_D is the lifetime of fluorescence of the donor and τ_{DA} in the presence of both the donor and acceptor, and Q_D and Q_{DA}, are donor fluorescent quantum yield in the absence and presence of the acceptor, respectively. Quantum yield is the probability of spontaneous emission of a photon by an excited molecule. Since the energy transfer depends on the interaction between molecular transition dipole moments, a 1/6th distance dependence is observed. The Förster distance is a number that defines the distance at which 50% of the energy absorbed by the donor is passed on to the acceptor. It is arrived at based on a derivation by Förster, which connects this distance to the spectral properties of the donor and the acceptor.

$$R_0 = \left(\frac{9000 \cdot ln10 \cdot \kappa^2 \cdot Q_D \cdot J}{128 \cdot \pi^5 \cdot \eta^4 \cdot N_A}\right)^{1/6} \tag{2.8}$$

where κ^2 is an orientation factor describing the angular orientation between the transition dipole moments of the donor and acceptor, relative to the physical angle between the two chromophores, Q_D is the intrinsic quantum yield of the donor, J is the spectral overlap integral between the donor and acceptor, η is the refractive index of the bulk solvent, and N_A is the Avogadro number. The overlap integral J can be calculated from first principles to arrive at a wavelength dependent expression below.

$$J = \int f_D(\lambda) \cdot \varepsilon_A(\lambda) \cdot \lambda^4 \cdot d\lambda \tag{2.9}$$

where $f_D(\lambda)$ is the normalized donor fluorescence signal and $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor at the wavelength λ. The Förster theory can be readily applied to a pair of carefully selected chromophores to serve as a 'spectroscopic ruler' to measure distances between the donor and acceptor.

Förster Resonance Energy Transfer (FRET) as a technique can thus be used to gauge the distance between chromophore pairs. Since, FRET is a highly distance-dependent phenomenon it has become a popular tool to measure the dynamic activities of biological molecules. One common application is simply to measure the distance between two positions of interest on a biomolecule. If the biomolecule only involves one donor and one acceptor group, the distance between the donor and the acceptor can be easily measured if there is no conformational change within this process. Besides, if the biomolecule has a huge conformational change, one can also measure the dynamical activities between two sites on this macromolecule. Today, this technique is widely applied in many fields such as single-molecule experiments, molecular motors, biosensors and DNA mechanical movements. Upon proper consideration in quantity of chromophores and precise positions in macromolecules, it can be used in-vivo.

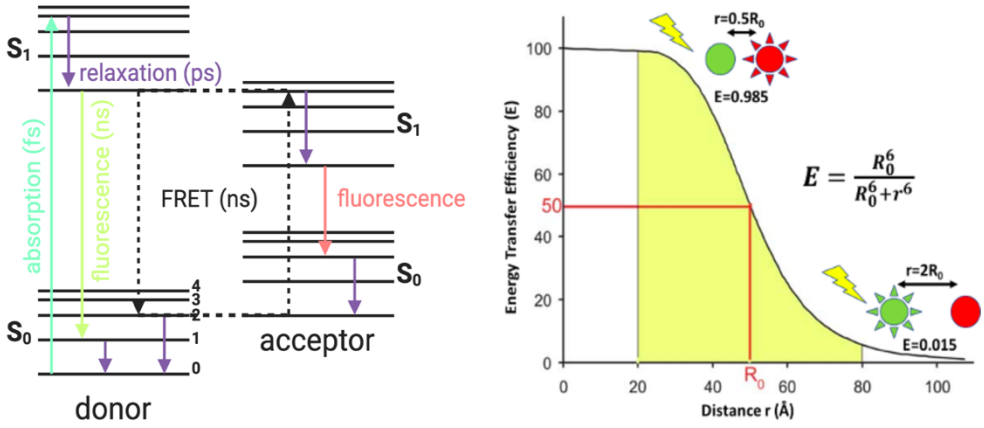

Figure 2.1 Jablonski diagram and Förster resonance energy transfer (FRET)

The limitations of FRET is that this transfer process is effective only when the separating distance of donor-acceptor pair is smaller than 10 nanometers, having a practical range of $r = 0.5R_0$ to $2R_0$. Since, FRET only gives the information about distances, if a dramatic change happens, such as aggregation or conformational kinks, it is hard to resolve the exact movement of dye pair. Apart from considerations concerning the distances, the refractive index of the solvent η and the orientational averaging of the donor and acceptor κ^2 cannot be overlooked. Fast averaging of the donor and acceptor transition dipoles, relative to fluorescence decay lifetimes of the donor and the rate of motion of the donor relative to the acceptor, results in a value of 2/3 for complete isotropic dynamic averaging of the donor and acceptor, which is assumed to hold true for applications to protein spectroscopy. This however, is not a trivial assumption to make and care needs to be taken in considering interactions of the donor or acceptor with its environment, and a comparison of the lifetime of fluorescence decay with anisotropic motions of the donor or acceptor.

2.1.3. Fluorescence Quenching: Photoinduced Electron Transfer

Quenching requires molecular contact between the fluorophore and a quencher. For quenching to occur the fluorophore and quencher have to come into molecular contact, allowing the electron clouds of both molecules to interact. Because the electron clouds are strongly localized, electron density falls off very rapidly with distance, quenching requires molecular contact at the van der Waals radii. This contact can be through diffusive encounters referred to as dynamic quenching, or

through complex formation, which is static quenching. Because of the requirement of close contact, the extent of quenching is sensitive to molecular factors that affect the rate and probability of contact, like steric shielding and charge-charge interactions. In contrast, resonance energy transfer is a through-space interaction that occurs over longer-range dipole-dipole interactions.

In photoinduced electron transfer (PET) a complex D^+A^- is formed between the electron donor D and the electron acceptor A upon excitation. This charged complex may emit photons as an exciplex or be quenched and return to the ground state without emission of a photon by transfer of the extra electron on the acceptor back to the electron donor. The direction of electron transfer in the excited state is determined by the reduction and oxidation potential of the ground and excited states. Donor and acceptor do not form a complex when both are in the ground state because it is unfavorable energetically. When the fluorophore is excited, it provides the energy required for charge separation and the energy decreases resulting in the ability to donate or accept electron changes. The important part of this process is the decrease in the total energy of the charge transfer complex.

The change is free energy of separation ΔG_{cs} can be estimated using the Rehm-Weller formula to determine the PET efficiency

$$\Delta G_{cs} = E_{ox} - E_{red} - E_{0,0} + C \tag{2.10}$$

In this equation the first one-electron oxidation potential E_{ox} describes the process $(D^+ + e \rightarrow D)$ and the first one-electron reduction potential E_{red} describes the process $(A + e \rightarrow A^-)$. $E_{0,0}$ is the energy of the zero-zero transition to the lowest singlet excited state of the fluorophore. The last term C is the solvent dependent Coulombic interaction energy experienced by the ion pair following the electron transfer reaction. The reaction rate at van der Waals contact follows an exponential distance dependence.

The Rehm-Weller equation has its origin in the free energy change in an electric potential. The energy released by a mole of electrons moving in a potential is given by

$$\Delta G = -n \cdot F \cdot \Delta E \tag{2.11}$$

where, n is the number of electrons in the reaction, F is the Faraday constant, which is the charge on 1 mole of electron and ΔE is the potential.

2.2. Single Molecule Techniques

Single-molecule methods have evolved into one of the mainstream techniques used to probe conformational dynamics and interactions of biomolecules. The ability of single molecule experiments to access information without ensemble averaging makes it an attractive technique to obtain distributions of molecular properties. There are two basic techniques used in SMF detection: confocal microscopy and total internal reflection microscopy.

2.2.1. Confocal Microscopy

The conception of confocal microscopy is often attributed to Minsky (Minsky 1988), who published an initial patent for a confocal microscope in 1961 (Patent: US3013467 A). This method enables in-effect optical sectioning through a sample. The word ‘confocal’, implying ‘having the same focus’, refers to the presence of two pinholes which are conjugated in the same image plane. One of them is used for spatial filtering of the excitation laser beam by removing side lobes at a position where the laser beam is focused, and the other for the emitted light path, which eliminates a significant proportion of stray signals coming from above or below the focal plane. The first practical working confocal microscope was built by the pioneering efforts of Eggar and Petran (Egger, Petran 1967), whose first biological application was reported in 1967 to visualize unstained nerve cells in the brain. Confocal microscopy can significantly improve the effective signal-to-noise ratio for detection by removing out-of-focus fluorescence. Many recent single-molecule studies utilize a confocal microscope primarily simply to generate a ‘confocal excitation volume’ (Shashkova, Leake 2017).

Experimental set-up for single-molecule FRET involves an excitation light source focused to a diffraction-limited spot by a high numerical aperture (N.A. = 1.49) objective to get a small confocal volume (femtoliter). In free diffusion-based experiments, sufficiently diluted sample molecules (concentration ~ 50 pM) ensure probability (<1%) of observing single molecule. Single molecules are then recorded from the photon arrival trajectories using time-correlated single photon counting detectors as they diffuse through the illumination volume. The simplest way to analyze SM-FRET data involves grouping donor and acceptor counts into fixed time intervals (bins). Experiments are done for long enough time to obtain a statistically relevant number of bins to get a distribution of transfer efficiency. The distribution of transfer efficiencies is called FRET efficiency histogram (FEH) and

gives information about various conformations of protein present under the experimental conditions. Bins are selected by applying a photon threshold to ensure the presence of a molecule and the binned data is the converted into E using the equation

$$E = \frac{n_A}{n_D + n_A} \tag{2.12}$$

where, n_D and n_A are the number of donor and acceptor photons per time bin. A FRET efficiency histogram (FEH) is plotted from which the number of molecules, their average E and relative populations are obtained. This demonstrates an important advantage of single-molecule technique over the ensemble techniques where averaged-out values of transfer efficiency are obtained. Despite their apparent simplicity, obtaining quantitative information from FEH is extremely challenging. FEH is a reduced representation of the whole-time stamped photon trajectory and includes shot noise, inevitable blinking and bleaching of the dyes as well as the stochastic fluctuations in measured count rate from the Brownian dynamics of the protein as it traverses the confocal volume. Additionally, when the conformational fluctuations occur in timescales comparably slower or faster than the bin times, FEH get dynamically averaged, and conformational dynamics cannot be resolved.

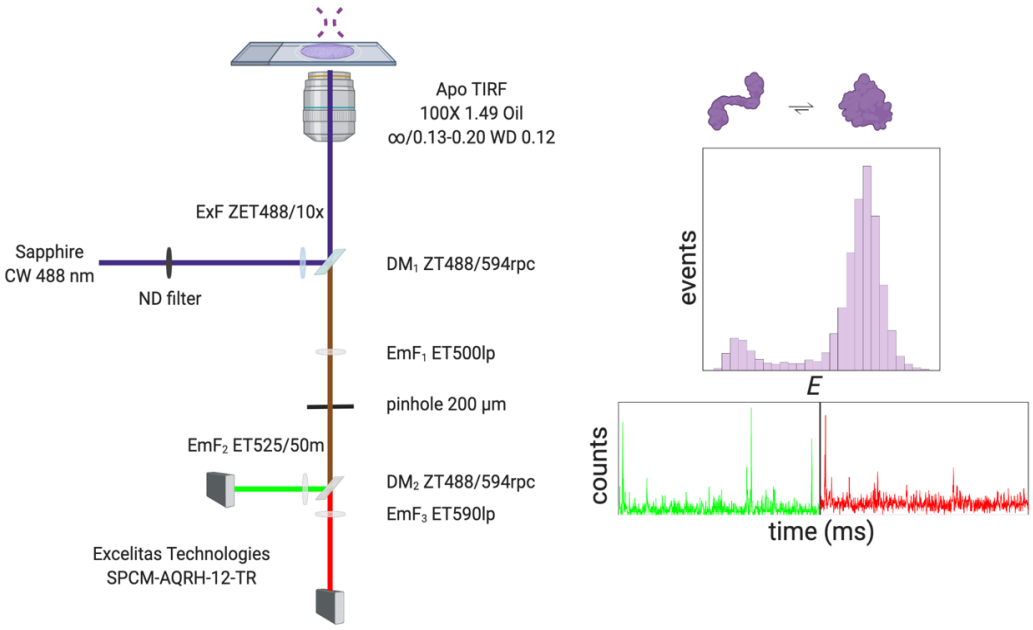

Figure 2.2 Schematic of the two-color single-molecule FRET confocal instrumentation set-up involves an excitation light source, confocal microscope, optics for collecting and separating donor and acceptor emission, single-photon detectors, and photon counting device. Time tagged data is collected and analyzed to generate FRET efficiency histograms (FEH).

Apart from the low collection efficiency of the setup, photon emission is a stochastic process and hence detection of the limited number of photons emitted under single-molecule conditions is accompanied with a fluctuation in the number of photons detected. This results in the broadening of the FEH even when the inter-dye distance is not changing. The variance caused by this stochastic fluctuation of photons is called shot noise and is given by

$$\sigma_{SN}^2 = \frac{E(1-E)}{N_T} \tag{2.13}$$

where E is the energy transfer efficiency and N_T is the number of photons in the bin. As a result of shot noise broadening, the temporal resolution of SM-FRET with binning analysis is set by the minimum time required to collect sufficient number of photons. So, to determine an E = 0.5 using 25 photons would result in an uncertainty of ±0.10. Conversely, the uncertainty can be reduced to ±0.01 if 2500 photons are collected.

2.2.2. Total Internal Reflection Fluorescence Microscopy

Total internal reflection fluorescence (TIRF) is now one of the most frequently used imaging methods in single-molecule fluorescence microscopy. The method is based on total internal reflection of incident excitation light from a glass–water interface, such as between a glass coverslip/slide and a water-based physiological buffer. TIRF in effect delimits the excitation field to result in selective illumination and excitation of fluorophores that are positioned close to the coverslip/slide surface (in practice, this 'close distance' can be approximated as being, very roughly, the depth of penetration itself). Thus, TIRF is particularly valuable for identifying single fluorescently labeled molecules immobilized on to a glass slide/coverslip. Due to the fact that TIRF detects only minimal signals from the out-of-focus regions, the signal-to-noise ratio is significantly improved, enabling better contrast for detecting single molecules (Axelrod 1981, Shashkova, Leake 2017).

The physical phenomenon of total internal reflection happens when a propagating light exceeding the critical angle θ_c of incidence encounters a boundary to a medium of lower refractive index. Its refractive behavior is governed by Snell's law. Total internal reflection does not occur as a new phenomenon at the critical angle, rather it is a continuous transition from a scenario in which refraction is predominant, and reflection corresponds to a very small fraction of the incident light, to a scenario in which all the light is reflected (total reflection) when the critical angle is exceeded. At the critical angle, refraction occurs at 90° and all the light is entirely reflected back into the incident medium. Snell's law is reduced to ($\eta_1 \cdot \sin\theta_c = \eta_2$) and the critical angle can be expressed as

$$\theta_c = \sin^{-1}\left(\frac{\eta_2}{\eta_1}\right) \tag{2.14}$$

For an incident light passing through a glass coverslip with a refractive index of 1.49 and encountering an aqueous solution (refractive index of 1.33), the critical angle is ~63°. When the angle of incident light is above 63° total internal reflection will occur. Even though the light no longer propagates into the buffer, there will be

a tiny amount of penetration of the reflected light across the interface, which propagates parallel to the surface of the glass, creating an electromagnetic field in the buffer. This field is referred to as the evanescent field and can be described as partially emerging from the glass and travelling some distance in the solution before re-entering back into the glass. This propagation distance is called the Goos-Hänchen shift. The size of this shift ranges from a fraction of a wavelength when the incident light is perpendicular to the interface, to nearly parallel at the critical angle, having a sinusoidal characteristic period which becomes longer with the increasing angle. The evanescent field intensity decays exponentially with increasing distance from the glass interface according to the following equation:

$$I(z) = I(0)e^{-z/d} \tag{2.15}$$

where, $I(z)$ represents the intensity from the interface perpendicularly along the z-axis and $I(0)$ is the intensity at the interface. The characteristic penetration depth d of the evanescent wave from a 488 nm laser excitation incident at 80° is approximately 56 nm.

$$d = \frac{\lambda}{4 \cdot \pi \cdot \sqrt{\eta_1^2 \cdot \sin^2\theta - \eta_2^2}} \tag{2.16}$$

The exponential decay of the evanescent field intensity makes it possible to confine the excitation of fluorophores to the penetration depth, and hence reduces background immensely and affords single-molecule detection. The minimal background results in a much higher signal-to-noise ratio compared to confocal fluorescence microscopy given that this optical sectioning is approximately one-tenth. This enhanced signal makes it great for single-molecule fluorescence detection using low illumination intensities. Apart from the enhanced SNR, the advantages of using TIRF is the high axial resolution and ability to observe multiple individual molecules at the same time (imaging). TIRF measurements require the molecules to be very close (i.e., within 100 nm), and this can be achieved by surface immobilization.

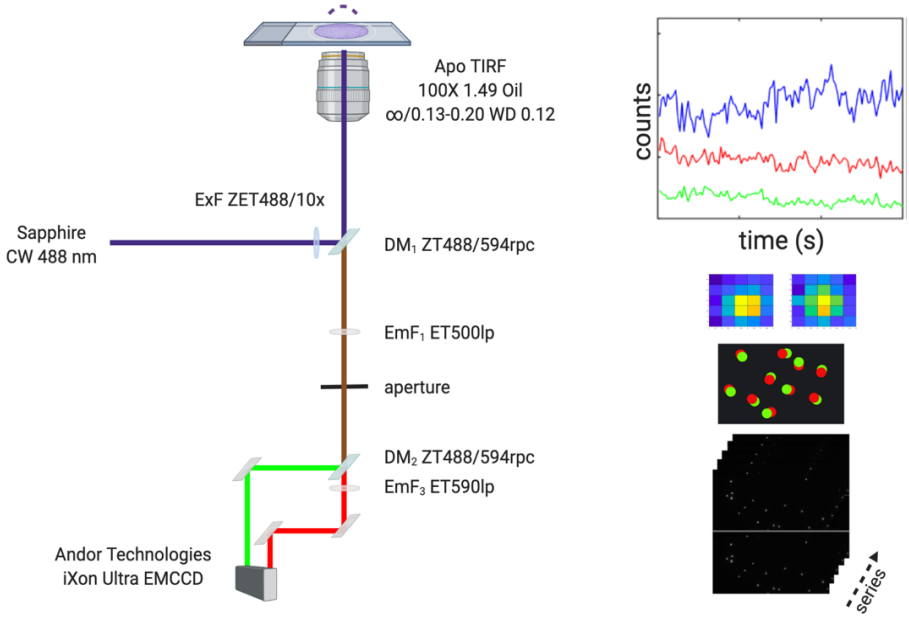

Figure 2.3 Schematic of the two-color single-molecule TIRF instrumentation set-up involves an excitation light source, confocal microscope, optics for collecting and separating donor and acceptor emission and an EMCCD camera. Images are analyzed to get the FRET efficiency.

Often times it is convenient to immobilize the labeled protein onto the surface to track it for as long as needed. The difficulty is to avoid nonspecific interactions between the very sticky labeled protein and the glass surface. The most straightforward procedure uses recombinant DNA technology to attach to one end of the protein an immobilization tag that binds to a chemical group present on a suitably derivatized glass surface. There are multiple solutions depending on the immobilization tag and how it is attached the biotin–streptavidin linkage, His tag binding, and use of antibodies (Chung, Eaton 2013, Pal, Lesoine et al. 2005, Chung, Louis et al. 2009). An alternative to a direct tether is to encapsulate single protein molecules into phospholipid vesicles, which are then tethered to the glass surface. This method, developed by Haran and coworkers (Boukobza, Sonnenfeld et al. 2001), has been used for studies in this thesis. Vesicle encapsulation is allegedly a much milder procedure and permits to probe the protein while is freely diffusing but contained in a volume much smaller than the confocal volume (the

average vesicle volume is about 2000 times smaller than the confocal volume), thus guaranteeing uniform irradiation (Wang, Z., Campos et al. 2016).

2.3. Protein Labeling with Fluorophores for Fluorescent Readouts

The application of SM-FRET to study protein folding is challenging. Labeling proteins with bulky extrinsic fluorophores without perturbing folding is not trivial and requires careful examination and its thoughtful placement in the amino acid sequence. Changes in the microenvironment around the fluorophores can affect the photosensitivity leading to decreased count rates before the dyes become inactive. Such low count rates normally result in millisecond time resolution of the technique (Roy, Hohng et al. 2008). A common preferred site-specific labeling approach is the use of maleimide-thiol conjugation using cysteine. Cysteine is a relatively rare amino acid, especially in non-secreted proteins. An existing cysteine can be used or alternatively it can be introduced at a carefully selected location by site-directed mutagenesis. Sometimes, one of the two cysteines is more reactive, allowing site-specific labeling with donor and acceptor dyes by two-step sequential conjugation reactions. More often, the conjugation reaction of proteins with two cysteines yields a mix of donor only, acceptor only, donor-acceptor, acceptor-donor or with none. Subsequent purifications can separate these, but the mix of labeling positions is often unavoidable. Nevertheless, if both dyes have sufficient rotational freedom when conjugated to the protein, the change in FRET should still be accurate as it is distance dependent.

The maleimide-thiol reaction is very specific when performed in mildly basic (pH < 7.5) conditions however, the thiol chemo-selectivity is lost and the maleimide moiety begins to react with free amines (e.g., lysine) above pH 7.5. Furthermore, at the single-molecule level one can run into problems that is not commonly observed when analyzing results from bulk experiments. For example, sample heterogeneity stemming from stickiness of dyes due to incomplete hydrolysis and its subsequent stereoisomerism. Although maleimide-thiol adducts can be stabilized through stretching (Huang, Wu et al. 2019), it is easier incubating overnight in mildly basic conditions (see Appendix B). For comprehensive experimental details on labeling and other protocols refer to Appendix A. Included below in Table 2.1 are the physical properties of the dyes that we used, and Figure 2.4 shows the pH dependent stability of the same as a control. It was necessary to confirm pH stability of the dyes as most fluorophores show huge pH dependence in their photo-physical behavior.

Table 2.1 Properties of fluorophores

Dye	Molecular Weight	λ_{abs} (nm)	λ_{fl} (nm)	τ_{fl} (ns)	Quantum Yield	ε_{max} ($M^{-1}.cm^{-1}$)
Alexa 488	698	493	516	4.1	0.92	72000
Alexa 594	886	588	612	3.9	0.66	92000
Atto 655	812	663	684	1.9	0.30	125000

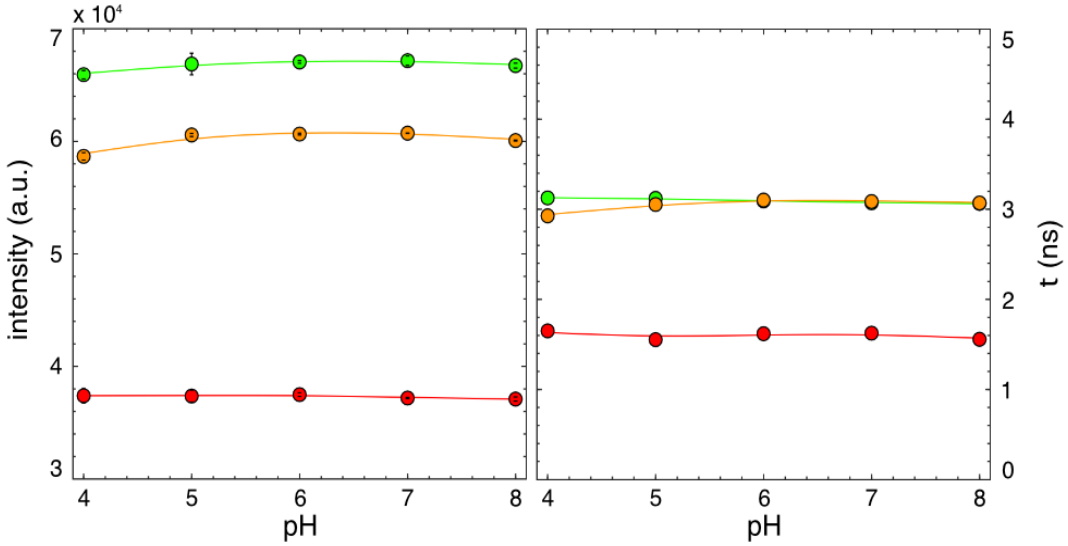

Figure 2.4 pH stability of extrinsic fluorophores used: Alexa 488 (green), Alexa 594 (orange) and Atto 655 (red). Intensity scaled to brightness (left) and lifetimes (right) of the dyes.

3. COILED-COIL BASED SCAFFOLDS

In this chapter we take on the challenge of demonstrating the design of conformational transducers based on the coupling of the folding of a scaffold protein upon ligand binding and their potential for broadband sensing (many orders of magnitude in ligand concentration). As protein scaffold we chose the antiparallel coiled-coil fold. We aimed to implement a transducer for pH sensing based on a histidine grafting strategy and characterized its sensitivity to pH. Finally, we introduced a high-sensitivity fluorescence readout to create a protein-based biosensor. The overall goal was to demonstrate general principles for biosensor design and contribute to develop engineering tools for the manipulation of the conformational properties of proteins to result in broadband (wide dynamic range) biosensors.

An antiparallel coiled-coil contains two helices that are super coiled around each other and connected by a loop. Coiled-coils are structural motifs that usually contain the repeating pattern of heptad repeats $[hxxhpxp]_n$ of hydrophobic (h) and polar (p) amino-acid residues. The positions in the heptad repeats are usually labeled abcdefg; where a and d generate the hydrophobic core, b and c are usually helix inducing, e and g favor inter-helical salt bridges and f can be helix inducing or used to control solubility of the protein as shown in Figure 3.1 (Malashkevich, Higgins et al. 2015, Zhang, W., Zheng et al. 2018, Higgins, Malashkevich et al. 2014, Pratap, Luisi et al. 2013, Bjelić, Wieser et al. 2013). The packing of the side chains of the hydrophobic amino acids in the coiled-coil interface is exceptionally tight, with almost perfect van der Waals contact (Crick 1952).

To design our pH sensor, we decided to use an antiparallel coiled-coil protein as our initial scaffold. This protein fold has two antiparallel α-helices connected by a loop, adopting a left-handed supercoil tertiary structure. Coiled-coils are easy to design and work with, the length of the helices can be tweaked to control distances between fluorophores. Monomeric coiled-coils are known to be fast folding proteins with microsecond folding times (Donten, Hassan et al. 2015). The high stabilization in coiled-coil proteins comes largely from local interactions and its gradual, heterogenous unfolding is probe dependent (Emerson Holtzer, Larry Bretthorst et al. 2001). This together with their small size (~70 residues) make coiled-coils likely candidates for downhill folding (Sancho, Muñoz 2011).

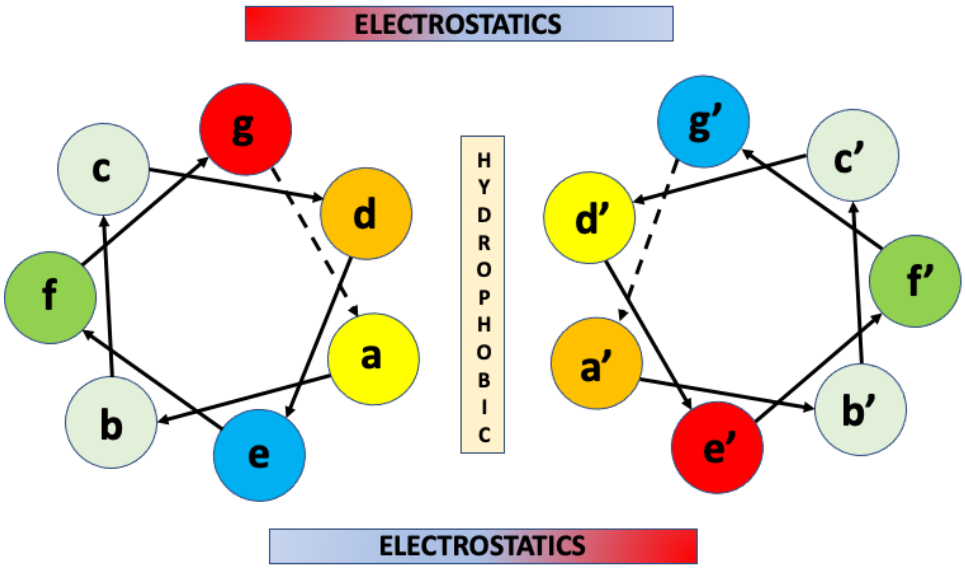

Figure 3.1 Schematic wheel for design of canonical anti-parallel coiled-coils.

3.1. Engineering the Coiled-Coil Scaffold

As starting point for our design, we selected the coiled-coil region of an RNA binding protein (pdb:1NT2). AGADIR (Muñoz, Serrano 1994) was used to calculate helical propensities. The original sequence showed high helical propensity on the second helix and low in the first, and the sequence also had relatively high helix propensity in the connecting loop.

Several sequence changes were then made to increase helicity in helix 1. Positive charges were eliminated, and a capping box was added to the beginning of the first helix. N-capping is used to nucleate helix wherein the CO group of the *i* residue is hydrogen-bonded to the NH group of residue *i+4* (or sometimes residue *i+3*) (Presta, Rose 1988). This was followed by replacing Ser (S23) and Asn (N25 and N28) in central positions of helix 1 with stereo-chemically similar amino acids with enhanced helical propensity: Ala (A) and Gln (Q), respectively. These changes led to a large, predicted increase in helicity for helix 1: from an average of about 4-5% to 40%. Val (V27) was replaced with Gly (G) to better delimit the helices from propagating into the loop. In helix 2, to improve the definition of the helix end, Gly (G) was incorporated for C-terminal capping by replacing Pro (P73). At lower pH (~4, AGADIR includes predictions as a function of pH (Muñoz, Serrano 1995)) helix 1 is more stable (rises to 60%) and helix 2 becomes less stable (decreases to 40%, especially at the end), both changes compensate one another. To get our final

coiled-coil scaffold, which we term Control, three additional mutations [helix 1: Ala (A16) to Lys (K) and Ile (I20) to Val (V); helix 2: Arg (R56) to Val (V)] were made to make more room for the bulky sidechain of His (H) residues that will be incorporated to introduce the pH transducer (see next section). These last mutations were introduced as one per heptad, and into the a or d positions of the coiled-coil. The theoretical pI of the Control was 4.64. Robetta (Raman, Vernon et al. 2009) was used to model the structure of the Control coiled-coil, as shown in Figure 3.2.

3.2. Introducing the pH Conformational Transducer

The rationale for introducing a pH conformational transducer into the coiled-coil was to couple the ionization of certain residues to the folding-unfolding equilibrium of the protein. Such coupling is achieved when residues with ionizable size chains experience a difference in pK_a between that in the unfolded state (when is similar to the intrinsic pK_a of the free amino acid) and the folded state. Such change of pK_a arises when the residue engages in non-covalent interactions with the rest of the folded protein. These interactions are basically the degree of solvation of the charge and electrostatic interactions with charged residues and dipoles. When the interactions are favorable, the pK_a shifts in response to the stabilization (higher population) of the ionized species. In contrast, destabilizing interactions shift the pK_a in the opposite direction. If present in a protein, ionizable amino acids with large enough pK_a shifts will shift the folding-unfolding equilibrium of the protein in response to the changes in the concentration of hydronium ions (protons).

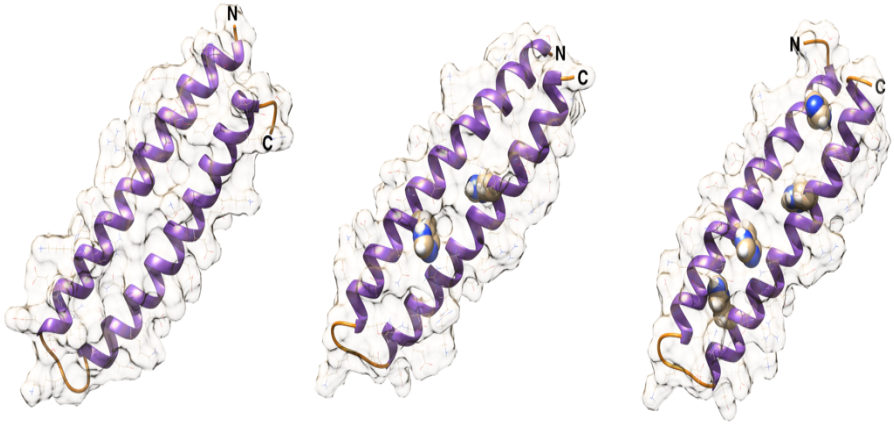

Figure 3.2 Modeled structures of coiled-coils: Control (left), 2-Histidine (middle) and 4-Histidine (right).

To introduce a conformational transducer sensitive to pH into the coiled-coil scaffold, we resourced to a histidine grafting strategy. The idea is to engineer histidine (His) grafts into its hydrophobic core to induce folding via His's sidechain ionization (Nagpal, Luong et al. 2020). His is useful for our application because the His sidechain has pK_a around 6, which is close to physiological pH. It is an essential amino acid and has a positively charged imidazole functional group. The unprotonated imidazole is nucleophilic and can serve as a general base, while the protonated form can serve as a general acid. A pK_a shift can occur when His is buried in the hydrophobic core of the protein depending upon its solvation. We made two subsequent designs to introduce buried His residues which we call 2-Histidine (I24H and V60H) and 4-Histidine (V10H, I24H, F46H and V60H). This resulted in slight changes in pI from 4.64 (Control), to 4.81 (2-Histidine) and 5.00 (4-Histidine) respectively. We also modeled the structures of the three scaffolds with Robetta (Raman, Vernon et al. 2009) as shown in Figure 3.2. It is to be noted that such models are structural models of the protein and not determined experimentally.

3.3. Characterization of the pH Conformational Transducer

The conformational behavior of the control and histidine-grafted coiled-coils as a function of pH was studied using far-UV circular dichroism (CD) to monitor the α-helix content of the scaffold as indicator of its degree of folding. We measured the complete far-UV CD spectrum, but given that these coiled-coils are either helical

or unfolded, we could use the intensity of the CD signal at 222 nm as proxy of the average helical content of the coiled-coil at each condition (Greenfield 2006). The results of these experiments are shown in Figure 3.3. To estimate the helical content of the coiled-coils as a function of temperature and pH we used the following equations:

$$Helicity\ (\%) = \frac{[\theta]_{222}}{[\theta]_{222}^{max}} \times 100 \tag{3.1}$$

$$[\theta]_{222}^{max} = -\left[1 - \left(\frac{2.5}{n}\right)\right] \times 40{,}000\ deg.cm^2.dmol^{-1} \tag{3.2}$$

where, $[\theta]_{222}$ is the mean residue ellipticity at 222 nm and n is the number of peptide bonds that are in helical conformation. Since the coiled-coil folded structures (see Figure 3.2) have about 30x2 helical amino acids , the $[\theta]_{222}^{max}$ was assumed to be about -38,500 deg.cm^2.dmol^{-1} for all percent helicity calculations.

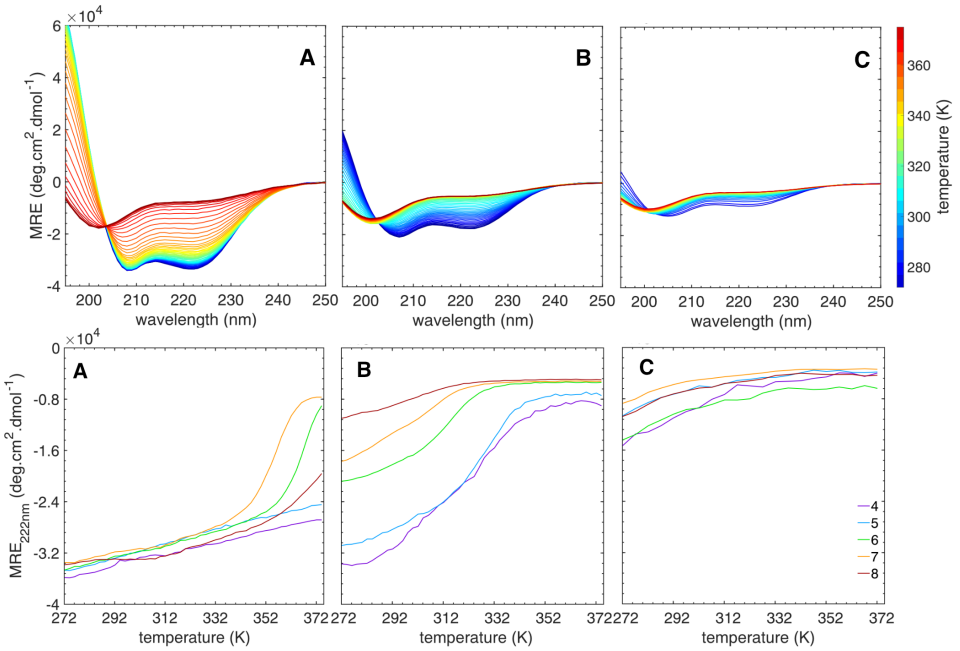

Figure 3.3 Thermal denaturation (ramping temperature from -1 ℃ to 101 ℃) of coiled-coils at pH 7 using circular dichroism (top); Control (A), 2-Histidine (B) and 4-Histidine (C). The mean residue ellipticity at 222 nm as a function of pH (bottom); pH 4 (purple), pH 5 (blue), pH 6 (green), pH 7 (yellow) and pH 8 (red). Experiments were done in 20 mM phosphate and citrate buffers (ionic strength adjusted to 100 mM with NaCl) at the different pH values using (5 μM and 20 μM, 400 μL) proteins.

From the coiled-coil CD thermal denaturation experiments as a function of pH we established the baselines for the folded (full helices supercoiled) and unfolded (random coil) states. The mean residue ellipticity baselines moved from approximately -34,000 deg.cm^2.dmol^{-1} at -1 °C for the folded Control to roughly -4,000 deg.cm^2.dmol^{-1} at 101 °C for the unfolded 4-Histidine. 4-Histidine was unfolded at all pH values and even at the lowest accessible temperatures its helical content did not increase significantly, indicating that histidine protonation was incapable to compensate the penalty of so many bulky sidechains in the core of the coiled-coil and thus failed to make this scaffold fold into a coiled-coil structure. As a result, temperature had no significant effect on the ability of 4-Histidine construct to fold nor did pH: it remained as a random coil at all conditions. In contrast, the Control was folded at all pH but showed some degree of helix to coil transition at the highest temperatures and lowest pH values, possibly owing to the

ionization of Glu (E) and Asp (D) residues below pH 5. At neutral pH; Control is mildly cooperative with a T_m (the midpoint of the unfolding transition) of about 85 °C. The most interesting behavior was observed in the 2-Histidine construct where the T_m changed from 25 °C at pH 8 to 54 °C at pH 4. The stability of the protein increased with decrease in pH. 2-Histidine was unfolded in basic condition and as the pH was made more acidic, the helical content increased and the protein folded back into a coiled-coil from a random coil. These coupled transitions due to pH and temperature showed little cooperativity, consistent with a conformational rheostat-like unfolding.

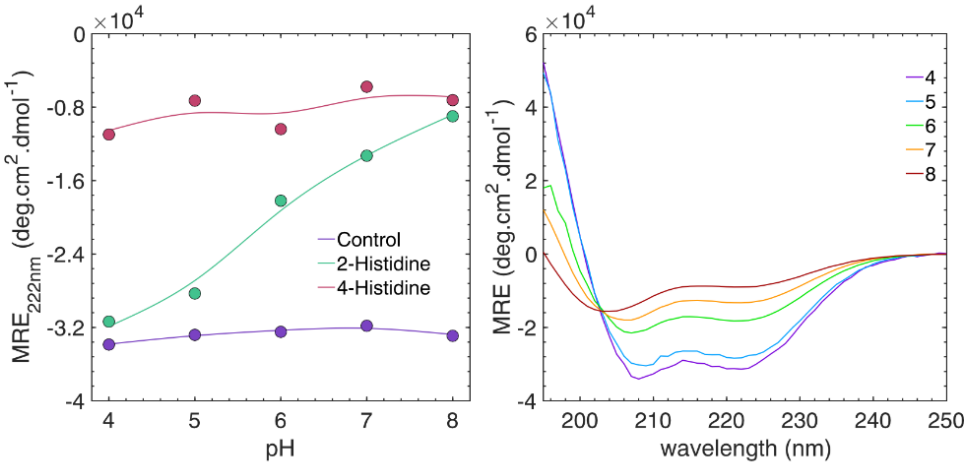

Figure 3.4 Steady state circular dichroism bulk experiments on the three coiled-coiled constructs. Effect of pH on the helical content as monitored by the mean residue ellipticity at 222 nm (left) and far UV spectra for 2-Histidine as a function of pH (right).

As shown in Figure 3.4, the Control was folded (with about 87% helicity) across the entire pH range, and 4-Histidine remained unfolded (about 20% helicity) throughout. However, for 2-Histidine, the helicity increased linearly as pH decreased. Particularly, the helical content of 2-Histidine was about 25% at pH 8; and it increased to 33% at pH 7. This trend follows as the conditions become increasingly acidic, with the helical content increasing to 46% at pH 6, 74% at pH 5 and to 83% at pH 4. This conformational behavior with an effective linear change in helicity of roughly 60% over 4 orders of magnitude in hydrogen ion concentration in solution demonstrates that the engineered coiled-coil scaffold together with the histidine grafting strategy results in an efficient broadband conformational pH transducer.

For a scaffold to be useful for sensing applications, it needs to exhibit a large change in conformation with pH at physiological temperatures. Room temperature is conventionally taken to be varying between 20-25 °C, but it could also be important to function in a broader temperature range, including lower temperature and 37 °C, which would be useful for in vivo applications. Figure 3.5 (left) shows the change in CD signal with pH of the 2-Histidine construct at various temperatures in the 10 to 37 °C range. These results demonstrate that this engineered protein is highly sensitive to pH in the entire temperature range, although the maximum signal change decreases slightly at physiological temperature. These results indicate that our engineered 2-Histidine coiled-coil is an excellent choice for pH transducing in the 4-8 pH range and from room to physiological temperatures. Furthermore, the sensitivity to temperature as well as pH will make it necessary to calibrate the sensor at each given temperature.

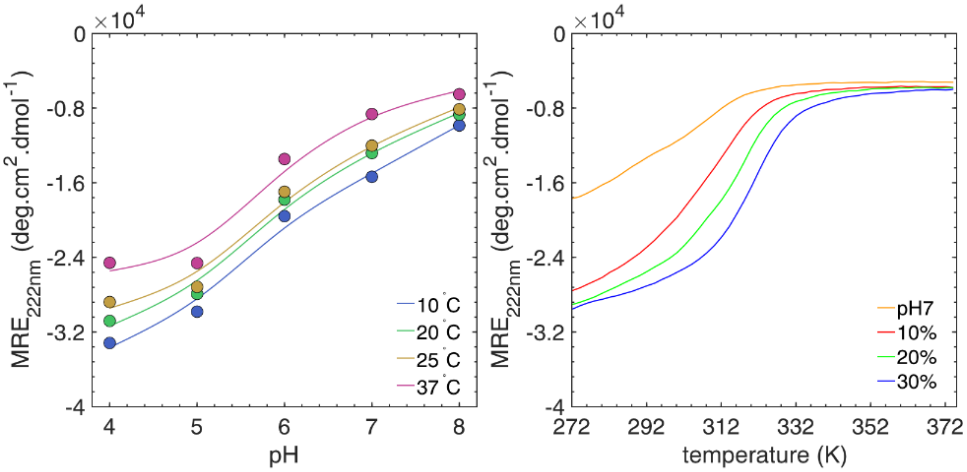

Figure 3.5 Stability of 2-Histidine at physiological temperatures (left) and changes in mean residue ellipticity at 222 nm as a function of temperature at neutral pH with addition of glycerol (right).

Additionally, we explored the stability of the 2-Histidine coiled-coil using glycerol as a cosolvent. Glycerol is known to shift the native protein ensembles to more compact states and is routinely used as protein stabilizer in aqueous solution (Doose, Neuweiler et al. 2005, Vagenende, Yap et al. 2009). Preferential interactions of proteins in glycerol/water mixtures mainly originates from electrostatic interactions. Furthermore, glycerol interacts with large patches of contiguous hydrophobic residues where glycerol acts as an amphiphilic interface between the hydrophobic surface and the polar solvent. These interactions shift

the native protein toward more compact conformations (Vagenende, Yap et al. 2009). As shown in Figure 3.5 (right), with increasing concentration of glycerol the T_m of 2-Histidine increases from 35 °C with no glycerol to 49 °C at 30% glycerol at neutral pH. Additionally, examination of T_m with 20% glycerol across the pH range resulted in an effective change of 20 °C between pH 8 (41 °C) and pH 4 (61 °C). These experiments demonstrate that we can tune the response of our coiled-coil pH transducer by simply modulating the native stability of the scaffold via cosolvents or further engineering its sequence.

3.4. Introducing a Förster Resonance Energy Transfer (FRET) Readout

In order to implement a high sensitivity fluorescence readout of pH into the 2-Histidine coiled-coil we implemented a FRET strategy by introducing the donor and acceptor fluorophores at the ends of the protein. The idea is that the folded protein (at low pH) will result in a very short distance between the two fluorophores (< 2 nm according to our structural model) and hence at very high transfer efficiency distances (very high acceptor and low donor emissions). On the other hand, as pH raises, the unfolding of the coiled-coil is expected to produce a significant increase in the end to end distance, and hence a drop in FRET efficiency. To explore this strategy, we introduced two Cys residues in specific locations of the coiled-coil using recombinant DNA technology. The thiol group of the Cys side-chain can be labeled specifically using commercially available maleimide derivatives of fluorophores (see protein labeling section in Appendix A). We designed two different versions of the FRET readout (Figure 3.6), one with Cys added to the two termini of the 2-Histidine coiled-coil (we call this version 2-Histidine-FRET, or 2H-F) and a second version including a short flexible linker (Lys-Lys-Gly) connecting the Cys to coiled-coil (we call this version 2-Histidine-linker-FRET, or 2H-l-F). The linker was introduced to make the Cys more accessible for conjugation with fluorescent dyes and provide the fluorophore with added flexibility to facilitate reaching the dynamic isotropic averaging conditions of the two fluorophores that results in an orientation factor of 2/3.

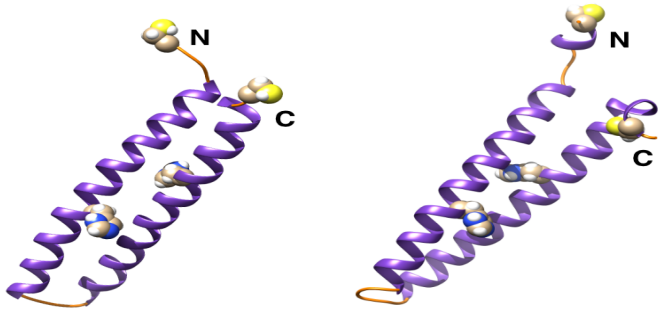

Figure 3.6 Modeled structures of 2-Histidine coiled-coils for FRET: 2H-F (left) and 2H-I-F (right).

We could not sucessfully label these variants uniformly with both fluorophores for proper FRET measurements, despite trying numerous chemical labeling strategies. Of all of the attempted labeling approaches, the best results were obtained using a sequential labeling strategy in the presence of 6M urea. Furthermore, these heterogeneously labeled 2H-F samples could not be efficiently separated, even after using analytical reverse phase high-performance liquid chromatography (HPLC).

Although the labeling was incomplete (a very large fraction of the molecules had donor but no acceptor), we could still observe a slight change in FRET efficiency with pH (Figure 3.7). If we focus on the changes in acceptor emission (not sensitive to the heterogeneous labeling), we could see for 2H-F the acceptor emission was more than duplicated between pH 8 and 4 in 2H-F, highlighting an increase in FRET efficiency that is consistent with the large structural/folding transition of the coiled-coil upon acidification that we observed by CD (Figures 3.3-5). However, the donor intensity was much, much higher at all pHs, and only decreased by 10% between pH 8 and 4. As a consequence, the estimated FRET efficiency changed by ~ 6%. The most likely explanation for these results is that there is a large fraction of the coiled-coil molecules labeled with only donor (one or two), which would obviously be unresponsive to pH as they cannot undergo FRET. A large donor-only fraction greatly increases the base donor fluorescence, hiding the true change in FRET efficiency. The implication is that the FRET readout could result in much larger signal changes if the sample were more uniformly labeled with donor and acceptor, i.e., by either attaining higher labeling efficiency, or higher sample homogeneity via purification. This labeling problem was similar for 2H-I-F, which showed virtually no change in FRET efficiency, as shown in Figure 3.7. One possible reason behind the poor dual labeling of this protein is that the Alexa dyes that we employed, which are negatively charged to increase their water solubility,

strongly bind to helix 2 which has a high positive charge at the neutral pH required for the reaction thus potentially interfering with the coupling reaction to the protein ends. The structure of the 2-Histidine coiled-coil showing its distribution of charged residues is provided in Appendix B.

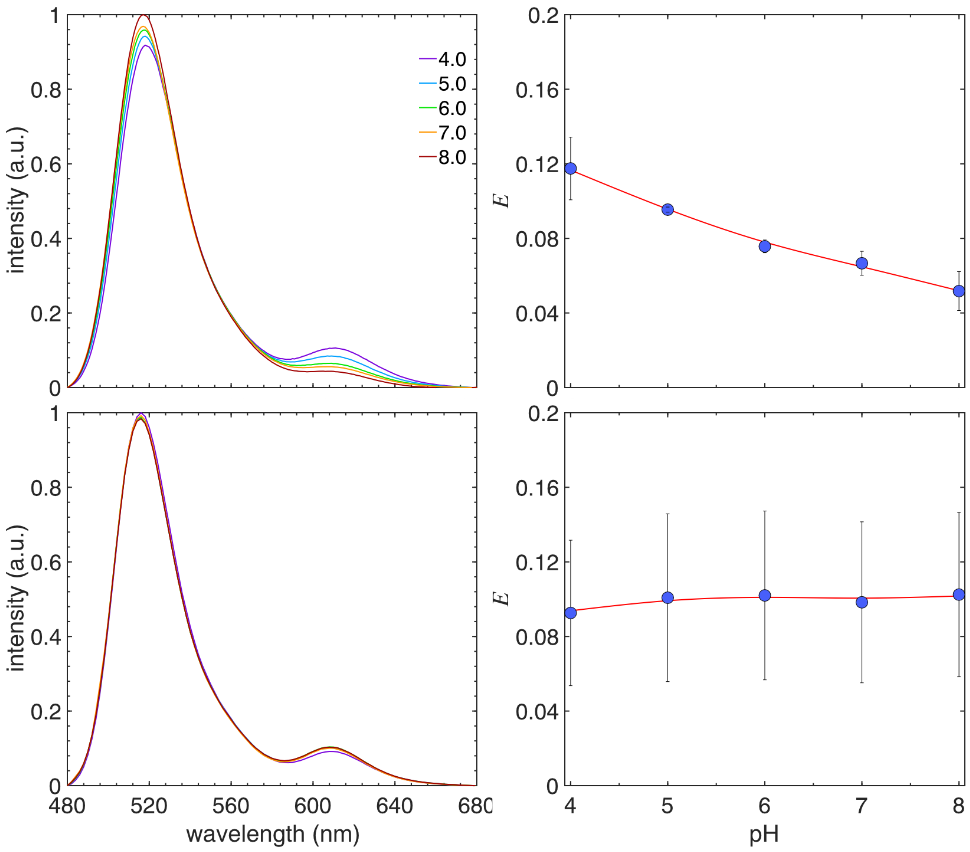

Figure 3.7 Ensemble steady state fluorescence intensity measurements (left) and its corresponding FRET (right) on 2-Histidine constructs, 2H-F (top) and 2H-I-F (bottom). All ensemble steady-state fluorescence experiments were done in 20 mM buffers (ionic strength adjusted to 150 mM with NaCl) using labeled protein (10 nM, 2 mL). Fluorescence emmission spectra was collected over a 480-700 nm range with 2 nm step size and integration of 1 s. The excitation wavelength was set to 460 nm and slit widths for excitation and emission were 5 nm and 10 nm.

3.5. Introducing a Photoinduced Electron Transfer (PET) Readout

Due to the difficulties in labeling of 2-Histidine constructs with two dyes (previous section), we decided to explore the use of photoinduced electron transfer (PET) as an additional approach. PET allows for observation of changes in distance at distances shorter than 2 nm (Lakowicz 2006). Implemented with a PET readout, the folding of the protein as pH decreases could result in an exponential decay in fluorescence intensity due to electron transfer to an engineered Trp. To implement a PET readout into the 2-Histidine coiled-coil, we designed a dye/PET-quencher pair at the ends of the protein. The idea being that the folded protein (at low pH) will result in a very short distance between the two ends (< 2 nm according to our structural model) and hence in a highly quenched state (very low fluorophore emission). On the other hand, the unfolding of the coiled-coil as pH increses is expected to produce a significant increase in the end to end distance, and hence an exponential increase in fluorescence intensity. To explore this strategy, we introduced a Cys residue in helix 2 and a Trp in helix 1 of the coiled-coil using recombinant DNA technology. The thiol group of the Cys side-chain was specifically labeled using commercially available maleimide derivatives of Atto 655 (see protein labeling section in Appendix A).

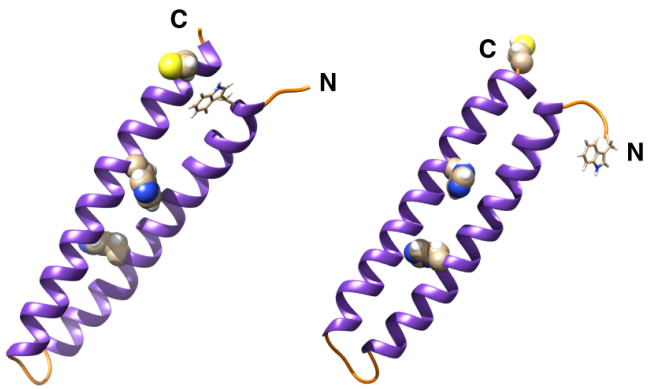

Figure 3.8 Modeled structures of 2-Histidine coiled-coils for PET: 2H-P (left) and 2H-P2 (right).

We designed three different versions of the PET readout. The first which we call 2-Hisitidine-PET (2H-P), replaced the Glu in the fourth position with Trp (E4W) and Lys at the seventieth position with Cys (K70C) as shown in Figure 3.8. These positions were chosen because they are expected to be closer together along the super-coil axis in the core of the protein when it folds. We also wanted to examine the effect of a longer loop on the behavior of the protein so we introduced eleven amino acids at the thirty ninth position to create four repeats of Ala-Gly-Ser (AGS) between Glu-Ser in the 2H-P construct and called it 2-Histidine-Loop-PET (2H-L-P). The final variant 2-Hisitidine-PET2 (2H-P2) was modification of the original 2-Histidine with Trp added to the N-terminal and Cys to the C-terminal as shown in Figure 3.8.

The bulk fluorescence photoinduced electron transfer quenching of Atto 655 dye by Trp results are shown in Figure 3.9. The 2H-P construct had a dynamic range between pH 4.0-7.5 with an effective quenching of about 6-fold with the largest drop in intensity (~ 2-fold) between pH 4.5-5.0 and about 7% error at pH 6.5. 2H-L-P with a longer loop also had a dynamic range between pH 4.0-7.5 but with an increased effective quenching of about 10-fold. The sensitivity of this construct concentrated in the 4.0-5.0 range, showing about 4.5-fold change between pH 5.0 and 4.5. The final variant with terminal probes, 2H-P2 had a very linear response between pH 4.5 and 7.0 and featured a 12.5-fold difference in signal intensity between the folded and the unfolded states.

It is interesting to observe that in the three PET constructs, the maximal change in fluorescence intensity is within the pH 5.5-4.5 range, and this can be attributed to dynamics of protonation/deprotonation of the two histidines (pK_a between 5.5-6.5) buried in the hydrophobic core of the coiled-coil scaffold. This pH region also results in the largest experimental errors, which could also reflect solubility/stickness problems that result in errors in the determination of the protein concentration given that the theoretical pI of these proteins is 4.78 (2H-P) and 4.81 (2H-P2).

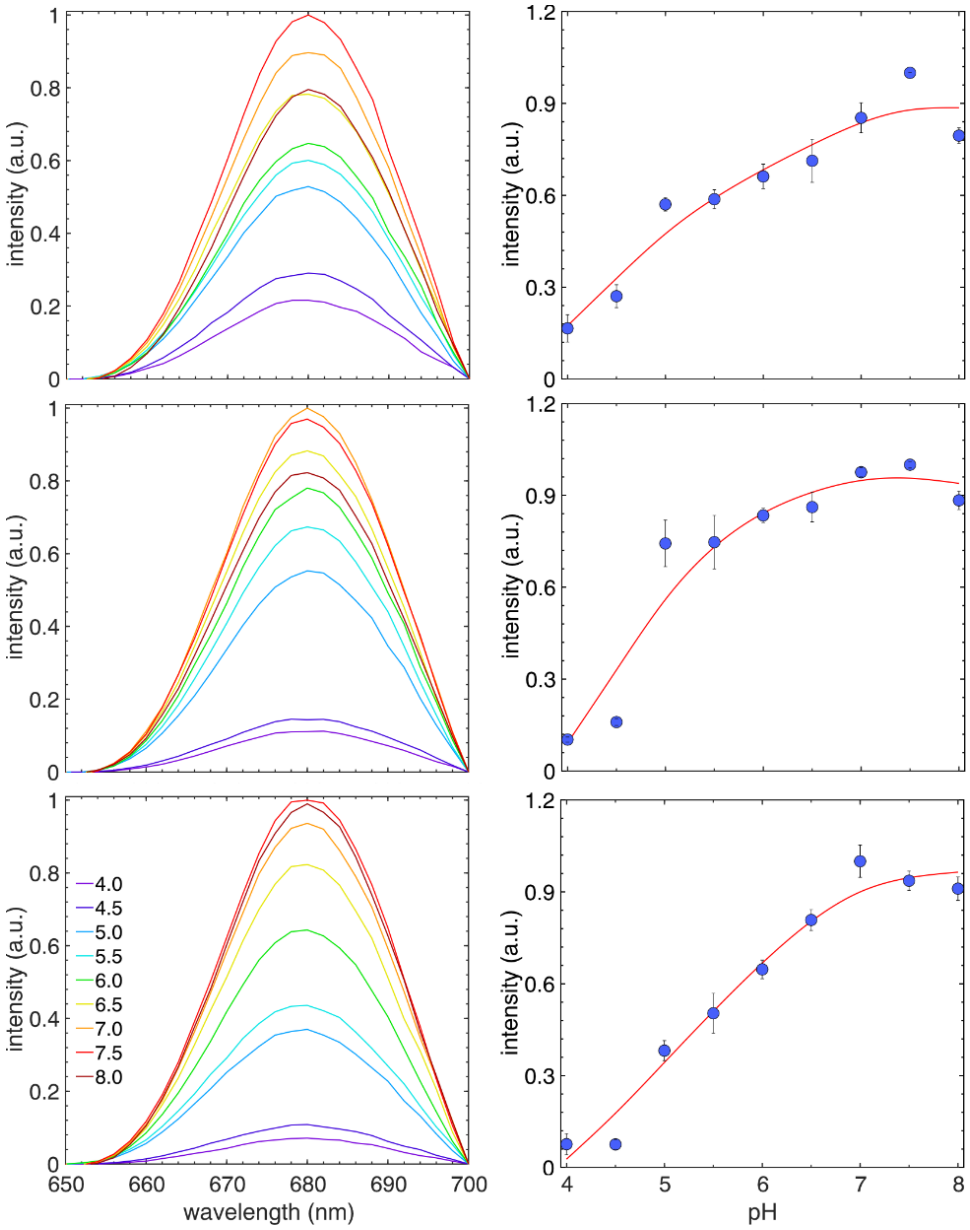

Figure 3.9 PET fluorescence quenching of 2-Histidine. 2H-P (top) 2H-L-P (middle) and 2H-P2 (bottom). All experiments were done using singly labeled protein sample (10 nM, 2 mL) in 20 mM buffers (having total ionic strength of 150 mM and 0.001% Tween-20). The proteins were excited at 600 nm and slits used were 5nm/10nm (Ex/Em).

Nevertheless, the results from the PET constructs demonstrate the effective implementation of an effective fluorescence readout onto the engineered coiled-coil pH transducer. They also indicate that these biosensors exhibit sensitivity over a very broad dynamic range (pH 4.0-7.5), capable of detecing changes of hydronium ion concentration over 3.5 orders of magnitude and with maximal signal changes of 10-fold.

3.6. Combining Two Fluorophores for Ratiometric PET Readouts

The major limitation of PET readouts is that they are reliant on a total change in intensity, and hence they are not ratiometric. Consequently, their readings of pH would require accurately knowing the sensor concentration or having a suitable external reference. To introduce a ratiometric readout into the successful PET-based pH sensors discussed in the previous section (see Figure 3.9), we designed a two-color strategy. Particularly, we designed a new 2H-P2 variant containing a site for labeling the protein specifically (the sensor needs each fluorophore to be in its correct place) with a second fluorophore to provide an internal signal independent of pH for calibration. Since specific labeling is essential here, we implemented an orthogonal chemistry approach (Evans, Millhauser 2015) to introduce by recombinant means the unnatural amino acid *p*-acetyl-L-phenylalanine (X) at the 39th position in the loop connecting both helices (replacing a Lys). *p*-acetyl-L-phenylalanine can be keto-labeled specifically (Evans, Millhauser 2015), hence providing an internal reference site for labeling with hydroxylamine derivative of the Alexa 488 green dye. In this case the strategy is that on excitation of the red dye Atto 655, the spatial proximity of the Trp that occurs when the protein is folded (low pH) will recapitulate the pH-sensitive fluorescence properties of Figure 3.9. In parallel, the excitation of Alexa 488 will produce green emission that should be insensitive to pH, the distance from the loop to the termini should not change upon separation of the two helices in the coiled-coil and thus will not result in a change in fluorescence intensity due to FRET. Hence, the green emission will provide an internal reference for determination of the concentration of sensor, and for ratiometric calibration of the red emission for pH determination.

IN-VIVO INCORPORATION OF UNNATURAL AMINO ACIDS IN PROTEINS

We then used orthogonal chemistry methods, which exploit the protein biosynthesis machinery of *E. coli* for site-specific incorporation of unnatural amino acid in genetically encoded recombinant proteins. The strategy requires the use of auxotrophs, having non-native orthogonal tRNA/aminoacyl-tRNA synthetase pairs. The tRNA transfers the unnatural amino acid and the synthetase aminoacylates the tRNA in response to amber (TAG) stop codon. The fidelity of translation depends on the transportation of the unnatural amino acid from the growth media to the cytoplasm.

The plasmids, pBAT-4 (ampicillin resistant) containing our gene of interest (2H-PF) and pEVOL (chloramphenicol resistant) containing genes for expression of tRNA/aminoacyl-tRNA synthetase were co-transformed into *E. coli* BL21 (DE3) competent cells. The transformed cells were screened by incubating overnight on a plate of LB agar with ampicillin (100 ng/mL) and chloramphenicol (34 ng/mL) at 37 °C as shown in Figure 3.10. For test of protein expression, a colony of bacteria was picked from the LB agar plate and incubated overnight with shaking at 37 °C in 10 mL of LB media with both antibiotics. The following day, 1 mL inoculum was added to 10 mL of LB media with antibiotics and *p*-acetyl-L-phenylalanine (0.4 mg/mL), it was incubated with shaking at 37 °C till an $O.D._{600}$ of ~0.8 was reached. The protein expression was induced on adding 1mM IPTG and arabinose (4 mg/mL). Post-induction incubation time, temperature and concentration of arabinose were tested as parameters for expression. SDS-PAGE was used on harvested cells and optimized parameters for protein expression were: 1X arabinose/IPTG for induction and 7 hours of post-induction incubation with shaking at 37 °C.

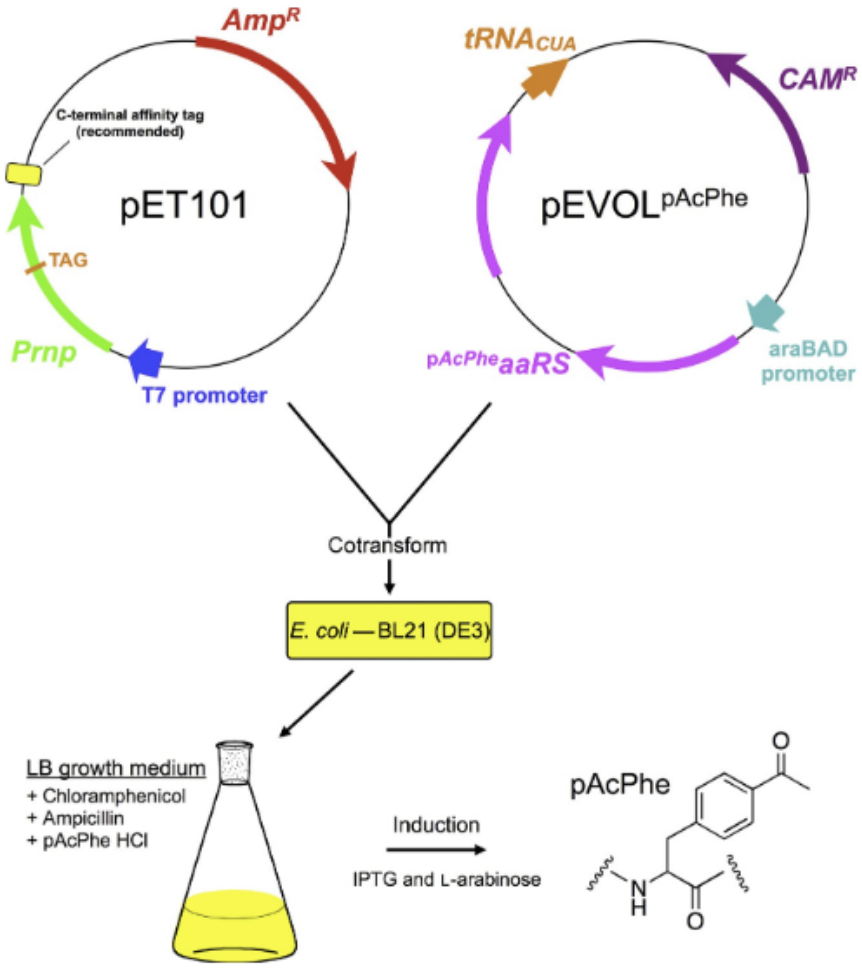

Figure 3.10 Unnatural amino acid p-acetyl-L-phenylalanine incorporation into recombinant protein (Evans, Millhauser 2015). Figure reproduced with permission.

With the expression conditions optimized, the protein was grown in a 2L batch and purified and labeled using standard procedures. We initially labeled it with Atto 655 to assess the integrity of the PET signal. As the pH is lowered from 8 to 4 and the protein folds with the ends coming closer, the fluorescence intensity dropped, as expected (Figure 3.11). However, the drop in fluorescence with pH was of about 35%, and hence lower than what we observed in the base construct (Figure 3.10). Moreover, upon subsequent labeling with Alexa 488 in the loop using the unnatural amino acid, the PET signal change with pH disappeared (see middle panel of Figure 3.11). Given that the second labeling step involved a long incubation in acidic pH, it is possible that the lost response of the PET readout was due to chemical modification of Atto 655 during the second labeling step. In addition, we

experienced similar problems with protein misfolding/solubility/aggregation, which resulted in large variance in the fluorescence measurement replicas in the pH 7.00 to 5.00 range. This led us to conclude that whereas PET is a promising approach to implement fluorescence readouts of local conformational changes, the future implementation of a two-color PET readout will need a different labeling strategy.

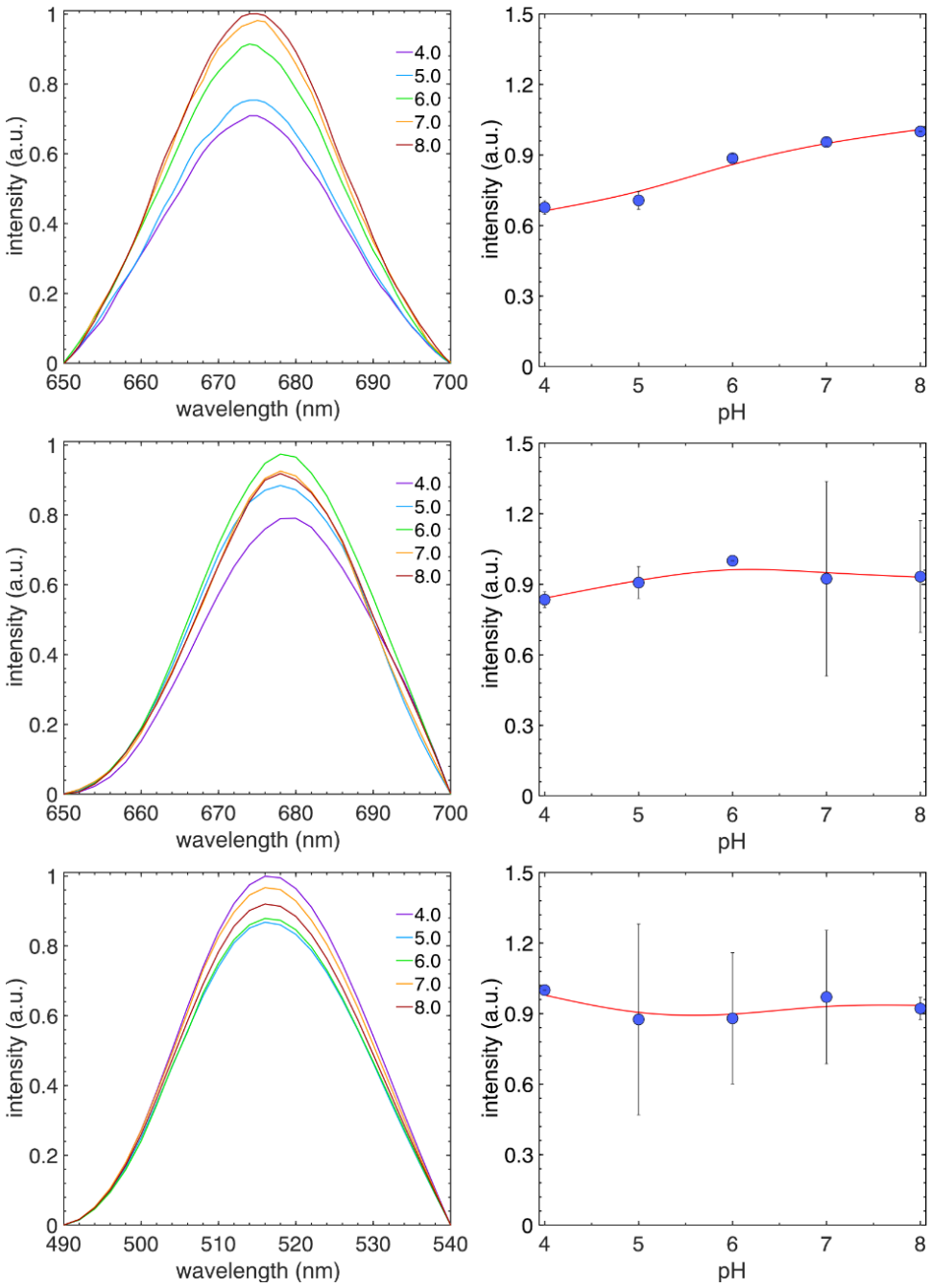

Figure 3.11 PET-Fluor of 2-Histidine with unnatural amino acid. Protein labeled with just Atto 655 as PET readout (top), protein labeled with both Alexa 488 and Atto 655 as ratiometric readout (middle and bottom). All experiments were done using singly labeled protein sample (10 nM, 2 mL) in 20 mM buffers (having total ionic strength of 150 mM and 0.001% Tween-20).

3.7. Conclusions

Coiled-coils emerge as excellent protein scaffold for engineering conformational transducers to be used in biosensor applications. From our work we foremost learnt that His grafting into the core or the protein can be used successfully for implementing pH conformational transducing. We showed that the combination of the His grafting approach and the engineering of an antiparallel, 70-residue coiled-coil out of a natural protein result on the implementation of conformational pH transducers that are sensitive to pH over an extremely broad range (broadband). We showed that the scaffold that we engineered is folded and stable as well as pH insensitive in the 8.00 to 4.00 range, whereas the version with four histidines introduced into the coiled-coil interface remains unfolded throughout this range and the version with two histidines results on optimal coupling between folding and His ionization and hence on an effective broadband transducer. The induced destabilization is additive and hence controllable: two His make the coiled-coil at the brink of stability, and hence responsive to histidine ionization, and four make the scaffold completely unfolded, regardless of pH. We also demonstrated that the engineering of a PET readout based on the quenching of the extrinsic dye Atto 655 placed in one end of the protein by a tryptophan placed at the other results in large changes in fluorescence intensity with pH. These fluorescence changes follow exactly the folding upon His ionization behavior observed by CD: quenching increases as the protein folds up when the pH drops. We produced PET readouts with a maximal change in signal of approximately 10-fold, which is a very promising result for implementing high-gain fluorescence sensors. These PET sensors also recapitulate the broadband pH response of the coiled-coil conformational transducer, which indicates that the effective coupling of local protein conformational changes to a fluorescence signal requires an approach that is sensitive to small changes in distance between the readout moieties. Because PET readouts are not ratiometric, and hence their use relies on external calibration of the intensity, we attempted to implement a two-color PET readout approach. Our results indicated that this type of readout requires an alternative labeling strategy, which may be pursued by our lab in the future. Finally, the true application potential of these sensors would be in fully genetically encoded versions. In this regard, the 2-His coiled-coil is an excellent choice for a FRET readout based on fusions with fluorescent proteins, where the labeling issues disappear. This issue is retaken in chapter 5 where we develop a plug and play platform for genetically encoded fluorescence biosensors.

4. BBL-TANDEM SCAFFOLD: TOWARDS SINGLE-MOLECULE ANALOG pH BIOSENSING

In this chapter we tackle the design of a proof of concept of single-molecule analog fluorescence biosensor. For this purpose, we aim to convert the intrinsic pH-induced conformational changes of the downhill folding protein BBL (Cerminara, Desai et al. 2012) into a high-sensitivity ratiometric fluorescence readout. The goal behind using a downhill folding protein is to induce a gradual change in fluorescent signal as a function of the ligand concentration (here protons) on an individual sensor molecule, which would enable instant analog reading of a single-molecule device. In addition, the use of a fast-folding reaction coupled to proton binding could also lead to an ultrafast (sub-millisecond) response time.

The protein BBL (pdb: 2CYU) is an ideal model system for this project. BBL is a 40-residue fragment of the E2 subunit of the 2-oxoglutarate dehydrogenase multienzyme complex of Escherichia coli, which comprises its peripheral E1-E3 subunit binding domain (Robien, Clore et al. 1992). BBL has been identified as a globally downhill folding protein (Sadqi, Fushman et al. 2006, Garcia-Mira, Sadqi et al. 2002). That is, BBL has no free energy barrier to folding in any thermodynamic condition. The thermal unraveling of BBL tertiary and secondary structure is broad and weakly coupled, with different structural/spectroscopic techniques such as CD and differential scanning calorimetry displaying different behaviors that highlight a low degree of cooperativity (Garcia-Mira, Sadqi et al. 2002). The analysis of the BBL thermal unfolding transition at atomic resolution using nuclear magnetic resonance showed a highly heterogeneous unfolding process in which individual atoms exhibit notable differences with a distribution of melting temperatures (T_m) that spans 60 K (Sadqi, Fushman et al. 2006). These results ruled out the possibility that BBL folded-unfolded through a binary transition and was the first experimental identification of a downhill folder, the existence of which was predicted by the analytical funnel landscape theory of Wolynes and coworkers (Bryngelson, Onuchic et al. 1995). The ensemble population in BBL shifts from fully native to fully unfolded by gradual melting of its structure (Garcia-Mira, Sadqi et al. 2002). Analysis of the thermal unfolding data with a structure-based statistical mechanical approach developed by Muñoz and Eaton (Muñoz, Thompson et al. 1997) yielded no barrier heights to folding under any condition, a globally downhill folding scenario. These earlier findings have been later confirmed by a wide range of techniques, including atomic resolution equilibrium unfolding experiments, multiprobe laser T-jump kinetics, differential scanning calorimetery, single-molecule fluorescence spectroscopy and large-scale molecular dynamics simulations (Li, P., Oliva et al. 2009, Sadqi, Fushman et al. 2006, Muñoz, Sanchez-Ruiz 2004, Liu, J., Campos et al. 2012, Lindorff-Larsen, Piana et al. 2011).

The BBL folding-unfolding behavior is thus a paradigm of conformational rheostat, a mechanism that was postulated to be of functional significance (Garcia-Mira, Sadqi et al. 2002), particularly when coupled to the folding upon binding of intrinsically disordered proteins (IDPs) (Muñoz, Campos et al. 2016). BBL also naturally exhibits gradual unfolding via mild acidification with an apparent pK_a of approximately 5 (Sadqi, Lapidus et al. 2003). The reason behind such sensitivity to pH in the 8-4 range is the presence of two buried histidines (H13 and H37), which are involved in several tertiary interactions in the native BBL structure that bring the two helices together, as shown in Figure 4.1. pH titration using NMR rendered pK_a values of ~6.6 for H13 and ~5.6 for H37 (Cerminara, Desai et al. 2012). The pK_a of H37 is 1 pH unit lower than that of free histidine, indicating that the native structure of BBL strongly favors the deprotonated state of H37. Additionally, BBL has five negatively charged residues (three E and two D) that could also contribute to the unfolding of BBL at more acidic conditions (Cerminara, Desai et al. 2012). Electrostatic repulsion at lower pH could be a source for long-range interactions that destabilize the protein. A calculation of the net charge on the protein considering a simple scenario of invariant pK_a of the ionizable amino acids on the protein (of which there are 14), indicates that the net charge of BBL should increase from 2.3 at pH 7 to 8.5 at pH 3.

Earlier work from our group explored the potential of using BBL's folding sensitivity to pH as gradual (rheostatic) conformational transducer of pH and ionic strength. In that work, it was shown that the coupling between proton binding and the gradual (un)folding of BBL resulted in a broadband conformational transducer with sensitivity to pH in the 4-8 range (Cerminara, Desai et al. 2012). The authors also showed that the unfolding upon binding response of BBL at room temperature was ultrafast (<20 μs), owing to its microsecond folding kinetics. At mildly acidic pH BBL's folding proved to be extremely sensitive to ionic strength, because the ions in solution shielded the electrostatic repulsions caused by the excess of positive charge in that pH range. This behavior was exploited to implement a bifunctional pH-ionic strength transducer based on the natural BBL protein, and an engineered BBL version that replaced all the negatively charged residues in BBL (D and E) by its uncharged counterparts (N and Q) to result on a protein with high positive charge at any pH below 9 (BBL+). BBL+ folding was insensitive to pH (always unfolded), but sensitive to ionic strength in broadband fashion (over 3 orders of magnitude in ionic strength). The combination of BBL and BBL+ thus afforded the possibility of determining both ionic strength and pH from the conformational changes of the two proteins (Cerminara, Desai et al. 2012) variants. This work demonstrated the potential of coupling downhill (un)folding to binding to implement high-performance conformational transducers capable of broadband and ultrafast responses. However, the pH-induced conformational changes of BBL are very local, and accordingly their monitoring needs the use of an optical technique sensitive to the backbone conformation, such as CD. CD is based on the differential absorbance of circularly polarized light, and hence it is low sensitivity, extremely reliant on highly transparent samples, and, most importantly, not specific

because every protein and nucleic acid have comparable CD signals. The challenge in converting such conformational rheostatic transducers into functional biosensors is to effectively couple them to high sensitivity and specificity readouts (i.e., fluorescence) that could be used for real-time monitoring in living cells. An extremely high sensitivity readout is also essential to demonstrate that rheostatic conformational transducers can give rise to biosensors capable of producing analog (quantitative) signals at the single-molecule level. This property, which has been theoretically predicted based on the conformational behavior of downhill folding (Campos, Sadqi et al. 2020), offers the highly desirable and unique feature of extreme biosensor miniaturization to reach real-time monitoring with nanoscale resolution.

In this chapter we tackle this challenge by: i) attempting to design a BBL-based biosensor that can effectively convert the pH sensitivity of BBL into a high-sensitivity ratiometric fluorescence readout; ii) demonstrating a proof of concept of analog real-time reporting of pH from a single-molecule device.

4.1. Design of a Conformational Rheostat Amplifier

Our goal was to convert the pH unfolding of BBL on a large enough change in conformation so that it results in a clear change in FRET efficiency. The first problem is that BBL is a very small domain (about 40 residues) that remains highly compact upon pH unfolding (Sadqi, Lapidus et al. 2003). In fact, the BBL native 3D structure (pdb:2CYU) shows the protein ends pointing in opposite directions (Garcia-Mira, Sadqi et al. 2002). It has been shown that, upon pH denaturation, the distance between the two ends of BBL becomes even shorter (Sadqi, Lapidus et al. 2003), which suggest that upon pH denaturation BBL remains compact but the two helices swivel thereby shortening the end-to-end distance to ~ 2nm. To produce a high sensitivity fluorescence readout based on FRET one must use fluorophores in the visible range with very high quantum yields and strong absorbance (Freidus, Pradeep et al. 2018). These properties result on FRET pairs with R_0 values (distance at which there is 50% FRET efficiency) around 6 nm, that is, sensitive to changes in the 3 to 9 nm range due to the six-power dependence of the FRET efficiency with the distance (see section 2.2). Therefore, conversion of the pH-induced changes in end-to-end distance of BBL (Figure 4.1) into a suitable range for high-sensitivity FRET measurements requires the design and implementation of an amplifier of the conformational change.

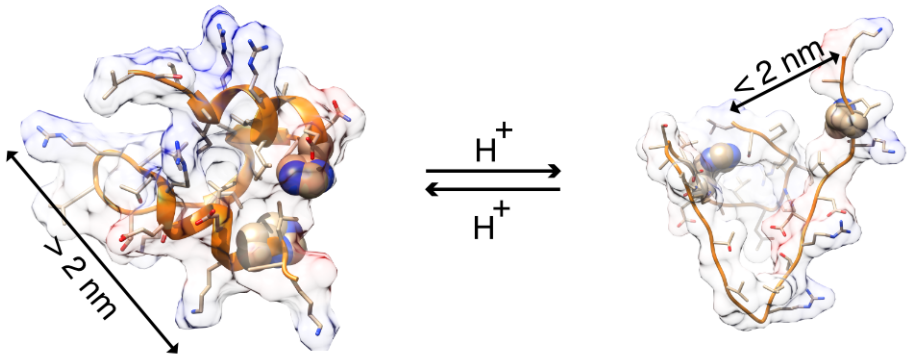

Figure 4.1 Protonation of two buried Histidines forms the basis of the pH-induced unfolding observed in BBL.

Our design strategy for such amplifier was to engineer a protein containing two BBL domains arranged in a tail to head tandem connected by a rigid structural spacer of potentially adjustable length. The combination of two BBL domains aims to amplify the magnitude of the 0.5 nm change that takes place in one domain. The spacer aims to increase the total end to end distance to shift the conformational changes closer to the maximum sensitivity range (close to R_0). As rigid spacer we used either a polyproline stretch, or a sequence designed to continuously connect the second helix of the first BBL with the first helix of the second BBL in the tandem (see Figure 4.2). Both types of rigid spacer can be made shorter or longer by simple extension of the sequence, providing an effective tool to optimize the average distance between the fluorophores so that is longer than R_0 at neutral pH (both BBL domains maintaining the native helix orientation and low FRET efficiency), and shorter at acidic pH (swiveling of the helices and high FRET efficiency). Based on this strategy, we designed three BBL-tandem scaffolds: one with an alpha-helix-forming sequence as spacer (B-hel-B), and two with polyproline stretches of 4 or 10 residues (B-P_4-B and B-P_{10}-B).

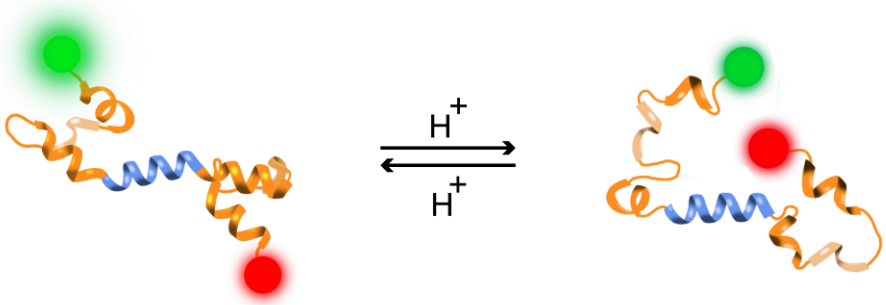

Figure 4.2 Two tandem BBLs were connected by a 12 amino acid helical linker, which we call B-hel-B.

The helix spacer in B-hel-B was 12 residues long (Figure 4.2). Its sequence was designed to form a highly stable α-helix using the program AGADIR (sequence information is provided in Appendix B). Given the structural parameters of α-helices (0.15 nm rise per residue and helical pitch of 0.55 nm), the helix spacer should increase the end-to-end distance in the tandem by about 1.8 nm.

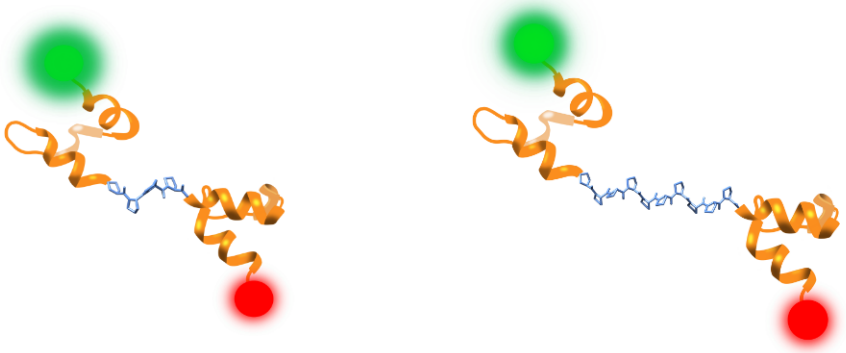

Figure 4.3 Two BBLs connected in tandem by a poly-Proline linker, 4-Proline which we call B-P4-B (left) and 10-Proline which we call B-P10-B (right).

The other type of spacer we decided to use was a stretch of prolines, which should form a polyproline II helix (Adzhubei, Sternberg et al. 2013), a type of protein secondary structure that is almost fully extended and rather rigid. A left-handed polyproline II helix (PPII) is formed when sequential residues all adopt backbone

dihedral angles φ and ψ of roughly -75° and 145°, respectively, and have all their peptide bonds as trans isomers. A full PPII helical turn requires 3.0 residues with 0.31 nm rise per residue (effective helical pitch of 0.93 nm), and hence it is rather extended. The PPII conformation is the preferred conformation of a stretch of prolines in aqueous solution (Mandel, Holzwarth 1973, Shi, L., Holliday et al. 2014), but it is also common in short segments of proteins and polypeptides in which prolines alternate with other amino acids. Additionally, a more compact right-handed polyproline I helix (PPI) can also be formed when sequential residues all adopt backbone dihedral angles of roughly -75° and 160° and have their peptide bonds as cis isomers (3.3 residues per helical turn and 0.17 nm rise per residue). Because of these properties, polyproline chains have been used as a "molecular ruler" in structural biology and biophysics, e.g., to calibrate FRET efficiency measurements (Stryer, Haugland 1967). Particularly, we designed two polyproline spacers, one with four prolines and the other with ten, expected to result in end-to-end distance extensions of 1.2 nm and 3.1 nm, respectively, when in a rigid PPII conformation.

Additionally, the three BBL tandem designs incorporated the sequence CKKNNDAL at the N-terminal of the protein, as a strategy to facilitate fluorophore conjugation and dynamic isotropic averaging of the fluorophore (see section 2.1.2).

4.2. Converting pH into a FRET Signal: Bulk Fluorescence of BBL-Tandems

To implement the FRET readout, we labeled the various BBL-tandem constructs on the cysteines placed at their ends with Alexa 488 and Alexa 594. After purification of the doubly labeled sample, we performed steady-state fluorescence experiments at various pH values to determine whether our approach to amplify the conformational rheostatic change of BBL worked. FRET efficiency (*E*) was calculated after normalizing the spectra from the intensity of the acceptor band divided by the total.

The results from these experiments are summarized in Figure 4.4. We observed that the FRET efficiency of B-P_4-B was the lowest of the three and that of B-P_{10}-B the highest, consistently with their long and short spacers. However, neither of these constructs convert the local conformational changes of the BBL domains into a FRET change. B-P_4-B has a completely flat profile with E~0.47. The B-P_{10}-B exhibits FRET efficiencies around 0.6, but it is also essentially unresponsive to pH. In addition, this protein exhibited some propensity to aggregate around its pI (~5.5), resulting in large errors around that pH region (see Figure 4.4 middle).

The FRET values for the two tandems with polyproline linkers are qualitatively consistent with the difference in the spacer's length. However, the difference in *E*

between B-P_4-B and B-P_{10}-B corresponds to an additional extension of ~0.6 nm for the latter, whereas the increase for a rigid PPII should be almost 2 nm. Of course, this calculation assumes that the PPII and its linkages to the two BBL domains are all completely rigid. It is very likely that the connections with the two BBL domains are rather flexible, resulting on a more collapsed globular structure with shorter end to end distance, as we see experimentally. In addition, more recent work suggests that polyproline segments are not as rigid as traditionally assumed (Doose, Neuweiler et al. 2007, Moradi, Babin et al. 2009, Ruggiero, Sibik et al. 2016). However, the most important result is that the polyproline linkers do not effectively amplify the local conformational changes of the BBL domains with pH.

In contrast, B-hel-B exhibits a significant change in *E* as a function of pH in bulk. In this protein, *E* changes monotonically from ~0.48 at pH 8 up to ~0.62 at pH 4 (Figure 4.4.). The total change is similar in magnitude to the difference in *E* of the two polyproline spacers, corresponding to a global expansion of ~0.6 nm. The change is most pronounced between pH 6 and 4, but it does continue up to pH 8, albeit in weaker fashion. The change in *E* of B-hel-B is equivalent to a change of 1.6 in the acceptor/donor ratio, which is comparable to the total change in signal of many commercial FRET-based sensors (Edwards 2021, Quijano-Rubio, Yeh et al. 2021, Sukumar, Natarajan et al. 2020). Additionally, we note that there is a slight red shift in the acceptor emission maximum of B-hel-B as the pH is lowered. Such an effect is however not observed in the donor emission. This could be due to the acceptor encountering a more hydrophobic environment. Based on these results, we conclude that the coupling of the second helix of BBL1 with the first helix of BBL2 via a helical spacer works as an effective amplifier of the local conformational changes that the BBL domains experience with pH. Hence, B-hel-B is a suitable candidate for fluorescence detection of pH changes in the 8-4 range.

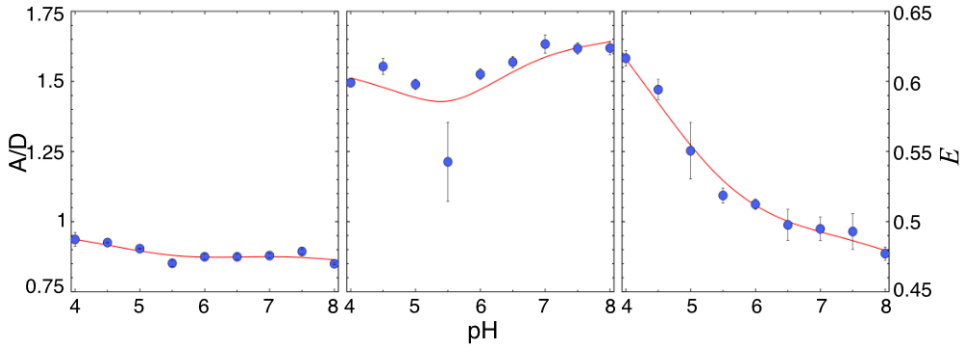

Figure 4.4 Steady-state bulk response of the three variants of BBL as a function of pH. B-P_4-B (left) B-P_{10}-B (center) and B-hel-B (right). All protein samples were prepared at 10 nM labeled protein concentration in 20 mM buffer (Table 4.2) by serial dilution in the same buffer with 0.001% TWEEN 20 from 5 μM stocks. The donor was excited at 460 nm and emission collected from 480 nm – 680 nm on a PTI QuantaMaster 400 (Horiba) instrument. Data were acquired using slit widths of 5 nm for excitation and 10 nm for emission, 2 nm step size and 1 s integration at each wavelength. All experiments were done in triplicates.

Another important factor to consider for the suitability of B-hel-B as fluorescence pH biosensor is the potential dependence on ionic strength. The wildtype BBL protein did refold from the acid denatured state by increasing ionic strength (Cerminara, Desai et al. 2012). The persistence of such an effect on B-hel-B could hamper the interpretation of the fluorescence changes in terms of pH change. We tested this possibility by performing bulk fluorescence measurements on B-hel-B at the two extreme pH values, and as a function of ionic strength in the 0.1 to 1 M range. Our results showed very little ionic strength dependence for B-hel-B at both extreme pH values (Figure 4.5). At pH 4, B-hel-B is essentially ionic strength insensitive up to 0.8 M, exhibiting a minor increase in E above that point (Figure 4.5). In contrast, at pH 8 the response is flat above 0.5 M, and there is a minor effect at the lowest ionic strength. Overall, the fluorescence sensitivity to pH and lack of sensitivity to ionic strength indicate that B-hel-B is a good fluorescence pH biosensor and an excellent candidate for exploring the implementation of single-molecule analog pH detection.

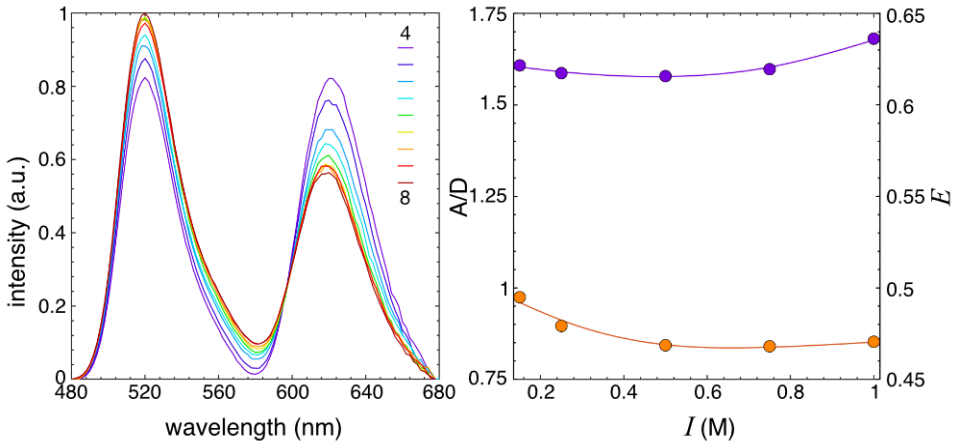

Figure 4.5 Steady-state bulk fluorescence intensity measurements of B-hel-B as a function of pH (left) and ionic strength dependence (right). Spectrum in purple is pH 4 and orange is pH 7.

4.3. Single Molecule Confocal Microscopy of Free Diffusion B-hel-B

Before performing single-molecule fluorescence experiments on the B-hel-B pH biosensor, it is important to consider the expected differences between a conformational switch (two-state folding) and a conformational rheostat (downhill folding) in the context of single molecule-FRET experiments. In the context of the detection of a limited number of photons from an individual molecule, the actual output of the experiment depends on the interplay between the experimental binning time (T_B), which sets the time resolution, and the folding relaxation time (τ_F) of the protein. In a two-state folding scenario where a protein folds over a significant free-energy barrier (> 3 *RT*), individual molecules populate either the folded or the unfolded state. When $T_B \ll \tau_F$, the FRET efficiency histogram shows two peaks with amplitudes dependent on the folding equilibrium constant. Since the conformational dynamics within each state are much faster than the overall folding relaxation, both peaks should have shot-noise limited widths, as defined by equation 2.25 (Gopich, Szabo 2007). However, when $T_B \gg \tau_F$ molecules undergo multiple folding-unfolding events during the observation time, and thus the FRET efficiency histogram becomes a single, shot-noise limited peak at a position reflective of the population weighted average. In a one-state (downhill) folding scenario, the FRET efficiency histogram will have a unimodal distribution at all conditions. When $T_B \gg \tau_F$ the FRET efficiency histogram will be shot noise limited, and indistinguishable from two-state folding. The difference arises when $T_B \ll \tau_F$, which for a downhill folder will still result on a unimodal distribution, but in this case

with potentially additional width over the shot noise, revealing the actual conformational distribution.

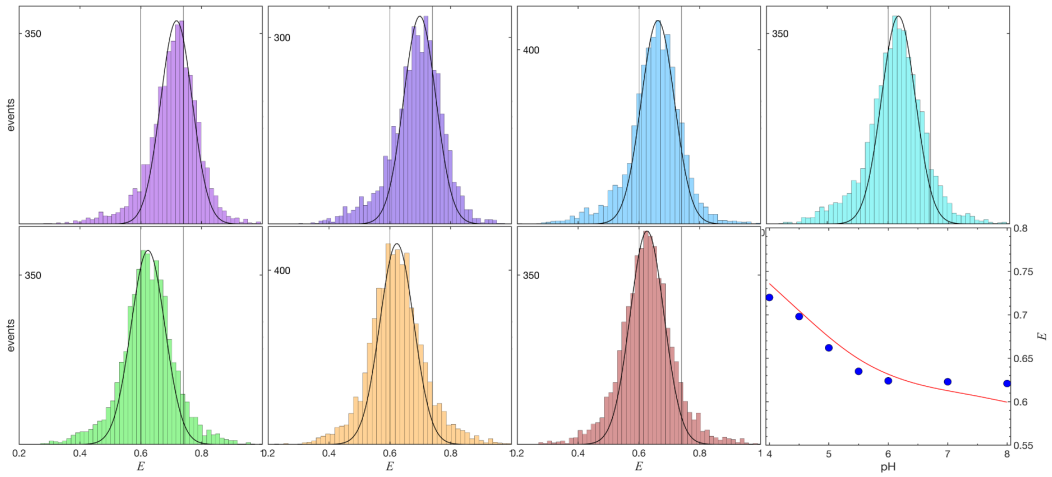

Figure 4.6 SM-FRET histograms of B-hel-B showing a unimodal distribution at various pH values. pH 4.0 – 5.5 (top) and pH 6.0 – 8.0 (bottom) as we move left to right. The two vertical lines at 0.60 and 0.72 signal the dynamic range of the overall (averaged) change in signal with pH. The bottom-right figure shows the mean FRET from single molecules (blue dots) and trend line from bulk experiments (red line). All experiments were done with labeled protein (50 pM, 400 µL) in buffers having 0.001% TWEEN 20 with 1mM 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) and 10 mM 2-mercaptoethylamine (cysteamine) for photoprotection.

The other important factor to consider for single-molecule experiments is the signal to noise of the experiment, as we will be detecting a limited number of photons per unit of time. The key parameter is the photon count rate that is obtainable from an individual molecule. In optimal conditions and using special photoprotection cocktails to reduce fluorophore blinking and bleaching (Campos et al. 2011), we can obtain count rates of 150-250 photons per millisecond from one molecule. Using the shot noise equation (equation 2.25), these count rates convert into a determination of the FRET efficiency with ± 0.057 shot noise accuracy for measurements made on a single molecule with E = 0.5 and 0.5 ms time resolution (T_B).

A conformational switching device (biosensor) can produce an analog readout by either performing the measurements using ensemble averaging (i.e., bulk

measurements) or by time averaging of one molecule, or by any combination of the two. In our case, to distinctly demonstrate that the B-hel-B pH sensor produces analog pH readouts at the single-molecule level we need to make fluorescence measurements of individual molecules with sufficiently short time-resolution to avoid systemic time averaging ($T_B \ll \tau_F$) and with enough photons (count rates) to reduce the shot noise in the determination of E to levels below the overall signal change due to pH.

For this purpose, we used confocal single-molecule FRET spectroscopy in our custom-built confocal fluorescence microscope (see section 2.2.1 in chapter 2). We performed SM-FRET measurements on B-hel-B in free diffusion experiments (molecules diffusing in and out of the illuminated confocal volume) at various pH values in the 8.0 to 4.0 range. The results of these experiments are presented in Figure 4.6. In our experimental conditions (50 pM protein concentration, 100 μW excitation power) we obtained count rates (1.8×10^4 photons per second for pH 4.0 and 2.6×10^4 photons per second for pH 8.0) from freely diffusing molecules that allowed to bin the experimental data into 0.5 ms bins using a photon threshold of 75. Donor only bursts were removed using cumulative Poisson distribution function as shown in supporting information, a statistically very stringent test to eliminate any bursts coming from molecules without an active acceptor. We collected data from thousands of molecules and calculated E for each single-molecule photon burst using equation 2.24 (section 2). Figure 4.6 shows the resulting FRET efficiency histogram at the various pH values. The figure panels also show a curve representing the width of the FRET histogram expected from shot-noise over the experimental mean FRET value given the photon threshold.

The first observation is that the single-molecule experiments recapitulate the average E as a function of pH measured in bulk. This is shown in the last panel of Figure 4.6, in which the average E from the FRET efficiency histograms is compared to the E curve measured in bulk (shown in red). Actually, the datapoints shown in that last panel correspond to the average signal from a few thousands of molecules measured with 0.5 ms time resolution relative to the steady-state bulk signal. The second observation is that the FRET efficiency histograms are always unimodal and have a maximum that gradually shifts from lower E at pH 8.0 to the highest value at pH 4.0. Particularly, the population moves monotonically from a mean FRET of 0.72 at pH 4 to 0.60 at pH 8. The standard deviation from shot noise was about 0.05 (0.0520 with 4,257 bursts for pH 4.0 and 0.0558 with 6,428 bursts for pH 8.0). These results are an indication that the protein changes gradually in analog fashion, or alternatively that it changes conformation due to histidine protonation so quickly that the histogram is time-averaged as discussed in the previous paragraphs. Given that the time resolution of the experiment is already 0.5 ms, the second reason is extremely unlikely. Moreover, comparison of the histograms with the expected shot noise (black curves in figure 4.6) shows that the experimental distribution is broader, particularly for the higher pH values. The extra

width on the histograms suggests that the conformational dynamics of the protein are not too much than the time resolution (i.e., >> 1/ (0.5 ms) or 2,000 s^{-1}).

To further investigate this issue, we performed a kinetic analysis of the SM-FRET data. Resolving biomolecular dynamics using SM-FRET experiments through FRET efficiency histogram analysis has practical limitations. In SM-FRET experiment, the probability that a photon is emitted by either the donor or acceptor depends on the inter-dye distance. Thus, the pattern of the color of photons contains information about the inter-dye distance, which is in turn reflects the conformational dynamics happening in the biomolecule. Gopich and Szabo (Gopich, Szabo 2009, Gopich, Szabo 2007) developed a likelihood analysis method to extract kinetic and dynamic parameters from the sequence of the color of photons and inter-photon times. GS-MLA model calculates the likelihood that a given set of parameters of a predefined model describes the observed time-stamped photon trajectory. This likelihood function is then maximized to obtain the parameters that best describe the data. In our case, we used a kinetic model that represents the conformational dynamics of a protein as diffusion on a 1D free energy surface (Ramanathan, Munoz 2015). This type of kinetic modelling is ideally suited for the analysis of minimally cooperative protein (un)folding transitions and hence to distinguish between two-state and downhill folding. The same type of analysis should be able to distinguish between a scenario in which the coupling between protein (un)folding and histidine ionization produces a shifting unimodal (analog) or bimodal (binary) distribution. It will also provide the time response of the biosensor from the slowest eigenvalue of the diffusive rate matrix (Ramanathan, Munoz 2015). The results we obtained with this method for the B-hel-B photon trajectory data as a function of pH are summarized in Table 4.1 and in Figure B4.3 (Appendix B). Interestingly, the kinetic analysis indicates that the (un)folding coupled to histidine ionization of B-hel-B is very fast at the highest pH values (sub-milliseconds) and gets even faster as pH becomes more acidic (about 50 μs at pH 4.0). As we discussed in section 1.3 these results imply that the response of the biosensor is limited by the proton (ligand) concentration at the highest pH values, and then speeds up as pH drops. However, the acceleration in response is of only a factor of ~10 for 4 orders of magnitude in $[H^+]$, which suggests that at the lowest pH values the biosensor's time response is limited by the folding rate constant of B-hel-B. The latter is consistent with previous work in which the time response of a single BBL domain was measured by laser-induced temperature jump kinetics (Cerminara, Desai et al. 2012).

These SM-FRET experiments and analysis demonstrated that B-hel-B changes its conformational ensemble gradually in response to pH, and that such gradual changes get efficiently converted onto a change in fluorescence signal (FRET efficiency) that operates as an analog single-molecule fluorescence biosensor of pH. Furthermore, the results shown in Table 4.1 also demonstrate

that the time response of this biosensor is ultrafast: from 0.5 ms at pH 8.0 to 50 μs at pH 4.0.

Table 4.1: Slowest non-zero eigenvalue (e), which represents the relaxation rate of B-hel-B conformational dynamics in response to histidine e ionization, and the mean FRET at each pH calculated using a maximum likelihood analysis of photon trajectories and a 1D diffusional free energy surface model.

pH	E	e (s^{-1})
4.0	0.7189	20,273
4.5	0.6989	10,636
5.0	0.6664	13,320
5.5	0.6404	6,772.8
6.0	0.6366	5,909.2
7.0	0.6340	3,238.4
8.0	0.6400	2,066.5

4.4. Recording pH from Single Molecule Fluorescence: Signal to Noise Analysis from Stochastic Simulations of B-hel-B

The experiments described in the previous section demonstrated that B-hel-B is a fluorescence biosensor for measuring pH that can operate in analog fashion at the single-molecule level. However, in practical terms, it is important to determine what are the specific properties and limits of such sensor, particularly in terms of signal to noise, pH accuracy and time response. Knowing these properties is essential to

ascertain the potential usability of the biosensor as a single-molecule pH sensing device. To tackle this key question, we performed a series of stochastic simulations based on the SM-FRET experimental data that we obtained for B-hel-B (see previous section).

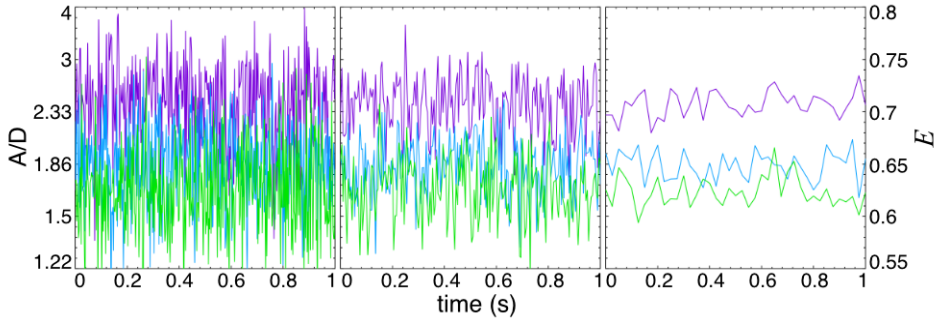

Figure 4.7 Simulations from experimental data to demonstrate the number of molecules required to get a FRET readout of pH with a given signal to noise ratio. (Left) 5 molecules (0.5 ms)/1 molecule (10 ms); (middle) 10 molecules (0.5 ms)/1 molecule (20 ms); (right) 50 molecules (0.5 ms)/1 molecule (100 ms). pH 4 (purple), pH 5 (blue) and pH 6 (green).

The SM-FRET data contains bursts of 0.5 ms because they come from free diffusion experiments. However, we can simulate longer time recordings and the effects of signal averaging (digital filtering) by stochastically appending bursts from different molecules into a simulated compounded trajectory. Such trajectories simulate data acquisition streams for different time averaging and number of molecules. These simulations show that for a time resolution of 0.5 ms, the simultaneous recording of a total of approximately 50 molecules is needed to produce a *E* signal with an accuracy of 0.01 standard deviation, whereas with 5 molecules the signal to noise decreases by a factor of 4 (shot noise of 0.04). As the simulations show, given the limited maximum signal of the B-hel-B sensor (Figure 4.7), distinguishing pH within 1 unit in the 6.00 to 4.00 range requires using a biosensor with at least 50 molecules. Alternatively, the simulation predicts that one could obtain the same signal to noise recording from a single molecule provided that its behavior as averaged over 100 ms.

4.5. Long Fluorescence Recordings from Single Molecules: TIRF Microscopy on Vesicle Immobilized B-hel-B

Inspired by the simulations described in the previous section, we aimed to achieve the recording of the fluorescence signal of B-hel-B single molecules over long times (seconds) and hence demonstrate the analog single-molecule sensing of pH in times and conditions relevant to its application as a biosensor. For this purpose, we resourced to TIRF microscopy for the individual visualization of multiple B-hel-B molecules while immobilized to the surface. For the immobilization we encapsulated B-hel-B into synthesized biotinylated liposomes that were subsequently tethered to the surface via biotin-streptavidin-biotin linkages (Boukobza et al. 2001). The liposome encapsulation was performed at specific pH, so that the liposome will a pH constant environment to record the B-hel-B signal. Encapsulation allows for the controlled environment of the single-molecule biosensor and for a better mimic of the biologically relevant environments and conditions. TIRF uses very mild illumination and permits the simultaneous imaging of multiple molecules present in the field of view (typically a few microns in each dimension using a 100x TIRF objective). The mild illumination conditions of TIRF together with the immobilization allow for the observation of single molecules over much longer periods of time (seconds), unlike free diffusion measurements using confocal microscopy (about 1 ms), albeit with lower temporal resolution. It has additional advantage of better signal to noise ratio owing to the fact that the excitation volume spans over a very narrow depth (about 100 nm) of the evanescent wave. This factor greatly reduces the fluorescence background and hence produces high contrast visualization of single molecules during long times. The caveat is the time resolution, which is limited by the readout time of the CCD camera used to produce the images.

In this case, we produced the liposome-encapsulated B-hel-B biosensor molecules at various pH values and characterized their fluorescence properties using our custom-built TIRF microscope. A total of 120 images were collected in series (13.407 Hz) with 50 ms exposure time to get a continuous stream of data corresponding to 6 s of continuous pH monitoring from individual molecules. To identify single molecules, the 512 × 512 pixel image was separated onto the donor and acceptor halves, and the two images were overlaid. A count threshold of 5,000 was applied to remove the background. Locations of single-molecules were identified from the pixels with maximum intensity, and areas of 5 × 5 pixels (16,000 nm pixel size) surrounding the maximum intensity pixel were used to encompass the complete image from each molecule (optical resolution with our 1.49 N.A. objective and light in the green-red range is about 200 nm). Once the molecules were identified, we calculated the FRET efficiency per frame (50 ms resolution) by integrating the intensities from the donor and acceptor image slices. Figure 4.8 shows examples of one molecule pH recordings for pH 4.0, 5.0 and 6.0 (range

corresponding to the maximal sensitivity of the B-hel-B sensor). The images (shown at 50 ms intervals for 1 second) show that at pH 4.0 the example molecule exhibits significantly higher emission in the red (high E) whereas at pH 6.0 it is close to 50-50. The results at pH 5.0 are in between. Calculation of the FRET efficiency (*E*) by integrating the intensity of the two channels over the 5 × 5 pixel area shows the change in *E* as a function of time from each molecule over a period of 6 seconds. The mean E values from these trajectories shows a change of about 10% between pH 6.0 and 4.0. These values are in close agreement with the free diffusion confocal SM-FRET data (Figure 4.6) and are also consistent with the calibration of the sensor in bulk (Figure 4.4: note here that the bulk data overestimates the overall E due to incomplete labeling with both fluorophores in the sample, whereas the donor-only molecules are easily identified and discarded in the single-molecule experiments). The data shown in Figure 4.8 represents a first direct demonstration of real-time analog recording of pH from a single-molecule biosensor with a 50 ms time response.

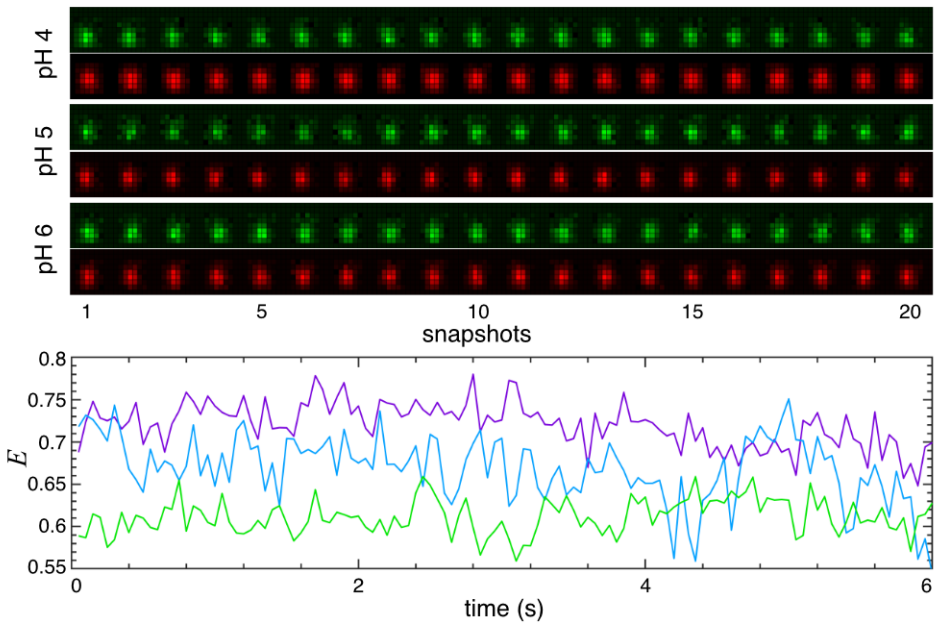

Figure 4.8 TIRF measurements on B-hel-B immobilized in vesicles. Top) two-color snapshots of an individual molecule as a function of time taken over a period of 1 s. Each snapshot was taken with an acquisition time of 50 ms. Bottom) recording of the FRET efficiency during 6 seconds for each molecule shown on the top. pH 4 (purple), pH 5 (blue) and pH 6(green). 250 µL egg L-α-phosphatidylcholine (20 mg/mL dissolved in chloroform) was mixed with 3 µL 1, 2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(cap biotinyl) (5 mg/mL dissolved in chloroform) and lyophilized. The lipid mixture (Avanti) was resuspended in 450 µL of buffers with Trolox/cysteamine and sonicated for 15 min. B-hel-B was added to 200 µL of lipid mixture to get a final concentration of 50 nM of labeled protein. The lipid-protein mixture was extruded through a filter syringe (0.1 µm) and cleaned using micro spin column (S400-HR) and 100 µL was loaded on streptavidin covered coverslip.

4.6. Conclusions

In this chapter we presented our results with the design, engineering and characterization of a fluorescence pH biosensor based on a conformational rheostat transducer and a fluorescence signal amplifier strategy. We found that the strategy of connecting two BBL domains via a helical linker does effectively convert the pH-induced conformational changes of the BBL domain into end-to-end changes that result in a FRET efficiency signal that can be used to monitor

pH. The FRET signal is comparable to those of other existing FRET biosensors, but it is on the low side to achieve efficient single-molecule detection. We then performed a detailed characterization of the pH response of the biosensor using single-molecule fluorescence techniques. From these experiments we demonstrated that individual B-hel-B molecules change their fluorescence signal gradually in response to pH, and that they do so with sub-millisecond kinetics throughout the entire pH 8.0 to 4.0 range. These results demonstrate that B-hel-B can be used as an ultrafast pH biosensor capable of measuring pH with <1 ms response times. Finally, we could also demonstrate the real-time analog monitoring of pH from individual B-hel-B molecules encapsulated in liposomes. These results constitute the first practical demonstration of a single molecule analog biosensor. B-hel-B provides a proof of concept that confirms the feasibility of the single-molecule analog sensing concept and can be used as basis to further develop and engineer other biosensors. The current limitation of this sensor is its modest maximal signal, which makes single molecule sensing technically difficult and relatively slow. To achieve more practical applications, it would be important in the future to maximize the signal outcome of this sensor. This could be attempted by optimizing the coupling of the fluorescence signal and conformational change, or by improving the methods to effectively produce a two-color-PET readout, which results in much larger signals. Efforts along these directions are currently underway in our laboratory. The current sensor could also be expanded by including other domains to create bifunctional sensors or cooperative/competitive sensors. Furthermore, single domain in the tandem arrangement can be specifically engineered to exhibit different affinities (increase the broadband range), or to be inhibitive thus creating a sharp response that could be used for applications that require accurate measurements of small changes in pH. This research thus opens immense opportunities for future developments in the field of fluorescence biosensors.

5. PLUG AND PLAY PLATFORM FOR GENETICALLY ENCODED BIOSENSORS

In this chapter we attempt to create genetically encoded fluorescent reporter systems in which any protein can be plugged to create biosensors. Genetically encoded fluorescent biosensors can be directly synthesized within cells, expressed in specific subsets of cell populations and easily fused with different protein tags for targeting specific subcellular microdomains. These advantages apart from lower cytotoxicity make their application useful compared to extrinsic fluorophores which need to be introduced inside cells artificially.

Living cells respond to stimuli at the cellular and subcellular level and it is where the foundation of signaling networks lies. The discovery of fluorescent proteins almost three decades ago has enabled development of diverse array of genetically encoded fluorescent biosensors capable of monitoring these networks. These biosensors have aided our understanding of spatiotemporal dynamics in signaling pathways at the cellular and subcellular levels. In the postgenomic era, wherein identities of most signaling pathway components are known, understanding the regulation and dynamics of signaling for physiological and pathological cellular functions has become even more critical. These signaling reactions are often discrete and transitory in nature, so recombinant biosensors are ideal for endogenous cellular monitoring and quantifying in real time.

5.1. Fluorescent Proteins

Green Fluorescent Protein (GFP) was first isolated from the jellyfish, *Aequorea victoria*. Its molecular mass is about 27 kDa and has 238 amino acids. The structure is cylindrical, about 42 Å in length and 24 Å in diameter having a beta barrel structure consisting of eleven β-strands with a coaxial helix containing a covalently bonded chromophore. In the native protein, inward-facing sidechains of the barrel induce specific cyclization reactions in Ser65-Tyr66-Gly67 that induce ionization of p-hydroxybenzylideneimidazolinone to the phenolate form and chromophore formation. The hydrogen-bonding network and electron-stacking interactions with these sidechains influence the color, intensity and photostability of GFPs (Tsien 1998).

There are numerous advantages of using FPs foremost amongst them is the good optical properties: visible excitation, emission and high photostability. Applicable to nearly all organisms without the need for permeation and hydrolysis. It can be targeted to specific tissues, cells, organelles, or proteins and is unlikely to diffuse

enough to blur spatial gradients. cDNAs or plasmids are cheap to replicate and distribute. Additionally, it is modular and can readily be modified and improved by mutagenesis. Consequently, it is generalizable to measure many bioactive species other than pH, as long as a conformationally sensitive transducer is available. Amongst the disadvantages of using FPs are the maximum change in emission ratio is currently less than for small-molecule dyes. Additionally, the binding kinetics are somewhat slower and gene transfections are required. Furthermore, it can also be cytotoxic.

5.2. Split Green Fluorescent Protein

The unassisted reconstitution of proteins from peptide fragments have been demonstrated for numerous proteins like ribonuclease, tRNA synthetases and inteins (Richards, Vithayathil 1959, Shiba, Schimmel 1992, Southworth, Adam et al. 1998). These reassembly processes are contingent upon the proper choice of a splitting site within the protein. In particular, circular permutation and protein insertion strategies have provided strategies to split GFPs that can fold, fluoresce and serve as biosensors despite rearrangement of the natural coding sequences (Ghosh, Hamilton et al. 2000, Kent, Oltrogge et al. 2009).

In our study we used a variant of avGFP that was split between residues 157 and 158, and our coiled-coil 2-Histidine construct was added between the two fragments. The larger fragment of 157 residues containing GFP fluorophore: Ser65, Tyr66 and Gly67 was attached to the N-terminal of 2-Histidine via a 6-residue linker (GGSGSG) and to the C-terminal the smaller fragment of 81 residues was attached via a 4-residue linker (GGSG). We call this chimera splitGFP-2H and its sequence is provided in Appendix B (see section on Chapter 5). At neutral pH, when 2-Histidine is unfolded, we expect very little fluorescence and as pH is decreased and the coiled-coil folds bringing the two fragments of the split GFP together, we expect to observe an increase in fluorescence intensity.

On measuring the bulk fluoescence the emission λ_{max} was observed to be at 520 nm similar to what was reported earlier (Ghosh, Hamilton et al. 2000). The pH dependance of the fluoresence intensity is shown in Figure 5.1. Contrary to our expectations, the intensity was highest at neutral pH and dropped by about 1.5 fold as pH was lowered to 4. At basic and neutral pH, 2-Histidine was expected to be unfolded and with increasing acidic conditions become more coiled-coil. The highest intensity at pH 4 and lowest at pH 8 should have been observed but, was not.

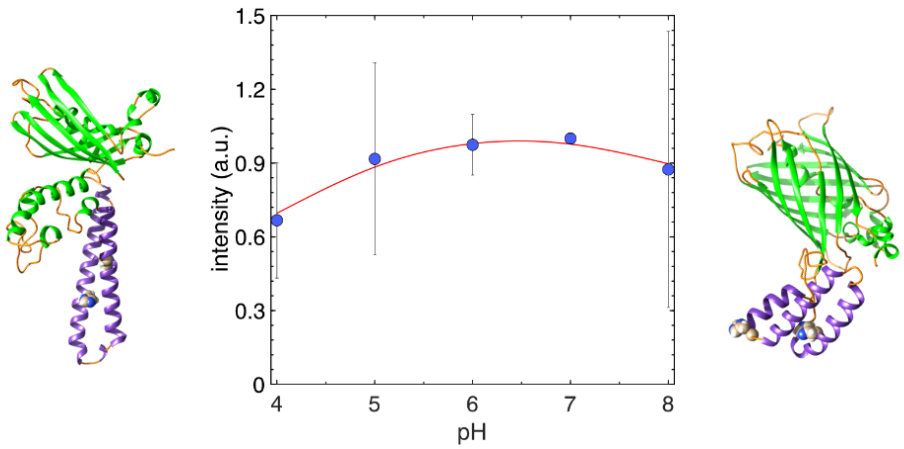

Figure 5.1 pH dependent fluorescence intensity measurements on chimeric splitGFP 2-Histidine. Experiments were done in 20 mM buffers (ionic strength adjusted to 150 mM with NaCl) using protein (10 nM, 2 mL). Fluorescence emmission spectra was collected over a 485-585 nm range with 2 nm step size and integration of 1 s. The excitation wavelength was set to 465 nm and slit widths for excitation and emission were 5 nm and 10 nm and all measurements were taken at room temperature.

The reason this could have happened is that adding 248 residues around 2-Histidine made it fold into a three-bundle helix bringing the splitGFP fragments together to fluoresce. As the pH is lowered, the helices snap back into a coiled-coil distorting the GFP thus decreasing the fluorescence intensity. Alternately, it is also possible that since the GFP is split and its fluorophore is exposed to the solvent, when pH is decreased the two fragments denature and do not come together as a complete barrel. These are mere speculations and we saw no reason to further investigate either scenarios. Furthermore, GFPs have been reported to show intrinsic pH dependence and this would complicate our data analysis (signal contributions from GFP maturation and signal from conformational change of transducer).

Table 5.1 Physical properties of fluorescent protein variants

Protein	No. of residues	Molecular Weight	pI	Charged residues (Asp+Glu) (Arg+Lys)	ε (M^{-1} cm^{-1}) at 280 nm
splitGFP-2H	319	35899.42	5.25	56 40	22015
Gamillus-mCherry2	513	57369.45	5.86	71 56	64540

5.3. pH insensitive Fluorescent Proteins

As an alternate and more robust approach we sought out FPs that were insensitive to pH. To design a pH insensitive ratio metric reporter system for pH biosensors, we selected Gamillus as donor and mCherry2 as acceptor. Gamillus is a green fluorescent protein, derived from jellyfish *Olindias formosus* and is a very rapidly maturing monomer. In comparison with conventional EGFP, Gamillus exhibits 1.8 times higher in vitro brightness, 1.6 times larger chromophore maturation rate, and 2.0 times higher photostability under 440 - 480 nm mercury arc light illumination (Shinoda, Ma et al. 2018). mCherry is currently the most widely used mRFP for live cell imaging due to its monomeric structure, high brightness, fast maturation, and good photostability. However, monomeric fluorescent proteins, such as mCherry, can be quite cytotoxic. mCherry2 is a variant of mCherry2 having 4 mutations and derived from corals belonging to Discosoma sp. Due to its decreased cytotoxicity, mCherry2 is a promising template for the development of biosensors. (Shen, Chen et al. 2017). These proteins were chosen because they both have very low acid sensitivity and are pH stable (4.0 - 8.0). Their physical and optical properties are listed in Tables 5.1-5.3.

Table 5.2 Optical properties of Gamillus and mCherry2

Protein	**Size (kDa)**	**pK_a**	**λ_{abs} (nm)**	**λ_{fl} (nm)**	**ε_{max} ($M^{-1}.cm^{-1}$)**	**QY**	**Brightness**
Gamillus	26.5	3.4	504	519	83,000	0.9	74.7
mCherry2	26.7	3.3	589	610	79,400	0.22	17.47

To design our plug and play system, the two fluorescent proteins were recombinantly cloned into pDream2.1; which is an excellent expression vector and can be used in any one of three major protein expression systems: bacteria, insect cells and mammalian cells. pDream2.1 carries the human, cytomegalovirus (CMV) promoter for high-level expression of genes in a variety of mammalian cell lines; the P10 baculovirus promoter for expression of genes in baculovirus-infected insect cells and the T7 promoter for convenient expression of genes in bacteria. There are seven restriction enzyme sites in the multiple cloning site (MCS) and a Flag tag for single column purification using anti-Flag antibodies and cleavable using enterokinase (EK) as shown in Figure 5.2.

Our design introduced a spacer-6xHis tag (GSGSS-HHHHHH) followed by a spacer-thrombin cleavage site (SSG-LVPR↓GS) after the FLAG tag in pDream2.1 plasmid for ease of purification as shown in Figure 5.2. Gamillus-MCS-mCherry2 follow the thrombin cleavage site. Restriction sites: BglII, EcoRI, KpnI, SacI and HindIII are available in the MCS (translating to amino acid residues EIWNSVPSSSL) to insert any protein between the two reporter FPs to create a biosensor.

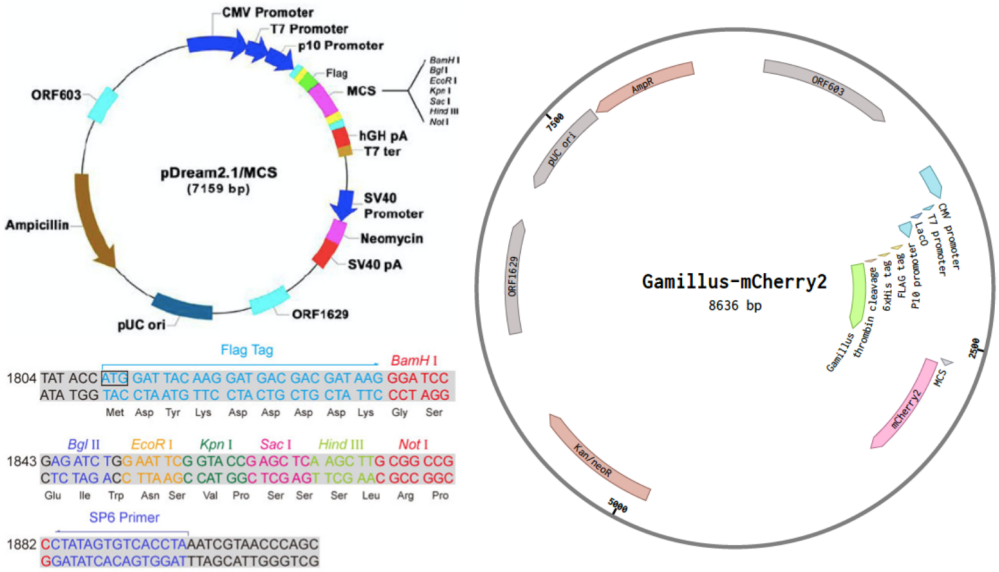

Figure 5.2 pDream2.1/MCS plasmid from GenScript (left). The plasmid containing Gamillus (green) and mCherry2 (pink) on the right.

The plug and play reporter system was transformed in BL21(DE3) bacterial cells and purified. To study its pH dependent behavior we measured FRET in bulk (shown in Figure 5.3). Its shows some pH dependence between pH 4 to pH 8, a change of about 4% in *E* across the dynamic range. This change in *E* could be attributed to change in angular orientation between the transition dipole moments of the donor and acceptor, proteins (see equation 2.8 in chapter 2). We expect this to correct itself on increasing the spacer length or by introducing a protein of interest in the MCS.

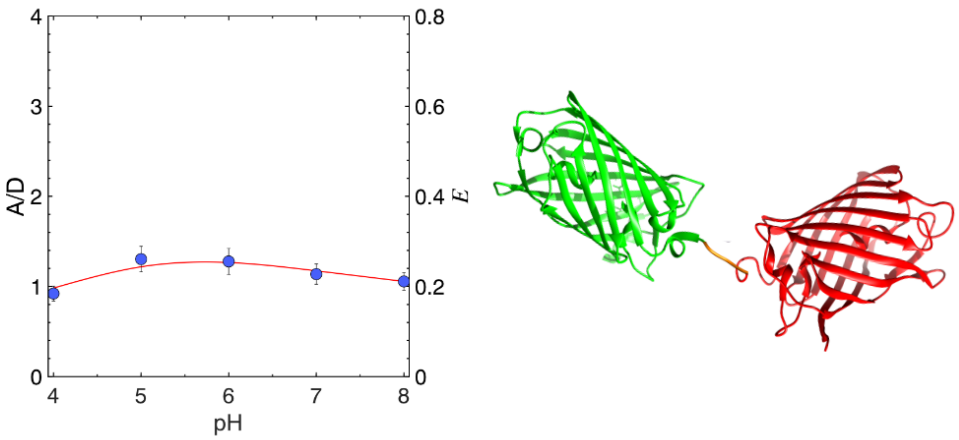

Figure 5.3 Low sensitivity pH response of the Plug-and-Play system. All experiments were done in 20 mM buffers (ionic strength adjusted to 150 mM with NaCl) using protein (10 nM, 2 mL). Fluorescence emission spectra was collected over a 490-690 nm range with 2 nm step size and integration of 1 s. The excitation wavelength was set to 470 nm and slit widths for excitation and emission were 5 nm and 10 nm; all measurements were taken at room temperature.

Table 5.3 Optical properties of Gamillus and mCherry2 as a FRET pair

Donor	**Acceptor**	**QY_D**	**EC_A** **(M^{-1} cm^{-1})**	**QY_A**	**R_0** **(Å)**	**$R_0 \times QY_A$**
Gamillus	mCherry2	0.9	79,400	0.22	60.28	13.26

There are numerous factors that need to be considered carefully for application of this plug and play system. Optimization of codon usage and inclusion of Shine-Dalgarno (AGGAGGT) 8 bases upstream of the start codon ATG in *E. coli* and use of Kozak (gccRccATGC) consensus sequences, where R is a purine (A or G) for mammalian cells. Efficiency of translation, unfortunate fusion to host proteins can

cause hindrance to folding and formation of inclusion bodies. Dimerization (homodimerization is more difficult to monitor than heterodimerization), protein degradation and export. Time, temperature, availability of O_2 and finally efficiency of posttranslational fluorophore maturation. Total amount of FPs and their molecular properties. Brightness and susceptibility to photoisomerization/bleaching along with autofluorescence of cells or culture media.

5.4. Conclusions

Although, all those forementioned challenges exist, this Plug-and-Play system can readily be used by introducing a protein of choosing between Gamillus and mCherry2. It can afford researchers a tool to study processes in-vitro and in-vivo that require a pH insensitive probe as shown in Figure 5.4. What makes our system practical and useful is the low cytotoxicity, enhanced brightness and quick maturation time for ratiometric optical readouts.

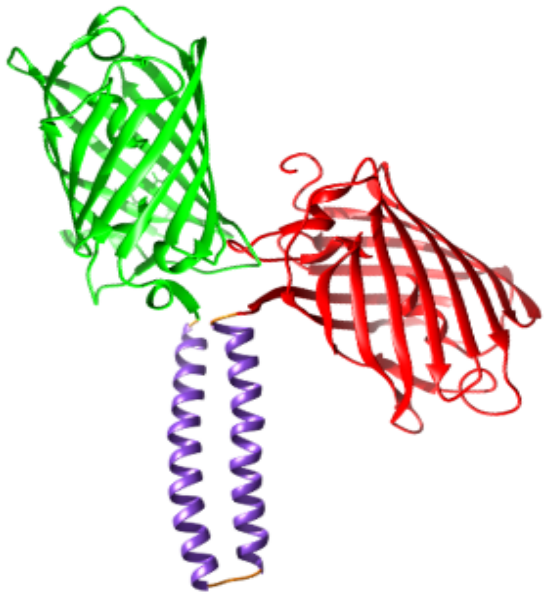

Figure 5.4 Plug-and-Play scaffold for pH insensitive FRET.

Appendix A: General Protocols with Material and Methods

Materials

1. *E. coli* competent cells (One Shot® BL21(DE3), Invitrogen).
2. LB-Agar (Fischer Scientific)
3. Luria-Bertani (LB) growth medium (Fischer Scientific)
4. Ampicillin (Fischer Scientific)
5. Isopropyl β-D-1 thiogalactopyranoside (IPTG) (Fischer Scientific)
6. EDTA-free protease inhibitor tablets (cOmplete Tablets, Roche)
7. Lamelli buffer (NuPAGE®, Invitrogen)
8. Pre-casted SDS-PAGE gel (Criterion™ TGX™, Bio-Rad)
9. SDS-PAGE molecular weight marker (Precision Plus Protein™, Bio-Rad)
10. SDS-PAGE running buffer (Novex® Tris-Glycine, Invitrogen)
11. Brilliant Blue (Fischer Scientific)
12. HiTrap™ Q HP ion exchange chromatography columns (GE Healthcare).
13. Reverse phase C-18 chromatography column (Higgins Analytical).
14. 5 kDa MWCO centricon centrifugal filter (Amicon® Ultracel, Merck Millipore).
15. tris(Hydroxymethyl)aminomethane (TRIS) (Fischer Scientific)
16. Disodium hydrogen phosphate (Na_2HPO_4) (Fischer Scientific)
17. Sodium dihydrogen phosphate (NaH_2PO_4) (Fischer Scientific)
18. Water (HPLC grade, Fischer Scientific)
19. Trifluoroacetic acid (TFA) (ACROS Organics)
20. Acetonitrile (HPLC Grade, Fischer Scientific)
21. Alexa 488 and Alexa 594 (Invitrogen)
22. 2-Mercaptoethanol (βME) (Fischer Scientific)
23. tris(2-Carboxyethyl)phosphine (TCEP) (Fischer Scientific)

24. Incubator (HERA THERM, Thermo Scientific)
25. Shaker incubator (MAXQ 5000, Thermo Scientific)
26. High-speed centrifuge (SORVALL LYNX 4000, Thermo Scientific)
27. Ultra high-speed centrifuge (SORVALL WX+ ULTRA Series, Thermo Scientific)
28. Homogenizer (FastPrep-24^{TM} 5G, MP Bio)
29. HPLC (1260 Infinity, Agilent Technologies)
30. Lyophilizer (FreeZone 4.5^{Plus}, LABCONCO)
31. Electrospray ionization mass spectrometer (Q EXACTIVE, Thermo Scientific)
32. UV-visible spectrophotometer (Cary 60 UV-Vis, Agilent Technologies)
33. Circular dichroism spectrometer (Chirascan Plus, Applied Photophysics)

Protocols

Cloning, transformation, protein expression and purification

Designed proteins were encoded for optimal expression in BL21(DE3) *E. coli* competent cells with the pBAT4 plasmid, which is ampicillin resistant and inducible with IPTG. The encoded and optimized plasmids was purchased from TOP Gene Technologies, Canada.

Plasmid transformation protocol by heat shock method

1. Prior to transformation, agar (containing 10 μg/ml ampicillin) plates were prepared and stored at 4 °C. Before use the agar plates were taken out from 4 °C storage, kept at room temperature for 30 minutes and then incubated at 37 °C.
2. Competent cells were taken out of -80 °C and thawed on ice for 20-30 minutes.
3. Plasmid DNA (5μL of ~100 ng stock solution) was transferred into 50 μL of competent cells in a micro centrifuge tube and mixed gently with finger tapping at the bottom of the tube. **(Note 1)**

4. The cell-DNA mixture was incubated on ice for 30 minutes.
5. The transformation mixture was heat shocked by dipping 2/3rd of the micro-centrifuge tube in a 42 °C pre-heated water bath for 45 seconds.
6. Immediately after the heat shock treatment the tube was placed on ice and kept for 2 minutes.
7. 250 μL of SOC media (without ampicillin, incubated at 37 °C) was added into the micro-centrifuge tube and the mixture was incubated in 37 °C shaker at 250 rpm for 45 minutes.
8. 50 μL of the mixture was plated in one agar plate (containing ampicillin) and remaining amount was plated in another agar plate (containing ampicillin) **(Note 2)**.
9. Plates were kept at 37 °C for overnight incubation.

Test of expression

1. A small, circular and isolated bacteria colony was picked out by a sterile loop and transferred in 10 mL sterilized LB containing 100 μg/mL ampicillin in a 50 mL falcon tube and kept at 37 °C shaker (225 rpm) for overnight incubation. **(Note 3)**
2. Next morning, the cell culture tube was centrifuged at 4000 rpm at 4 °C and the supernatant (LB with some enzymes secreted from bacterial cells) was discarded. The cell-pellet were re-suspended in 10 mL sterilized LB containing 100 μg/mL ampicillin. This exchange of LB removes β-lactamase for all transfers and growth inoculations and ensures that ampicillin is not degraded.
3. 1 mL of the re-suspended mixture was added to 10 mL of sterilized LB with 100 μg/mL ampicillin in duplicates and kept at 37 °C shaker (250 rpm) for incubation. Absorbance at 600 nm (OD_{600nm}) of these samples were monitored at 1-hour intervals. **(Note 4)**

4. At $OD_{600nm} \geq 0.7$, protein expression was induced by addition of 100 μg/mL IPTG (+IPTG). One falcon tube was incubated overnight at 37°C and another at 20 °C shaker (250 rpm).
5. After overnight incubation samples with and without IPTG from each temperature culture were collected, normalized to the minimum absorbance, centrifuged at 13X g and the supernatants were discarded. Then 40 μL Lamelli buffer was added to each of the precipitates, mixed well and cells were lysed by heating at 95 °C for 15 minutes.
6. After lysis, the samples were centrifuged and ran in SDS-PAGE denaturing gel, with an initial voltage of 100 V, which was increased to 120 V after 20 minutes.
7. Brilliant blue staining for an hour followed by de-staining (2-3 hours to overnight) of the gels were done and compared with the standard marker to identify expression of protein.

Bulk expression and purification

1. A fresh plating from the glycerol stock was used for preparation of the starter culture and incubated overnight at 37 °C shaker (250 rpm).
2. Similar as step-2 in test of expression protocol.
3. The whole volume (10 mL) of the re-suspended cells were added in a 6 L conical flask containing freshly prepared 1 L sterilized LB with 100 μg/mL ampicillin and kept overnight at 37 °C shaker (250 rpm) for incubation. Absorbance at 600 nm (OD_{600nm}) of the incubated sample was monitored at every 1-hour intervals. **(Note 5)**
4. At $OD_{600nm} \geq 0.7$, protein expression was induced by addition of 100 μg/mL IPTG and the conical flask was incubated overnight in a shaker (250 rpm) at 37 °C.
5. After overnight incubation samples were collected in centrifugation bottles, centrifuged at 4 °C, 8000 rpm for 1 hour and the supernatants (LB) were discarded.

6. The cells (pellet after centrifugation) were collected in a 50 mL falcon tube and a protease inhibitor tablet was added and mixed well with the cells using 15-20 mL of lysis buffer.
    a. Lysis buffer for BBL based scaffolds: 20 mM TRIS and 5 mM TCEP at pH 7.2.
    b. Lysis buffer for coiled-coil based scaffolds: 20 mM phosphate buffer and 5 mM TCEP at pH 7.
7. The cells were lysed in a homogenizer using 5 cycles of 30 seconds.
8. After lysis, the lysate was transferred in centrifugation tube and centrifuged at 4 °C, 35000 rpm for 1 hour. The supernatant was separated from the pellet. **(Note 6)**
9. The pH of the medium was adjusted depending on the pI of the protein. The pH used in the ion exchange was kept ± 2 pH units about the pI of the protein.
10. Using a 0.2 μm syringe filter supernatant was filtered to remove any precipitates or agglomerates.
11. The supernatant was then loaded in an ion exchange HiTrap Q-column (positively charged static phase and negatively charged mobile phase) at a flow rate of 5 mL/min of the lysis buffer A. Once a stable baseline is reached an increasing salt gradient (buffer B) was applied to elute the column bound proteins. The eluents were collected peak wise in HPLC glass vials by a robotic collection device. **(Note 7)**
    a. Buffer A: 20 mM phosphate buffer with 5 mM TCEP at pH 7.
    b. Buffer B: 20 mM phosphate buffer with 1 M NaCl and 5 mM TCEP at pH 7
12. Eluent from each of the vials was used for SDS-PAGE denaturing gel electrophoresis to identify the protein of interest.
13. The ion exchange column fractions containing the protein of interest were then loaded into a C18 reverse phase column at a flow rate of 1 mL/min using the HPLC. The gradient of acetonitrile was then applied to elute

proteins.

a. Buffer A: 95 % water + 5 % acetonitrile + 0.01 % of TFA, pH ~2.
b. Buffer B: 5 % water + 95 % acetonitrile + 0.01 % of TFA, pH ~2.

14. The collected fractions were analyzed using denaturing SDS-PAGE gel electrophoresis.
15. Successful purification was confirmed by ESI mass spectrometry. **(Note 8)**
16. The fractions containing the protein of interest was snap frozen using liquid nitrogen, lyophilized and then stored as dry powder in micro-centrifuge tubes at 4 °C.

Protein characterization by CD spectroscopy

1. Determination of proteins structural content and measurements conformational changes of the purified protein were done by using circular dichroism (CD) spectrometer. **(Note 9)**
2. The pH and temperature responses of the purified proteins were also studied by circular dichroism (CD) before conjugating with extrinsic fluorophores for high gain fluorescence signal read-outs.

Fluorophore labeling and purification protocols

1. An appropriate concentration (50-500 μM, 1 mL) of the protein in a suitable buffer was prepared at room temperature. The buffer was de-oxygenated by nitrogen purging to maintain the reduced thiols prior to use. **(Note 10)**
2. 10-fold molar excess TCEP **(Note 11)** was added to the protein solution to reduce the disulfide bonds in the protein.
3. After 30 minutes, excess TCEP was filtered using the centricon filter (cutoff 3kDa) and washed thoroughly with the reaction buffer. **(Note 12)**
4. The pH of the reduced protein solution was then adjusted to 6.5-7.2, making the protein thiol groups sufficiently nucleophilic to react. **(Note 13)**

5. Alexa 594 maleimide (A594) and Alexa 488 maleimide dyes were dissolved separately in degassed reaction buffer immediately prior to use and were protected from light.
6. 1.5 molar excess of the each of the dyes were added dropwise with continues stirring at room temperature. The whole process of mixing was performed within 30 minutes.
7. The reaction mixture was kept at room temperature for 3 hours under continues stirring. It was observed that overnight reaction at 4 °C also produces similar yield.
8. Upon completion of the reaction, 2 μL βME was added consume excess thiol-reactive reagent.
9. Sephadex G-25 column was used to separate the protein-dye conjugate from the unreacted dyes.
10. The Sephadex G-25 eluent was then passed through 10 μm - C18 reverse phase column to separate the differently labeled species from each other. **(Note 14)**

Notes

1. Mixing with pipette tip is not recommended as the shear stress from pipetting may decrease the transformation efficiency.
2. The comparative plating of small and large amount of the LB-cells were crucial to compare the growth of the bacterial colonies by visual inspection. The number of colonies in the 250 μL plating should be ~5 times of the 50 μL plating. Liquids were dried out before incubation. As too much liquid media on the agar plate helps the bacteria to diffuse through the liquid and hinder the formation of colonies.
3. The 30 % glycerol stocks were made from this overnight culture and stored at -80 °C for subsequent uses.

4. The parameters, which can be tested are: OD_{600} for inducing the protein expression (should be kept between 0.7-1.2); temperature for growth (37 °C, 30 °C, 25 °C and 20 °C) and time of expression (3 hours, 5 hours, 8 hours and/or overnight) after inducing with IPTG.
5. 5 L of free space was left empty in the conical flask to provide required oxygen for cell growth. Decreasing the volume of the free space reduces the cell count.
6. As the pK_a of the coiled-coil scaffold is 4.8, the protein should be soluble in the supernatant of pH=7.2. After centrifugation, the presence of coiled-coil in the supernatant was confirmed by SDS-PAGE denaturing gel electrophoresis. We observed more than ~95% of the expressed coiled-coil was solubilized in the supernatant and the cell debris contained negligible amount of the protein.
7. pK_a of coiled-coil scaffold was estimated as 4.8. At pH 7.2 coiled-coil would be negatively charged. Hence an ion exchange column of positive static phase was used for purification of coiled-coil scaffold. It's important to note that DNAs extracted from cells after lysis were also negatively charged and stuck to Q-columns. We have observed a high absorbance at 260nm, which resulted OD_{280nm}/ OD_{260nm}~1. To remove the protein-bound DNA we have used the reverse phase chromatography (RP-HPLC) and ammonium sulfate precipitation methods.
8. It was confirmed that ESI-MS samples were void of salts.
9. CD signals were monitored at 222 nm, 208 nm and 190 nm to estimate the helical contents of the proteins. It is also suggested to prepare buffers using HF to adjust the pH and NaF to adjust the ionic strength since Cl has an absorbance at wavelengths lower that 200 nm.
10. Alexa fluorophores are not stable in phosphate buffer for long duration; hence TRIS is a better choice as reaction buffer. If phosphate is used, then the reaction should be completed within few hours.

11. TCEP stock solution in 20 mM TRIS-HCl (pH 6.5) was prepared in advance and the pH of the solution was adjusted to 6.5 by dropwise addition of 0.1M NaOH solution.
12. The removal of TCEP before labeling reaction is crucial as excess TCEP affects the yield of the cysteine-maleimide reaction. It is already reported that TCEP strongly interacts with maleimide and consumes the maleimide functionalized dyes. To prevent the unwanted consumption of maleimide functionalized dyes the excess TCEP was removed by filtration.
13. pH of the reaction medium plays crucial role in extrinsic fluorophore labeling efficiency. The maximum yield was obtained at pH range 6.5-7.0, for most of the proteins. pH close to 8 is strictly avoidable as amine protonation starts at this pH and drastically decreases the labeling efficiency.
14. Addition of two fluorophores at a time in the reaction mixture would lead to 6 different species: free A488, free A594, protein-A488-A488, protein-A594-A594, protein-A488-A594 and unreacted protein with equal probability for each. Among these, the target species, protein-A488-A594 was separated through RP-HPLC by changing the gradient of solvent B. It was observed that 0.5 mL/min increment of B (in a 10 μm - C18 reverse phase column) helped in separation of overlapping peaks of different species.

Appendix B: Optimized Conditions for Proteins Expression and Engineering

Chapter 3: Coiled-Coils

Recombinant Protein Expression. All coiled-coils variants were produced by recombinant means as full genes, cloned in the bacterial expression vector pBAT4. Plasmids containing the various genes were transformed into *E. coli* BL21(DE3) competent cells. Cells were grown in LB broth with 50 μg/mL ampicillin at 293 K to an OD_{600} 0.7-0.9, followed by induction with 1 mM isopropyl-β-D-thio-galactopyranoside (IPTG). The cells were kept in growing conditions overnight and then harvested and centrifuged at 8000 rpm for 30 min.

Protein Purification. The pellets were resuspended in 20 mM phosphate buffer at pH 7.5 with protease inhibitor cocktail and 2 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP). Cell lysis was performed by homogenizer (5 cycles of 15 s) and the lysate was ultracentrifuged at 35000 rpm for 40 min at 277 K. The obtained supernatant was pH adjusted to 7.5 and syringe filtered with 0.2 μm filter and loaded onto a HPLC HiTrap Q cation exchange column (GE Healthcare) and eluted with a gradient from 0 M to 1 M NaCl in 20 mM phosphate buffer with 2 mM TCEP at pH 7.5. The fractions containing the protein variant were pooled and subjected to a second round of purification using PROTO 300 C18 reverse phase column (Higgins Analytical) using a 5 % - 95 % acetonitrile gradient with 0.1 % trifluoroacetic acid (TFA) for elution. Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis, and confirmed by electrospray mass spectrometry followed by pure variants being pooled, lyophilized and stored at 277 K.

Protein Labeling. The purified proteins were sequentially labeled with extrinsic fluorophores via thiol-maleimide conjugation. The protein was dissolved in labeling buffer (20 mM phosphate buffer with 150 mM NaCl) and TCEP was added (protein:TCEP 1:1), pH of the solution was adjusted to 7.2 and allowed to equilibrate for 30 min. Alexa 594 C_5 maleimide dye (acceptor) was dissolved in the labeling buffer and added dropwise to the protein solution (protein:acceptor 1:1.1) and mixed. The reaction mixture was left overnight at 277 K and next morning 1 μL of β-mercaptoethanol (BME) was added to the reaction mixture to quench any unreacted dye. The sample was run through HPLC reverse phase column to remove excess dye and isolate singly labeled protein. Buffer exchange was done to remove acetonitrile via centricon and similar process was followed as for the first labeling reaction with Alexa 488 C_5 maleimide dye (donor) except (protein:donor 1:1.4).

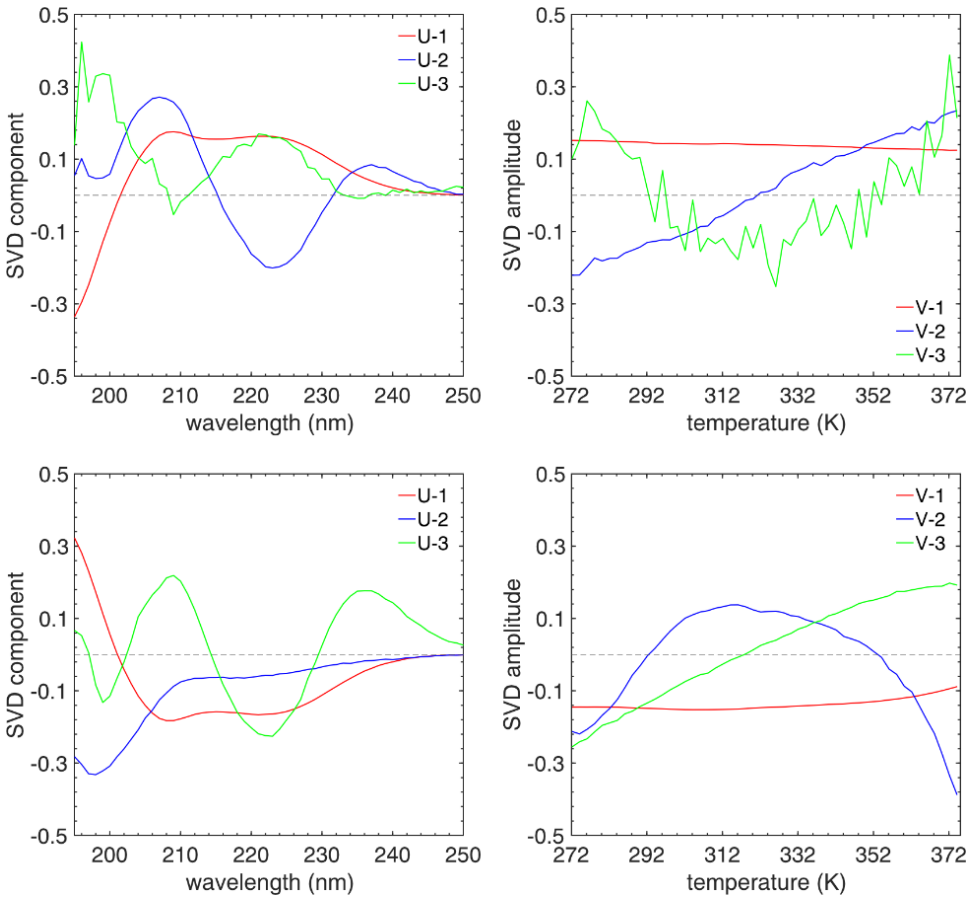

Figure B3. 1 SVD analysis of Control at pH 4 (top) and at pH 8 (bottom), structural components (left) and their corresponding amplitude (right).

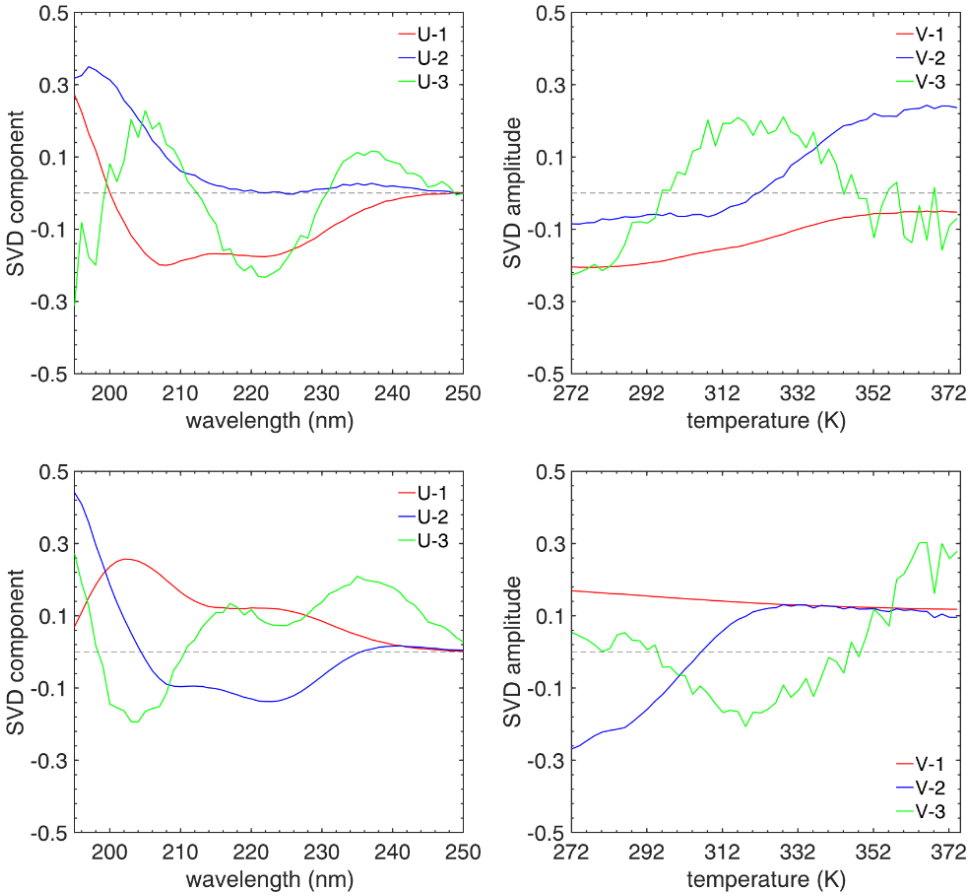

Figure B3. 2 SVD analysis of 2-Histidine at pH 4 (top) and at pH 8 (bottom), structural components (left) and their corresponding amplitude (right).

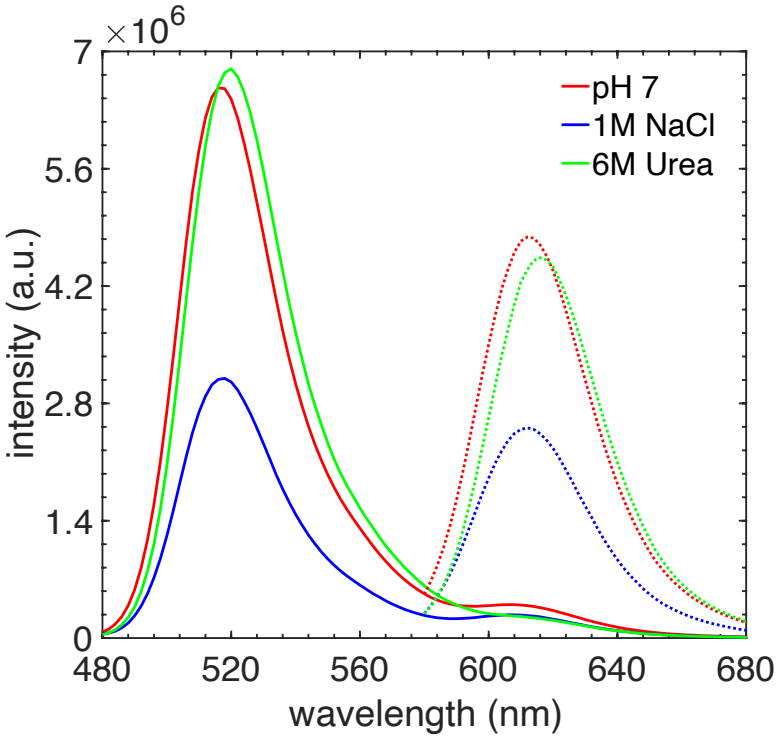

Figure B3. 3 2H-F fluorescence with donor and acceptor excitations.

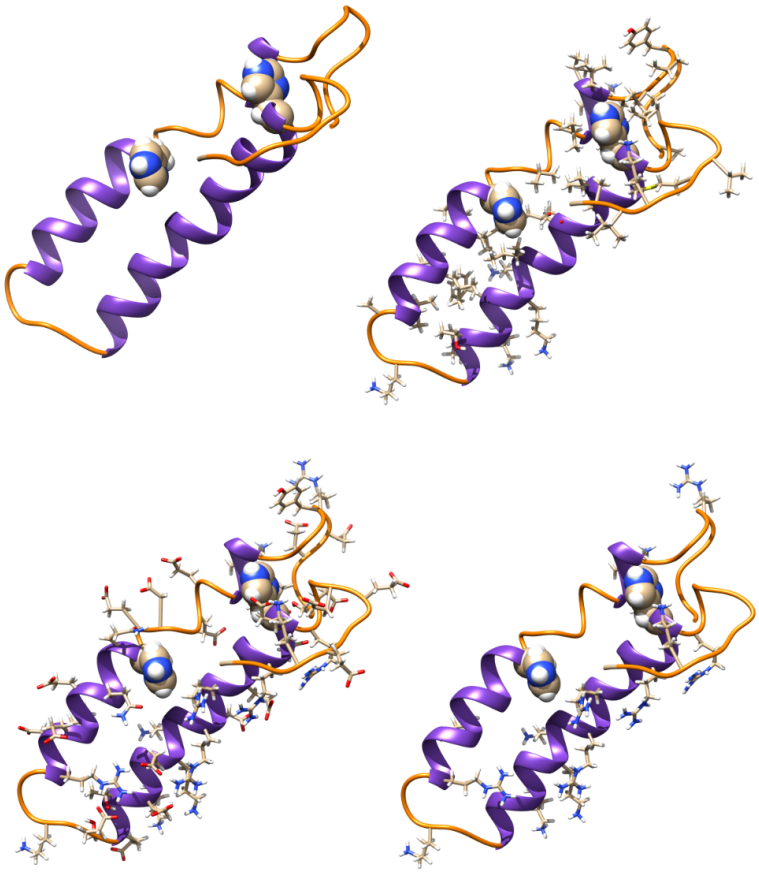

Figure B3. 4 Partially unfolded 2-Hisitidine (top-left), hydrophobic sidechains (top-right), polar sidechains (bottom-left) and positively charged sidechains (bottom-right).

Simulations on 2-Histidine

MD simulations were performed in the GROMACS suite using the OPLS all-atom force field. Water molecules were modeled with the TIP3P representation. Periodic boundary conditions were used, and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) summation using a grid spacing of 0.16 nm combined with a fourth-order cubic interpolation to derive the potential and forces in-between grid points. The real space cutoff distance was set to 1.0 nm and the van der Waals cutoff to 1.0 nm. The bond lengths were fixed, and a time step of 2 fs was used for numerical integration of the equations of motion. Coordinates were recorded every 10 ps. The simulations were performed at 310 K. We altered the protonation state of all histidine residues before the MD run based on their estimated pK_a values and relative to a nominal pH of 5. The ionization states were kept constant for the entire MD run. Histidine protonation was carried out using the pdb2gmx tool (HISE type). The protein was placed in a dodecahedral water box large enough to contain protein and at least 1.0 nm of solvent on all sides (volume ~ 335 nm^3). The structure was solvated with 10,570 water molecules, and 8 Na^+ ions were added to neutralize the system. The starting structure was subjected to energy minimization using the steepest descent method. The simulations were subjected to the modified Berendsen thermostat to maintain the exact temperature, followed by Parrinello-Rahman for pressure coupling at 1 bar before the production run was started. All the simulations were run on the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputing center (SDSC).

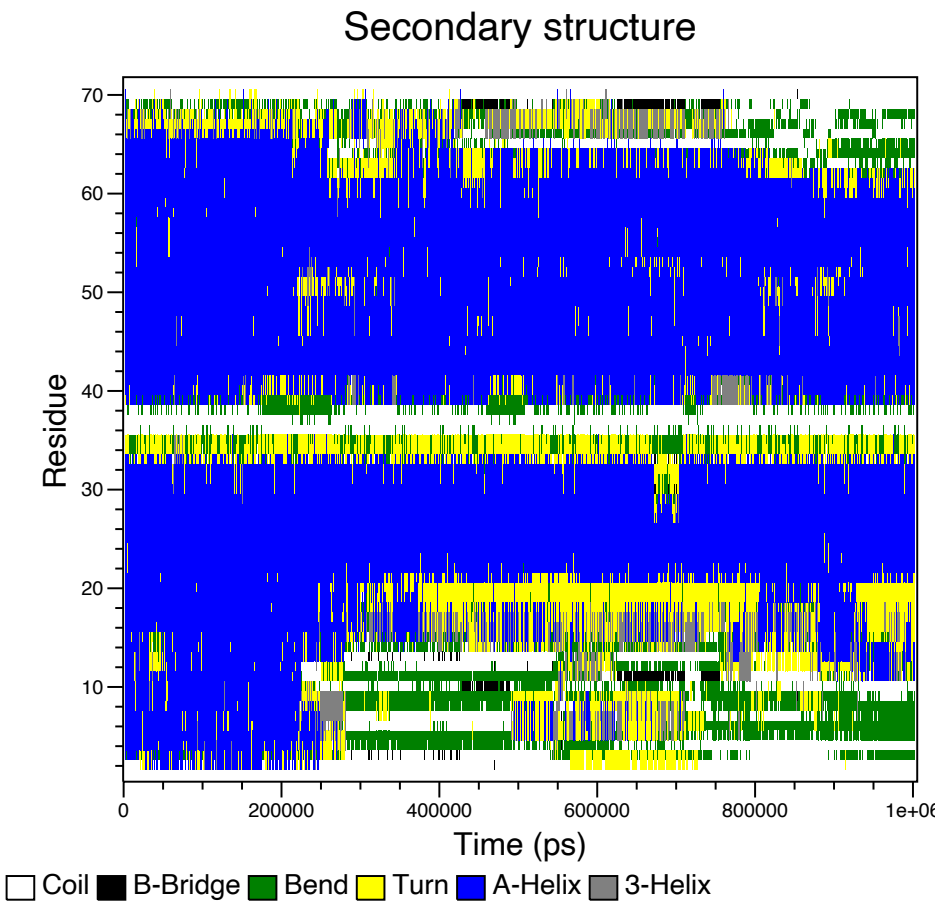

Figure B3. 5 Changes in secondary structure of 2-Histidine as a function of time

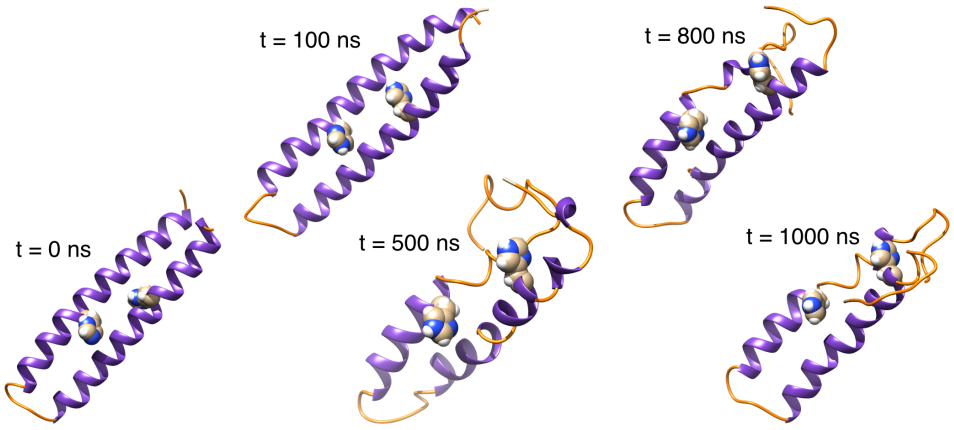

Figure B3. 6 Time dependent change in helical structure from simulations on 2-Histidine.

Table B3. 1 Physical properties of coiled-coil variants.

Protein	**No. of residues**	**Molecular Weight**	**pI**	**Charged residues (Asp+Glu) (Arg+Lys)**	**ε (M^{-1} cm^{-1}) at 280 nm**
1NT2 (Coiled-coil segment)	73	8870.19	4.93	22 17	1490
Control (C)	73	8578.81	4.64	22 14	1490
2-Histidine (2H)	73	8640.80	4.81	22 14	1490
4-Histidine (4H)	73	8668.77	5.00	22 14	1490
C-F	76	8842.14	4.64	22 14	1615
2H-F	75	8847.08	4.81	22 14	1615
2H-I-F	81	9473.88	5.32	22 18	1615

2H-P2	75	8930.15	4.81	22 14	6990
2H-P	72	8615.81	4.78	21 13	6990
2H-L-P	83	9389.57	4.78	21 13	6990

To further understand the challenges faced with 2-Histidine designs, Control was labeled, and it had about 30% change in FRET as shown in Figure 3.8. This was not expected and also contradicts our CD results. We suspect that the dye affects the structure of the protein, and this contradictory behavior is a consequence of that. This is one of the biggest challenges in fluorescence wherein the dyes interact with the protein and affects its behavior. Although if we just wanted a pH sensor capable of giving a 30% change in FRET, control labeled with Alexa 488 and Alexa 594 is a great candidate for it.

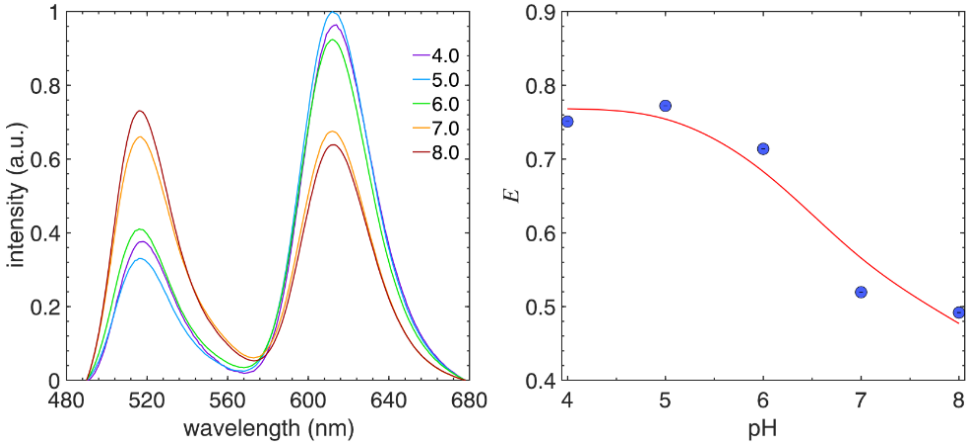

Figure 5.5 Ensemble steady state fluorescence intensity measurements (left) on Control and its corresponding change in FRET (right) as a function of pH.

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{3.3}$$

Furthermore, dynamic anisotropy was done as a function of pH to verify whether the dyes were free to swivel and not trapped within the protein. Anisotropy r is used to measure the average angular displacement of the fluorophore that occurs between absorption and emission of a photon. It is the ratio of the difference between the intensities and the total intensity. $I_{\parallel}$ is the intensity collected when the emission polarizer is parallel to the polarized excitation and $I_{\perp}$ is the intensity when

the polarizer is perpendicular to the polarized excitation. Anisotropy is dependent on the rate and degree of rotational diffusion during the life time of the excited state. The rate of rotational diffusion in turn depends on the viscosity of the solvent and the size and shape of the rotating molecule. Since we were using buffers with no cosolvents, the viscosity contribution of the solvent could be ignored and changes in anisotropy would depend on the shape and size of the probe. The lifetimes were measured and anisotropy data was fitted to a single exponential decay and time from the decay was approximated as time of tumbling for the dyes and reported in Figure 3.9. The donor was excited at 495 nm and emission was collected at 519 nm, the acceptor was excited at 590 nm and emission collected at 617 nm to a photon count of 1000. Slit widths were 10 nm and time range of 200 ns were used with a pulsed laser at 100% power and frequency of 20.00 MHz was used.

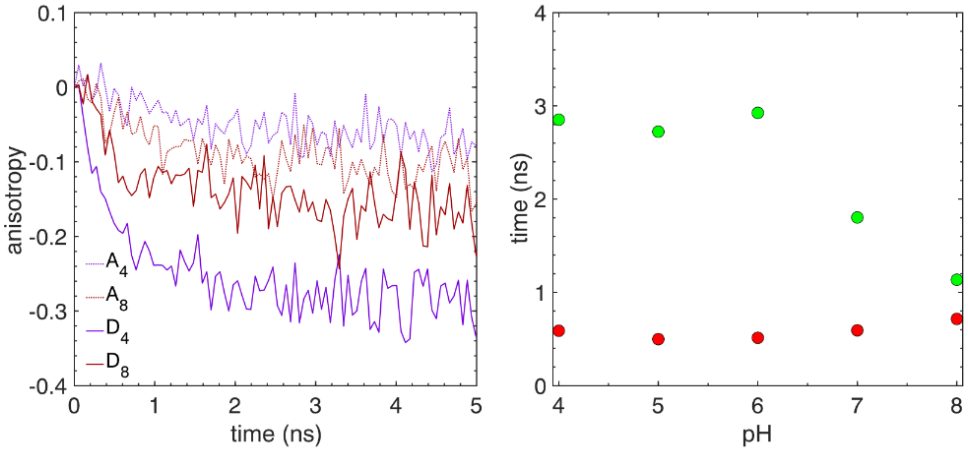

Figure 5.6 Dynamic anisotropy (left) of Control at pH 4 (violet) and pH 8 (dark red), donor (solid line) and acceptor (dashed line) and their corresponding time decays as a function of pH (right); donor (green) and acceptor (red).

The anisotropy of acceptor does not change drastically but that of donor changes by more than 2-fold. This can be interpreted as the acceptor having lesser degree of freedom and whereas the donor had much more degrees of freedom. As the pH is increased, the helix melts and the dye collapses into a more compact or globular conformation of the protein. This confirmed the suspicion about dye affecting the behavior of the protein.

Table B3. 2 Summary of MARCOIL predictions of coiled-coil domains at different levels of stringency

	Threshold 50%	Threshold 90%	Threshold 99%
C	From 11-72 (62)	From 15-34 (20) From 46-69 (24)	From 22-30 (9) From 49-65 (17)
2H	From 11-72 (62)	From 16-33 (18) From 46-69 (24)	From 49-65 (17)
4H	From 13-72 (60)	From 16-33 (18) From 46-69 (24)	From 49-65 (17)

List of amino acid sequence with [highest coiled-coil probability in percent] and heptad phase.

Control (C)

01M[00.4]f 02G[01.1]g

03S[04.4]a 04A[13.7]b 05E[20.4]c 06E[22.7]d 07D[26.3]e 08R[31.0]f 09Y[33.8]g

10V[45.4]a 11V[51.6]b 12A[66.6]c 13L[80.4]d 14V[83.8]e 15K[91.9]f 16K[95.9]g

17L[97.8]a 18E[98.6]b 19E[98.8]c 20V[98.9]d 21D[99.0]e 22E[99.1]f 23A[99.1]g

24I[99.2]a 25Q[99.2]b 26M[99.2]c 27L[99.2]d 28Q[99.2]e 29E[99.2]f 30K[99.0]g

31L[98.7]a 32E[98.0]b 33D[95.6]c 34I[91.5]d 35R[86.9]e 36A[79.2]f 37G[69.5]g

38K[70.5]a 39E[70.5]b 40S[68.2]c 41E[68.7]f 42I[68.3]g

43T[72.9]a 44E[86.1]b 45K[89.5]c 46F[91.8]d 47E[97.3]e 48K[98.6]f 49K[99.3]g

50I[99.6]a 51R[99.8]b 52E[99.9]c 53L[99.9]d 54R[100.]e 55E[100.]f 56L[100.]g

57V[100.]a 58R[100.]b 59D[100.]c 60V[99.9]d 61E[99.9]e 62R[99.9]f 63E[99.9]g

64I[99.8]a 65E[99.5]b 66E[98.5]c 67V[95.7]d 68M[92.9]e 69E[91.0]f 70K[86.1]g

71I[75.5]a 72A[50.1]b 73G[15.4]c

2-Histidine (2H)

01M[00.4]f 02G[01.0]g

03S[04.3]a 04A[13.3]b 05E[19.9]c 06E[22.1]d 07D[25.6]e 08R[30.2]f 09Y[32.9]g

10V[44.0]a 11V[50.1]b 12A[64.6]c 13L[78.0]d 14V[81.4]e 15K[89.2]f 16K[93.2]g

17L[95.2]a 18E[96.1]b 19E[96.4]c 20V[96.5]d 21D[96.6]e 22E[96.7]f 23A[96.7]g

24H[96.8]a 25Q[96.8]b 26M[96.9]c 27L[96.9]d 28Q[97.0]e 29E[96.9]f 30K[96.8]g

31L[96.5]a 32E[95.7]b 33D[93.5]c 34I[89.5]d 35R[85.2]e 36A[77.7]f 37G[68.3]g

38K[69.4]a 39E[69.5]b 40S[67.4]c 41E[68.2]f 42I[67.9]g

43T[72.6]a 44E[85.9]b 45K[89.4]c 46F[91.8]d 47E[97.2]e 48K[98.6]f 49K[99.3]g
50I[99.6]a 51R[99.8]b 52E[99.9]c 53L[99.9]d 54R[99.9]e 55E[99.9]f 56L[99.9]g
57V[99.9]a 58R[99.9]b 59D[99.9]c 60H[99.9]d 61E[99.9]e 62R[99.9]f 63E[99.8]g
64I[99.7]a 65E[99.4]b 66E[98.5]c 67V[95.7]d 68M[92.8]e 69E[91.0]f 70K[86.1]g
71I[75.5]a 72A[50.1]b 73G[15.4]c

4-Histidine (4H)

01M[00.2]f 02G[00.5]g
03S[02.1]a 04A[06.2]b 05E[09.3]c 06E[10.3]d 07D[11.9]e 08R[14.0]f 09Y[15.2]g
10H[20.2]a 11V[28.9]b 12A[49.6]c 13L[68.9]d 14V[73.7]e 15K[84.9]f 16K[90.6]g
17L[93.4]a 18E[94.7]b 19E[95.1]c 20V[95.2]d 21D[95.4]e 22E[95.5]f 23A[95.6]g
24H[95.6]a 25Q[95.8]b 26M[95.8]c 27L[95.9]d 28Q[95.9]e 29E[95.9]f 30K[95.7]g
31L[95.5]a 32E[94.8]b 33D[92.7]c 34I[89.1]d 35R[85.1]e 36A[78.1]f 37G[69.3]g
38K[70.4]a 39E[70.5]b 40S[68.5]c 41E[69.2]f 42I[68.8]g
43T[73.2]a 44E[85.4]b 45K[88.5]c 46H[90.5]d 47E[96.7]e 48K[98.1]f 49K[99.0]g
50I[99.4]a 51R[99.6]b 52E[99.8]c 53L[99.8]d 54R[99.9]e 55E[99.9]f 56L[99.9]g
57V[99.9]a 58R[99.9]b 59D[99.9]c 60H[99.8]d 61E[99.8]e 62R[99.8]f 63E[99.8]g
64I[99.6]a 65E[99.4]b 66E[98.4]c 67V[95.6]d 68M[92.8]e 69E[90.9]f 70K[86.0]g
71I[75.4]a 72A[50.1]b 73G[15.4]c

Table B3. 3 Amino acid sequence information of coiled-coil variants

Protein	AA Sequence
1NT2 (Coiled-coil segment)	~~KELRR~~EDRYV VALVK~~A~~LEE~~I~~ DE~~S~~I~~N~~ML~~N~~EK LEDIRA~~V~~KES EITEKFEKKI RELREL~~R~~RDV EREIEEVMEK IA~~P~~
C	MGSAEEDRYV VALVKKLEEV DEAIQMLQEK LEDIRAGKES EITEKFEKKI RELRELVRDV EREIEEVMEK IAG
2H	MGSAEEDRYV VALVKKLEEV DEA***H***QMLQEK LEDIRAGKES EITEKFEKKI RELRELVRD***H*** EREIEEVMEK IAG
4H	MGSAEEDRY***H*** VALVKKLEEV DEA***H***QMLQEK LEDIRAGKES EITEK***H***EKKI RELRELVRD***H*** EREIEEVMEK IAG
C-F	MGCGSAEEDR YVVALVKKLE EVDEAIQMLQ EKLEDIRAGK ESEITEKFEK KIRELRELVR DVEREIEEVM EKIAGC
2H-F	MCGSAEEDRY VVALVKKLEE VDEA***H***QMLQE KLEDIRAGKE SEITEKFEKK IRELRELVRD ***H***EREIEEVME KIAGC
2H-I-F	MGCKKGSAEE DRYVVALVKK LEEVDEA***H***QM LQEKLEDIRA GKESEITEKF EKKIRELREL VRD***H***EREIEE VMEKIAGKKC G
2H-PF	MWGSAEEDRY VVALVKKLEE VDEA***H***QMLQE KLEDIRAG*E SEITEKFEKK IRELRELVRD ***H***EREIEEVME KIAGC
2H-P2	MWGSAEEDRY VVALVKKLEE VDEA***H***QMLQE KLEDIRAGKE SEITEKFEKK IRELRELVRD ***H***EREIEEVME KIAGC

2H-P	MSAEWDRYVV ALVKKLEEVD EA***H***QMLQEKL EDIRAGKESE ITEKFEKKIR ELRELVRD***H***E REIEEVMECI AG
2H-L-P	MSAEWDRYVV ALVKKLEEVD EA***H***QMLQEKL EDIRAGKEAG SAGSAGSAGS EITEKFEKKIR ELRELVRD***H***E REIEEVMECI AG

Table B3. 4 Deoxyribonucleic acid sequence information of coiled-coil variants.

Protein	DNA Sequence
C	ATGGGCAGCGCGGAAGAAGATCGCTATGTGGTGGCGCTG GTGAAAAAACTGGAAGAAGTGGATGAAGCGATTCAGATGC TGCAGGAAAAACTGGAAGATATTCGTGCGGGCAAAGAAAG CGAAATTACCGAAAAATTTGAAAAAAAAATTCGCGAACTGC GCGAACTGGTGCGCGATGTGGAACGCGAAATTGAAGAAGT GATGGAAAAAATTGCGGGCTAA
2H	ATGGGCAGCGCGGAAGAAGATCGCTATGTGGTGGCGCTG GTGAAAAAACTGGAAGAAGTGGATGAAGCGCATCAGATGC TGCAGGAAAAACTGGAAGATATTCGTGCGGGCAAAGAAAG CGAAATTACCGAAAAATTTGAAAAAAAAATTCGCGAACTGC GCGAACTGGTGCGCGATCATGAACGCGAAATTGAAGAAGT GATGGAAAAAATTGCGGGCTAA
4H	ATGGGCAGCGCGGAAGAAGATCGCTATCATGTGGCGCTGG TGAAAAAACTGGAAGAAGTGGATGAAGCGCATCAGATGCT GCAGGAAAAACTGGAAGATATTCGTGCGGGCAAAGAAAGC GAAATTACCGAAAAACATGAAAAAAAAATTCGCGAACTGCG CGAACTGGTGCGCGATCATGAACGCGAAATTGAAGAAGTG ATGGAAAAAATTGCGGGCTAA
C-F	ATGGGCTGCGGCAGCGCGGAAGAAGATCGCTATGTGGTG GCGCTGGTGAAAAAACTGGAAGAAGTGGATGAAGCGATTC AGATGCTGCAGGAAAAACTGGAAGATATTCGTGCGGGCAA AGAAAGCGAAATTACCGAAAAATTTGAAAAAAAAATTCGCG AACTGCGCGAACTGGTGCGCGATGTGGAACGCGAAATTGA AGAAGTGATGGAAAAAATTGCGGGCTGCTAA
2H-F	ATGTGCGGCAGCGCGGAAGAAGATCGCTATGTGGTGGCG CTGGTGAAAAAACTGGAAGAAGTGGATGAAGCGCATCAGA TGCTGCAGGAAAAACTGGAAGATATTCGTGCGGGCAAAGA AAGCGAAATTACCGAAAAATTTGAAAAAAAAATTCGCGAAC TGCGCGAACTGGTGCGCGATCATGAACGCGAAATTGAAGA AGTGATGGAAAAAATTGCGGGCTGCTAA

2H-l-F	ATGGGCTGCAAAAAAGGCAGCGCGGAAGAAGATCGCTATG TGGTGGCGCTGGTGAAAAAACTGGAAGAAGTGGATGAAGC GCATCAGATGCTGCAGGAAAAACTGGAAGATATTCGTGCG GGCAAAGAAAGCGAAATTACCGAAAAATTTGAAAAAAAAAT TCGCGAACTGCGCGAACTGGTGCGCGATCATGAACGCGAA ATTGAAGAAGTGATGGAAAAAATTGCGGGCAAAAAATGCG GCTAA
2H-PF	ATGTGGGGCAGCGCGGAAGAAGATCGCTATGTGGTGGCG CTGGTGAAAAAACTGGAAGAAGTGGATGAAGCGCATCAGA TGCTGCAGGAAAAACTGGAAGATATTCGTGCGGGCTAGGA AAGCGAAATTACCGAAAAATTTGAAAAAAAAATTCGCGAAC TGCGCGAACTGGTGCGCGATCATGAACGCGAAATTGAAGA AGTGATGGAAAAAATTGCGGGCTGCTAA
2H-P2	ATGTGGGGCAGCGCGGAAGAAGATCGCTATGTGGTGGCG CTGGTGAAAAAACTGGAAGAAGTGGATGAAGCGCATCAGA TGCTGCAGGAAAAACTGGAAGATATTCGTGCGGGCAAAGA AAGCGAAATTACCGAAAAATTTGAAAAAAAAATTCGCGAAC TGCGCGAACTGGTGCGCGATCATGAACGCGAAATTGAAGA AGTGATGGAAAAAATTGCGGGCTGCTAA
2H-P	ATGAGCGCGGAATGGGATCGCTATGTGGTGGCGCTGGTGA AAAAACTGGAAGAAGTGGATGAAGCGCATCAGATGCTGCA GGAAAAACTGGAAGATATTCGTGCGGGCAAAGAAAGCGAA ATTACCGAAAAATTTGAAAAAAAAATTCGCGAACTGCGCGA ACTGGTGCGCGATCATGAACGCGAAATTGAAGAAGTGATG GAATGCATTGCGGGCTAA
2H-L-P	ATGAGCGCGGAATGGGATCGCTATGTGGTGGCGCTGGTGA AAAAACTGGAAGAAGTGGATGAAGCGCATCAGATGCTGCA GGAAAAACTGGAAGATATTCGTGCGGGCAAAGAAGCGGGC AGCGCAGGCTCTGCCGGTAGCGCCGGCAGCGAAATTACC GAAAAATTTGAAAAAAAAATTCGCGAACTGCGCGAACTGGT GCGCGATCATGAACGCGAAATTGAAGAAGTGATGGAATGC ATTGCGGGCTAA

Chapter 4: BBL-Tandem

Recombinant Protein Expression. All tandem BBL variants were produced by recombinant means as full genes, cloned in the bacterial expression vector pBAT4. Plasmids containing the various genes were transformed into *E. coli* BL21(DE3) competent cells. Cells were grown in LB broth with 50 μg/mL ampicillin at 293 K to an OD_{600} 1.0-1.2, followed by induction with 1 mM isopropyl-β-D-thio-galactopyranoside (IPTG). The cells were kept in growing conditions overnight and then harvested and centrifuged at 8000 rpm for 30 min.

Protein Purification. The pellets were resuspended in 20 mM phosphate buffer at pH 7.5 with protease inhibitor cocktail and 2 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP). Cell lysis was performed by freeze-thaw method (6 cycles) and the lysate was ultracentrifuged at 35000 rpm for 40 min at 277 K. The obtained supernatant was pH adjusted to 7.5 and syringe filtered with 0.2 μm filter and loaded onto a HPLC HiTrap SP cation exchange column (GE Healthcare) and eluted with a gradient from 0 M to 1 M NaCl in 20 mM phosphate buffer with 2 mM TCEP at pH 7.5. The fractions containing the protein variant were pooled and subjected to a second round of purification using PROTO 300 C18 reverse phase column (Higgins Analytical) using a 5 % - 95 % acetonitrile gradient with 0.1 % trifluoroacetic acid (TFA) for elution. Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis, and confirmed by electrospray mass spectrometry followed by pure variants being pooled, lyophilized and stored at 277 K.

Protein Labeling. The purified proteins were sequentially labeled with extrinsic fluorophores via thiol-maleimide conjugation. The protein was dissolved in labeling buffer (20 mM phosphate buffer with 150 mM NaCl) and TCEP was added (protein:TCEP 1:1), pH of the solution was adjusted to 7.2 and allowed to equilibrate for 30 min. Alexa 594 C_5 maleimide dye (acceptor) was dissolved in the labelling buffer and added dropwise to the protein solution (protein:acceptor 1:1.1) and mixed. The reaction mixture was left overnight at 277 K and next morning 1 μL of β-mercaptoethanol (BME) was added to the reaction mixture to quench any unreacted dye. The sample was run through HPLC reverse phase column to remove excess dye and isolate singly labeled protein. Buffer exchange was done to remove acetonitrile via centricon and similar process was followed as for the first labeling reaction with Alexa 488 C_5 maleimide dye (donor) except (protein:donor 1:1.4).

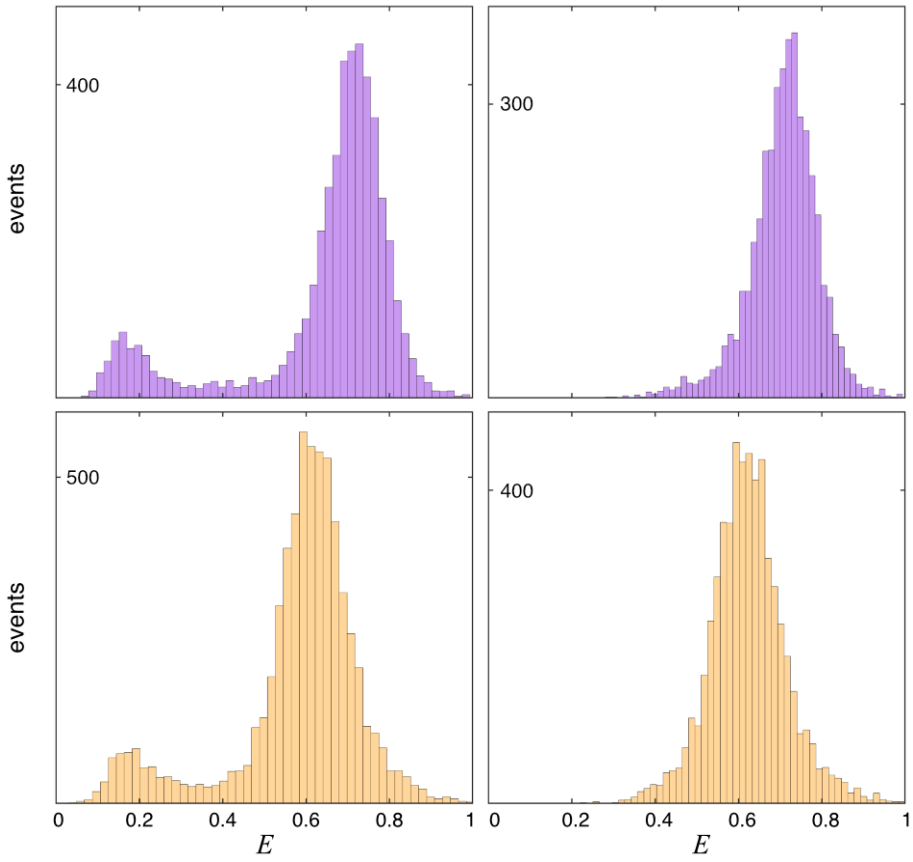

Figure B4. 1 Donor only bursts were removed using cumulative Poisson distribution function from the 0.5 ms binned data, pH 4 (top) and pH 7 (bottom).

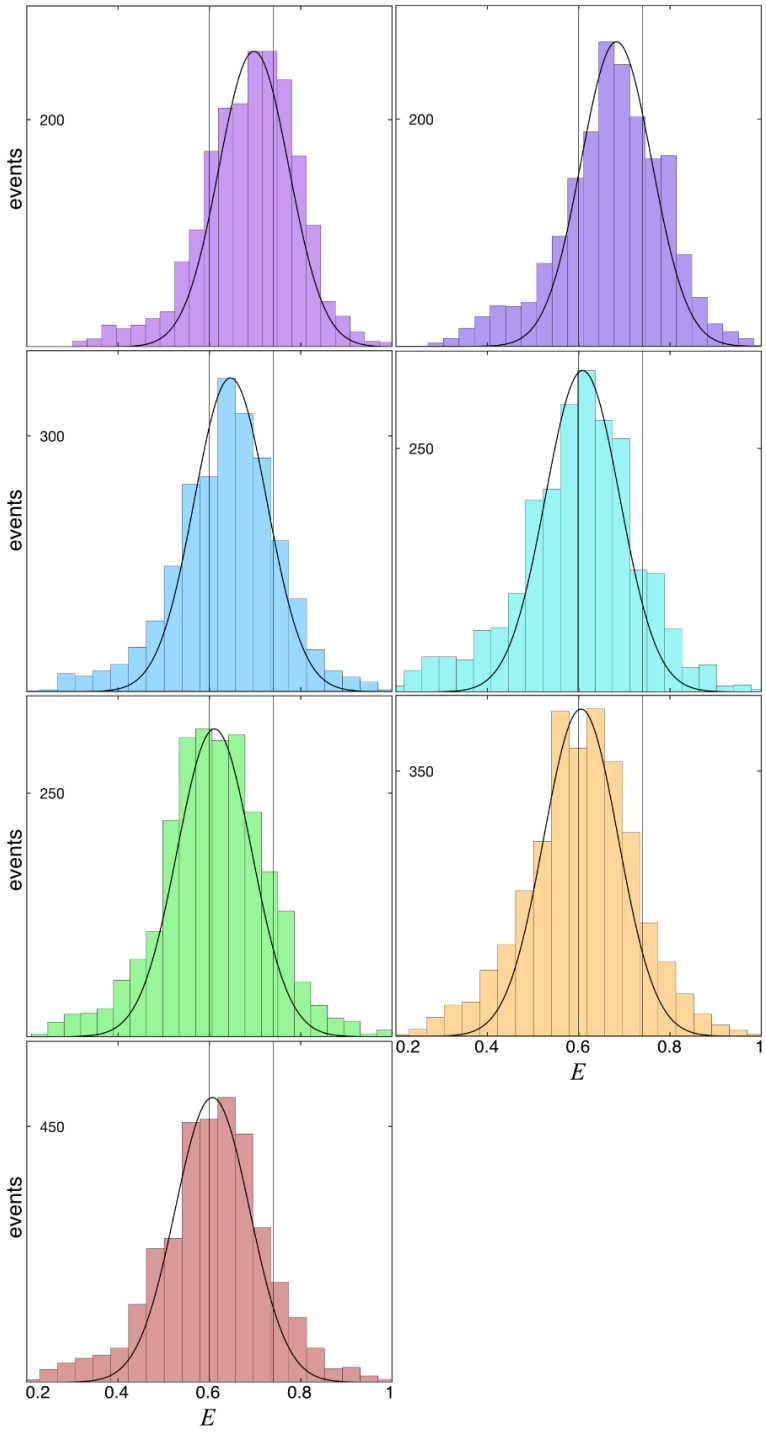

Figure B4. 2 150 micro-second binned data with a photon threshold of 36. The Gaussian curve over the histogram represents the shot noise.

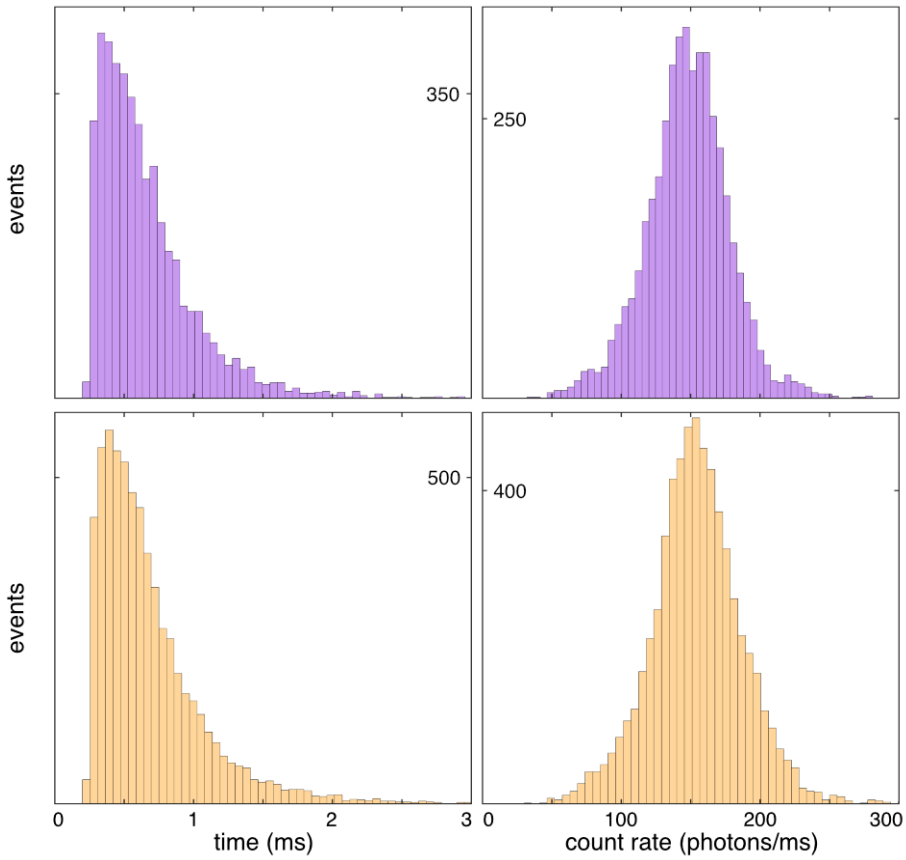

Figure B4. 3 Burst analysis data with a photon threshold of >40 for pH 4 (top) and pH 7 (bottom). Histogram of diffusion time of protein through the confocal volume (left) and histogram of count rate of photons (right).

Table B4. 1 Physical properties of BBL variants.

Protein	**No. of residues**	**Molecular Weight**	**pI**	**Charges (Asp+Glu) (Arg+Lys)**	**ε (M^{-1} cm^{-1}) at 280 nm**
BBL	40	4264.90	9.69	5 7	NA
B-hel-B	102	10913.63	9.56	13 18	125
B-P_4-B	94	10007.58	9.62	11 16	125
B-P_{10}-B	100	10590.28	9.62	11 16	125
B-B^+-B	143	15339.7	10.28	14 27	11125

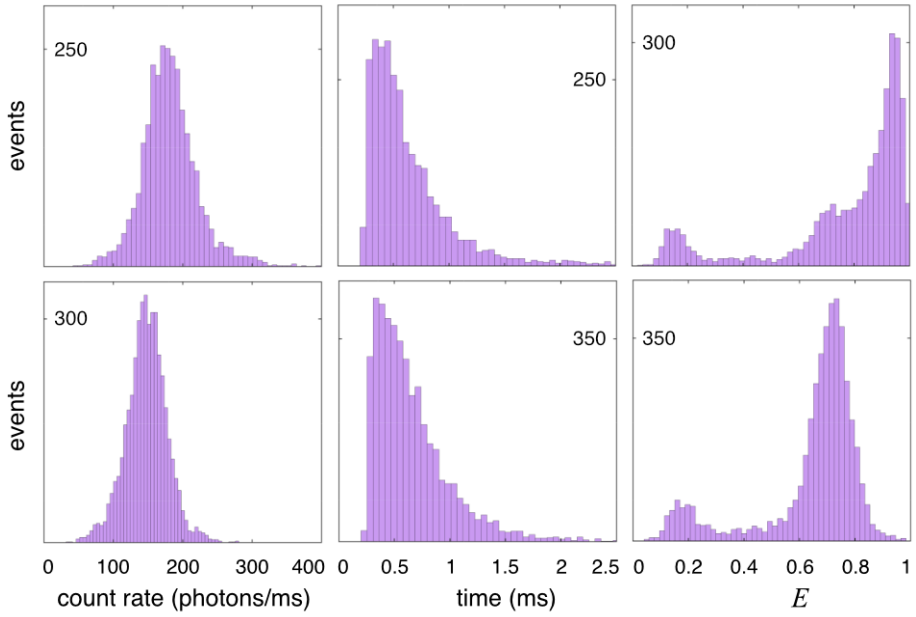

Figure B4. 4 Single molecule photon statistics of protein show heterogeneity of sample at pH 4.

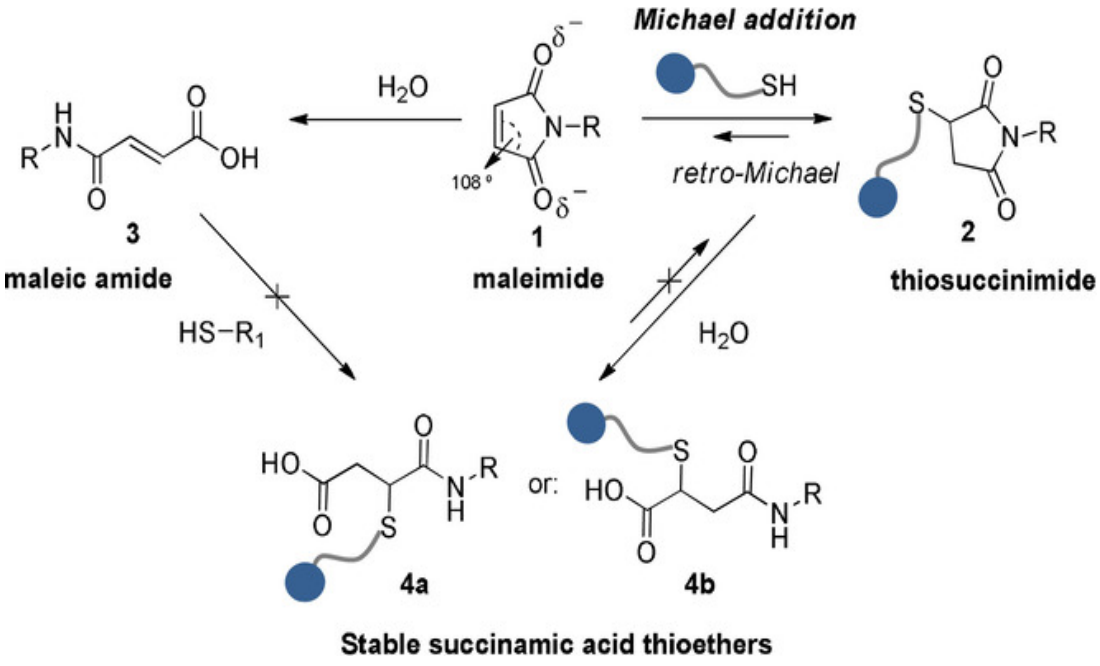

Figure B4. 5 General Michael addition and hydrolytic pathways of maleimides and thiosuccinimides (Ravasco, João M. J. M., Faustino et al. 2019). Figure reproduced with permission.

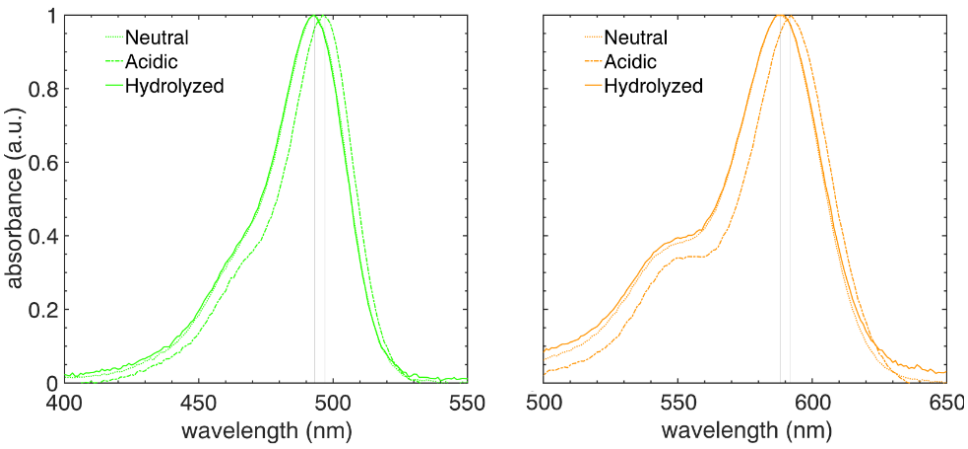

Figure B4. 6 Hydrolysis of maleimide in the dyes makes the sample less sticky and can be achieved by incubating overnight in neutral buffer. Lack of hydrolysis shows a red shift in acidic buffers and leads to heterogeneity of sample. Alexa 488 (Green) and Alexa 594 (Red).

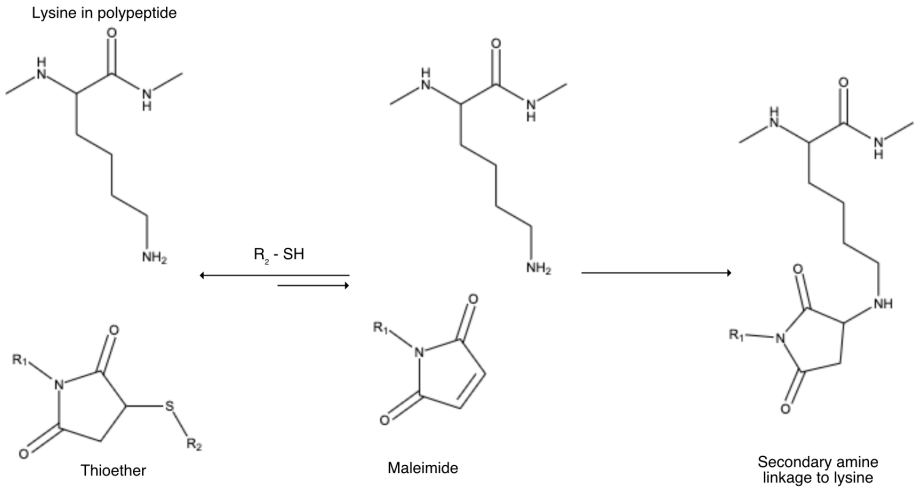

Figure B4. 7 Maleimide from the dye can link to lysine.

Tandem BBL based Ionic Strength Sensors

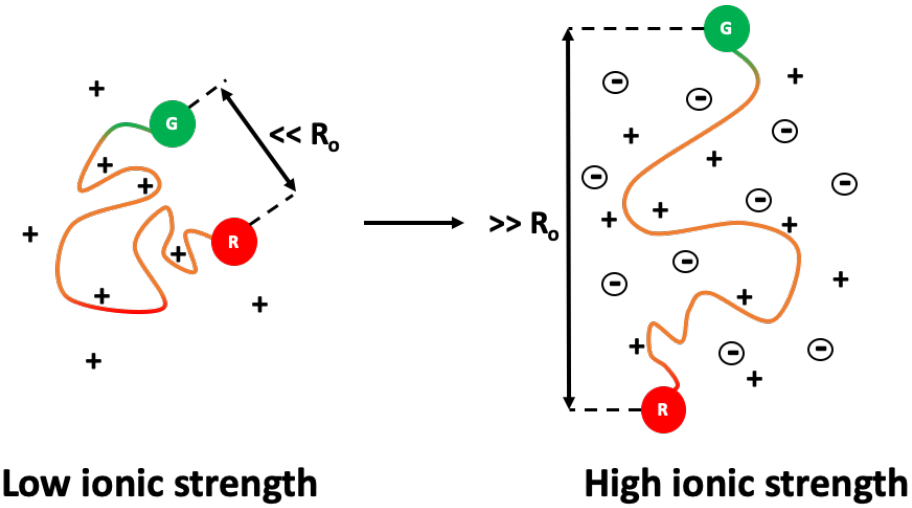

Figure B4. 8 Scheme for FRET occurring in BB+B engineered to be unfolded and refold as ionic strength is increased.

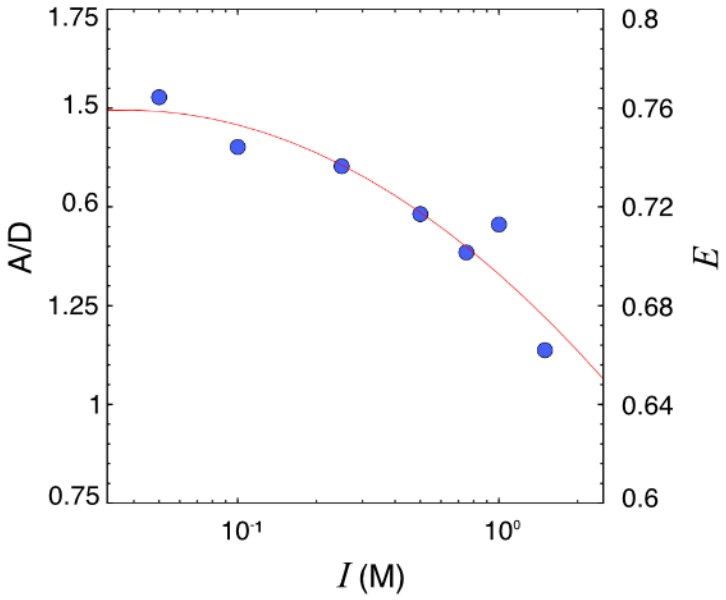

Figure B4. 9 Steady state bulk fluorescence intensity measurements on BB+B as a function of ionic strength.

Table B4. 2 Buffers used at different pH.

pH	Monoprotic	Diprotic
4.0	Acetate	Citrate
4.5	Acetate	Citrate
5.0	Acetate	Citrate
5.5	Acetate	Citrate
6.0	MES	Citrate
6.5	MES	Citrate
7.0	MOPS	Phosphate
7.5	MOPS	Phosphate
8.0	HEPES	Phosphate

Table B4. 3 Amino acid sequence information of BBL variants.

Protein	AA Sequence
BBL (pdb: 2CYU)	ALSPAIRRLL AEHNLDASAI KGTGVGGRLT REDVEKHLAK
B-hel-B	MGCKKNNDAL LSPAIRRLLA EHNLDASAIK GTGVGGRLTR EDVEKHLAKL AEALAKRLAE QLSPAIRRLL AEHNLDASAI KGTGVGGRLT REDVEKHLAK AC
B-P_4-B	MGCKKNNDAL LSPAIRRLLA EHNLDASAIK GTGVGGRLTR EDVEKHLAKP PPPLSPAIRR LLAEHNLDAS AIKGTGVGGR LTREDVEKHL AKAC
B-P_{10}-B	MGCKKNNDAL LSPAIRRLLA EHNLDASAIK GTGVGGRLTR EDVEKHLAKP PPPPPPPPPL SPAIRRLLAE HNLDASAIKG TGVGGRLTRE DVEKHLAKAC
B-B^+-B	MDGKKNNDAL SPAIRRLLAE *W*NLDASAIKG TGVGGRLTRE DVEKHLAKAP AKKCLSPAIR RLLA*Q*HNLNA SAIKGTGVGG RLTR*QN*V*Q*KH LAKDCKKNND ALSPAIRRLL AE*W*NLDASAI KGTGVGGRLT REDVEKHLAK APA

Table B4. 4 Deoxyribonucleic acid sequence information of BBL variants.

Protein	DNA Sequence
B-hel-B	ATGGGCTGCAAGAAAAACAATGATGCACTGAGCCCGGCCATT CGCCGTCTGCTCGCGGAACACAATTTAGACGCAAGTGCGATC AAAGGTACCGGTGTTGGCGGTCGCCTGACGCGTGAGGATGTT GAAAAACATTTAGCGAAAGCACTGGCGGAAGCACTGGCCAAA CGTCTGGCGGAACAGGCGATCCGTCGCCTGCTCGCTGAACAT AACCTGGATGCGAGCGCCATTAAGGGCACGGGCGTTGGTGG CCGTCTTACCCGTGAAGATGTGGAAAAACACTTAGCCAAGGC ATGTTAA
B-P_4-B	ATGGGCTGCAAAAAAAACAACGATGCGCTGAGCCCGGCGATT CGTCGTCTGCTGGCGGAACATAACCTGGATGCGAGCGCGATT AAAGGCACCGGCGTGGGCGGCCGTCTGACCCGTGAAGATGT GGAAAAACATCTGGCGAAAGCGGGCCCGCCGCCGCCGAAAA AAAACAACGATGCGCTGAGCCCGGCGATTCGTCGTCTGCTGG CGGAACATAACCTGGATGCGAGCGCGATTAAAGGCACCGGC GTGGGCGGCCGTCTGACCCGTGAAGATGTGGAAAAACATCTG GCGAAAGCGTGCTAA
B-P_{10}-B	ATGGGCTGCAAGAAAAACAATGATGCACTGAGCCCGGCCATT CGCCGTCTGCTCGCGGAACACAATTTAGACGCAAGTGCGATC AAAGGTACCGGTGTTGGCGGTCGCCTGACGCGTGAGGATGTT GAAAAACATTTAGCGAAAGCAGGTCCGCCTCCGCCGCCGCCA CCGCCACCGCCGAAAAAGAATAACGATGCCTTGTCTCCGGCG ATCCGTCGCCTGCTCGCTGAACATAACCTGGATGCGAGCGCC ATTAAGGGCACGGGCGTTGGTGGCCGTCTTACCCGTGAAGAT GTGGAAAAACACTTAGCCAAGGCATGTTAA
B-B^+-B	ATGGATGGCAAAAAGAACAATGATGCCCTGAGCCCCGCCATC CGTCGCCTGCTGGCCGAGTGGAACCTGGATGCCAGCGCCAT CAAAGGCACCGGCGTGGGTGGTCGCCTGACCCGCGAAGATG TGGAGAAACATCTGGCCAAAGCTCCGGCAAAGAAGTGCCTGA GTCCGGCGATCCGCCGCTTGTTAGCTCAGCACAACCTCAATG CATCTGCTATCAAGGGCACTGGCGTCGGGGGTCGTCTCACCC GTCAGAACGTGCAGAAACACCTGGCAAAAGATTGTAAGAAAAA CAACGACGCGTTATCTCCGGCTATTCGCCGTCTGTTGGCGGA ATGGAATCTCGACGCTTCCGCAATTAAGGGTACGGGTGTTGG CGGCCGTCTTACGCGTGAGGACGTCGAAAAGCACTTGGCTAA GGCACCGGCCTAA

Chapter 5: PLUG and PLAY

splitGFP-2H Protein Expression. splitGFP-2H was produced by recombinant means as full gene, cloned in the bacterial expression vector pBAT4. Plasmids containing the genes were transformed into *E. coli* BL21(DE3) competent cells. Cells were grown in LB broth with 50 μg/mL ampicillin at 303 K to an OD_{600} 1.2, followed by induction with 1 mM isopropyl-β-D-thio-galactopyranoside (IPTG). The cells were kept in growing conditions overnight and then harvested and centrifuged at 8000 rpm for 30 min.

splitGFP-2H Protein Purification. The pellets were resuspended in 20 mM phosphate buffer at pH 7.3 with protease inhibitor cocktail. Cell lysis was performed by homogenizer (5 cycles of 15 s) and the lysate was ultracentrifuged at 35000 rpm for 40 min at 277 K. The obtained supernatant was subjected to ammonium sulfate precipitation, cutting between 30-60 % at 277 K with equilibration time of 1 hr and constant stirring. The precipitate was resuspended in 20 mM phosphate at pH 7.3 and syringe filtered with 0.2 μm filter and loaded onto a HPLC HiTrap Q anionic exchange column (GE Healthcare) and eluted with a gradient from 0 M to 1 M NaCl in 20 mM phosphate buffer at pH 7.3. The fractions containing the protein variant were pooled and protein purity was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis and stored at 277 K.

Gamillus-mCherry2 Protein Expression. Gamillu-mChery2 was produced by recombinant means as full gene, cloned in the bacterial expression vector pDream2.1. Plasmids containing the genes were transformed into *E. coli* BL21(DE3) competent cells. Cells were grown in LB broth with 50 μg/mL ampicillin at 310 K to an OD_{600} 0.8, followed by induction with 1 mM isopropyl-β-D-thio-galactopyranoside (IPTG). The cells were kept in growing conditions for 6 hr and then harvested and centrifuged at 8000 rpm for 30 min.

Gamillus-mCherry2 Protein Purification. The pellets were resuspended in 20 mM phosphate buffer with 150 mM NaCl, 10 mM imidazole and protease inhibitor cocktail at pH 7.3. Cell lysis was performed by homogenizer (5 cycles of 15 s) and the lysate was ultracentrifuged at 35000 rpm for 40 min at 277 K. The obtained supernatant was syringe filtered with 0.2 μm filter and loaded onto a HPLC HisTrap FastFlow Crude immobilized ion affinity column (GE Healthcare) and washed with 20 mM phosphate buffer with 150 mM NaCl and 20 mM imidazole. It was eluted with a gradient from 20 mM to 500 mM imidazole in 20 mM phosphate buffer and 150 mM NaCl at pH 7.3. The fractions containing the protein variant were pooled and protein purity was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis and stored at 277 K.

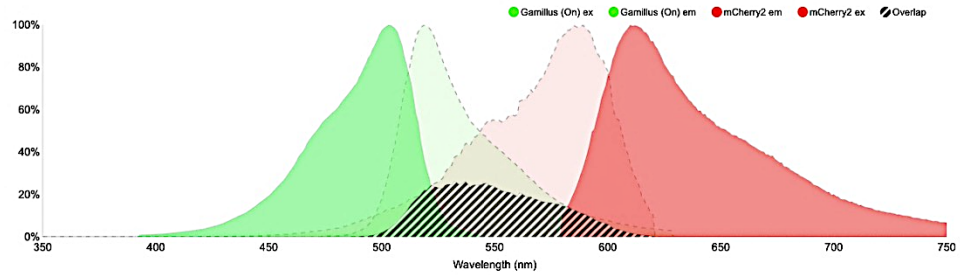

Figure B5. 1 Excitation and emission spectra of Gamillus and mCherry2.

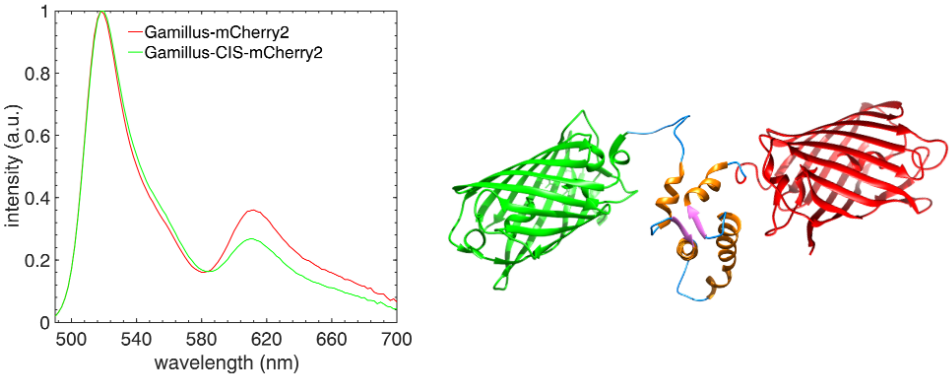

Figure B5. 2 A variant of calnuc introduced between Gamillus and mCherry2.

Table B5. 1 Amino acid sequence information of fluorescent protein variants.

Protein	AA Sequence
splitGFP-2H	MASKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFICT TGKLPVPWPT LVTTLCYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNYNHNV LIMADKQGGS GSGGSAEEDR YVVALVKKLE EVDEAHQMLQ EKLEDIRAGK ESEITEKFEK KIRELRELVR DHEREIEEVM EKIAGGSGKN GIKVNFKTRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHMV LLEFVTAAGI THGMDELYN
Gamillus-mCherry2	MDYKDDDDKG SGSSHHHHHH SSGLVPRGSV SKGEEASGRA LFQYPMTSKI ELNGEINGKK FKVAGEGFTP SSGRFNMHAY CTTGDLPMSW VVIASPLQYG FHMFAHYPED ITHFFQECFP GSYTLDRTLR MEGDGTLTTH HEYSLEDGCV TSKTTLNASG FDPKGATMTK SFVKQLPNEV KITPHGPNGI RLTSTVLYLK EDGTIQIGTQ DCIVTPVGGR KVTQPKAHFL HTQIIQKKDP NDTRDHIVQT ELAVAGNLWH GMDELYKEIW NSVPSSSLVS KGEEDNMAII KEFMRFKVHM EGSVNGHEFE IEGEGEGRPY EGTQTAKLKV TKGGPLPFAW DILSPQFMYG SKAYVKHPAD IPDYLKLSFP EGFNWERVMN FEDGGVVTVT QDSSLQDGEF IYKVKLRGTN FPSDGPVMQC RTMGWEASTE RMYPEDGALK GEIKQRLKLK DGGHYDAEVK TTYKAKKPVQ LPGAYNVDIK LDILSHNEDY TIVEQYERAE GRHSTGGMDE LYK

Table B5. 2 Deoxyribonucleic acid sequence information of fluorescent protein variants.

Protein	DNA Sequence
splitGFP-CC2H	ATGGCGAGCAAAGGCGAAGAACTGTTTACCGGCGTG GTGCCGATTCTGGTGGAACTGGATGGCGATGTGAAC GGCCATAAATTTAGCGTGAGCGGCGAAGGTGAAGGC GATGCGACCTATGGCAAACTGACCCTGAAATTTATTTG CACCACCGGCAAACTGCCGGTGCCGTGGCCGACCCT GGTGACCACCCTGTGCTATGGCGTGCAGTGCTTTAGC CGCTATCCGGATCATATGAAACGCCATGATTTTTTTAA AAGCGCGATGCCGGAAGGCTATGTGCAGGAACGCAC CATTTTTTTTAAAGATGATGGCAACTATAAAACCCGTG CGGAAGTGAAATTTGAAGGCGATACCCTGGTGAACCG CATTGAACTGAAAGGCATTGATTTTAAAGAAGATGGCA ACATTCTGGGCCATAAACTGGAATATAACTATAACCAT AACGTGCTGATTATGGCGGATAAACAGGGTGGCAGC GGCTCTGGCGGTAGCGCGGAAGAAGATCGCTATGTG GTGGCGCTGGTGAAAAAACTGGAGGAAGTGGATGAA GCGCATCAGATGCTGCAGGAAAAACTGGAAGATATTC GTGCGGGCAAAGAAAGCGAAATTACCGAAAAATTTGA AAAAAAAATTCGCGAACTGCGCGAACTGGTGCGCGAT CATGAACGCGAAATTGAAGAAGTGATGGAAAAAATTG CGGGCGGCAGCGGCAAAAACGGCATTAAAGTGAACT TTAAAACCCGCCATAACATTGAAGATGGCAGCGTGCA GCTGGCGGATCATTATCAGCAGAACACCCCGATTGGC GATGGCCCGGTGCTGCTGCCGGATAACCATTATCTGA GCACCCAGAGCGCGCTGAGCAAAGATCCGAACGAAA AACGCGATCATATGGTGCTGCTGGAATTTGTGACCGC GGCGGGCATTACCCACGGCATGGATGAACTGTATAAC TAA
Gamillus-mCherry2	ATGGATTACAAGGATGACGACGATAAGGGATCCGGCA GCAGCCATCATCATCATCATCACAGCAGCGGCCTGGT GCCGCGCGGCAGCGTGAGCAAAGGCGAAGAAGCCA GCGGCCGCGCCCTGTTTCAGTATCCCATGACCAGCAA AATCGAACTGAATGGCGAAATCAATGGCAAAAAATTCA AAGTGGCCGGCGAAGGCTTCACCCCGAGCAGCGGCC GCTTCAACATGCATGCCTACTGCACCACCGGCGATCT GCCAATGAGCTGGGTGGTGATTGCCAGCCCACTGCA

	GTATGGCTTTCACATGTTTGCCCACTATCCAGAAGATA TTACCCACTTCTTTCAGGAATGCTTTCCAGGCAGCTAC ACCCTGGATCGCACCCTGCGCATGGAAGGCGATGGC ACCCTGACCACCCATCACGAATACAGCCTGGAAGATG GCTGCGTGACCAGCAAAACCACCCTGAATGCCAGCG GCTTCGATCCCAAAGGCGCCACCATGACCAAAAGTTT TGTGAAACAGCTGCCCAACGAAGTGAAAATCACCCCA CATGGCCCGAATGGCATTCGCCTGACCAGCACCGTG CTGTATCTGAAAGAAGATGGCACCATCCAGATTGGCA CCCAGGATTGCATCGTGACCCCGGTGGGCGGCCGCA AAGTGACCCAGCCCAAAGCCCACTTCCTGCACACCCA GATCATTCAGAAAAAAGATCCAAACGATACCCGCGAT CACATCGTGCAGACCGAACTGGCCGTGGCCGGCAAC CTGTGGCATGGCATGGATGAACTGTACAAAGAGATCT GGAATTCGGTACCGAGCTCAAGCTTGGTGAGCAAAG GCGAAGAAGATAACATGGCCATCATCAAAGAATTTAT GCGCTTCAAAGTGCACATGGAAGGCAGCGTGAATGG CCATGAATTTGAAATTGAAGGCGAAGGCGAAGGCCGC CCCTATGAAGGCACCCAGACCGCCAAACTGAAAGTGA CCAAAGGCGGCCCCCTGCCGTTTGCCTGGGATATCC TGAGCCCCCAGTTCATGTATGGCAGCAAAGCCTACGT GAAACATCCAGCCGATATTCCGGATTATCTGAAACTG AGCTTTCCCGAAGGCTTCAACTGGGAACGCGTGATGA ACTTCGAAGATGGCGGCGTGGTGACCGTGACCCAGG ATAGCAGCCTGCAGGATGGCGAATTCATCTACAAAGT GAAACTGCGCGGTACCAACTTTCCCAGCGATGGCCCA GTGATGCAGTGCCGCACCATGGGCTGGGAAGCCAGC ACCGAACGCATGTATCCGGAAGATGGCGCCCTGAAA GGCGAAATCAAACAGCGCCTGAAACTGAAAGATGGC GGCCACTACGATGCCGAAGTGAAAACCACCTACAAAG CCAAAAAACCCGTGCAGCTGCCAGGCGCCTACAATGT GGATATCAAACTGGATATTCTGAGCCACAATGAAGATT ACACCATTGTGGAACAGTATGAACGCGCCGAAGGCC GCCATAGCACCGGCGGCATGGATGAGCTGTATAAATA A

Appendix C: Instrumentation

SINGLE-MOLECULE TECHNIQUE

When fluorescent dyes are in the excited singlet state (S_1), apart from emitting photons they can also transition to a nonfluorescent dark state by conversion to a triplet state (T_1) via intersystem crossing, formation of radical states or other nonradiative processes. Although these transitions are rare, their probability increases with higher irradiance, and this is referred to as photo-blinking. Photo-blinking renders dyes unavailable for subsequent excitations since T_1 decays are slow (i.e., microseconds to hours). Dyes in T_1 also encounter reactive oxygen species (ROS) like superoxide anion (O_2^-), hydrogen peroxide (H_2O_2) and hydroxyl radicals (OH^-). The singlet state of O_2, which is a byproduct of interactions between two triplets, is also highly reactive. These encounters lead to formation of radical ion state and irreversible loss of fluorescence, and this is referred to as photo-bleaching. It is therefore necessary to minimize triplet buildup to maximize photon emission for any single molecule technique (Grewer, Brauer 1994, Wang, Campos et al. 2016).

The mechanisms involved in T_1 quenching are energy transfer, charge transfer, reduction or oxidation. The former two, energy transfer and charge transfer return the dye to S_0, making them available to excitation. The latter two require an additional redox process for recovery. Reducing and oxidizing agents quench T_1 by transforming the dye into ionic radicals and these ionic radicals can cross-react with other ionic radicals to bleach the fluorophores. Therefore, they must be combined as a redox pair for successful recovery of the dye ground state S_0 (Vogelsang, Kasper et al. 2008). Although singlet O_2 can be highly reactive but without O_2 the dyes remain in the triplet for very long times, resulting in 40- to 50-fold reduction in photon count rate relative to experiments in presence of dissolved oxygen. Oxygen is one of the most efficient triplet quenchers available. So, for fast SMF applications it is better to keep dissolved oxygen as is. The combination of dissolved O_2, the triplet quencher trolox and the radical scavenger cysteamine is extremely efficient for rhodamine-based dyes like Alexa Fluor 488 and Alexa Fluor 594, yielding count rates near 1 MHz (Campos, Luis A., Liu et al. 2011, Chung, McHale et al. 2012).

CONFOCAL FLUORESCENCE MICROSCOPY

Proteins have relatively low stability (~10-30 kJ/mol) and even weak interactions with dyes can drastically alter its folding behavior. Therefore, minimizing possible protein-dye interactions is critical. The rule of thumb for protein modification with

dyes require that the dyes not be placed in the hydrophobic core of the protein. Solvent exposed, surface positions with potential steric clashes or interactions with neighboring charged residues can also perturb the folded state and should be avoided. The best approach is to attach the dyes at the N- and C-termini and potentially use a short flexible polypeptide linker (4-6 small polar amino acids like glycine, serine, alanine and asparagine). Alternatively, it can also be placed in long unstructured loops, provided there is enough flexibility, and it points outward. If the design requires insertion of the dye within an α-helices, it can be placed in *i, i+3* and *i*, *i+4* (or *i*, *i-3* and *i*, *i-4*) positions such that they are fully solvent-exposed. However, pitfalls like perturbing native structural interactions (like electrostatic or hydrophobic interactions) should be avoided.

The temporal resolution of single-molecule fluorescence resonance energy transfer (SM-FRET) pivots on how efficiently photons emitted from individual molecules are collected in sufficient numbers as to minimize statistical shot noise. The timescales of protein folding vary widely, with secondary structure like α-helix formation and loop closure happening in nanoseconds and complete folding taking sometimes hours and days (Muñoz, Cerminara 2016). However, the key timescale for protein folding appears to be in the microseconds range (Jones, Henry et al. 1993, Eaton 1999, Naganathan, Muñoz 2005, Sancho, Muñoz 2011). Conformational dynamics and binding coupled to folding processes also take place in microseconds (Prigozhin, Gruebele 2013, Shammas, Crabtree et al. 2016, Muñoz, Campos et al. 2016).

Fluorescence emission rates of dyes and the instrument collection efficiency dictate the temporal resolution of SMF techniques. To probe ultra-fast folding proteins, it is imperative to optimize instrumental parameters and to do that lets have a look at how various factors play out in the overall photon collection in a confocal setup. To explore that lets take Alexa Fluor 488 as an example.

The relationship between extinction coefficient and cross-section of absorption can be derived from the Beer-Lambert equation (Lakowicz 2006).

$$\sigma = \frac{1000 \cdot ln10 \cdot \varepsilon}{N_A} \qquad (B.1)$$

where, ε is molar extinction coefficient (72,000 M^{-1} cm^{-1}) and N_A is Avogadro number. From equation 5 we can calculate the light absorption cross-section σ of the dye as 2.75 $Å^2$.

$$p = \frac{\sigma}{\pi \cdot r^2} \qquad (B.2)$$

Assuming the dye occupies a spherical volume with radius r, which is the Stokes radius (~5.8 × 10^{-8} cm) (Heyman, Burt 2008). Using equation 6 gives probabilty of excitation p of the dye as ~0.026.

$$E = \frac{h \cdot c}{\lambda} \qquad (B.3)$$

Upon using Planck's equation, when the wavelength λ is 488 nm, the photons have energy E of 4.07 × 10^{-19} J.

$$N_{photons} = \frac{P}{E} \qquad (B.4)$$

Since we are using a continuous wave (CW) laser set to a power P of 100 × 10^{-6} W. Using equation 8 we get 2.457 × 10^{5} photons per nanosecond being delivered to the sample. Only photons that cross the tiny fraction of the area f_{area} occupied by the dye have a chance at absorption.

$$f_{area} = \frac{\pi \cdot r^2 \cdot N_A \cdot C}{1000} \qquad (B.5)$$

Assuming the protein concentration C used in free diffusion SM-FRET experiments is 50 pM, the fractional area thus is ~3.2 × 10^{-4} results in about 79 photons ($N_{photons}$ × f_{area}) or ~2.04 excitations (79 × 0.026). We know the lifetime of the dye is 4.1 ns and the quantum yield is 0.92, so it can be excited at ~324 MHz (79 × 4.1) and the actual photon emission rate from a single molecule will be about 298 MHz (324 × 0.92).

The NA of the objective determines the maximum angle at which the impinging light can be focused according to the following equation.

$$NA = \eta \cdot \sin\theta \qquad (B.6)$$

We used a 100 × oil immersion objective with a NA of 1.49 and 0.12 mm working distance, having an absolute numerical aperture of 0.987 (1.49/1.51) relative to a theoretical maximum of 1 (90° half-angle). The refractive index η of oil used was 1.51, giving the maximal half-angle of the objective θ as 80.66°.

The excited molecules emit photons spherically from the focal point, however the objective can only collect light from the cap that intersects with the collection light cone.

$$A_{sphere} = 4 \cdot \pi \cdot r^2 \qquad (B.7)$$

$$A_{cap} = 2 \cdot \pi \cdot r^2 \cdot (1 - \cos\theta) \qquad (B.8)$$

The ratio between cap of intersection A_{cap} (1.771×10^{-14}) and total emitted spherical surface A_{sphere} (4.227×10^{-14}) gives a collection efficiency of 0.42. A single molecule instrument has many other losses, transmittance of dichroics and filters, reflections within the multi-lens configurations in objective (~ 0.8 for Apo TIRF 100X 1.49 Oil) and the quantum efficiency of the photoavalanche diode detector (~ 0.7 for SPCM-AQRH-12-RH). When all these losses are computed, the collection efficiency of a SMF confocal microscope is abysmal (~ 0.02-0.04).

In a perfect diffraction-limited optical setup, the lateral resolution along the xy-axis (r_{xy} ~ 200 nm) is defined by the Rayleigh limit to be

$$r_{xy} = \frac{0.61 \cdot \lambda}{NA} \qquad (B.9)$$

$$r_z = \frac{2 \cdot \eta \cdot \lambda}{NA^2} \qquad (B.10)$$

and the axial resolution along the z-axis (r_z ~ 600 nm) is also dependent on the refractive of the mounting media. κ is a geometric factor usually equal to the optical

resolution in the z direction divided by the optical resolution in the xy plane but its lower limit (κ = 2.08) is defined to be

$$\kappa = \frac{2.33 \cdot \eta}{NA} \qquad (B.11)$$

$$V_c = \pi^{3/2} \kappa \omega^3 \qquad (B.13)$$

The mathematical formula for the confocal volume V_c can be used to find the volume to be <1 fL ($1\ fL = 1 \times 10^{-18}\ m^3$).

BIBLIOGRAPHY

A. HALILOVIĆ, E. MERDAN, Ž. KOVAČEVIĆ and L. G. POKVIĆ, 2019. Review of Biosensors for Environmental Field Monitoring, - *2019 8th Mediterranean Conference on Embedded Computing (MECO)* 2019, pp. 1-8.

ABAD, M.F.C., DI BENEDETTO, G., MAGALHÃES, P.J., FILIPPIN, L. and POZZAN, T., 2004. Mitochondrial pH Monitored by a New Engineered Green Fluorescent Protein Mutant*. *Journal of Biological Chemistry,* **279**(12), pp. 11521-11529.

ADZHUBEI, A.A., STERNBERG, M.J.E. and MAKAROV, A.A., 2013. Polyproline-II Helix in Proteins: Structure and Function. *Journal of Molecular Biology,* **425**(12), pp. 2100-2132.

AKKILIC, N., GESCHWINDNER, S. and HÖÖK, F., 2020. Single-molecule biosensors: Recent advances and applications. *Biosensors and Bioelectronics,* **151**, pp. 111944.

AKMAL, A. and MUÑOZ, V., 2004. The nature of the free energy barriers to two-state folding. *Proteins: Structure, Function, and Bioinformatics,* **57**(1), pp. 142-152.

ASHWORTH, J., HAVRANEK, J.J., DUARTE, C.M., SUSSMAN, D., MONNAT, R.J., STODDARD, B.L. and BAKER, D., 2006. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature,* **441**(7093), pp. 656-659.

AXELROD, D., 1981. Cell-substrate contacts illuminated by total internal reflection fluorescence. *Journal of Cell Biology,* **89**(1), pp. 141-145.

AZAD, T., TASHAKOR, A. and HOSSEINKHANI, S., 2014. Split-luciferase complementary assay: applications, recent developments, and future perspectives. *Analytical and Bioanalytical Chemistry,* **406**(23), pp. 5541-5560.

BAKER, D., 2000. A surprising simplicity to protein folding. *Nature,* **405**(6782), pp. 39-42.

BALAJI, A. and ZHANG, J., 2017. Electrochemical and optical biosensors for early-stage cancer diagnosis by using graphene and graphene oxide. *Cancer Nanotechnology,* **8**(1), pp. 10.

BALZAROTTI, F., EILERS, Y., GWOSCH, K.C., GYNNÅ, A.H., WESTPHAL, V., STEFANI, F.D., ELF, J. and HELL, S.W., 2017. Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science,* **355**(6325), pp. 606-612.

BJELIĆ, S., WIESER, M., FREY, D., STIRNIMANN, C.U., CHANCE, M.R., JAUSSI, R., STEINMETZ, M.O. and KAMMERER, R.A., 2013. Structural Basis for the Oligomerization-State Switch from a Dimer to a Trimer of an Engineered Cortexillin-1 Coiled-Coil Variant. *PLOS ONE,* **8**(5), pp. e63370.

BONNOT, C., NUSSAUME, L. and DESNOS, T., 2018. Identification of Chemical Inducers of the Phosphate-Starvation Signaling Pathway in A. thaliana Using Chemical Genetics. In: F. FAUSER and M. JONIKAS, eds, *Plant Chemical Genomics: Methods and Protocols.* New York, NY: Springer New York, pp. 65-84.

BOUKOBZA, E., SONNENFELD, A. and HARAN, G., 2001. Immobilization in Surface-Tethered Lipid Vesicles as a New Tool for Single Biomolecule Spectroscopy. *The Journal of Physical Chemistry B,* **105**(48), pp. 12165-12170.

BRUNO DA SILVA, F., CONTESSOTO, V.G., DE OLIVEIRA, V.M., CLARKE, J. and LEITE, V.B.P., 2018. Non-Native Cooperative Interactions Modulate Protein Folding Rates. *The Journal of Physical Chemistry B,* **122**(48), pp. 10817-10824.

BRYNGELSON, J.D., ONUCHIC, J.N., SOCCI, N.D. and WOLYNES, P.G., 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics,* **21**(3), pp. 167-195.

CALZINI, M.A., MALICO, A.A., MITCHLER, M.M. and WILLIAMS, G.J., 2021. Protein engineering for natural product biosynthesis and synthetic biology applications. *Protein Engineering, Design and Selection,* **34**, pp. gzab015.

CAMPOS, L.A., LIU, J., WANG, X., RAMANATHAN, R., ENGLISH, D.S. and MUÑOZ, V., 2011. A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nature Methods,* **8**(2), pp. 143-146.

CAMPOS, L.A., SADQI, M. and MUÑOZ, V., 2020. Lessons about Protein Folding and Binding from Archetypal Folds. *Accounts of Chemical Research,* **53**(10), pp. 2180-2188.

CARO, J.A., HARPOLE, K.W., KASINATH, V., LIM, J., GRANJA, J., VALENTINE, K.G., SHARP, K.A. and WAND, A.J., 2017. Entropy in molecular recognition by proteins. *Proc Natl Acad Sci USA,* **114**(25), pp. 6563.

CASEY, J.R., GRINSTEIN, S. and ORLOWSKI, J., 2010. Sensors and regulators of intracellular pH. *Nature Reviews Molecular Cell Biology,* **11**(1), pp. 50-61.

CASTILLO, J., GÁSPÁR, S., LETH, S., NICULESCU, M., MORTARI, A., BONTIDEAN, I., SOUKHAREV, V., DORNEANU, S.A., RYABOV, A.D. and CSÖREGI, E., 2004. Biosensors for life quality: Design, development and applications. *Sensors and Actuators B: Chemical,* **102**(2), pp. 179-194.

CERMINARA, M., CAMPOS, L.A., RAMANATHAN, R. and MUÑOZ, V., 2013. Slow Proton Transfer Coupled to Unfolding Explains the Puzzling Results of Single-Molecule Experiments on BBL, a Paradigmatic Downhill Folding Protein. *PLOS ONE,* **8**(10), pp. e78044.

CERMINARA, M., DESAI, T.M., SADQI, M. and MUÑOZ, V., 2012. Downhill Protein Folding Modules as Scaffolds for Broad-Range Ultrafast Biosensors. *Journal of the American Chemical Society,* **134**(19), pp. 8010-8013.

CHARBGOO, F., SOLTANI, F., TAGHDISI, S.M., ABNOUS, K. and RAMEZANI, M., 2016. Nanoparticles application in high sensitive aptasensor design. *TrAC Trends in Analytical Chemistry,* **85**, pp. 85-97.

CHEN, C. and WANG, J., 2020. Optical biosensors: an exhaustive and comprehensive review. *Analyst,* **145**(5), pp. 1605-1628.

CHEN, P., ZHOU, X., ANDOY, N.M., HAN, K., CHOUDHARY, E., ZOU, N., CHEN, G. and SHEN, H., 2014. Spatiotemporal catalytic dynamics within single nanocatalysts revealed by single-molecule microscopy. *Chemical Society Reviews,* **43**(4), pp. 1107-1117.

CHUNG, H.S. and EATON, W.A., 2013. Single-molecule fluorescence probes dynamics of barrier crossing. *Nature,* **502**(7473), pp. 685-688.

CHUNG, H.S., LOUIS, J.M. and EATON, W.A., 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proceedings of the National Academy of Sciences,* **106**(29), pp. 11837-11844.

CHUNG, H.S., MCHALE, K., LOUIS, J.M. and EATON, W.A., 2012. Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science,* **335**(6071), pp. 981-984.

CRICK, F.H.C., 1952. Is alpha-keratin a coiled coil? *Nature,* **170**(4334), pp. 882-883.

D'OELSNITZ, S. and ELLINGTON, A., 2018. Continuous directed evolution for strain and protein engineering. *Current opinion in biotechnology; Chemical Biotechnology • Pharmaceutical Biotechnology,* **53**, pp. 158-163.

DAY, R.N., TAO, W. and DUNN, K.W., 2016. A simple approach for measuring FRET in fluorescent biosensors using two-photon microscopy. *Nature Protocols,* **11**(11), pp. 2066-2080.

DIMISE, E.J., WIDBOOM, P.F. and BRUNER, S.D., 2008. Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile *Thermobifida fusca*. *Proc Natl Acad Sci USA,* **105**(40), pp. 15311.

DOBSON, C.M., 2003. Protein folding and misfolding. *Nature,* **426**(6968), pp. 884-890.

DONTEN, M.L., HASSAN, S., POPP, A., HALTER, J., HAUSER, K. and HAMM, P., 2015. pH-Jump Induced Leucine Zipper Folding beyond the Diffusion Limit. *The Journal of Physical Chemistry B,* **119**(4), pp. 1425-1432.

DOOSE, S., NEUWEILER, H., BARSCH, H. and SAUER, M., 2007. Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proceedings of the National Academy of Sciences,* **104**(44), pp. 17400-17405.

DOOSE, S., NEUWEILER, H. and SAUER, M., 2005. A Close Look at Fluorescence Quenching of Organic Dyes by Tryptophan. *ChemPhysChem,* **6**(11), pp. 2277-2285.

EATON, W.A., 1999. Searching for "downhill scenarios" in protein folding. *Proceedings of the National Academy of Sciences,* **96**(11), pp. 5897-5899.

EDINGER, A.L. and THOMPSON, C.B., 2004. Death by design: apoptosis, necrosis and autophagy. *Current opinion in cell biology,* **16**(6), pp. 663-669.

EDWARDS, K.A., 2021. Periplasmic-binding protein-based biosensors and bioanalytical assay platforms: Advances, considerations, and strategies for optimal utility. *Talanta Open,* **3**, pp. 100038.

EGGER, M.D. and PETRAN, M., 1967. New Reflected-Light Microscope for Viewing Unstained Brain and Ganglion Cells. *Science,* **157**(3786), pp. 305-307.

EL GHAZOUANI, A., BASLÉ, A., FIRBANK, S.J., KNAPP, C.W., GRAY, J., GRAHAM, D.W. and DENNISON, C., 2011. Copper-Binding Properties and Structures of Methanobactins from Methylosinus trichosporium OB3b. *Inorganic chemistry,* **50**(4), pp. 1378-1391.

EMERSON HOLTZER, M., LARRY BRETTHORST, G., ANDRÉ D'AVIGNON, D., HOGUE ANGELETTI, R., MINTS, L. and HOLTZER, A., 2001. Temperature Dependence of the Folding and Unfolding Kinetics of the GCN4 Leucine Zipper via 13Cα-NMR. *Biophysical journal,* **80**(2), pp. 939-951.

ERNSTER, L. and SCHATZ, G., 1981. Mitochondria: a historical review. *Journal of Cell Biology,* **91**(3), pp. 227s-255s.

EVANS, E.G.B. and MILLHAUSER, G.L., 2015. Chapter Nineteen - Genetic Incorporation of the Unnatural Amino Acid p-Acetyl Phenylalanine into Proteins for Site-Directed Spin Labeling. *Methods in enzymology,* **563**, pp. 503-527.

FACCIO, G., 2019. Chapter 7 - Proteins as Nanosized Components of Biosensors. In: O.V. ZENKINA, ed, *Nanomaterials Design for Sensing Applications.* Elsevier, pp. 229-255.

FAN, Y., ANDERSON, T.J. and ZHANG, B., 2018. Single-molecule electrochemistry: From redox cycling to single redox events. *Current Opinion in Electrochemistry,* **7**, pp. 81-86.

FREIDUS, L.G., PRADEEP, P., KUMAR, P., CHOONARA, Y.E. and PILLAY, V., 2018. Alternative fluorophores designed for advanced molecular imaging. *Drug discovery today,* **23**(1), pp. 115-133.

GALBÁN, S., JEON, Y.H., BOWMAN, B.M., STEVENSON, J., SEBOLT, K.A., SHARKEY, L.M., LAFFERTY, M., HOFF, B.A., BUTLER, B.L., WIGDAL, S.S., BINKOWSKI, B.F., OTTO, P., ZIMMERMAN, K., VIDUGIRIS, G., ENCELL, L.P., FAN, F., WOOD, K.V., GALBÁN, C.J., ROSS, B.D. and REHEMTULLA, A., 2013. Imaging Proteolytic Activity in Live Cells and Animal Models. *PLOS ONE,* **8**(6), pp. e66248.

GANGULY, D., OTIENO, S., WADDELL, B., ICONARU, L., KRIWACKI, R.W. and CHEN, J., 2012. Electrostatically Accelerated Coupled Binding and Folding of Intrinsically Disordered Proteins. *Journal of Molecular Biology,* **422**(5), pp. 674-684.

GARCIA-MIRA, M.M., SADQI, M., FISCHER, N., SANCHEZ-RUIZ, J.M. and MUÑOZ, V., 2002. Experimental Identification of Downhill Protein Folding. *Science,* **298**(5601), pp. 2191-2195.

GERSTEIN, M. and KREBS, W., 1998. A database of macromolecular motions. *Nucleic acids research,* **26**(18), pp. 4280-4290.

GHOSH, I., HAMILTON, A.D. and REGAN, L., 2000. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *Journal of the American Chemical Society,* **122**(23), pp. 5658-5659.

GOPICH, I.V. and SZABO, A., 2009. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *The Journal of Physical Chemistry B,* **113**(31), pp. 10965-10973.

GOPICH, I.V. and SZABO, A., 2007. Single-Molecule FRET with Diffusion and Conformational Dynamics. *The Journal of Physical Chemistry B,* **111**(44), pp. 12925-12932.

GREENFIELD, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols,* **1**(6), pp. 2876-2890.

GREWER, C. and BRAUER, H., 1994. Mechanism of the Triplet-State Quenching by Molecular Oxygen in Solution. *The Journal of physical chemistry,* **98**(16), pp. 4230-4235.

GRINSTEIN, S., FURUYA, W. and BIGGAR, W.D., 1986. Cytoplasmic pH regulation in normal and abnormal neutrophils. Role of superoxide generation and Na+/H+ exchange. *Journal of Biological Chemistry,* **261**(2), pp. 512-514.

GRUEBELE, M., 2008. Chapter 6:Fast Protein Folding. In: V. MUÑOZ, ed, *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches.* Cambridge: Royal Society of Chemistry, pp. 106-138.

GU, L., LI, Y., ZHANG, S., XUE, Y., LI, W., LI, D., XU, T. and JI, W., 2019. Molecular resolution imaging by repetitive optical selective exposure. *Nature Methods,* **16**(11), pp. 1114-1118.

GU, Q., NANNEY, W., CAO, H.H., WANG, H. and YE, T., 2018. Single Molecule Profiling of Molecular Recognition at a Model Electrochemical Biosensor. *Journal of the American Chemical Society,* **140**(43), pp. 14134-14143.

GUNNARSSON, A., SNIJDER, A., HICKS, J., GUNNARSSON, J., HÖÖK, F. and GESCHWINDNER, S., 2015. Drug Discovery at the Single Molecule Level: Inhibition-in-Solution Assay of Membrane-Reconstituted β-Secretase Using Single-Molecule Imaging. *Analytical Chemistry,* **87**(8), pp. 4100-4103.

HAJNÓCZKY, G., CSORDÁS, G., DAS, S., GARCIA-PEREZ, C., SAOTOME, M., SINHA ROY, S. and YI, M., 2006. Mitochondrial calcium signalling and cell death: Approaches for assessing the role of mitochondrial Ca2+ uptake in apoptosis. *Cell calcium,* **40**(5), pp. 553-560.

HARA-CHIKUMA, M., WANG, Y., GUGGINO, S.E., GUGGINO, W.B. and VERKMAN, A.S., 2005. Impaired acidification in early endosomes of ClC-5 deficient proximal tubule. *Biochemical and biophysical research communications,* **329**(3), pp. 941-946.

HARRIS, J.L. and CRAIK, C.S., 1998. Engineering enzyme specificity. *Current opinion in chemical biology,* **2**(1), pp. 127-132.

HEYMAN, N.S. and BURT, J.M., 2008. Hindered Diffusion through an Aqueous Pore Describes Invariant Dye Selectivity of Cx43 Junctions. *Biophysical journal,* **94**(3), pp. 840-854.

HIGGINS, C.D., MALASHKEVICH, V.N., ALMO, S.C. and LAI, J.R., 2014. Influence of a heptad repeat stutter on the pH-dependent conformational behavior of the central coiled-coil from influenza hemagglutinin HA2. *Proteins: Structure, Function, and Bioinformatics,* **82**(9), pp. 2220-2228.

HOHNG, S., LEE, S., LEE, J. and JO, M.H., 2014. Maximizing information content of single-molecule FRET experiments: multi-color FRET and FRET combined with force or torque. *Chemical Society Reviews,* **43**(4), pp. 1007-1013.

HOLZMEISTER, P., ACUNA, G.P., GROHMANN, D. and TINNEFELD, P., 2014. Breaking the concentration limit of optical single-molecule detection. *Chemical Society Reviews,* **43**(4), pp. 1014-1028.

HOM, R.A., VORA, M., REGNER, M., SUBACH, O.M., CHO, W., VERKHUSHA, V.V., STAHELIN, R.V. and KUTATELADZE, T.G., 2007. pH-dependent Binding of the Epsin ENTH Domain and the AP180 ANTH Domain to PI(4,5)P2-containing Bilayers. *Journal of Molecular Biology,* **373**(2), pp. 412-423.

HUANG, W., WU, X., GAO, X., YU, Y., LEI, H., ZHU, Z., SHI, Y., CHEN, Y., QIN, M., WANG, W. and CAO, Y., 2019. Maleimide–thiol adducts stabilized through stretching. *Nature Chemistry,* **11**(4), pp. 310-319.

INUI, T., KOHNO, H., KAWASAKI, Y., MATSUURA, K., UEDA, H., TAMURA, Y., WATANABE, M., INAGE, Y., YAKITA, Y., WAKABAYASHI, Y. and MATSUMIYA, G., 2020. Use of a Smart Watch for Early Detection of Paroxysmal Atrial Fibrillation: Validation Study. *JMIR Cardio,* **4**(1), pp. e14857.

ISHII, N., 2017. GroEL and the GroEL-GroES Complex. In: J.R. HARRIS and J. MARLES-WRIGHT, eds, *Macromolecular Protein Complexes: Structure and Function.* Cham: Springer International Publishing, pp. 483-504.

JANSSEN, K.P.F., DE CREMER, G., NEELY, R.K., KUBAREV, A.V., VAN LOON, J., MARTENS, J.A., DE VOS, D.E., ROEFFAERS, M.B.J. and

HOFKENS, J., 2014. Single molecule methods for the study of catalysis: from enzymes to heterogeneous catalysts. *Chemical Society Reviews,* **43**(4), pp. 990-1006.

JOHANSSON, A., APPELQVIST, H., NILSSON, C., KÅGEDAL, K., ROBERG, K. and ÖLLINGER, K., 2010. Regulation of apoptosis-associated lysosomal membrane permeabilization. *Apoptosis,* **15**(5), pp. 527-540.

JONES, C.M., HENRY, E.R., HU, Y., CHAN, C.K., LUCK, S.D., BHUYAN, A., RODER, H., HOFRICHTER, J. and EATON, W.A., 1993. Fast events in protein folding initiated by nanosecond laser photolysis. *Proceedings of the National Academy of Sciences,* **90**(24), pp. 11860-11864.

KEMBRO, J.M., CORTASSA, S. and AON, M.A., 2014. Complex oscillatory redox dynamics with signaling potential at the edge between normal and pathological mitochondrial function. *Frontiers in Physiology,* **5**, pp. 257.

KENT, K.P., OLTROGGE, L.M. and BOXER, S.G., 2009. Synthetic Control of Green Fluorescent Protein. *Journal of the American Chemical Society,* **131**(44), pp. 15988-15989.

KOCH, M., PANDI, A., BORKOWSKI, O., BATISTA, A.C. and FAULON, J., 2019. Custom-made transcriptional biosensors for metabolic engineering. *Current opinion in biotechnology; Tissue, Cell and Pathway Engineering,* **59**, pp. 78-84.

KOHN, J.E. and PLAXCO, K.W., 2005. Engineering a signal transduction mechanism for protein-based biosensors. *Proceedings of the National Academy of Sciences of the United States of America,* **102**(31), pp. 10841-10845.

KONG, L., GAN, Y., LIANG, T., ZHONG, L., PAN, Y., KIRSANOV, D., LEGIN, A., WAN, H. and WANG, P., 2020. A novel smartphone-based CD-spectrometer for high sensitive and cost-effective colorimetric detection of ascorbic acid. *Analytica Chimica Acta,* **1093**, pp. 150-159.

LAKOWICZ, J.R., 2006. *Principles of Fluorescence Spectroscopy* |. 3 edn. Singapore: Springer.

LAN, L., YAO, Y., PING, J. and YING, Y., 2019. Ultrathin transition-metal dichalcogenide nanosheet-based colorimetric sensor for sensitive and label-free detection of DNA. *Sensors and Actuators B: Chemical,* **290**, pp. 565-572.

LARSON, J.D., RODGERS, M.L. and HOSKINS, A.A., 2014. Visualizing cellular machines with colocalization single molecule microscopy. *Chemical Society Reviews,* **43**(4), pp. 1189-1200.

LI, P., OLIVA, F.Y., NAGANATHAN, A.N. and MUÑOZ, V., 2009. Dynamics of one-state downhill protein folding. *Proceedings of the National Academy of Sciences,* **106**(1), pp. 103-108.

LI, Q., NANCE, M.R., KULIKAUSKAS, R., NYBERG, K., FEHON, R., KARPLUS, P.A., BRETSCHER, A. and TESMER, J.J.G., 2007. Self-masking in an intact ERM-merlin protein: an active role for the central alpha-helical domain. *Journal of Molecular Biology,* **365**(5), pp. 1446-1459.

LIAO, Z., ZHANG, Y., LI, Y., MIAO, Y., GAO, S., LIN, F., DENG, Y. and GENG, L., 2019. Microfluidic chip coupled with optical biosensors for simultaneous detection of multiple analytes: A review. *Biosensors and Bioelectronics,* **126**, pp. 697-706.

LINDORFF-LARSEN, K., PIANA, S., DROR, R.O. and SHAW, D.E., 2011. How Fast-Folding Proteins Fold. *Science,* **334**(6055), pp. 517-520.

LIU, J., CAMPOS, L.A., CERMINARA, M., WANG, X., RAMANATHAN, R., ENGLISH, D.S. and MUÑOZ, V., 2012. Exploring one-state downhill protein folding in single molecules. *Proceedings of the National Academy of Sciences,* **109**(1), pp. 179-184.

LIU, X., YAO, Y., YING, Y. and PING, J., 2019. Recent advances in nanomaterial-enabled screen-printed electrochemical sensors for heavy metal detection. *TrAC Trends in Analytical Chemistry,* **115**, pp. 187-202.

LUZAROWSKI, M. and SKIRYCZ, A., 2019. Emerging strategies for the identification of protein–metabolite interactions. *Journal of experimental botany,* **70**(18), pp. 4605-4618.

MALASHKEVICH, V.N., HIGGINS, C.D., ALMO, S.C. and LAI, J.R., 2015. A switch from parallel to antiparallel strand orientation in a coiled-coil X-ray structure via two core hydrophobic mutations. *Peptide Science,* **104**(3), pp. 178-185.

MANDEL, R. and HOLZWARTH, G., 1973. Ultraviolet circular dichroism of polyproline and oriented collagen. *Biopolymers,* **12**(3), pp. 655-674.

MARVIN, J.S. and HELLINGA, H.W., 2001. Conversion of a maltose receptor into a zinc biosensor by computational design. *Proceedings of the National Academy of Sciences,* **98**(9), pp. 4955-4960.

MILO, R., 2013. What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays,* **35**(12), pp. 1050-1055.

MINSKY, M., 1988. Memoir on inventing the confocal scanning microscope. *Scanning,* **10**(4), pp. 128-138.

MORADI, M., BABIN, V., ROLAND, C., DARDEN, T.A. and SAGUI, C., 2009. Conformations and free energy landscapes of polyproline peptides. *Proceedings of the National Academy of Sciences,* **106**(49), pp. 20746-20751.

MOSKVIN, O.V., KAPLAN, S., GILLES-GONZALEZ, M. and GOMELSKY, M., 2007. Novel Heme-based Oxygen Sensor with a Revealing Evolutionary History *. *Journal of Biological Chemistry,* **282**(39), pp. 28740-28748.

MOTT, H.R. and OWEN, D., 2018. Allostery and dynamics in small G proteins. *Biochemical Society transactions,* **46**(5), pp. 1333-1343.

MUNEER, S., AYOKO, G.A., ISLAM, N. and IZAKE, E.L., 2020. Utilizing the thiol chemistry of biomolecules for the rapid determination of anti-TNF-α drug in blood. *Talanta,* **208**, pp. 120411.

MUÑOZ, V. and SERRANO, L., 1994. Elucidating the folding problem of helical peptides using empirical parameters. *Nature Structural Biology,* **1**(6), pp. 399-409.

MUÑOZ, V., 2007. Conformational Dynamics and Ensembles in Protein Folding. *Annual Review of Biophysics and Biomolecular Structure,* **36**(1), pp. 395-412.

MUÑOZ, V., CAMPOS, L.A. and SADQI, M., 2016. Limited cooperativity in protein folding. *Current Opinion in Structural Biology,* **36**(Supplement C), pp. 58-66.

MUÑOZ, V. and CERMINARA, M., 2016. When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochemical Journal,* **473**(17), pp. 2545-2559.

MUÑOZ, V. and SANCHEZ-RUIZ, J.M., 2004. Exploring protein-folding ensembles: A variable-barrier model for the analysis of equilibrium unfolding experiments. *Proceedings of the National Academy of Sciences of the United States of America,* **101**(51), pp. 17646-17651.

MUÑOZ, V., THOMPSON, P.A., HOFRICHTER, J. and EATON, W.A., 1997. Folding dynamics and mechanism of β-hairpin formation. *Nature,* **390**(6656), pp. 196-199.

MUNSHI, S., SUBRAMANIAN, S., RAMESH, S., GOLLA, H., KALIVARATHAN, D., KULKARNI, M., CAMPOS, L.A., SEKHAR, A. and NAGANATHAN, A.N.,

2019. Engineering Order and Cooperativity in a Disordered Protein. *Biochemistry,* **58**(19), pp. 2389-2397.

NAGANATHAN, A.N. and MUÑOZ, V., 2010. Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proceedings of the National Academy of Sciences,* **107**(19), pp. 8611-8616.

NAGANATHAN, A.N. and MUÑOZ, V., 2005. Scaling of Folding Times with Protein Size. *Journal of the American Chemical Society,* **127**(2), pp. 480-481.

NAGPAL, S., LUONG, T.D.N., SADQI, M. and MUÑOZ, V., 2020. Downhill (Un)Folding Coupled to Binding as a Mechanism for Engineering Broadband Protein Conformational Transducers. *ACS Synthetic Biology,* **9**(9), pp. 2427-2439.

ONG, J.J., POLLARD, T.D., GOYANES, A., GAISFORD, S., ELBADAWI, M. and BASIT, A.W., 2021. Optical biosensors - Illuminating the path to personalized drug dosing. *Biosensors and Bioelectronics,* **188**, pp. 113331.

PAL, P., LESOINE, J.F., LIEB, M.A., NOVOTNY, L. and KNAUF, P.A., 2005. A Novel Immobilization Method for Single Protein spFRET Studies. *Biophysical journal,* **89**(2), pp. L11-L13.

PASSAM, F.J. and CHIU, J., 2019. Allosteric disulphide bonds as reversible mechano-sensitive switches that control protein functions in the vasculature. *Biophysical Reviews,* **11**(3), pp. 419-430.

PIWON, N., GÜNTHER, W., SCHWAKE, M., BÖSL, M.R. and JENTSCH, T.J., 2000. ClC-5 Cl--channel disruption impairs endocytosis in a mouse model for Dent's disease. *Nature,* **408**(6810), pp. 369-373.

PRATAP, J.V., LUISI, B.F. and CALLADINE, C.R., 2013. Geometric principles in the assembly of α-helical bundles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* **371**(1993), pp. 20120369.

PRESTA, L.G. and ROSE, G.D., 1988. Helix signals in proteins. *Science,* **240**(4859), pp. 1632-1641.

PRIGOZHIN, M.B. and GRUEBELE, M., 2013. Microsecond folding experiments and simulations: a match is made. *Physical Chemistry Chemical Physics,* **15**(10), pp. 3372-3388.

QUIJANO-RUBIO, A., YEH, H., PARK, J., LEE, H., LANGAN, R.A., BOYKEN, S.E., LAJOIE, M.J., CAO, L., CHOW, C.M., MIRANDA, M.C., WI, J., HONG, H.J.,

STEWART, L., OH, B. and BAKER, D., 2021. De novo design of modular and tunable protein biosensors. *Nature,* **591**(7850), pp. 482-487.

RAMAN, S., VERNON, R., THOMPSON, J., TYKA, M., SADREYEV, R., PEI, J., KIM, D., KELLOGG, E., DIMAIO, F., LANGE, O., KINCH, L., SHEFFLER, W., KIM, B., DAS, R., GRISHIN, N.V. and BAKER, D., 2009. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function, and Bioinformatics,* **77**(9), pp. 89-99.

RAO, H., LIU, W., HE, K., ZHAO, S., LU, Z., ZHANG, S., SUN, M., ZOU, P., WANG, X., ZHAO, Q., WANG, Y. and LIU, T., 2020. Smartphone-Based Fluorescence Detection of Al3+ and H2O Based on the Use of Dual-Emission Biomass Carbon Dots. *ACS Sustainable Chemistry & Engineering,* **8**(23), pp. 8857-8867.

RE, F., SESANA, S., BARBIROLI, A., BONOMI, F., CAZZANIGA, E., LONATI, E., BULBARELLI, A. and MASSERINI, M., 2008. Prion protein structure is affected by pH-dependent interaction with membranes: A study in a model system. *FEBS Letters,* **582**(2), pp. 215-220.

REN, B., ROBERT, F., WYRICK, J.J., APARICIO, O., JENNINGS, E.G., SIMON, I., ZEITLINGER, J., SCHREIBER, J., HANNETT, N., KANIN, E., VOLKERT, T.L., WILSON, C.J., BELL, S.P. and YOUNG, R.A., 2000. Genome-Wide Location and Function of DNA Binding Proteins. *Science,* **290**(5500), pp. 2306-2309.

RICHARDS, F.M. and VITHAYATHIL, P.J., 1959. The Preparation of Subtilisin-modified Ribonuclease and the Separation of the Peptide and Protein Components. *Journal of Biological Chemistry,* **234**(6), pp. 1459-1465.

RIVINOJA, A., KOKKONEN, N., KELLOKUMPU, I. and KELLOKUMPU, S., 2006. Elevated Golgi pH in breast and colorectal cancer cells correlates with the expression of oncofetal carbohydrate T-antigen. *Journal of cellular physiology,* **208**(1), pp. 167-174.

RIZZINO, A., 2008. Transcription factors that behave as master regulators during mammalian embryogenesis function as molecular rheostats. *Biochemical Journal,* **411**(2), pp. e5-e7.

RIZZUTO, R., DE STEFANI, D., RAFFAELLO, A. and MAMMUCARI, C., 2012. Mitochondria as sensors and regulators of calcium signalling. *Nature Reviews Molecular Cell Biology,* **13**(9), pp. 566-578.

ROBIEN, M.A., CLORE, G.M., OMICHINSKI, J.G., PERHAM, R.N., APPELLA, E., SAKAGUCHI, K. and GRONENBORN, A.M., 1992. Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide

succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of Escherichia coli. *Biochemistry,* **31**(13), pp. 3463-3471.

RODA, A., ARDUINI, F., MIRASOLI, M., ZANGHERI, M., FABIANI, L., COLOZZA, N., MARCHEGIANI, E., SIMONI, P. and MOSCONE, D., 2020. A challenge in biosensors: Is it better to measure a photon or an electron for ultrasensitive detection? *Biosensors and Bioelectronics,* **155**, pp. 112093.

ROOS, A. and W, F.B., 1981. Intracellular pH. *Physiological Reviews,* **61**(2), pp. 296-434.

ROSSMANN, M., J. GREIVE, S., MOSCHETTI, T., DINAN, M. and HYVÖNEN, M., 2017. Development of a multipurpose scaffold for the display of peptide loops. *Protein Engineering, Design and Selection,* **30**(6), pp. 419-430.

ROY, R., HOHNG, S. and HA, T., 2008. A practical guide to single-molecule FRET. *Nature Methods,* **5**(6), pp. 507-516.

RUGGIERO, M.T., SIBIK, J., ORLANDO, R., ZEITLER, J.A. and KORTER, T.M., 2016. Measuring the Elasticity of Poly-l-Proline Helices with Terahertz Spectroscopy. *Angewandte Chemie International Edition,* **55**(24), pp. 6877-6881.

SADQI, M., FUSHMAN, D. and MUÑOZ, V., 2006. Atom-by-atom analysis of global downhill protein folding. *Nature,* **442**(7100), pp. 317-321.

SADQI, M., LAPIDUS, L.J. and MUÑOZ, V., 2003. How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences,* **100**(21), pp. 12117-12122.

SANCHO, D.D. and MUÑOZ, V., 2011. Integrated prediction of protein folding and unfolding rates from only size and structural class. *Physical Chemistry Chemical Physics,* **13**(38), pp. 17030-17043.

SCOGNAMIGLIO, V., ARDUINI, F., PALLESCHI, G. and REA, G., 2014. Biosensing technology for sustainable food safety. *TrAC Trends in Analytical Chemistry,* **62**, pp. 1-10.

SHAMMAS, S.L., CRABTREE, M.D., DAHAL, L., WICKY, B.I.M. and CLARKE, J., 2016. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies*. *Journal of Biological Chemistry,* **291**(13), pp. 6689-6695.

SHARMA, R., SANCHO, D.D. and MUÑOZ, V., 2017. Interplay between the folding mechanism and binding modes in folding coupled to binding processes†

Raj. *Phys. Chem. Chem. Phys.,* **19**, pp. 28512-28516.

SHASHKOVA, S. and LEAKE, M.C., 2017. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience reports,* **37**(4), pp. BSR20170031.

SHEN, Y., CHEN, Y., WU, J., SHANER, N.C. and CAMPBELL, R.E., 2017. Engineering of mCherry variants with long Stokes shift, red-shifted fluorescence, and low cytotoxicity. *PLOS ONE,* **12**(2), pp. e0171257.

SHI, L., HOLLIDAY, A.E., SHI, H., ZHU, F., EWING, M.A., RUSSELL, D.H. and CLEMMER, D.E., 2014. Characterizing Intermediates Along the Transition from Polyproline I to Polyproline II Using Ion Mobility Spectrometry-Mass Spectrometry. *Journal of the American Chemical Society,* **136**(36), pp. 12702-12711.

SHI, S., ANG, E.L. and ZHAO, H., 2018. In vivo biosensors: mechanisms, development, and applications. *Journal of Industrial Microbiology and Biotechnology,* **45**(7), pp. 491-516.

SHIBA, K. and SCHIMMEL, P., 1992. Functional assembly of a randomly cleaved protein. *Proceedings of the National Academy of Sciences,* **89**(5), pp. 1880-1884.

SHINODA, H., MA, Y., NAKASHIMA, R., SAKURAI, K., MATSUDA, T. and NAGAI, T., 2018. Acid-Tolerant Monomeric GFP from Olindias formosa. *Cell Chemical Biology,* **25**(3), pp. 330-338.e7.

SHOEMAKER, B.A., PORTMAN, J.J. and WOLYNES, P.G., 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences,* **97**(16), pp. 8868-8873.

SOSNICK, T.R., KRANTZ, B.A., DOTHAGER, R.S. and BAXA, M., 2006. Characterizing the Protein Folding Transition State Using ψ Analysis. *Chemical reviews,* **106**(5), pp. 1862-1876.

SOUTHWORTH, M.W., ADAM, E., PANNE, D., BYER, R., KAUTZ, R. and PERLER, F.B., 1998. Control of protein splicing by intein fragment reassembly. *The EMBO journal,* **17**(4), pp. 918-926.

SPITZBERG, J.D., ZREHEN, A., VAN KOOTEN, X.F. and MELLER, A., 2019. Plasmonic-Nanopore Biosensors for Superior Single-Molecule Detection. *Advanced Materials,* **31**(23), pp. 1900422.

STRYER, L. and HAUGLAND, R.P., 1967. Energy transfer: a spectroscopic ruler. *Proceedings of the National Academy of Sciences,* **58**(2), pp. 719-726.

SUKUMAR, U.K., NATARAJAN, A., MASSOUD, T.F. and PAULMURUGAN, R., 2020. Applications of Fluorescent Protein-Based Sensors in Bioimaging. In: Z. CHENG, ed, *Fluorescent Imaging in Medicinal Chemistry.* Cham: Springer International Publishing, pp. 149-183.

TALLEY, K. and ALEXOV, E., 2010. On the pH-optimum of activity and stability of proteins. *Proteins: Structure, Function, and Bioinformatics,* **78**(12), pp. 2699-2706.

TEKPINAR, M. and ZHENG, W., 2013. Coarse-grained and all-atom modeling of structural states and transitions in hemoglobin. *Proteins: Structure, Function, and Bioinformatics,* **81**(2), pp. 240-252.

TROMBETTA, E.S., EBERSOLD, M., GARRETT, W., PYPAERT, M. and MELLMAN, I., 2003. Activation of Lysosomal Function During Dendritic Cell Maturation. *Science,* **299**(5611), pp. 1400-1403.

TSIEN, R.Y., 1998. THE GREEN FLUORESCENT PROTEIN. *Annual Review of Biochemistry,* **67**(1), pp. 509-544.

TURK, B. and TURK, V., 2009. Lysosomes as “Suicide Bags” in Cell Death: Myth or Reality?*. *Journal of Biological Chemistry,* **284**(33), pp. 21783-21787.

TYAGI, S. and KRAMER, F.R., 1996. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nature Biotechnology,* **14**(3), pp. 303-308.

URANO, Y., ASANUMA, D., HAMA, Y., KOYAMA, Y., BARRETT, T., KAMIYA, M., NAGANO, T., WATANABE, T., HASEGAWA, A., CHOYKE, P.L. and KOBAYASHI, H., 2009. Selective molecular imaging of viable cancer cells with pH-activatable fluorescence probes. *Nature medicine,* **15**(1), pp. 104-109.

VAGENENDE, V., YAP, M.G.S. and TROUT, B.L., 2009. Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. *Biochemistry,* **48**(46), pp. 11084-11096.

VALLÉE-BÉLISLE, A. and PLAXCO, K.W., 2010. Structure-switching biosensors: inspired by Nature. *Current Opinion in Structural Biology,* **20**(4), pp. 518-526.

VALLÉE-BÉLISLE, A., RICCI, F. and PLAXCO, K.W., 2009. Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proceedings of the National Academy of Sciences,* **106**(33), pp. 13802-13807.

VAUGHAN-JONES, R.D., SPITZER, K.W. and SWIETACH, P., 2009. Intracellular pH regulation in heart. *Journal of Molecular and Cellular Cardiology,* **46**(3), pp. 318-331.

VOGELSANG, J., KASPER, R., STEINHAUER, C., PERSON, B., HEILEMANN, M., SAUER, M. and TINNEFELD, P., 2008. A Reducing and Oxidizing System Minimizes Photobleaching and Blinking of Fluorescent Dyes. *Angewandte Chemie International Edition,* **47**(29), pp. 5465-5469.

WAINIO-THEBERGE, S., WOLFF, A. and NORTHOFF, G., 2021. Dynamic relationships between spontaneous and evoked electrophysiological activity. *Communications Biology,* **4**(1), pp. 741.

WANG, D., JIN, Y., DING, X., WANG, W., ZHAI, S., BAI, L. and GUO, Z., 2017. Gene regulation and signal transduction in the ICE–CBF–COR signaling pathway during cold stress in plants. *Biochemistry (Moscow),* **82**(10), pp. 1103-1117.

WANG, Y., WANG, Y., DU, X., YAN, S., ZHANG, P., CHEN, H. and HUANG, S., 2019. Electrode-free nanopore sensing by DiffusiOptoPhysiology. *Science Advances,* **5**(9), pp. eaar3309.

WANG, Z., CAMPOS, L.A. and MUÑOZ, V., 2016. Chapter Fourteen - Single-Molecule Fluorescence Studies of Fast Protein Folding. *Methods in enzymology,* **581**, pp. 417-459.

WANG, Z., DOSHI, A., CHOWDHURY, R., WANG, Y., MARANAS, C.D. and CIRINO, P.C., 2020. Engineering sensitivity and specificity of AraC-based biosensors responsive to triacetic acid lactone and orsellinic acid. *Protein Engineering, Design and Selection,* **33**, pp. gzaa027.

YU, K., HOU, J., LI, K., YAO, Q., YANG, J., WU, M., XIE, Y. and YU, X., 2015. A single design strategy for dual sensitive pH probe with a suitable range to map pH in living cells. *Scientific Reports,* **5**(1), pp. 15540.

ZACHARCHENKO, T., VON CASTELMUR, E., RIGDEN, D.J. and MAYANS, O., 2015. Structural advances on titin: towards an atomic understanding of multi-domain functions in myofilament mechanics and scaffolding. *Biochemical Society transactions,* **43**(5), pp. 850-855.

ZHANG, W., ZHENG, Q., YAN, M., CHEN, X., YANG, H., ZHOU, W. and RAO, Z., 2018. Structural characterization of the HCoV-229E fusion core. *Biochemical and biophysical research communications,* **497**(2), pp. 705-712.

ZHANG, X., CUI, Z. and WANG, D., 2016. Sensing of biomolecular interactions using fluorescence complementing systems in living cells. *Biosensors and Bioelectronics; 30th Anniversary Issue,* **76**, pp. 243-250.

ZHU, Y. and CHEN, S., 2014. Many-body effect in ion binding to RNA. *The Journal of chemical physics,* **141**(5), pp. 055101.

ZRIMSEK, A.B., CHIANG, N., MATTEI, M., ZALESKI, S., MCANALLY, M.O., CHAPMAN, C.T., HENRY, A., SCHATZ, G.C. and VAN DUYNE, R.P., 2017. Single-Molecule Chemistry with Surface- and Tip-Enhanced Raman Spectroscopy. *Chemical reviews,* **117**(11), pp. 7583-7613.