

Protein Engineering Design Principles for the Development of Biosensors

by

Seymour de Picciotto

B.S., Swiss Federal Institute of Technology (2007)

M.S., Swiss Federal Institute of Technology (2009)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

© Massachusetts Institute of Technology 2015. All rights reserved.


Signature redacted

Author
.....
Department of Biological Engineering

May 7, 2015


Signature redacted

Certified by...
.....
K. Dane Wittrup

C. P. Dubbs Professor of Chemical and Biological Engineering

Signature redacted Thesis Supervisor

Certified by...
.....
Linda G. Griffith

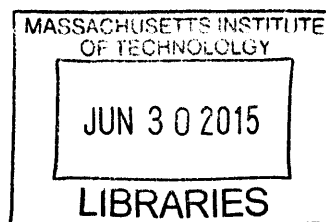
S. E. T. I. Professor of Biological and Mechanical Engineering

Signature redacted Thesis Supervisor

Accepted by
.....
Forest M. White

Chairman, Graduate Program Committee

ARCHIVES



Protein Engineering Design Principles for the Development of Biosensors

by

Seymour de Picciotto

Submitted to the Department of Biological Engineering
on May 7, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

Investigating protein location and concentration is critical to understanding function. Reagentless biosensors, in which a reporting fluorophore is conjugated to a binding scaffold, can detect analytes of interest with high temporal and spatial resolution. However, because these biosensors require laborious empirical screening to develop, their adoption has been limited. Hence, we establish design principles that will facilitate development.

In this thesis, we first develop a kinetic model for the dynamic performance of a reagentless biosensor. Using a sinusoidal signal for ligand concentration, our findings suggest that it is optimal to use a binding moiety whose equilibrium dissociation constant matches that of the average predicted input signal, while maximizing both the association rate constant and the dissociation rate constant at the necessary ratio to create the desired equilibrium constant. Although practical limitations constrain the attainment of these objectives, the derivation of these design principles provides guidance for improved reagentless biosensor performance and metrics for quality standards in the development of biosensors.

Following these guidelines, we use the human tenth type III fibronectin domain to engineer new binders against several ligands of the EGFR receptor. Using these binders and others, we design and characterize biosensors based on various target analytes, scaffolds and fluorophores. We observe that analytes can harbor specific binding pockets for the fluorophore, which sharply increase the fluorescence produced upon binding. Furthermore, we demonstrate that a fluorophore conjugated to locally rigid surfaces possesses lower background fluorescence. Based on these newly identified properties, we design biosensors that produce a 100-fold increase in fluorescence upon binding to analyte, about a 10-fold improvement over the previous best biosensor.

In order to improve the methodology of reagentless biosensor design, we establish a method for site-specific labeling of proteins displayed on the surface of yeasts. This procedure allows for the screening of libraries of sensors for binding and fluorescence enhancement simultaneously. Finally, we explore an alternative sensor design, based on competitive inhibition of fluorescence quenching.

Thesis Supervisor: K. Dane Wittrup

Title: C. P. Dubbs Professor of Chemical and Biological Engineering

Thesis Supervisor: Linda G. Griffith

Title: S. E. T. I. Professor of Biological and Mechanical Engineering

Acknowledgments

In his book *The Warrior of the Light*, Paulo Coelho shares his teachings in the form of a manual describing the challenges faced by the warrior and their solutions. The warrior is a metaphor for a person in the pursuit of their dream.

The warrior of light studies the two columns on either side of the door he is trying to open.

One is called Fear and the other is called Desire. “The Warrior looks at the column of Fear and on it is written: You are entering a dangerous, unfamiliar world where everything you have learned up until now will prove useless.”

The Warrior looks at the column of Desire and on it is written: “You are about to leave a familiar world wherein are stored all the things you ever wanted and for which you struggled long and hard.”

The warrior smiles because nothing frightens him and nothing holds him. With the confidence of one who knows what he wants, he opens the door.

This thesis began with an ambitious idea: develop sensors for EGFR ligands and use these tools to study that extracellular pathway. As it is often the case in science, the road turned out to be a lot more challenging than we had anticipated. Rather than a straight line between point A and B, this Ph.D. experience was extremely rich because of all the turns, detours, side roads, and stops as the destination took on new forms. Just like the warrior of light, this thesis started with me opening a door and I am extremely grateful to everyone who has believed in me and supported me in this adventure.

First and foremost, I am beholden to my thesis advisors Linda Griffith and Dane Wittrup. Linda’s incredible enthusiasm and her brilliant mind was everything I was hoping to find at MIT. She has believed in my potential since the first day, she has helped me navigate through challenges and has given me unconditional support even as the project deviated greatly from her research interests. Dane’s creativity, intellectual power, enthusiasm and guidance has shaped me professionally and personally. He has been supportive of all decisions I have taken, many of them thanks to his scientific advice and personal recommendations.

I am also thankful to Barbara Imperiali, who has been almost a third advisor. She hosted me in her lab, challenged my critical thinking, and provided invaluable insights into these projects, all of which has helped improve this research and has made me a better scientist. I am grateful to my other thesis committee members, Ernest Fraenkel, Gaudenz Dauser and Douglas Lauffenburger for their support and scientific advice. Also, without my collaborators Michael Traxlmayr, Elke Socher, Stephanie Cheung and Silvano Sanchini, none of this work would have been possible. I had the privilege to mentor extremely talented undergraduate students, Paige Dickson, Sixing Zhao, Daniel Sosa, whose contributions to this thesis were essential.

The Griffith Lab and Wittrup Lab have been amazing places to call home, where I found friendship, assistance and advice. In the early days, Joel Wagner and Edgar

Sanchez, were an amazing source of laughter and scientific discussions. As I transitioned to the Wittrup lab's office, Jordi Mata-Fink was a sounding board for ideas and experiments. Nicole Yang, Tiffany Chen, Alessandro Angelini, Michael Traxlmayr and James van Deventer were always willing to listen and offer advice. Xiaosai Yao, Cary Opel and Eric Zhu were a great source of non-scientific intellectual stimulation. John Rhoden and Byron Kwan gave the lab a fun flavor.

While research took place on campus, the motivation to open new doors was greatly stimulated by outside friends. Most of them are clueless about my doings, yet they have always believed in me. This experience would not have been anywhere as enjoyable without the many friendships I have maintained and built. Alain Baruh, Romain Kohler and Nicolas de Weck have been amazing friends and always made me feel like we had never been separated by the Atlantic Ocean. Alex Chaim made the Boston life incredibly rich and fun. Many other friends, room-mates, and classmates made this experience unique and cultural. Of course, Joyce has been by my side through the past three years adding new hues to colors.

My parents and brother offered great advice, whether solicited or not, and have been extremely encouraging. My grand mother, Anneanne'm, whose amazing cuisine and support made me always look forward to the next summer. Finally, to my grand mother, Nora de Picciotto, who would have loved to see this day come true, I dedicate this thesis.

Ever tried.

Ever failed.

No matter.

Try Again.

Fail again.

Fail Better.

- Samuel Beckett

This doctoral thesis has been examined by a Committee as follows:

Professor Douglas Lauffenburger
Chairman, Thesis Committee
Ford Professor of Biological Engineering, Chemical Engineering, and
Biology

Signature redacted

Professor Barbara Imperiali
Member, Thesis Committee
Class of 1922 Professor of Biology and Professor of Chemistry

Professor Gaudenz Danuser
Member, Thesis Committee
Patrick E. Haggerty Distinguished Chair in Basic Biomedical Science

Contents

1	Introduction	23
1.1	Protein engineering by yeast surface display	24
1.2	Scaffolds for binder engineering	25
1.3	Solvatochromic fluorophores	26
1.4	State of the art	28
1.5	Thesis Summary	29
2	Equilibrium and Dynamic Design Principles for Binding Molecules Engineered for Reagentless Biosensors	37
2.1	Abstract	38
2.2	Introduction	38
2.3	Model formulation	40
2.4	Results	42
2.4.1	Dynamic consideration reveals the crucial importance of kinetic rates optimization.	42
2.4.2	Varying input characteristics	43
2.4.3	Single sensor analysis	46
2.5	Discussion	47
3	Design principles for reagentless biosensors: specific fluorophore/analyte binding and minimization of fluorophore/scaffold interactions.	59
3.1	Abstract	60
3.2	Introduction	60

3.3	Results	62
3.3.1	Successfully designed reagentless biosensors are rare	62
3.3.2	A tailored binding pocket for NBD is responsible for the strong fluorescence activation	63
3.3.3	Rigid scaffolds lower background fluorescence	67
3.3.4	Combination of rigid scaffold with a specific fluorophore	68
3.4	Discussion	69
3.5	Future Perspectives	74
3.6	Material & Methods	75
3.6.1	Fibronectin engineering by yeast surface display	75
3.6.2	Protein expression and purification	75
3.6.3	Protein Labeling	77
3.6.4	Protein biotinylation	77
3.6.5	Size exclusion chromatography	78
3.6.6	Fluorescence Measurements	78
3.6.7	Biolayer Interferometry	78
3.6.8	YSD for enhanced MBP mutant isolation	78
3.6.9	Sequence Data analysis	79
3.6.10	Flow cytometry	79
3.6.11	EGF epitope mapping	80
4	Sortase A Catalyzed Site Specific Labeling of Yeast Surface Dis- played Proteins	101
4.1	Abstract	102
4.2	Introduction	102
4.3	Results	104
4.4	Discussion	106
4.5	Methods	107
4.5.1	Plasmid Construction	107
4.5.2	Sortase Expression and Purification	108

4.5.3	Yeast Transformation	108
4.5.4	Sortase Mediated Labeling	109
4.5.5	Flow cytometry	109
5	Competitive Inhibition of Fluorescence Quenching Based Sensors	115
5.1	Background	116
5.2	Results	117
5.3	Discussion	122
5.4	Conclusion & Future perspectives	123
5.5	Material & Methods	124
5.5.1	4-4-20 de-maturation	124
5.5.2	Protein Expression & Purification	125
5.5.3	Protein Labeling	126
5.5.4	Model formulation	126
A	Archived Plasmids	139
B	Sequences	143
C	MatLab code	151

List of Figures

1-1	Schematic of yeast surface display	31
1-2	Scaffolds for protein engineering	32
1-3	Mechanism of solvatochromism	33
1-4	Structures of common solvatochromic fluorophores	34
2-1	Mean concentration and time scales for biological processes	52
2-2	Relationship between mean signal intensity and amplitude	53
2-3	Sensor signal with time-variant ligand concentration	54
2-4	Bode Diagram for three different sensors	55
2-5	Bode Diagram for individual and combinatorial sensor deployment . .	56
2-6	Damkohler #	57
3-1	Reagentless biosensor workflow and overview.	81
3-2	Mapping the interaction between NBD and MBP in the Off7(N45C)*NBD and Off7(T46C)*NBD sensors	82
3-3	Off7(T46C)*NBD Sensor is regulated by MBP conformation.	83
3-4	The balance between scaffold rigidity and specific binding pockets . .	84
3-5	Characterization of ErbB ligand binders	85
3-6	Epitope mapping of Fn3 EGF 5.1.07 on hEGF	86
3-7	Chemical structures of the solvatochromic fluorophores used in this study	87
3-8	Supplemental: Epitope mapping of Off7(N45C)*NBD and Off7(T46C)*NBD on MBP	88
3-9	Characterization of MBP cavity mutants	89
3-10	MBP maturation for sensor by YSD	90

3-11	Epitope mapping of Off7(M114C)*NBD on MBP	91
3-12	Balancing self and antigen interactions	92
4-1	Overview of sortagging on yeast	110
4-2	Sortagging labeling optimization	111
5-1	Competitive inhibition of quenching	129
5-2	Proof of concept for competitive inhibition of quenching	130
5-3	Competitive inhibition of quenching modeling	131
5-4	4-4-20 affinity de-maturation	132
5-5	Characterization of the anti-fluorescein quenchers	133
5-6	Specificity of anti-fluorescein antibodies	134
5-7	Titration of MBP to 5 nM of Off7(D112C)*5BrF	135

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

2.1	Relationship between kinetic rate constants and input signal characteristics	51
3.1	Fluorescence measurements: excitation wavelength and integrated signal range	93
3.2	MiSeq past filter reads	94
3.3	Sequences of EGFR ligand Fn3 binders	95
3.4	List of selected binders, labeling sites and coupled fluorophores	96
4.1	Oligonucleotide sequences for Sortagging on yeast	112
5.1	Anti-fluorescein antibody characterization	136
A.1	Archived plasmids	140

Abbreviations

4-DMN-1	4-N,N-dimethylamino-1,8-naphthalimide
5BrF	5-bromomethylfluorescein
5IAF	5-iodoacetamidofluorescein
6IAF	6-iodoacetamidofluorescein
A_L	Normalized Amplitude
AR	Amphiregulin
BTC	Betacellulin
DARPin	Designed Ankyrin Repeat Protein
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EPR	Epiregulin
Fn3	Human tenth domain III fibronectin
HB-EGF	Heparin-Binding Epidermal Growth Factor
IANBD	N-((2-(iodoacetoxy)ethyl)-N-methyl)amino-7-nitrobenz-2-oxa-1,3-diazole
IgG2a	Immunoglobulin G subclass 2a
K_D	Equilibrium binding constant
k_{obs}	Observed binding rate
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
L₀	Initial ligand concentration
M5F	Maleimide-5-fluorescein
MACS	Magnetic-Activated Cell Sorting
MBP	Maltose Binidng Protein
MG	Malachite Green
MSA	Mouse Serum Albumin
NBD	nitrobenz-2-oxa-1,3-diazole
Sac7d	A hpyerthermophile protein derived from <i>Sulfolobus acidocaldarius</i>
S_B	Bound Sensor
scFv	Single chain variable fragment
S_F	Free sensor
Sso7d	A hpyerthermophile protein derived from <i>Sulfolobus solfataricus</i>
S_{Tot}	Sensor concentration
τ_{1/2}	Half-time of equilibration
YSD	Yeast Surface Display

Chapter 1

Introduction

...

...

Traditional biochemical methods have provided valuable insights into the function of proteins in their biological context, but these methods fail to noninvasively provide critical information about the spatial and temporal regulation of proteins. Traditional protein detection is done in two steps: recognition and signal transduction. The recognition step refers to the detection of the protein of interest, usually carried by binding to a receptor or antibody. The binding interaction needs to be translated into a measurable signal; this transduction step is most frequently carried by means of fluorescence. Due to separation between these steps, most methods fail to provide spatial and temporal information. Methods that couple the recognition and signal transduction event into one single step, such as Förster Resonance Energy Transfer (FRET) sensors, provide temporal and spatial resolution [1]. However, these sensors require manipulation of the biological system and are therefore invasive. In 1964 Buhr and Koshland introduced the novel concept of using an environment sensitive fluorophore coupled to an antigen-specific binder [2]. Changes in the environment of the fluorophore upon ligand binding convert the binding signal into a fluorescence read-out, offering the potential for noninvasive temporal and spatial resolution. Over the past two decades, several groups have reported the development and application of this class of sensor, referred to as reagentless biosensors [3, 4, 5, 6]. Here we will describe the components of this system and the established state of the art.

1.1 Protein engineering by yeast surface display

The first step in the design of reagentless biosensors is the engineering of an analyte specific binder. Since a binder with the appropriate affinity (see Chapter 2) is not always present in nature, a method to isolate and affinity mature binders is needed. Yeast surface display (YSD), first described by Boder & Wittrup in 1997 [7], is an ideal method for binder engineering. In YSD the protein of interest is expressed as a fusion protein to the yeast agglutinin protein Aga2p, a component used by yeast

cells to mediate cell-cell contacts during mating. Once brought to the extracellular domain, Aga2p forms a covalent double disulfide bond with Aga1p. The protein of interest to be displayed is surrounded by two small tags, the hemagglutinin (HA) and c-myc tags (Fig. 1-1). YSD has been used to isolate de novo binders to an analyte of interest [8], mature the affinity of an analyte/binder interaction [8], identify an epitope [9], increase thermal stability [10], solubility and expression [11], as well as improve the catalytic kinetic rates of an enzyme [12]. Yeast libraries can be larger than one billion members, limited in size by the logistics of cell culture. The library of mutants can be screened using both magnetic-activated cell sorting (MACS) and flow cytometry approaches. MACS has been shown to be a potent strategy to extract very low affinity binders from a naive library due to highly avid interactions [13]. Selection by fluorescence-activated cell sorting (FACS) is performed by incubation of the yeast library with the antigen of interest and then sorting clones possessing the properties of interest. Often, several rounds of selection and mutagenesis are required to achieve the properties of interest. This technique has been long established and shown successful for a great variety of targets [14].

1.2 Scaffolds for binder engineering

In a reagentless biosensor, the recognition unit consists of either the native complex partner or an engineered binder by directed evolution, generally defined as a scaffold. These recognition units are selected for their affinity and specificity and most approaches today rely on polynucleotide and polypeptide recognition units. Here, the drawbacks and advantages of some scaffolds that have been used in biosensor approaches or binder engineering will be discussed.

There have been several studies of biosensors based on aptamers, a polynucleotide scaffold. Tuleuova *et al.* developed an aptamer based Interferon γ (IFN γ) sensor [15].

In the initial state, the fluorescently labeled aptamer is conjugated to its complement, conjugated to a quencher. In the presence of IFN γ , the complementary DNA strand is replaced by the protein which removes the quenching. Similarly, Xu & Lu developed an aptamer based adenosine sensor using the environment sensitive fluorophore malachite green with up to 10-fold increase in fluorescence upon binding [16]. *In vivo*, aptamers are susceptible to degradation by nucleases, rendering them less attractive than protein based scaffolds.

Several protein scaffolds such as single chain variable fragments (scFv) [3], the human 10th fibronectin domain 3 (Fn3) [5], designed ankyrin repeat proteins (DARPin) [6, 4, 17], and a small DNA binding scaffold (Sac7d) [4] have been investigated for the development of reagentless biosensors. Except scFv, these scaffolds do not possess intramolecular disulfide bonds. This allows facilitated protein expression in *E. coli*, and introduction of cysteines for site-specific labeling with a fluorophore. Other established cysteine free protein engineering scaffolds are the B domain of staphylococcal protein A (affibody) [18], heavy chain domain (V_HH) of llama antibodies (nanobody) [19] and the *Sulfolobus solfataricus* DNA binding domain (Sso7d) [20]. The structure of these molecules is shown in Fig. 1-2. For reagentless biosensor applications, the absence of cysteines is a key advantage for the fluorophore conjugation, and a priori all the cysteine-free scaffolds would be equally suitable (see Chapter 3).

1.3 Solvatochromic fluorophores

In reagentless biosensor systems, a fluorophore is conjugated to the scaffold for signal transduction. Most fluorophores used today have been engineered for robustness of their fluorescence properties. However, sensitivity to environment factors such as temperature, solvent polarity, pH, and viscosity, can be an advantage in some situations. A specific class of fluorophores that are highly sensitive to local polarity changes dis-

play a property known as solvatochromism. Within this family of fluorophores, the solvatochromic indole of tryptophan has had a wide-spread use in the monitoring of protein/ligand interactions [21]. In recent decades, with an enhanced understanding of the solvatochromism mechanism, several new synthetic fluorophores have been discovered and used for the investigation of protein interactions and structural dynamics [22].

Solvatochromic fluorophores possess an electron donor and acceptor group, forming a dipole. Upon excitation by a photon of the appropriate wavelength, electronic rearrangement occurs in the fluorophore and results in a change in the dipole magnitude (Fig. 1-3A). The solvent molecules neighboring the fluorophore then reorient to accommodate this new dipole moment, a process referred to as solvent relaxation (Fig. 1-3B). Essentially, two outcomes are possible from this point: 1) a charge transfer resulting in a red-shifted photon emission and thus restoring the initial dipole or 2) an internal molecular rotation decoupling the electron orbitals, resulting in a kinetic loss of energy. The path taken will depend on the energy level of each state, which is a function of the solvent polarity, viscosity, the volume of the rotating groups, and the environment temperature (Fig. 1-3C) [23, 24, 25]. Thus, typically the fluorescence of solvatochromic fluorophores will increase in intensity and will undergo a blue shift (hypsochromic shift) from polar to non-polar solvents. There exists a large number of solvatochromic fluorophores and the choice of one over another relies on different factors such as extinction coefficient, wavelengths of excitation and emission, quantum yields of fluorescence, size, hydrophobicity, pH-stability, and others. The chemical structures of several common solvatochromic fluorophores are shown in Fig. 1-4.

1.4 State of the art

The design principles for reagentless biosensors rely on the hypothesis that one can label a binder with a solvatochromic fluorophore, which will see a change in environment polarity upon ligand binding. It has been thought that changes in environment are mediated by a combination of solvent shielding and hydrophobic patches on a protein. While binder engineering and site-specific labeling are well established methods, the challenge in reagentless biosensor design has been the careful choice of the labeling site. Three design rules have been formulated: the fluorophore must be labeled on a residue which is 1) solvent-exposed, 2) non-interfering with binding, and 3) proximal to the binding epitope [2].

The rational for the positioning of the fluorophore often relies on existing crystallographic structures of the protein bound to its ligand. Venkatraman *et al.* have used the crystal structure of the class II Major Histocompatibility Complex (MHC) protein in complex with a peptide fragment derived from influenza (HA tag) and identified the presence of a small hydrophobic pocket [26]. Positioning of the fluorophore at a location where it would be in contact with this pocket yielded increases by two to three orders of magnitude in fluorescence intensity upon binding. Similarly, calmodulin based sensors were developed by the same group: analysis of its crystal structure allowed the identification of specific residues both directly on the calmodulin and on a calcium dependent M13 binding peptide that would be exposed to changes in solvent polarity upon complex formation[27, 28]. Again, the increase in fluorescence upon calcium addition varied by up to two orders of magnitude. Brient-Litzler *et al.* have used the DARPin clone Off7, binding to Maltose Binding Protein (MBP), introducing cysteines at several positions that were identified thanks to the complex crystal structure [6]. They identified a clone labeled with IANBD whose fluorescence increase upon incubation with saturating concentrations of MBP was 15-fold.

Alternatively, computational methods have been used to predict protein-ligand interaction. De Lorimier *et al.* have developed a series of biosensors based on bacterial periplasmic binding proteins which are specific for a variety of small molecule ligand and undergo a conformational change upon ligand binding [29]. Environment sensitive fluorophores were attached near the hinge region between the two domains that reorient to accommodate the ligand binding, showing in rare instances up to two-fold increase in signal.

Often, the lack of structural data forces the identification of labeling position empirically. Renard *et al.* have evaluated several labeling position on the scFv D1.3 directed against hen egg white lysozyme, with the best sensor achieving a 1.9-fold increase in the fluorescence of NBD in presence of lysozyme [3]. Gulyani *et al.* used the Fn3 scaffold engineered for binding to Src SH3 domain and achieved 1.5-fold increase in fluorescence with a merocyanine fluorophore [5].

Regardless of the approach used to determine labeling sites, selection has always undergone a trial-and-error approach, with a low success rate. Despite its inception five decades ago, this sensor method has not seen wide spread adoption, most likely due to the medicinal chemistry nature of its development. Therefore, establishment of more robust design rules could significantly affect the adoption of reagentless biosensors.

1.5 Thesis Summary

This thesis employs the tool of protein engineering by YSD to develop and characterize reagentless biosensors. The design principles established can be used to generate reagentless biosensors reliably, with greater sensitivity, and with adequate resolution for real-time measurements.

- In chapter 2, we establish a mathematical model describing the faithfulness with which a sensor reports the concentration of an analyte. We show how binding kinetic rates affect the temporal resolution and signal to noise ratio of a sensor.
- In chapter 3, we generate binders against several Epidermal Growth Factor Receptor ligands, design sensors based on various scaffolds and characterize a model system. Based on this large data set, we identify two new rules for reagentless biosensors: local scaffold rigidity and fluorophore binding pockets on antigens. We validate these rules by designing the best reported sensor to date.
- In chapter 4, we describe a yeast surface display platform that can be used for high-throughput selections of reagentless biosensors via flow cytometry. More broadly, this method can be applicable to the engineering of post-translationally modified proteins displayed on the yeast surface.
- In chapter 5, we present another sensor design based on competitive inhibition of quenching. This approach relies on steric hindrance which is conceptually and experimentally easier to implement. We design proof of concept examples for two analytes based on two distinct scaffolds, highlighting the generality of this approach. We build a mathematical framework describing how this sensor system can be easily fine-tuned for different analyte concentration conditions. Finally, we lay ground for experimental validation of this theoretical model.

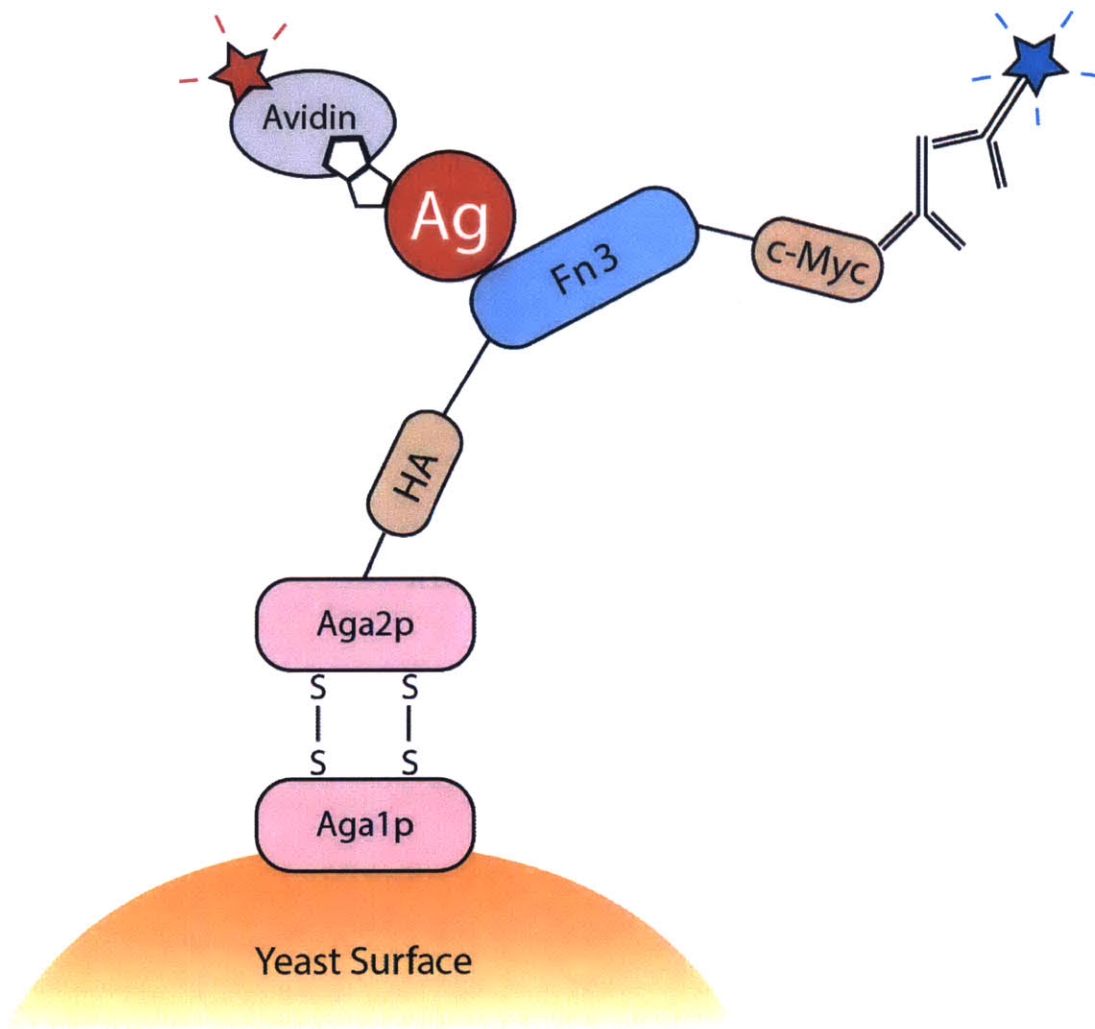


Figure 1-1: **Schematic of yeast surface display [30].** In this example, a human 10th fibronectin domain 3 is expressed as a fusion to Aga2p, in between the HA and c-myc tags. The tags can be detected with specific antibodies and biotinylated antigen can be detected with a fluorescently labeled avidin or anti-biotin antibody.

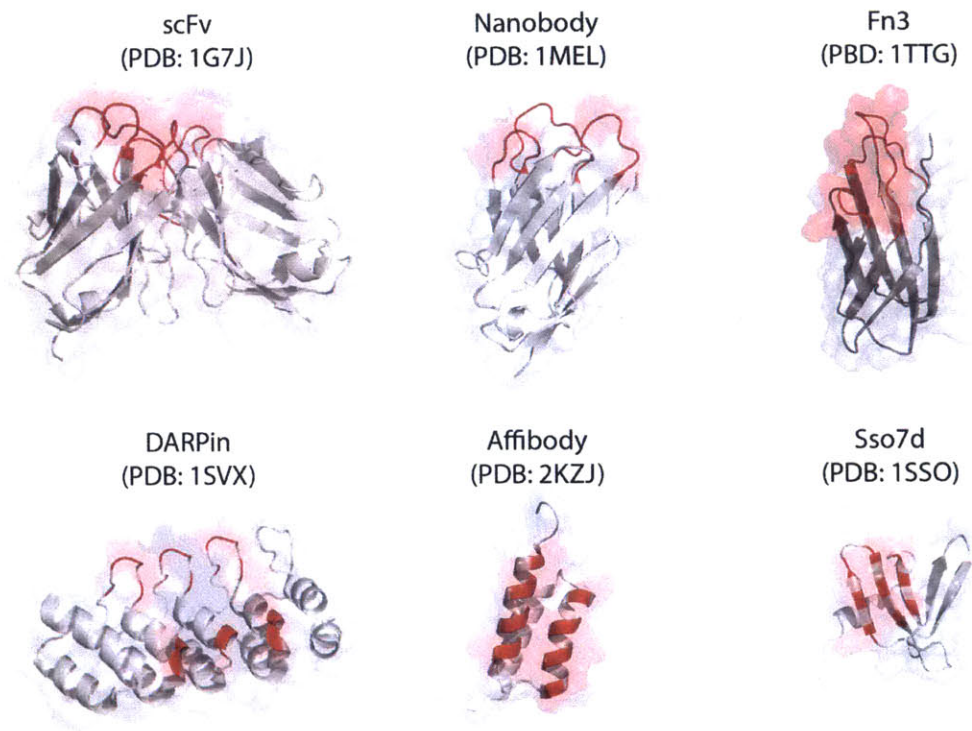


Figure 1-2: **Scaffolds for protein engineering.** In red are highlighted the regions usually randomized in library designs. All scaffolds are drawn approximately to scale.

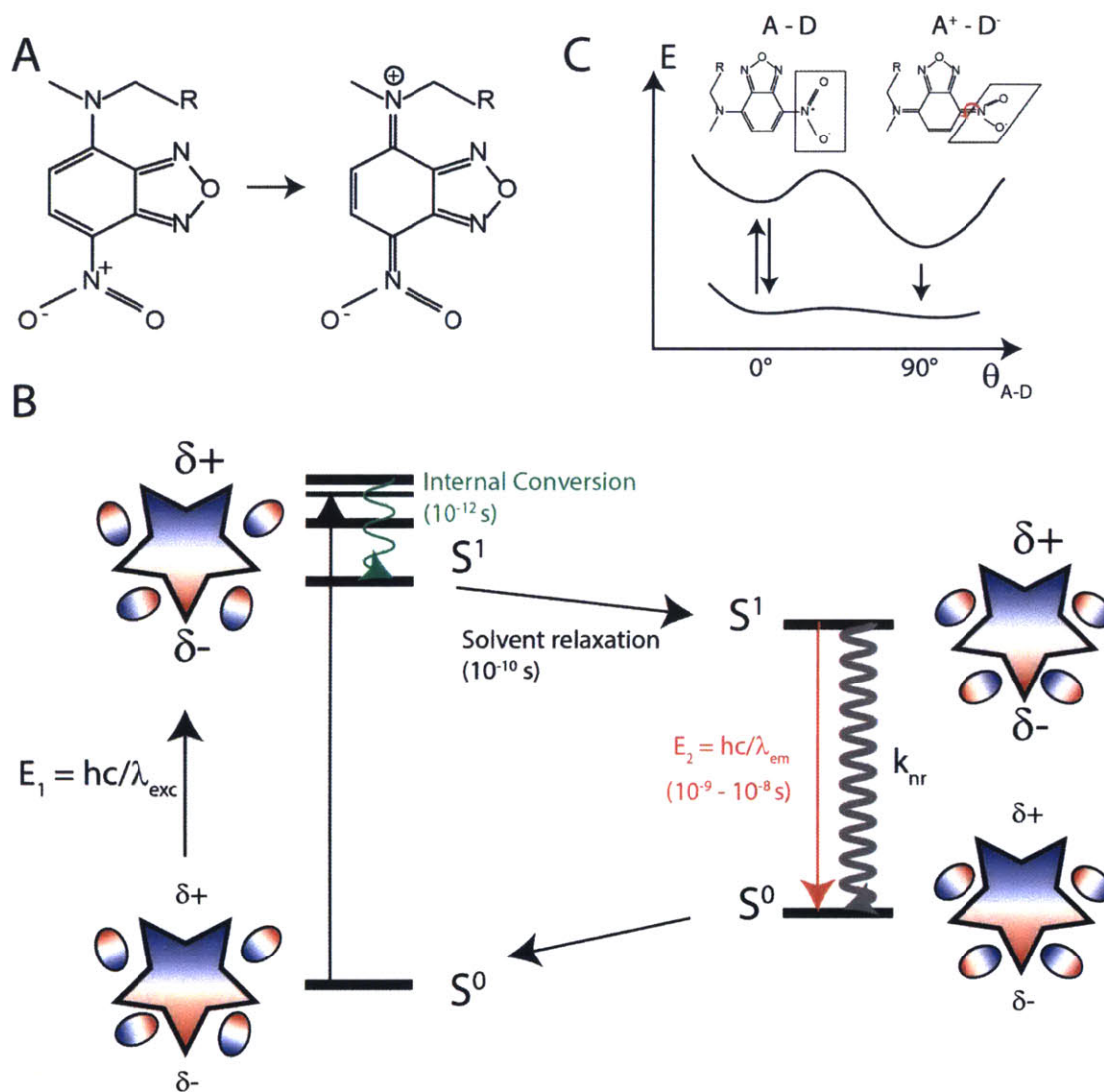


Figure 1-3: **Mechanism of solvatochromism.** **A)** Excitation of fluorophore can cause an increase in the dipole by rearrangement of electrons, the fluorophore NBD is depicted as an example. **B)** Jablonski diagram describing the effect of solvent polarity on the dipole moment of a fluorophore. **C)** Twisted-intramolecular charge transfer. A and D represent the electron acceptor and donor respectively. The energy of the system is plotted against the reaction coordinate, the rotation angle between A and D.

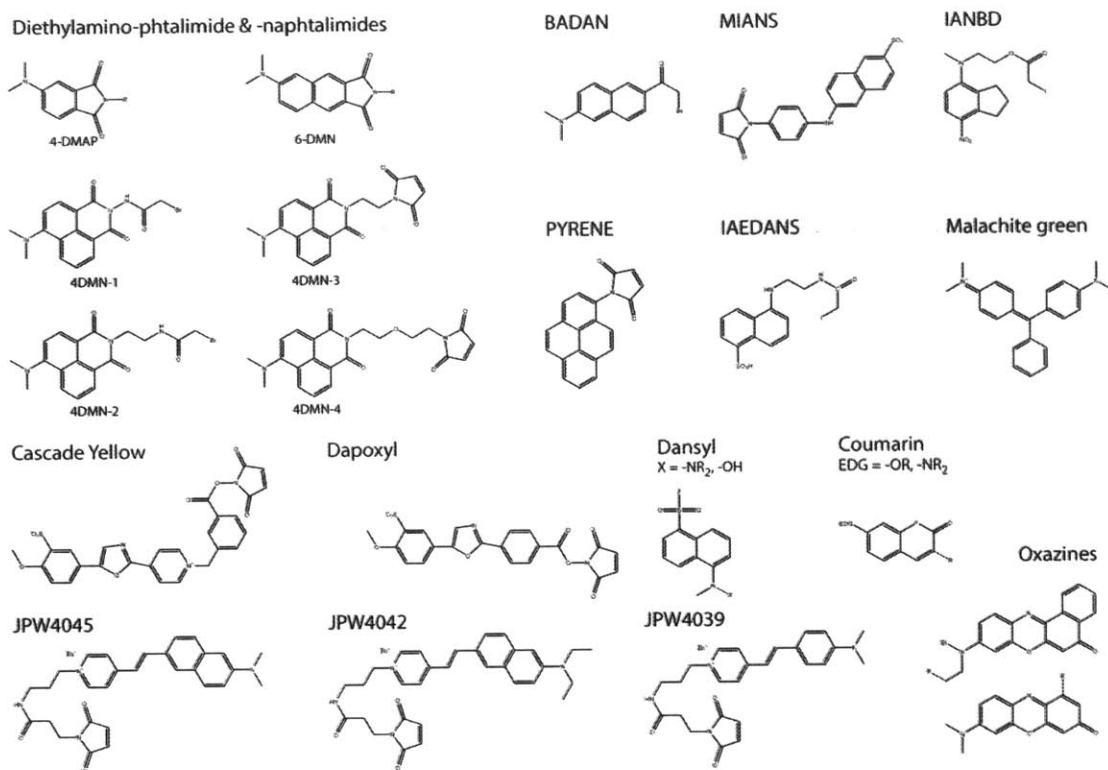


Figure 1-4: Structures of common solvatochromic fluorophores.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

Equilibrium and Dynamic Design Principles for Binding Molecules Engineered for Reagentless Biosensors

1

¹This chapter has been previously published in integrality: de Picciotto *et al.*, Analytical Biochemistry, Volume 460, 1 September 2014, Pages 9-15

2.1 Abstract

Reagentless biosensors rely on the interaction of a binding partner and its target to generate a change in fluorescent signal using an environment sensitive fluorophore or Förster Resonance Energy Transfer. Binding affinity can exert a significant influence on both the equilibrium and the dynamic response characteristics of such a biosensor. We here develop a kinetic model for the dynamic performance of a reagentless biosensor. Using a sinusoidal signal for ligand concentration, our findings suggest that it is optimal to use a binding moiety whose equilibrium dissociation constant matches that of the average predicted input signal, while maximizing both the association rate constant and the dissociation rate constant at the necessary ratio to create the desired equilibrium constant. Although practical limitations constrain the attainment of these objectives, the derivation of these design principles provides guidance for improved reagentless biosensor performance and metrics for quality standards in the development of biosensors. These concepts are broadly relevant to reagentless biosensor modalities.

2.2 Introduction

The field of biosensors has seen in the last decade a multitude of new approaches for the application of reagentless sensors. The overall strategy is the combination of a recognition unit and a signal-transducing unit into one molecular entity. The most commonly-used signal is change in sensor fluorescence, arising either from Fluorescence Resonance Energy Transfer (FRET) or from solvatochromism. FRET was first described over half a century ago, and its application in biology has grown with design and implementation of myriad biosensors (reviewed in [31]). Solvatochromism is a more recent development, but is becoming more widely used as

new scaffolds (affinity molecules) and dyes (environmentally-sensitive fluorophores) are developed and become available (reviewed in [32]). Several groups have successfully developed solvatochromic-based biosensors using DNA aptamers [33, 34, 35], native protein receptors [28, 36], peptides [37, 38] or engineered binders using protein scaffolds [3, 4, 5, 39, 17, 6]. However, to our knowledge, none of these groups have purposely engineered binders with affinities specified for optimal performance as a sensor, relying instead on previously-described proteins. Selecting an existing binder with an affinity above the detection threshold is likely an adequate approach for categorical detection of the presence or absence of an analyte. However in a complex biological system, analyte concentrations may vary rapidly on the timescale of seconds to minutes. For dynamic measurement of time-varying analyte levels, the biophysical characteristics of the binding event can significantly impact biosensor accuracy and sensitivity. Given the availability of directed evolution protein engineering methodology to create binding molecules of almost arbitrary affinity and widely varying association and dissociation rates [14, 40], these variables are available degrees of freedom for improvement of biosensor performance. Recently Haugh developed a reaction-diffusion model to investigate biosensor signal interpretation in live cell imaging, with an emphasis on capturing intracellular and membrane-localized phenomena [41]. This analysis resulted in the identification of an important trade-off between robust signal and perturbation of the biological system or signal saturation. Here, we perform a theoretical analysis of biosensor dynamics, delineating time and length scales important in observation of intracellular as well as extracellular phenomena (*e. g.* detecting autocrine loops). Using a sinusoidal signal as an input ligand concentration, as biological signal do vary, we present new important considerations for the appropriate implementation of a biosensor. Further, we propose metrics for quality standards in the development of biosensors by direct comparison between the input signal and measured signal and, thereby derive design criteria for improved performance.

2.3 Model formulation

The system consists of three state variables: the concentrations of ligand (L), unbound sensor (S_F) and bound sensor (S_B). By virtue of mass balance, the sum of the concentration of unbound and bound sensor is always equal to the total sensor concentration constant (S_{Tot}). A linear correlation between bound sensor and the output signal intensity is assumed. The two rate constants governing this process are the association (k_{on}) and the dissociation (k_{off}) rate constant.

The mathematical description of this interaction, a reversible bimolecular reaction, is well documented from the perspective of dynamic steady state equilibrium; however, it has generally been investigated in an environment of constant ligand concentration [42, 43]. To determine the optimal design criteria in a dynamic system where the input (*i.e.* L) is time-varying, we apply a frequency response approach by sinusoidally varying the analyte input, L, and characterizing the dynamic fluorescence intensity response of the sensor, which is proportional to the concentration of bound sensor S_B . A range of physiological behaviors can be modeled by systematic variation of the mean (L_0), amplitude (A_L) and period (T) of the time-variant ligand concentration. With these parameter definitions, the input function L is defined as:

$$[L] = L_0 + A_L \cdot \sin \frac{2\pi}{T} t \quad (2.1)$$

To score a given set of design parameters of a sensor, we choose three signal properties: mean signal intensity (M), normalized amplitude (A) and phase delay (Φ), as defined in equations 2.2, 2.3 and 2.4.

$$M = \frac{S_B^{\max,eq} + S_B^{\min,eq}}{2 \cdot S_{Tot}} \quad (2.2)$$

$$A = \frac{S_B^{\max,eq} - S_B^{\min,eq}}{S_{Tot}} \quad (2.3)$$

$$\Phi = \frac{t(S_B = S_B^{\max, \text{eq}}) - t(L = L_{\max})}{T} \quad (2.4)$$

We assume that the system is reaction limited. Indeed the Damkohler number, defined as the ratio between the characteristic time for diffusion and that of reaction (complex formation in this context, see Eqn. 2.5), will be much smaller than 1 for all relevant k_{on} , k_{off} , ligand concentration ($[L]$), diffusion coefficient (D) so long as the distance (r_b) over which the measurement must be spatially resolved is less than 1 micron (Fig. 2-6).

$$\text{Damkohler\#} = \frac{\tau_{\text{diff}}}{\tau_{\text{rxn}}} = \frac{\frac{r_b^2}{D}}{(k_{\text{on}} \cdot [L] + k_{\text{off}})^{-1}} \quad (2.5)$$

The described system is now fully governed by the three differential equations:

$$\frac{d[L]}{dt} = \frac{2\pi}{T} A_0 \cos\left(\frac{2\pi}{T} \cdot t\right) - k_{\text{on}}[L][S_F] + k_{\text{off}}[S_B] \quad (2.6)$$

$$\frac{d[S_F]}{dt} = -k_{\text{on}}[L][S_F] + k_{\text{off}}[S_B] \quad (2.7)$$

$$\frac{d[S_B]}{dt} = +k_{\text{on}}[L][S_F] - k_{\text{off}}[S_B] \quad (2.8)$$

To simplify this system further, we assume that the ligand is in excess. Of course, as shown in the analysis by Haugh [41], this is a constraint that must be calculated for any real system since this assumption can often break down. Systematic use of a sensor concentration of one twentieth that of the minimum ligand signal guarantees excess ligand concentration. By substituting equation 2.1 into 2.8, and using conservation of mass for the sensor species, we obtain the 1-D governing equation:

$$\frac{d[S_B]}{dt} = k_{\text{on}}(S_{\text{Tot}} - S_B) \left(L_0 + A_L \cdot \sin \frac{2\pi}{T} t \right) - k_{\text{off}} S_B \quad (2.9)$$

A convenient analytical solution to this system is not available. Therefore, we solved this equation for a variety of parameter conditions by numerical Euler integration (performed in MatLab, see Appendix C). The results are shown in the next section.

2.4 Results

2.4.1 Dynamic consideration reveals the crucial importance of kinetic rates optimization.

Intuitively, a sensor that has a very high affinity for its ligand might be expected to perform as a weak dynamic sensor since the characteristic time for complex dissociation would likely be much greater than the period of the signal. Relevant input signal conditions depend greatly on the system under study. In Fig. 2-1 we show approximate concentrations and time scales for concentration variation for various classes of biological events. Many physiological processes result in great variation of analyte concentration such as cell-cycle related proteins, signaling cascades, immune response activation among many others. Mathematically, the sinusoidal function is a benchmark for representing time-variant signals. For example the well-known Bode plot uses a sinusoidal signal to characterize a system's frequency response [44]. We first investigated how output signal differs with varying the dissociation rate constant (k_{off}). As an initial input signal, we chose a mean ligand concentration of 3 nM with sinusoidal oscillation between 1 and 5 nM with a period of 100 minutes. This signal is shown as a solid gray curve on Fig. 2-2A. In black are shown four different sensors with varying k_{off} but identical association rate constant ($k_{\text{on}} = 10^5 \text{ M}^{-1}\text{s}^{-1}$). In this first approach, we show the signal for the first 4 periods (400 minutes). An initial condition corresponding to $S_B = 0$ was chosen for the analysis depicted in Fig. 2-2A, hence an initial transient in signal response is observed. The signals progress toward their dynamic steady state, where higher signal intensities are reached with decreasing k_{off} as expected given the equation for complex concentration under the pseudo-first order approximation (see equation 2.10).

$$[S_B] = S_{\text{tot}} \frac{L_0}{L_0 + \frac{k_{\text{off}}}{k_{\text{on}}}} \quad (2.10)$$

This transient behavior is followed by a dynamic steady state characterized by a constant value of the mean signal. The equilibrium half time ($\tau_{1/2}$) is defined as:

$$\tau_{1/2} = \frac{\ln(2)}{k_{\text{on}}L_0 + k_{\text{off}}} \quad (2.11)$$

The periodic signal is assumed stable for $t > 100 \cdot \tau_{1/2}$, when the transient conditions would have decayed substantially. In the subsequent analysis, we report signals when $t > 100 \cdot \tau_{1/2}$. We therefore plotted the amplitude and mean signal value for the signals in Fig. 2-2B as a function of k_{off} . In both the transient and dynamic steady state regimes, the mean signal value increases with decreasing k_{off} . The oscillation amplitude appears very small for both extreme k_{off} values. However, there is an optimum at intermediate values as seen when the amplitude and mean signal value are plotted as a function of k_{off} (Fig. 2-2B). Interestingly, the optimum k_{off} value for this particular example is a physically realistic rate of $2 \cdot 10^{-4} \text{ s}^{-1}$.

2.4.2 Varying input characteristics

We next investigated how this optimum region behaves with different input signal characteristics across three orders of magnitude of mean ligand concentration for various sensor properties (k_{on} and k_{off} values) and a constant oscillation period of $T = 100 \text{ min}$ (Fig. 2-3). The amplitude was also held constant at $A_L = 2/3 \cdot L_0$ (see Eqn. 2.1). Recognizing that any particular sensor will be most sensitive over about a 10-fold concentration range, we divided the analysis into 3 concentration regimes representing anticipated physiological values of interest for oscillation (0.1-0.5 nM; 1-5 nM; 10-50 nM). Within each regime, we created heat maps to reflect normalized values of mean signal, mean amplitude, and phase shift for combinations of k_{off} and k_{on} spanning 6 orders of magnitude (Fig. 2-3).

The general features of the heat map are illustrated by Fig. 2-3A, which depicts

the mean output signal intensity (% max signal) for oscillation of L in the low concentration regime (0.1-0.5 nM). Regions in white at the bottom left indicate a high mean signal and are associated with the regime of tightest equilibrium binding affinity ($K_D = k_{\text{off}}/k_{\text{on}}$); *i.e.*, the regime where the receptor is saturated so that it is insensitive to variations in k_{on} and k_{off} . The diagonal at which the normalized mean signal intensity is 50% of the maximum corresponds to values where $K_D = L_0$. The position of this diagonal is naturally shifted upward as the mean ligand concentration is increased 10-fold to 3 nM (Fig. 2-3E) and 100-fold to 30 nM (Fig. 2-3I). By this criterion alone, the desire for robust signal detection would favor biosensors that have $K_D < L_0$, with greatest sensitivity for concentration discrimination (versus just threshold concentration detection) in the range $K_D \sim L_0$. However, consideration of dynamic response introduces additional constraints as discussed below.

Despite strong mean signal intensity, the highest affinity binders, with k_{on} and k_{off} represented on the bottom right of each panel of Fig. 2-3A, E, and I, are insensitive to time varying concentrations of the ligand. This phenomenon can be appreciated by examining the amplitude of the output signal as defined in Eqn. 2.3 and plotted as heat maps in Fig. 2-3B (L variation of 0.1-0.5 nM), f (L variation of 1-5 nM), and j (L variation of 10-50 nM). The region of greatest output signal amplitude is shown in white and is obtained for fast kinetic rate constants. Fast association rate constants mean that the biosensor is quickly able to capture ligand and therefore rapidly report the signal, while fast dissociation is crucial to adapt the variations in ligand concentration. This optimum area lines up again with an equilibrium dissociation constant $K_D = L_0$, however, fast kinetic rates are crucial to prevent the binding dynamics from obscuring the input (ligand oscillation) dynamics. As the mean ligand concentration rises from values in Fig. 2-3B, to those in 2-3F and J the abundance of ligand makes it easier for the sensor to capture and report the signal. This is illustrated by the shift to the left of the optimal region. Fig. 2-3C shows the phase shift, which is a measure of the delay in signal reporting. Here the favored regions are those of low shift, shown in white. It can be seen that again faster kinetic rates are beneficial in

reporting accurately the input signal. This optimal region expands to slower kinetics as L_0 increases from Fig. 2-3C, G to K.

In order to evaluate these three criteria simultaneously, we first define for illustrative purposes an arbitrary design threshold for each: $M > 20\%$, $A > 10\%$ and $\Phi < 0.1$. We then plotted on the fourth column of Fig. 2-3 the region of k_{on} and k_{off} where all criteria are satisfied. We observe that the criteria are easier to meet for conditions of high mean ligand concentration. For the example with the lowest ligand concentration (oscillating between 0.1 and 0.5nM), only a narrow range of rate constant combinations results in adequate performance by the chosen criteria. Therefore, depending on the ligand conditions, it may be necessary to sacrifice some characteristic of the output.

In Fig. 2-3 we have shown the dependency of the kinetic rate constants on the location of optimal regions for the three different biosensor design criteria and investigated how the sensors characteristics are affected for different mean ligand concentrations. The same approach can be undertaken by varying the other two parameters - in ranges that are biologically relevant - of the input signal (period and amplitude); we have summarized the effects on the position of optimal $\langle k_{on}, k_{off} \rangle$ region in Table 2.1. The output mean signal is insensitive to changes in input frequency or amplitude. However, for extreme conditions the changes in concentration occur very rapidly and the effective mean concentration rises. This phenomenon is present only for dynamics that are orders of magnitude faster than any biological process. Finally, it is interesting to note that increase in the input amplitude has no effect on the position of the optimum.

2.4.3 Single sensor analysis

So far we have been exploring how a variety of sensors perform for a circumscribed set of input signal conditions. We next analyzed how a set of three individual sensors, with combinations of properties that span the spectrum of physically possible values, performs in a variety of different input conditions. To do so, we use a Bode diagram representation since it is the canonical approach for representing dynamic systems responses. Oscillation cycle times spanning the range 10 min - 40 hr (frequencies of $3 \cdot 10^{-6}$ to $3 \cdot 10^{-3}$ Hz) were examined as it encompasses a great range of physiologically relevant systems (Fig. 2-1). For a signal oscillating between 1 and 5 nM and a sensor concentration of 0.05 nM, we have represented the normalized amplitude, phase shift and mean signal in Fig. 2-4 for three different sensors over this: In solid black we have represented a tight binder ($k_{\text{on}} = 10^6 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 10^{-6} \text{ s}^{-1}$), in dashed black a utopian binder ($k_{\text{on}} = 10^8 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 10^{-1} \text{ s}^{-1}$) with ideal characteristics but physically non feasible, and in dotted black a feasible binder ($k_{\text{on}} = 10^6 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 3 \cdot 10^{-3} \text{ s}^{-1}$).

Without a dynamic analysis of this system, the tight binder would likely be chosen as the most adequate sensor because of its high mean signal (fractional saturation). However, the normalized amplitude of this sensor is null for the relevant input system frequencies. The utopian sensor performs well in a variety of conditions, the phase shift is less than 10% for signal frequencies $< 0.01 \text{ s}^{-1}$. However, the k_{on} of this sensor, $10^8 \text{ M}^{-1}\text{s}^{-1}$, is physically not achievable in general; a more typical protein/protein $k_{\text{on}} = 10^5\text{-}10^6 \text{ M}^{-1}\text{s}^{-1}$. In the dotted black line is represented a feasible sensor, named so because it performs well in wide range of frequencies and it is technically feasible. The normalized mean signal for the feasible sensor increases with faster frequencies as the biosensor inaccurately reflects a higher signal.

2.5 Discussion

We have shown that there exists an optimal combination of the design parameters k_{on} and k_{off} for a reagentless biosensor and that these vary depending on the nature of the signal. What our results indicate is that the careful determination of binding kinetics is crucial for successful application of biosensors. As a general rule, the K_D of the interaction must match that of the expected mean ligand concentration to ensure greatest sensitivity. Biosensors with a K_D lower than the mean ligand concentration will yield a binary output. Faster association and dissociation rate constants combinations provide greater correlation to the input signal. Therefore, efforts to increase both k_{on} and k_{off} while maintain a K_D near the expected ligand concentration will generally improve dynamic response time while maintaining sensitivity.

While the presence of an optimal k_{off} , k_{on} combination may exist for a given input signal, these rate constants may not be either physically relevant or in the range of what can be engineered. The dissociation rate constant is very amenable to changes and most library selection strategies rely on optimizing this parameter. For the association rate constant, the scenario is quite different: two molecules come together by diffusion which is governed by Brownian motion. Smoluchowski calculated that if proteins were smooth spheres and they formed a complex every time they would collide, the association rate dictated by Brownian motion would thus be $7 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$. However, complex formation requires stringent orientation constraints and the observed rate of protein-protein complex formation is 10^5 - $10^6 \text{ M}^{-1}\text{s}^{-1}$ [45, 46]. Nevertheless, some outliers for protein-protein and DNA-protein have been identified with association rates up to $10^9 \text{ M}^{-1}\text{s}^{-1}$ [47, 48, 49, 50] due to favorable electrostatic interactions, and very slow association rates ($10^3 \text{ M}^{-1}\text{s}^{-1}$) due to high energy barriers to complex formation [51]. Within the gray box of Fig. 2-3A are the regions considered generally accessible to protein engineering. A large range of k_{on} , k_{off} couples with high performance are well outside the range of physical possibilities.

While the optimum set of constants cannot always be achieved, there are some alternatives that can be used. Renard & Bedouelle successfully demonstrated the use of three sensors with various affinities against lysozyme in combination in order to titrate lysozyme concentration over a three log range [52]. Another group also demonstrated the possibility of narrowing the titration range using a depletion strategy [53]. Moreover a combination of sensors with different output signal characteristic can be beneficial also for dynamic scenarios. We exemplify here the possibility of using a combination of different sensors at equimolar concentrations each with different dissociation rate constants to form a more robust sensing system as illustrated in Fig. 2-5. The three sensors have an association rate constant of $10^5 \text{ M}^{-1}\text{s}^{-1}$ and equilibrium dissociation constants of 100, 10 and 1 nM. Suppose for a given application that the tightest binder offers a more than satisfactory mean signal intensity but suffers from profound signal delays, not acceptable for that application. The intensity of the signal can only be correlated to a concentration of target antigen if appropriate calibrations controls have been established. While, in general, greater signal intensities are preferred as they are facilitate detection and have lower signal to noise ratios. In some cases, it is possible to sacrifice some of the signal intensity in order to improve other characteristics of the signal. Indeed, the weaker binders do offer better signal correlation as indicated by the lower phase shift, but the signal intensity may be too weak. By combining these sensors together, one obtains a signal profile that now offers improved time correlation and satisfactory mean signal intensity. The ratios and properties of these sensors can be fine tuned to obtain the desired output signal properties. This approach is advantageous as it allows manipulation of the sensor signal properties without any additional engineering, given the condition of having at least two sensors. The Matlab code provided online allows the reader to quickly assess any desired combinations (Appendix C).

While this model can be applied to any particular ligand concentration and variation dynamics, we have here often focused on the low nanomolar concentration range. This concentration range was motivated by the ErbB extracellular signaling network.

The ErbB receptor family is activated by growth factor shedding in an autocrine or paracrine manner. Previous mathematical modeling based on experimental evidence by the Lauffenburger and Wiley groups has allowed the determination of the effective ligand concentration at the cell surface [54, 55, 56]. By controlling ligand production, shedding and receptor levels, Dewitt *et al.* were able to determine a direct relationship between ligand shedding rate and effective concentration, shown to be in the 1-10 nM range [54]. Hence, we have chosen the range of 1-5 nM for most of our demonstration. As shown in Fig. 2-1, there is a great variety in the concentration and dynamics of physiologically relevant molecules. Our model strongly indicates that careful optimization of a biosensor is critical for its appropriate deployment to investigate the biological system of interest.

Previous biosensor modeling efforts have ranged from finite element methods for microelectromechanical systems [57], to partial differential equation systems of enzymatic reactions [58]. In the field of reagentless biosensors, Haugh developed a reaction-diffusion model and showed the importance of binding parameter optimization to prevent signal saturation or system perturbation [41]. Our analysis distinguishes itself by revealing that kinetic rate constants are crucial to the proper identification of signal fluctuations. We have provided guidelines for the optimization of these parameters for a desired application. Furthermore, from this dynamic analysis, we derived metrics which we suggest should become standard for the characterization of biosensors.

Often biosensors detection mechanisms rely on tethering receptor onto microchips. In these systems, convection and diffusion are essential properties that were not investigated in our model since they were assumed negligible. Squires *et al.*, investigated design constraints imposed by transport in surface-based biosensors [59]. Through finite element methods, they discussed the time scales and collection rates for these systems as a function of the channel dimensions, flow rate, ligand diffusion and binding kinetics. Although not discussed in their publication, faster kinetics for the ligand-

receptor interaction would likely improve the correlation between the input and output signal in these systems as well. Furthermore, they highlight the importance of tethered receptor density as higher density would allow lower detection limits and also increased signal to noise ratio. Thus, we suggest that reagentless biosensors could be clustered on the surface of beads. This would greatly improve the signal to noise ratio in fluorescent microscopy read-out. But as underlined in the model published by Squires *et al.*, increased binder density correlates with greater depletion effects which ultimately could have adverse local effect on the physiology of the system.

To our knowledge, we present here the first analysis of a sensor system for kinetic rate constants optimization under dynamic conditions. Furthermore we have specifically identified three criteria for scoring the applicability of reagentless biosensors: mean, amplitude and phase shift. Based on the analysis of these criteria, we have presented here the limitations and trade-offs in the design of biosensors. While several reviews have been published for the design principles of reagentless sensors [31, 60], they fail to acknowledge the importance of sensor-ligand kinetic binding parameters optimization for their particular application. Only through a dynamic analysis, as presented here, can the importance of the kinetic constants be highlighted.

The results described here hold true if the sensor concentration is appropriate to guarantee no ligand depletion. Ligand depletion would not only affect the founding assumption of this model but also severely perturb the biological sample to be studied. In this context, this model can significantly help scientists in choosing the adequate binding parameters of their sensor and the biological system to be studied to maximize their efficacy.

Table 2.1: Summary of the effect of input modification on optimum $\langle k_{\text{on}}, k_{\text{off}} \rangle$ couples. NE = no effect

Input Property	Output Mean Signal	Output Normalized Amplitude	Output Phase Shift
Increasing Frequency	NE	$k_{\text{on}} \nearrow$ and/or $k_{\text{off}} \nearrow$	$k_{\text{on}} \nearrow$ and $k_{\text{off}} \nearrow$
Increasing Amplitude	NE	NE	NE
Increasing Mean	$k_{\text{on}} \nearrow$ and/or $k_{\text{off}} \searrow$	$k_{\text{on}} \searrow$ and $k_{\text{off}} \nearrow$	$k_{\text{on}} \searrow$ and $k_{\text{off}} \nearrow$

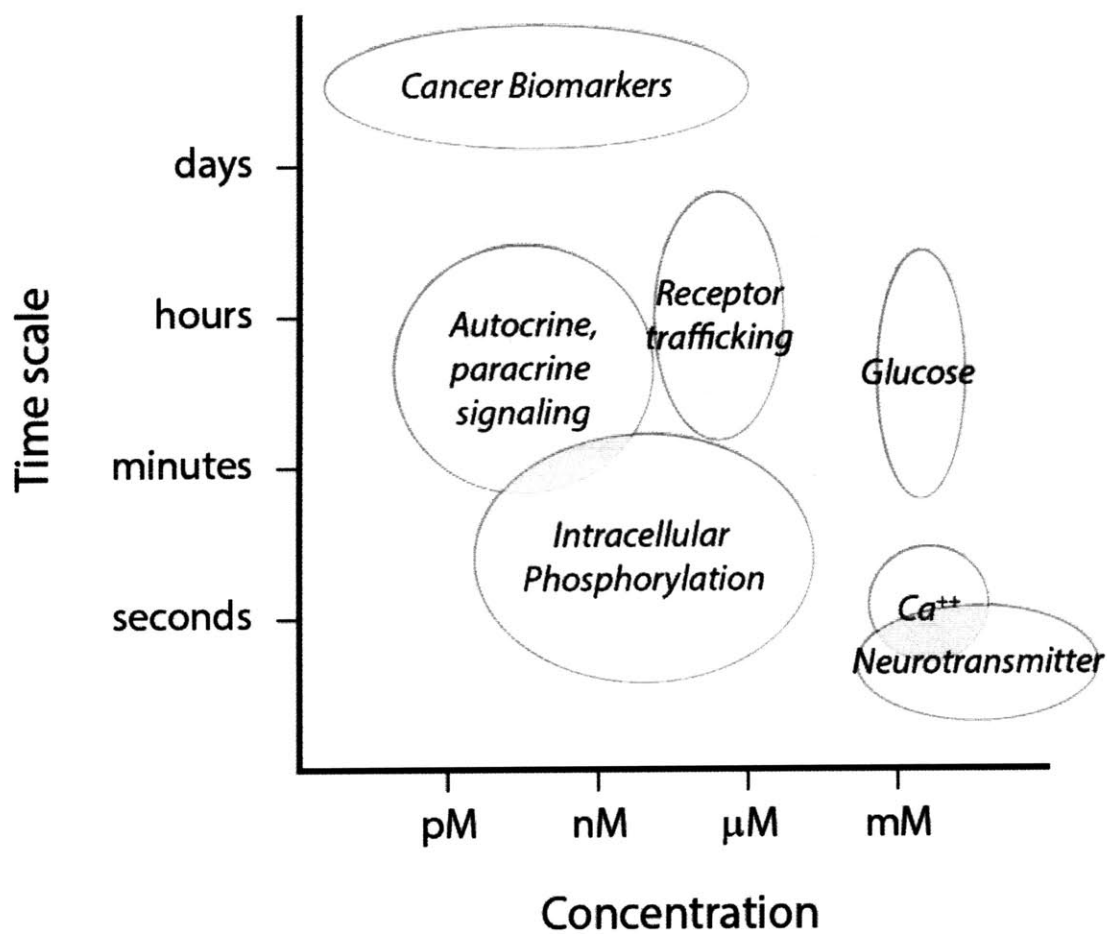


Figure 2-1: Various biologically relevant molecules and processes are depicted. Typical mean concentration is shown on the horizontal axis ranging from picomolar (pM) to millimolar (mM) against expected time scales for variation in the ligand concentrations.

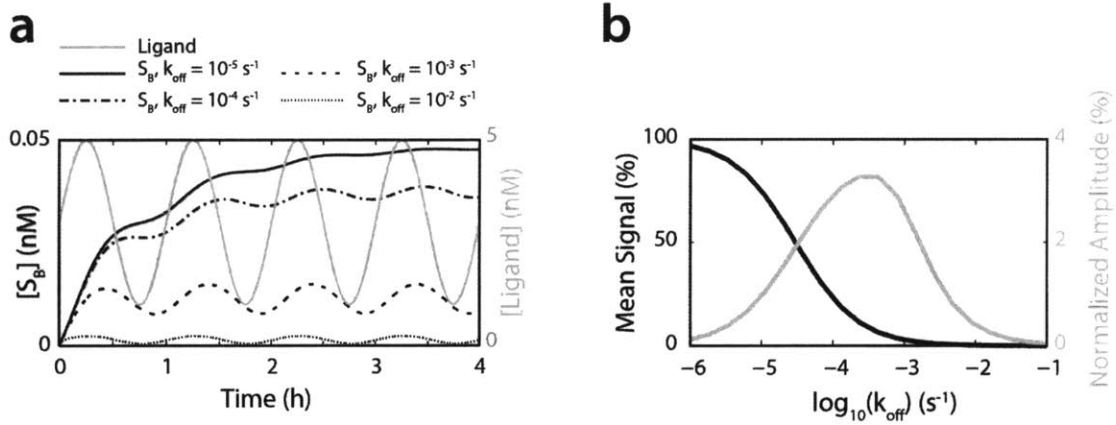


Figure 2-2: **A)** Output profiles generated from an input varying from 1-5 nM with time period of 100 minutes. The output signals are read on the left vertical axis, the input signal is read on the right vertical axis shown in gray. **B)** Once the output oscillations are stabilized, the mean concentration and normalized amplitude are given as a function of k_{off} .

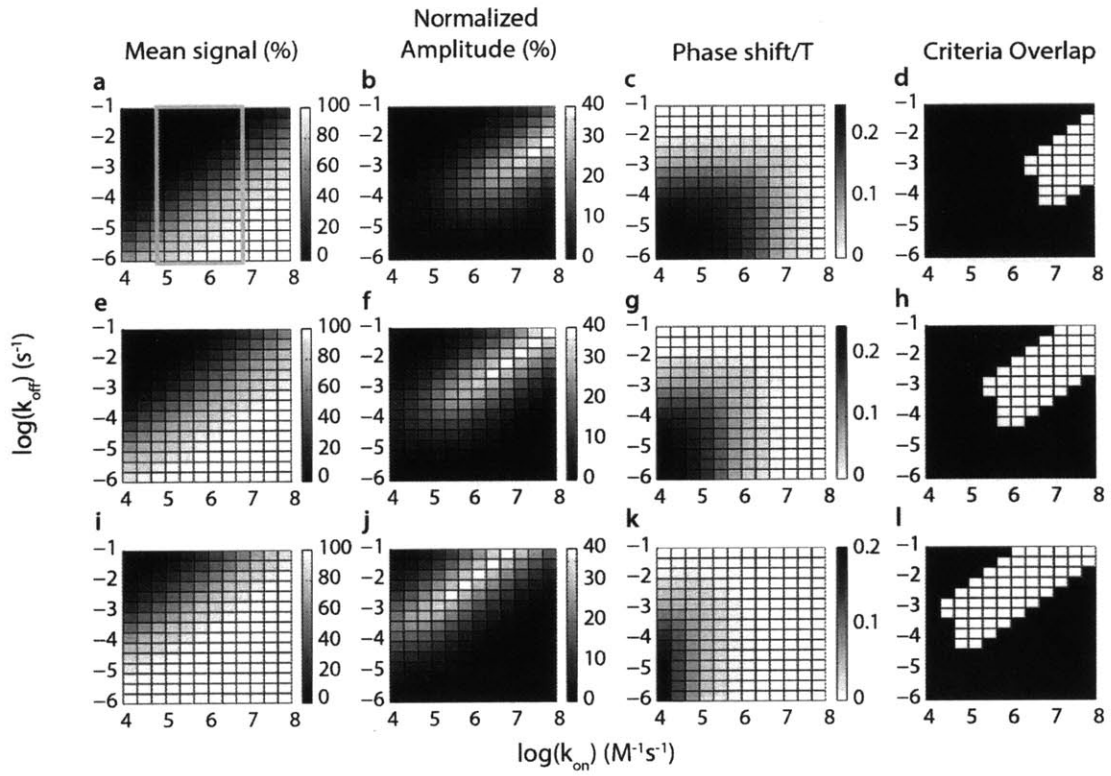


Figure 2-3: Input signal of period of $T = 100$ minutes varying from 0.1-0.5 nM (top row), 1-5 nM (middle row), 10-50 nM (bottom row). The first column represents the mean signal as a percentage of the total sensor, the second column shows the normalized amplitude, the third column the phase delay and finally the last column is an overlap of the optimal regions for all three criteria. The areas accessible to engineering are shown in the gray box.

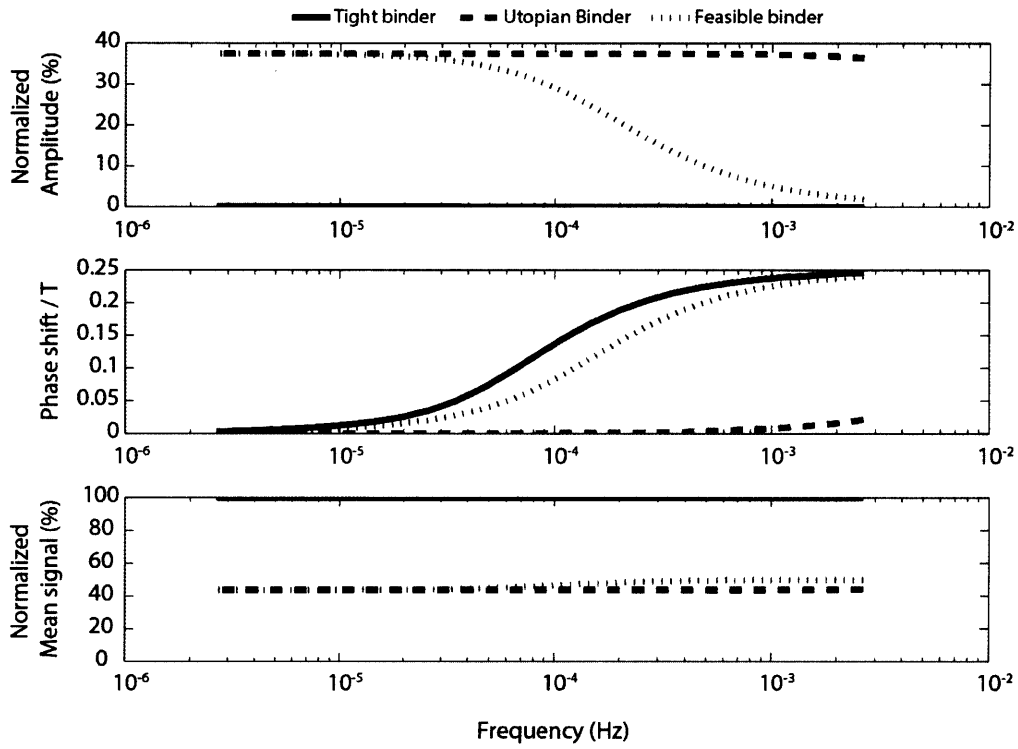


Figure 2-4: **Bode Diagram for three different sensors:** In solid black we have represented a tight binder ($k_{\text{on}} = 10^6 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 10^{-6} \text{ s}^{-1}$), in dashed black a utopian binder given our model ($k_{\text{on}} = 10^8 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 3 \cdot 10^{-1} \text{ s}^{-1}$) and in dotted black a feasible binder ($k_{\text{on}} = 10^6 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 3 \cdot 10^{-3} \text{ s}^{-1}$). The input signal varies between 1-5 nM and the sensor concentration is 0.05nM for all cases.

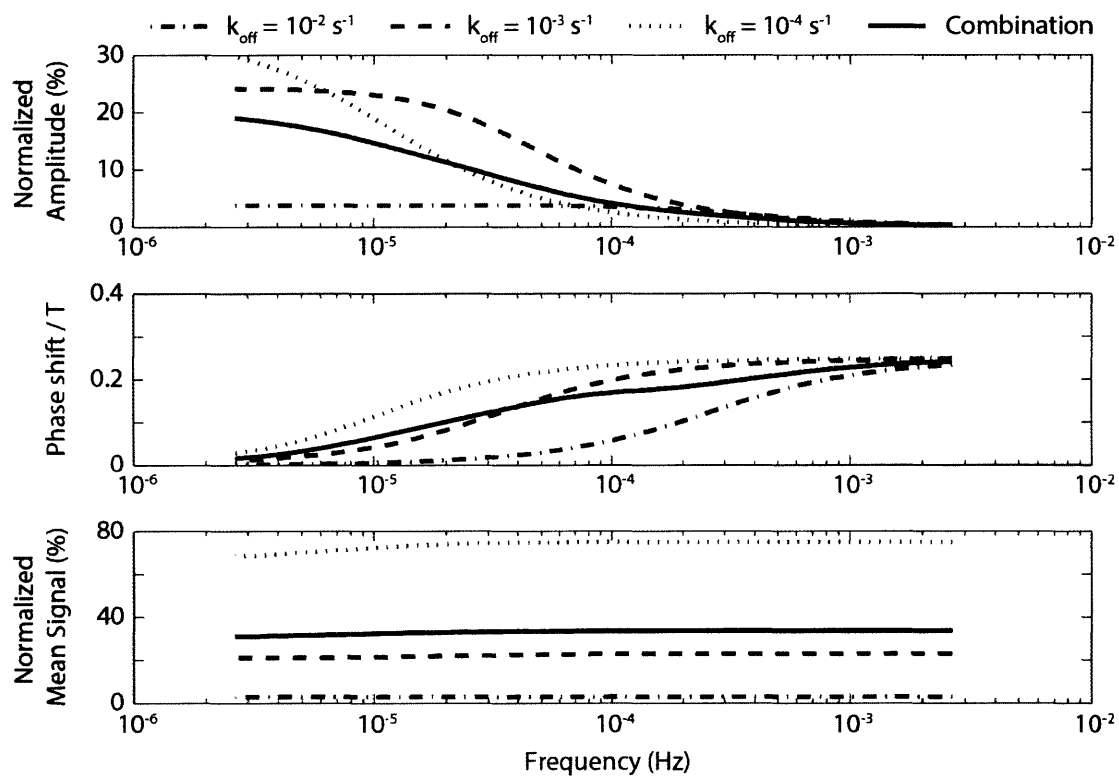


Figure 2-5: **Bode Diagram for individual and combinatorial sensor deployment.** The input signal varies from 1-5 nM, in all cases the total concentration of sensor(s) is equal to 0.05 nM. All sensors have a k_{on} of $10^5 \text{ M}^{-1}\text{s}^{-1}$.

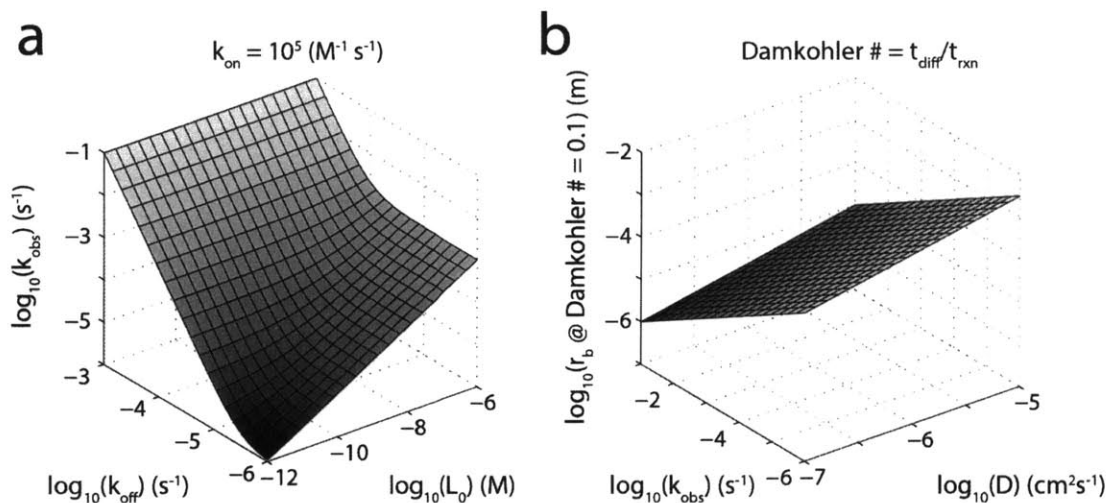


Figure 2-6: **A)** k_{obs} as a function of k_{off} and ligand concentration for $k_{\text{on}} = 10^5 \text{ M}^{-1}\text{s}^{-1}$. **B)** The diffusion distance rb for which the system is reaction limited (Damkohler # < 0.01) is plotted for a wide range of relevant observed rate constant (k_{obs}) and a relevant range for the diffusion coefficient of molecules in solutions such as water ($D = 10^{-5}$ to $10^{-7} \text{ cm}^2\text{s}^{-1}$).

Chapter 3

Design principles for reagentless biosensors: specific fluorophore/analyte binding and minimization of fluorophore/scaffold interactions.

3.1 Abstract

Quantifying protein location and concentration is critical for understanding function *in situ*. Reagentless biosensors, in which a reporting fluorophore is conjugated to a binding scaffold, can in principle detect analytes of interest with high temporal and spatial resolution. However, their adoption has been limited due to the extensive empirical screening required for their development. We sought to establish design principles for this class of biosensor by characterizing over 400 biosensors based on various protein analytes, protein scaffolds and fluorophores. We found that the brightest readouts are attained when a specific binding pocket for the fluorophore is present on the analyte. Also, interaction of the fluorophore with the binding protein it is conjugated to can raise background fluorescence and considerably limit sensor dynamic range. Exploiting these two concepts, we designed biosensors with a 100-fold increase in fluorescence upon binding to analyte, an order of magnitude improvement over the previously best reported reagentless biosensor. These design principles should facilitate the development of improved reagentless biosensors.

3.2 Introduction

Spatial and temporal fluctuations in the level of proteins in living systems contain valuable functional information that is difficult to obtain at high resolution in a non-destructive fashion. An ideal sensor would report the concentration and localization of a target analyte in real-time without interfering with its function. Sensors such as those based on Förster Resonance Energy Transfer (FRET) couple the analyte recognition and signal transduction event into a single step and provide temporal and spatial resolution [1]. However, such sensors require the introduction of two unique fluorophores into either two binders or into proteins which undergo large conforma-

tional changes upon binding. In 1964 Burr and Koshland introduced a novel concept in which an environment-sensitive fluorophore is coupled to a single analyte-specific binder [2]. Changes in the environment of the fluorophore (such as solvent shielding) upon analyte binding transduce the binding signal into a fluorescence read-out. Several groups have reported the development and application of this class of biosensors, referred to as reagentless biosensors [3, 4, 5, 6, 17].

Currently, reagentless biosensors are designed in three steps. First, a binder is engineered against the intended target using established display technologies [61, 62, 63, 64] and multiple scaffolds [65, 66, 18], ideally with the desired kinetic binding rate constants [67]. The second step is the careful choice of the labeling site. Three rules have been established for this process: the residue must be 1) exposed to solvent, 2) irrelevant to the analyte/scaffold interaction, and 3) close to the binding epitope. This last step remains challenging as it is somewhat at odds with the second rule. Finally, a fluorophore is conjugated to a specific site on the scaffold using established chemical [68] or enzymatic [69] techniques. Typically, cysteines are introduced at the desired labeling position via site-directed mutagenesis and conjugated to thiol-reactive fluorophores. Unfortunately the growth in numbers of reagentless biosensors has not been commensurate with their promise, perhaps due to as-yet unexplicated limitations in the design process.

In this work we have designed over 400 reagentless biosensors against various analytes using multiple scaffolds and fluorophores, but found that successful designs with significant dynamic range (above two-fold) were quite rare, quantitatively consistent with a literature survey of reagentless biosensors. In this study, we investigated both the analyte-bound and unbound states and investigated the mechanisms accounting for bright fluorescence in the bound state and dim fluorescence in the unbound state. First, from an in-depth characterization of the strongest sensor previously reported (dynamic range 10-15-fold from bound to unbound [6]), we discovered that a specific binding pocket for the fluorophore on the analyte greatly enhanced its fluo-

rescence in the bound state. Second, we found that fluorophores conjugated to locally stable scaffold structures generate lower background in the unbound state. Combining these observations, we devised sensors with up to 100-fold dynamic range, the strongest sensors reported to date for protein analytes. These two design principles of minimizing unbound fluorescence and maximizing bound fluorescence via specific interactions could facilitate the development of such sensors to enable the study of dynamic fluctuations of proteins in living cells and tissues.

3.3 Results

3.3.1 Successfully designed reagentless biosensors are rare

We evaluated the dynamic range of 426 different binder/fluorophore/analyte combinations to establish the underlying distribution of performance for this class of reagentless biosensors. We generated numerous novel binders based on the tenth type III domain of human fibronectin (Fn3) [70] and also reproduced the published results for a previously described reagentless biosensors, the Designed Ankyrin Repeat Protein (DARPin) clone Off7 against Maltose Binding Protein (MBP) [6]. To obtain Fn3-based binders, the G4 yeast library was selected against biotinylated Betacellulin (BTC) and Epidermal Growth Factor (EGF) following standard procedures¹ [30]. The lead binders showed single-digit nanomolar affinities (Fig. 3-5A-E, Table 3.3), and the EGF binder was specific (Fig. 3-1F) and competitive with EGFR binding (Fig. 3-6). In addition, based on literature precedent, we investigated four additional Fn3s, each recognizing, MBP, the Src SH3 domain, the Abl SH2 domain, and Hen Egg Lysozyme (Fig. 3-1A). Next, we selected conjugation sites for various solvatochromic fluorophores used previously in reagentless biosensors: IANBD,

¹Additional Fn3 binders were engineered with single-digit nanomolar affinity for human Amphiregulin and Epiregulin, but were not used in this study (see Fig. 3-5 and Table 3.3)

Badan, IAEDANS, MIANs, 4-DMN-1,2,3,4 and 5, and 4-DMAP (Table 3.4 and Fig. 3-8) on each of these scaffolds. The choice of fluorophore conjugation site was intended to place the solvatochromatic fluorophore at the edge of the interface without interfering with binding. Finally, the purified constructs were tested for their fluorescence properties with and without the target analyte (Fig. 3-1B). Reagentless biosensors are evaluated based on the fold-change in fluorescence between the background (F_0) and bound state (F), where F/F_0 ratios much greater than 1 are desired. Overall, the frequency of sensors showing a change in signal (F/F_0) greater than two-fold was 4.0% (17 out of 426, Fig. 3-1C). Our results are quantitatively consistent with the success rate found in literature (Fig. 3-1C). Overall, these results clearly demonstrate that the current state of the art approach rarely generates reagentless biosensors with a dynamic range over two-fold.

3.3.2 A tailored binding pocket for NBD is responsible for the strong fluorescence activation

Given these disappointing results summarized in Fig. 3-1C, we undertook detailed structure/function studies of a previously reported successful biosensor in an attempt to elucidate the underlying molecular mechanism that differentiates a successful example from the great majority of marginally-viable constructs. We turned our attention to the NBD-labeled DARPin Off7 recognizing MBP, which has been demonstrated to exhibit a 15-fold difference in fluorescence upon MBP-binding [6]. To our knowledge, this represents the biosensor with the largest F/F_0 ratio published for a protein/protein interaction to date. We found that the Off7/MBP reagentless biosensor showed fluorescence changes only when conjugated with the fluorophore NBD (Fig. 3-2A), and not with nine other solvatochromic fluorophores, suggestive of a specific interaction between NBD and MBP as opposed to a nonspecific solvent-shielding mechanism. When labeled at positions N45 or T46, Off7 shows the greatest increase in fluores-

cence (Fig. 3-2A), in accordance with previously reported data [6]. Because these two labeling sites are adjacent, we hypothesized the fluorescence activation of NBD might involve a specific interaction with a binding pocket on MBP. To elucidate the nature of this interaction, we examined the crystal structure of the Off7/MBP complex (1SVX) and based on the dimensions of IANBD (Supplementary Fig. 3-9A) we identified 14 amino acids within likely contact distance of NBD when labeled at position N45 or T46. There is a small cavity at this site in the wild-type MBP (wtMBP) structure that could potentially serve as a binding site for NBD (Supplementary Fig. 3-9B). We constructed individual alanine mutants of MBP at each of these positions. All mutants showed similar expression levels and size exclusion chromatography profiles as the wtMBP (Fig. 3-9C, 3-9D), indicating structural integrity was maintained. We then measured the fluorescence signal of 300nM NBD-labeled Off7 in the presence of 10 μ M of each of these MBP mutants. We used the M114C sensor as a control because the fluorophore is on the opposite side of MBP; none of the mutations showed any effect on this sensor (Fig. 3-2D). For the other two labeling sites (N45, T46), we observed changes in the fluorescence intensity (Fig. 3-2D) and emission wavelength (Supplementary Fig. 3-9E) compared to wtMBP; the mutations with the greatest reduction of activation are found around a cavity in MBP (Fig. 3-2E). In order to investigate the role of this cavity in more detail, we performed a more detailed analysis of the seven residues (Fig. 3-2F) forming this pocket by mutating them to smaller and larger side chains. We observed that most mutations in those cavity-forming residues resulted in strongly reduced biosensor activity (Fig. 3-2G). Fluorescence reductions were most pronounced for mutations to tryptophan. Importantly, structural integrity (Supplementary Fig. 3-10A and B) and binding was retained (Supplementary Fig. 3-10C). We hypothesize that when a bulky tryptophan sidechain is present at residues 347, 348, or 351, it fills the cavity and thereby sterically interferes with NBD insertion into the cavity. We also identified two mutations, V347F and V347A, which enhanced the activity of the sensor, with the latter yielding a sensor with a remarkable 75-fold increase in fluorescence over background, possibly resulting from improved MBP cavity binding to the fluorophore. Together, these data strongly sug-

gest that the Off7-conjugated fluorophore NBD binds to a hydrophobic cavity in the target analyte MBP, resulting in strong activation of fluorescence.

We next sought to validate the presence of an NBD binding pocket by a library screening strategy. In an inversion of the usual binder maturation process, we matured mutants of the analyte (MBP) for increased fluorescence of the fixed Off7*NBD biosensor (Fig. 3-3A). We assembled four libraries for yeast surface display, three with mutations focused on the cavity, and one with mutations introduced over the entire gene (Supplementary Fig. 3-11A and B). These libraries showed strong display and binding to Off7 (Supplementary Fig. 3-11C). The yeast displayed libraries were incubated with 1 μ M Off7(T46C)*NBD and sorted by flow cytometry for NBD fluorescence (Fig. 3-3B). After two rounds of selections, we analyzed the libraries by deep sequencing. The majority of mutations were within the hypothesized NBD binding pocket (Fig. 3-3C, D, 3-11D, E). However, multiple mutations were also enriched in the maltose binding pocket (Fig. 3-3D, 3-11D, E). From this we selected the six most frequently observed mutations (three around the NBD cavity, three around the maltose pocket) and expressed the single point mutants solubly (Supplementary Fig. 3-11F). All mutations enhanced fluorescence activation on their own (Fig. 3-3E). Having observed several mutations near the maltose binding pocket, we investigated whether the sensor activity was affected by MBP conformational changes upon maltose binding, in the presence of a saturating concentration of maltose (20mM). The wtMBP is sensitive to maltose, which causes the sensor to lose 5-fold activity (Fig. 3-3E), suggesting the NBD pocket adopts a different configuration in the maltose bound form of MBP. Finally, we investigated how these mutations affect the interaction of MBP with the fluorophore. Using an iodide quenching assay to determine the solvent exposure of NBD, we found that the change in solvent exposure upon complex formation with wtMBP is reduced in the presence of maltose (Fig. 3-3F). However for the mutation V347A, the fluorophore is further shielded from the solvent and addition of maltose has no effect. In contrast, the activity abrogating mutation V347W results in solvent exposure levels comparable to those in the absence of the MBP analyte.

These results suggest that the NBD fluorophore binds to a conformationally sensitive hydrophobic pocket which is responsible for high reagentless biosensor fluorescence.

To determine whether other DARPin-based biosensors exhibited evidence for a specific fluorophore-binding pocket on MBP, we next sought to understand the mechanism of activated fluorescence of NBD upon MBP binding when Off7 is labeled at position M114C, which is distant from the cavity identified above. Structural analysis of the complex identified 12 amino acids within interaction distance of NBD-labeled M114C (Supplementary Fig. 3-12A, B). Alanine-scanning mutagenesis identified two residues (T193, D197) facing toward the labeling site on an α -helix as sensitive to mutation (Supplementary Fig. 3-12C). Next, we generated mutants of these two positions, as well as one position one turn N-terminal to the α -helix, into amino acids with a range of hydrophobic side chain sizes. We observed that mutation A190F improved the sensor by a factor of three (supplemental Fig. 3-12D). Furthermore, all mutations to tryptophan dramatically reduced the fluorescence activation of the sensor (Fig. 3-12D, E). These mutants were shown to retain binding to the DARPin by size exclusion chromatography and biolayer interferometry (Supplementary Fig. 3-12F, 3-12G). We attempted to mature MBP for increased sensor activity with our YSD screen, however interestingly we did not achieve any additional enhancement, in strong contrast to the Off7(T46C)*NBD biosensor. Furthermore, this sensor was not sensitive to maltose (Fig. 3-3E). These results suggest that the Off7(M114C)*NBD biosensor, with a lower F/F_0 than N45C and T46C, does not depend on a specific binding interaction between the fluorophore and a binding pocket on MBP, perhaps relying on a more nonspecific solvent shielding mechanism for fluorescence enhancement.

3.3.3 Rigid scaffolds lower background fluorescence

Two different types of reagentless biosensors are apparent among those described so far: the first (exemplified by Off7(N45C)*NBD and Off7(T46C)*NBD) is defined by the apparent presence of a specific fluorophore binding pocket in the analyte, yielding strong fluorescence enhancements. The second and far more common class, exemplified by Off7(M114C)*NBD, shows increase in fluorescence due to more nonspecific shielding, achieving a F/F_0 ratio greater than 1 in large part because the background fluorescence intensity of the unbound biosensor is low (Fig. 3-4A). We hypothesized that high background fluorescence may arise from an intramolecular interaction between the conjugated fluorophore and the binding protein. In order to investigate this interaction, we measured the background fluorescence (*i.e.* in the absence of the analyte) of NBD labeled Off7 and Fn3 with and without a fusion partner. We found that fusing Off7 to a Sumo coexpression tag resulted in a shift of the background fluorescence toward shorter wavelengths (*i.e.* to a more activated state) (Fig. 3-4B), suggesting that interactions with the fusion partner can raise the background fluorescence. When labeled with NBD, we found that Fn3s show a significantly lower emission wavelength than the Off7-based mutants (Fig. 3-4B), possibly resulting from the flexibility of the Fn3 loops, which may increase the potential for intramolecular interaction. In light of these observations, we hypothesized that undesirable fluorophore pre-activation might be attenuated by employing a small and rigid protein scaffold. We identified the scaffold Sso7d as a candidate; Sso7d is a small (~ 7 kDa) protein consisting of an incomplete β -barrel with five β -strands and a C-terminal α -helix. Recently, our group developed a yeast library based on this scaffold, using three β -strands as the binding interface. We hypothesized that the rigid conformation of the β -strands would reduce the likelihood of fluorophore interactions with the scaffold, thereby reducing unwanted background fluorescence. The Sso7d scaffold has recently been demonstrated to be applicable to reagentless biosensors[4]. We used a murine serum albumin binder, clone M11.1.3, and labeled it at seven sites of

the engineered planar binding surface, as well as three sites in the loops connecting the strands, the C-terminal, and one near N-terminal position (Supplementary Fig. 3-7B). When labeled with IANBD, we observed that the maximum emission wavelength for unbound biosensors was significantly longer than for Off7 or Fn3, directly demonstrating reduced background activation with this scaffold. Furthermore, the fluorescence intensity was also lower (Supplementary Fig. 3-7C). These promising results were further validated using another mutant M18.2.5 (Supplementary Fig. 3-7B-C).

3.3.4 Combination of rigid scaffold with a specific fluorophore

Having established Sso7d as a promising scaffold for reagentless biosensors given the lower fluorescence intensity of unbound conjugates, we sought to validate the hypothesis that specific fluorophore binding pockets enhance fluorescence signals. To do so, we screened several analytes for the presence of an NBD binding pocket. The fluorescence of NBD is unchanged upon addition of 10 μ M lysozyme or MBP. However, in the presence of 10 μ M mouse serum albumin, the maximum emission is shifted to shorter wavelengths and the intensity increases 8-fold (Supplementary Fig. 3-7D), strongly suggesting that NBD binds MSA. Based on these results, we chose the MSA-binding clone M11.1.3; when labeled with NBD, we found that positions I23C, W25C and G26C, showed greater than 50-fold increases in fluorescence upon addition of 10 μ M MSA (Fig. 3-4C). To validate this cluster, we used another MSA binding clone: in agreement with the results with M11.1.3, labeling M18.2.5 at positions 23, 25 and 26 resulted in the largest F/F_0 ratio (Supplementary Fig. 3-7E). We measured the K_D for NBD/MSA interaction to be 19.7 μ M, and found that M18.2.5(L25C)*NBD had a 30-fold lower equilibrium binding constant, $K_D = 0.38 \mu$ M (Fig. 3-4D). Importantly, titration experiments demonstrated that the affinities of M11.1.3(W25C)*NBD and M11.1.3(G26C)*NBD and M18.2.5(L25C)*NBD were at least 10-fold higher, respectively, when compared to free NBD. Furthermore, addition of non-labeled M11.1.3

or M18.2.5 binders blocked biosensor activity of the labeled binders, but not of free NBD, confirming that binding of the Sso7d-based proteins to MSA is necessary for the biosensor activity and that the binding mode of those Sso7d-NBD fusions is different from free NBD. (Supplementary Fig. 3-7F).

3.4 Discussion

Altogether, we investigated three types of scaffolds, eight analytes, and ten fluorophores in 448 combinations as components of a reagentless biosensor system (to our knowledge, the largest such dataset available to date). Other studies have been mostly limited to one class of fluorophore or one scaffold and have not elucidated general design principles. Excluding the Sso7d results, we obtained an overall, unbiased success rate similar to what has been reported previously in the literature (Fig. 3-1C). Surprised by the rarity of successful sensor design following the consensus approach, we decided to investigate the mechanistic basis for a successful example. The complexes Off7(N45C)*NBD/MBP and Off7(T46C)*NBD/MBP showed the brightest signal intensity, and the greatest F/F_0 ratio previously reported. Also, the emission wavelength of NBD was strongly blue-shifted upon complex formation, indicating a reduced polarity in the binding pocket of the fluorophore [71].

Several factors can be responsible for fluorescence blue-shift and increased intensity upon complex formation of NBD. The important feature characterizing the fluorescence properties of NBD is the charge transfer occurring between the amino group (electron donor) and the nitro group (electron acceptor). Excitation by a photon of the appropriate energy accentuates the dipole between the acceptor and donor groups. Stabilization of the dipole will reduce the energy of the excited state and thus, the energy of the released photon [32]. Stabilization can arise from multiple factors. 1) Polar solvents stabilize the charged configuration by dipole-dipole interactions,

which lowers the energy of the excited state and increases the non radiative decay (k_{nr}) [25]. 2) H-bonding between the solvent and an oxygen of the nitro group, which would reinforce the electronic attraction of the acceptor [25]. 3) Dynamic relaxation with internal twisting of the dimethylamino group, coupled with an electron transfer. Twisted intramolecular charge transfer (TICT) is the result of an intramolecular rotation where the acceptor and donor are orbitally decoupled. TICT states are non luminescent [23]. Both static approaches, by rigidifying the amino group via covalent bonds, and dynamic approaches, by increasing the volume of the groups attached to the nitrogen, have supported this theory [72]. Decrease in the k_{nr} with increased solvent viscosity and lower temperatures also support the involvement of a TICT dependent fluorescence [25]. Hence, our data supports the hypothesis of a specific interaction of NBD with a hydrophobic cavity - blue-shifted emission shift - which restricts the motion of the fluorophore and prevents it from accessing a TICT state, and thus exhibits a greater quantum yield.

We have further confirmed the importance of this binding pocket by identifying sensor enhancing mutations using a yeast surface display screen (Fig. 3-3), evolving MBP to become a better analyte for the Off7(T46C)*NBD sensor. We found that the majority of mutations accumulated within the NBD binding pocket. We also observed that mutations along the alpha helix mutated between amino acids 150-158, constituting part of the NBD binding pocket, were intolerant to mutations. These residues may be critical for the interaction with NBD. Furthermore, Glu 153, Tyr 155 and Tyr 156 make van der Waals contact with maltose [73] and may be important for the cavity structural integrity. Szent-Gyorgyi *et al* have used yeast surface display to isolate single chain variable fragments binding to and activating the fluorescence of malachite green [74] in a mechanism also involving TICT [23, 75]. Thus, the fluorescence of both NBD and Malachite Green can be strongly activated upon the presence of a specific binding pocket.

We also found that maltose acts as an allosteric regulator of the sensor activity

(Fig. 3-3E). The binding site for maltose is located at the cleft between the two domains [73]. Upon maltose binding by wtMBP, the activity of the Off7(T46C)*NBD sensor is affected by a 5-fold factor. Through our yeast surface display analyte maturation screen, we identified a multitude of mutations located in domain I or at the hinge between domain I and II. These mutations are not involved in direct contacts with maltose [73], yet they may affect MBP conformation or dynamics. It is known that in the analyte-free form, MBP exists 95% in an open-state, and 5% in a partially closed state [76]. The transition between major (open) and minor forms of apo MBP involves a hinge rotation of $33.3 \pm 6.7^\circ$ in comparison with 35.2° between open and closed holo MBP. However, the apo minor and closed holo states of MBP are distinct and related by a domain reorientation of 18° accompanied by a 6 Å translation [76]. The mutations we observed may affect this equilibrium away from the partially closed state, or a distinct conformation that favors interaction of NBD with its pocket. This is supported by our observation of activity enhancing mutations and that the NBD interaction with the cavity is impeded in the maltose-bound form of MBP. Finally, NBD has been employed as a maltose sensor when labeled to MBP, and TICT has been hypothesized as the mechanism for the slight 1.8-fold increase in fluorescence [77].

While we identified specific binding pockets as a mechanism for very large activation of NBD, fortunately, the presence of a binding pocket is not an absolute requirement for useful activity. M114C undergoes a five-fold increase in fluorescence with only a minor blue-shift (2nm) without a hydrophobic pocket; the polarity of the environment of NBD is unaffected. Our epitope analysis did not identify the presence of an NBD-binding pocket on MBP for this biosensor and we thus speculate that its fluorescence intensity increase is mostly due to motion restriction, leading to an increase in fluorescence intensity by increasing the energy of the TICT state.

Our data highlight the importance of the fluorophore interactions with the analyte for fluorescence activation. The nature of these interactions are also important, as

exemplified in the literature where NBD-dependent binders were selected and yet showed little responsivity to analyte [78]. The involvement of binding pockets has been exploited previously with NBD derivatives binding to trypsinogen and ribosome subunit 50S as well as Nile Red binding to several other large proteins [79, 80]. Binding sites for small molecules are not uncommon and there are strategies to identify them [81, 82, 83, 84]. However, designing binders whose epitopes lie sufficiently near the fluorophore binding pocket further complicate the process, and thus we sought to establish a more generalizable approach to designing reagentless biosensors.

Next, we analyzed the fluorescence intensity and emission wavelength of three Off7 sensors in the presence of all MBP mutants designed in this study (Fig. 3-4A). At the far right, we can navigate through increases in fluorescence without changes in the emission wavelength, resulting from shielding of the fluorophore by MBP: the sensor Off7(M114C)*NBD relies on this mechanism. In the middle range, there is a stronger relationship between emission wavelength and intensity, resulting from an interaction with a fluorophore binding pocket: sensors Off7(N45C)*NBD and Off7(T46C)*NBD rely on this mechanism.

A generic solution to developing reagentless biosensors is to start with very low fluorescence activation in the background state. Prevention of intramolecular interactions in the unbound biosensor plays a major yet previously unappreciated role in sensor performance. The interactions between fluorophores and scaffold are difficult to predict and we did not identify trends relating emission wavelengths and properties of surrounding amino acid side changes amongst the various constructs investigated. The presence of a fusion partner such as SUMO to simplify expression led to higher background fluorescence. It is noteworthy that peptide-based reagentless biosensors have been found to exhibit both very low background fluorescence and high dynamic range, most likely due to the absence of strong intramolecular fluorophore interactions in the unbound state [36, 27, 26].

We further hypothesized that structural rigidity of DARPin over Fn3 scaffolds may be a contributing factor to the success of this Off7/MBP system. In order to test this hypothesis, we used the hyperthermostable Sso7d scaffold whose binding paratope is constituted by three beta-strands. For two Sso7d derived mutants (M11.1.3 and M18.2.5) labeled at various positions on and around the engineered binding epitope, we observed that the fluorescence emission maxima was significantly longer than that of Fn3 or Off7, resembling more the emission of the non-conjugated fluorophore (Fig. 3-4B). The fluorescence intensity for these constructs was also lower (Fig. 3-10C). Our results support the hypothesis that local conformational rigidity can reduce fluorophore pre-activation due to intramolecular interactions. Based on the low background fluorescence of NBD labeled Sso7d, we developed sensors for MSA as a proof of concept, and obtained up to a 100-fold increase in fluorescence on analyte binding. This represents the greatest fold increase in fluorescence ever reported for a protein-protein reagentless biosensor. When evaluating the best sensor from a DARPin, Fn3 and Sso7d scaffold (Fig 4E), it stands out that the success of Sso7d as a scaffold for reagentless biosensor can be attributed to its reduced background fluorescence.

We have demonstrated that combining low background intensity with enhanced activated state, a reagentless biosensor can be built with greater than 100-fold increase in fluorescence upon analyte binding. Reagentless biosensor design fall within three categories (Fig. 3-4F), two of which can lead to large F/F_0 ratios. In one scenario, the successful design is driven by achieving very strong fluorescence intensity mediated by a fluorophore specific binding pocket, as we have demonstrated for Off7(T46C)*NBD. A panel of fluorophores could be screened against an analyte of interest to identify the presence of a binding pocket through methods such as fluorescence, NMR or computational docking. Once a fluorophore is selected, the development of several binders is necessary to target the correct epitope. This medicinal-chemistry like fragment-based discovery approach is intensive in time and resources. On the other hand, one can build sensors with less large F/F_0 ratios (~ 10 -fold for NBD), drawing

on the second scenario of lower unbound fluorescence rather than dramatically higher bound fluorescence. The use of rigid scaffolds like Sso7d with lower background can significantly increase the odds of design success. Combination of both scenarios yields, as we have demonstrated, sensors with up to 100-fold increase in fluorescence. Beyond protein-protein interactions, these guidelines could help in the design of any improved sensor using solvatochromic fluorophores.

3.5 Future Perspectives

Several challenges lie ahead for the development of reagentless biosensors that can both show dramatic fluorescence increases in fluorescence upon analyte binding, and also be useful in *in vivo* applications. One drawback of the various fluorophores investigated in this study is their short emission wavelengths. For *in vivo* applications, fluorophores with longer emission wavelengths (>600 nm) demonstrate lower background noise. To achieve this goal, two approaches can be undertaken: the first one would involve the coupling of a second fluorophore that would act as a FRET acceptor and thus allow the emission of a longer wavelength photon [85]. The second approach would be to directly use solvatochromic fluorophores that have long emission wavelengths, such as merocyanine derivatives [5, 17].

In order to avoid artifacts from sensor localization, it is important to use ratio-metric sensors: the sensor is coupled to a second fluorophore of constant fluorescence properties, the ratio between the two fluorophores is thus used to accurately determine the fluctuation in analyte concentration.

In vivo low emission wavelengths can lead to poor signal to noise ratio and fluctuations in sensor localization can mislead the interpretation of analyte concentration fluctuations. To address these potential drawbacks, reagentless biosensors can be

coupled to microspheres, loaded with an infra-red marker dye. The presence of the dye allows ratiometric measurements and multiplexing. In addition, the localized concentration of sensors can improved the signal to noise ratio. These guidelines will help the implementation of reagentless biosensors for *in vivo* applications.

3.6 Material & Methods

3.6.1 Fibronectin engineering by yeast surface display

We choose the human tenth Fibronectin III domain as a scaffold to engineer binders as it has been a well validated protein engineering scaffold and it is free of cysteines [86, 87]. Our group has published a detailed protocol for fibronectin engineering by yeast surface display[30]. Briefly, binders were isolated by magnetic bead sorts from the G4 library [70]. Initial clones were then affinity matured by error-prone polymerase chain reaction (PCR) and selected by fluorescence activated cell sorting (FACS). Biotinylated analyte was detected with either Streptavidin Alexa Fluor 647 (Invitrogen) or Anti-Biotin antibody (eBiosciences) in alternation. For determination of K_D values lead clones were titrated on the surface of yeasts, and data were fitted to a monovalent binding isotherm.

3.6.2 Protein expression and purification

We found that cysteine containing fibronectins were poorly soluble as opposed to their cysteine-free counterparts. In order to increase yield and solubility, crucial for subsequent fluorophore conjugation, we chose to express fibronectins fused to a N-terminal hexahistidine tag and small ubiquitylation modified (SUMO) tag. Thus,

all protein encoding genes were cloned in a pE-sumo expression vector (LifeSensors) by restriction digest using the XbaI and BsaI sites. The sequence confirmed plasmid were transformed into Rosetta 2 (DE3) cells (Millipore) or XL1-Blue (Agilent) and grown overnight in 5mL LB media + Kanamycin (50 μ g/ml) at 37°C. Subsequently, the cultures were diluted 200-fold in Terrific Broth + kanamycin. When cells had reached an OD600 = 1.0-2.0, expression was induced by addition of 500 μ M Isopropyl β -D-1-thiogalactopyranoside IPTG (AMRESCO). After overnight expression at 20°C, cultures were centrifuged and the pellets were resuspended in 50 mM Sodium Phosphate (AMRESCO), 300 mM Sodium Chloride (AMRESCO), 1% Glycerol (AMRESCO), 3% Triton-X-100 (AMRESCO) at pH 8.0. Cells were sonicated and centrifuged, followed by filter sterilization of the supernatant and purification on a Ni-NTA resin (QIAGEN). Purification fractions were ran on a 4-12% bis-tris acrylamide gel to confirm expression and purity (Life Technologies). The elution fractions were dialyzed into PBS using Snakeskin dialysis tubing (Thermo Scientific), samples were concentrated to >100 μ M with an amicon 3K concentrator, flash frozen in liquid nitrogen, and stored at 80°C.

DARPinS were either expressed as N-terminal His6-Sumo fusions or as N-terminal His6 only fusions on a pQE vector (gift from the Plückthun laboratory). Expression and purification methods were the same as described above.

Sso7d proteins were expressed with an N-terminal histidine tag in Rosetta 2 (DE3) cells on a custom pE based vector (GH₆GG-Sso7d, see Appendix B). Proteins were purified under denaturing conditions in the presence of 2mM β -mercaptoethanol. Following labeling, the proteins were purified by cation exchange in 20 mM Tris pH 8.0 with a gradient from 0 to 1 M NaCl on a SP FF column (GE Healthcare).

All Maltose Binding Protein mutants were expressed with an N-terminal His6 tag (pQE vector, gift from Plückthun laboratory). Abl-SH2 and Src-SH3 were also cloned into the pE-sumo expression vector, expressed and purified like other sumo

fusion proteins in this study. Chicken egg lysozyme was purchased from Sigma and dissolved in PBS. For yeast surface display screenings, we purchased Amphiregulin, Betacellulin, Epidermal Growth Factor and Epiregulin from Peprotech. For further characterization and fluorescence assays, BTC and EGF were expressed as His6-Sumo N-terminal fusion as described previously[88]. Fatty acid free mouse serum albumin was purchased from Alpha Diagnostic International.

3.6.3 Protein Labeling

For labeling, the protein sample was diluted to a final concentration of 100 μ M in PBS and the reducing agent (tris(2-carboxyethyl)phosphine (TCEP) was added to a final concentration of 1mM. After 60 minute incubation at room temperature for thiol reduction, the thiol reactive fluorophore was added to a final concentration of 1 mM and the sample was reacted overnight on a rotating wheel at room temperature. Free fluorophore was removed on a G-25 desalting resin (GE Healthcare). Labeling efficiencies were in the range of 0.4 - 0.8 fluorophore/protein molar ratio, as derived from the A280 and fluorophore absorbance [6]. IANBD, IAEDANS, MIANS, Badan were purchased (Life Technologies), 4-DMN and 4-DMAP were synthesized by Elke Socher and Stephanie Cheung [28].

3.6.4 Protein biotinylation

Samples were biotinylated with Sulfo-NHS-LC-Biotin (Thermo) in PBS for 1h at room temperature and quenched with Tris buffer at pH 8.0. Excess biotin was removed on a zeba spin 7K desalting column. Extent of biotinylation was confirmed by MALDI-TOF mass spectrometry.

3.6.5 Size exclusion chromatography

All size exclusion chromatography was performed on an AKTA system (GE Healthcare). Analytical runs were performed in PBS at pH 7.4 on a Superdex 75 10/300 GL column at 0.75 mL/min. Preparative runs were performed in PBS on a Superdex 75 16/60 column at 1 mL/min.

3.6.6 Fluorescence Measurements

Spectroscopy studies were performed using a FluoroMax-2 Fluorometer with an integration time of 0.100 s, excitation and emission slits set at 5 and 10 nm respectively. All excitation wavelengths and emission scans are listed in Table 3.1. Stern Volmer constants from the iodide quenching experiment were extract from titration with 0-200 mM potassium iodide as described previously [89].

3.6.7 Biolayer Interferometry

Experiments were performed on a Blitz or an Octet RED instrument in PBS, 1 mg/ml bovine serum albumin, 0.002% v/v Tween-20. Equilibrium dissociation constant and kinetic binding rates were extracted from a global fit.

3.6.8 YSD for enhanced MBP mutant isolation

The gene for Maltose Binding Protein was cloned between the NheI and BamHI restriction sites in the pCTcon2 vector for display as an N-terminal fusion to Aga2p. To generate the MBP 1.0 library, mutations were introduced in the MBP gene as

described for fibronectin domains [30]. For the cavity targeted libraries, regions covering amino acids 150-158 and 344-352 were amplified with primers allowing yeast homologous recombination. The error prone PCR condition were similar to those of fibronectin loops [30]. The other fragments were amplified without error-prone PCR. The gene fragments were combined and transformed into EBY100 yeasts, allowing the yeasts to perform homologous recombination. Two NBD conjugated Off7 mutants (T46C and M114C) were added at 1 μ M to each library independently and sorted by flow cytometry. The fluorescence of NBD was detected under the FITC channel. Two rounds of sorting were performed to obtain a clean polyclonal population.

3.6.9 Sequence Data analysis

177 sequences were obtained by sanger sequencing. MiSeq yielded 8.4 million past filter reads (see table 3.2). All sequences from the MBP display screen were analyzed by a user-developed MatLab script, allowing the extraction of several parameters such as amino acid frequency. This code is available upon request.

3.6.10 Flow cytometry

Samples were analyzed on an Accuri C6 cytometer (Becton Dickinson, BD), 96- well plates were run on a FACSCalibur HTS with a high throughput plate sampler (BD), and sorting was performed on a MoFlo instrument (Beckman Coulter) or a FACS Aria instrument (BD). Protein display was detected using anti epitope tag antibodies, either mouse anti HA (clone 16B12, Covance), chicken anti c-myc (Gallus Immunotech) or mouse anti c-myc (clone 9E10, Covance). Secondary labeling was performed using goat anti mouse or chicken antibody conjugated to AlexaFluor488 or 647 (Life Technologies). Biotinylated analytes were detected with Streptavidin Alexa Fluor 647 or

488 (Life technologies) or anti-biotin-PE antibody (eBiosciences).

3.6.11 EGF epitope mapping

To identify the binding epitope of Fibronectin clone EGF-A, we chose an epitope mapping approach. Human EGF was fused to Aga2p with a $(G_4S)_3$ linker. Of the 53 amino acids in the protein, we mutated all 36 solvent exposed, non cysteine, proline or glycine residues to an amino acid with opposite side chain properties [90]. Binding of biotinylated fibronectin EGF 5.1.07 to yeast displayed EGF-mutants was tested at 10 nM and detected with Streptavidin Alexa 647.

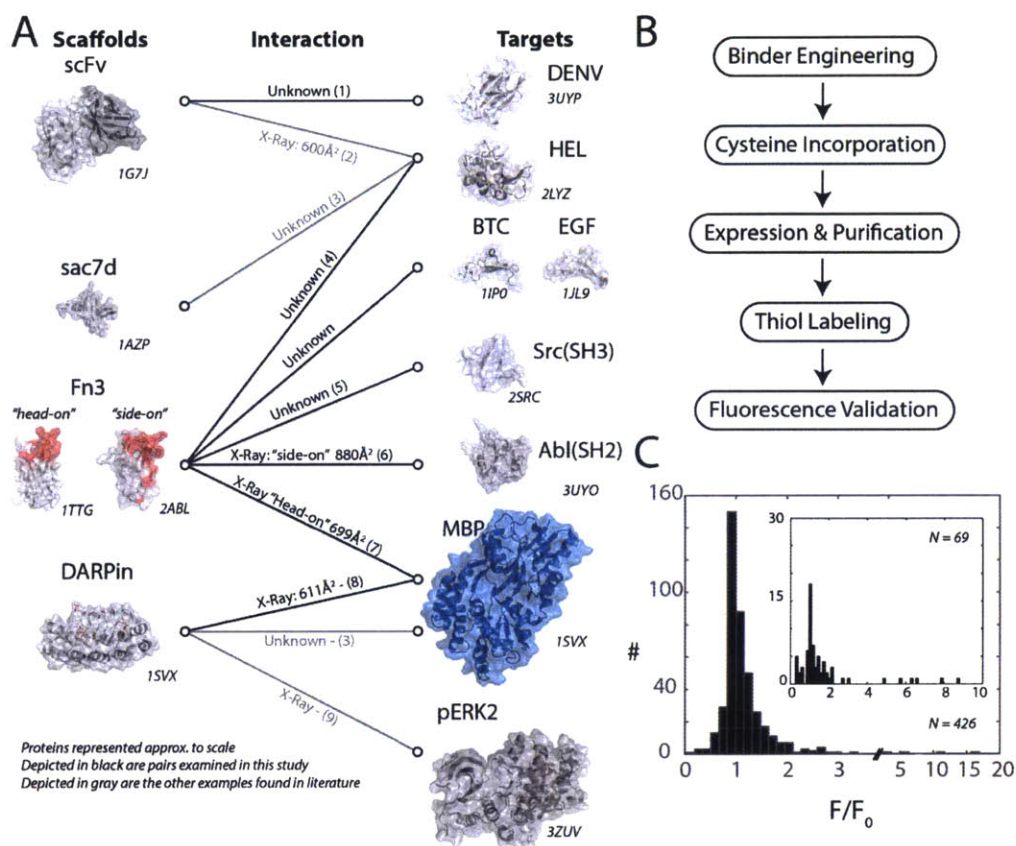


Figure 3-1: Reagentless biosensor workflow and overview. **A)** Exhaustive overview of all reagentless biosensors built from a protein scaffold recognizing another protein (as of April 2015). On the left side the various protein scaffolds that have been used and coupled with an environment sensitive dye are listed. On the right side the target proteins are shown. Dark arrows represent pairs evaluated in this study. The connecting lines indicate the information about the nature of the interaction as well as the reference citation: 1-9 = [91, 3, 4, 87, 5, 92, 86, 6, 17]. **B)** Reagentless biosensor engineering is a five step process. First, a binder must be engineered with desired specificity and affinity. Second, a chemical moiety for site-specific labeling must be incorporated. This is typically achieved by directed mutagenesis for a cysteine at the desired location. Next, the protein construct is expressed, purified and labeled with a thiol reactive derivative of the fluorescent dye. Finally the construct is evaluated for its fluorescence characteristic in the absence and presence of its target. **C)** Histogram of the sensitivities of sensors to their target reported as the fluorescence intensity ratio between the conditions in presence versus in absence of the target protein. The histogram showed in the inlet is from previously published results by other groups [3, 4, 91, 5].

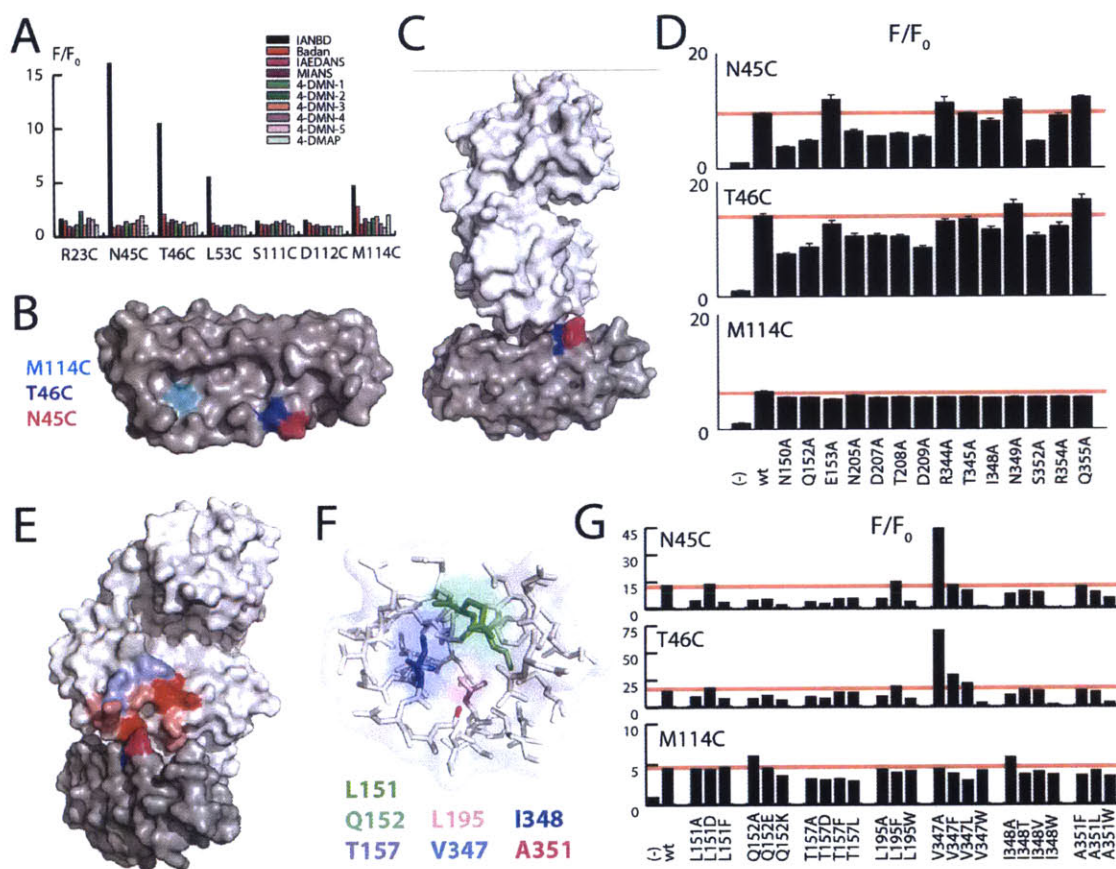


Figure 3-2: Mapping the interaction between NBD and MBP in the Off7(N45C)*NBD and Off7(T46C)*NBD sensors. **A)** Increase in fluorescence signal when adding MBP to Off7 mutants labeled with ten different fluorophores at seven positions. **B)** Top view of the Off7 protein (PDB: 1SVX). **C)** Side view of the Off7/MBP complex (PDB: 1SVX). **D)** F/F_0 ratios of three Off7*NBD sensors with MBPwt and mutants with a single alanine mutation. **E)** Effects from alanine scanning are plotted against the Off7/MBP structure, red indicates reduced activity, blue increased activity. Deleterious mutations are surrounding a cavity in MBP. **F)** Residues mutated in the cavity analysis. **G)** F/F_0 ratios for three Off7*NBD sensors with MBP cavity mutants.

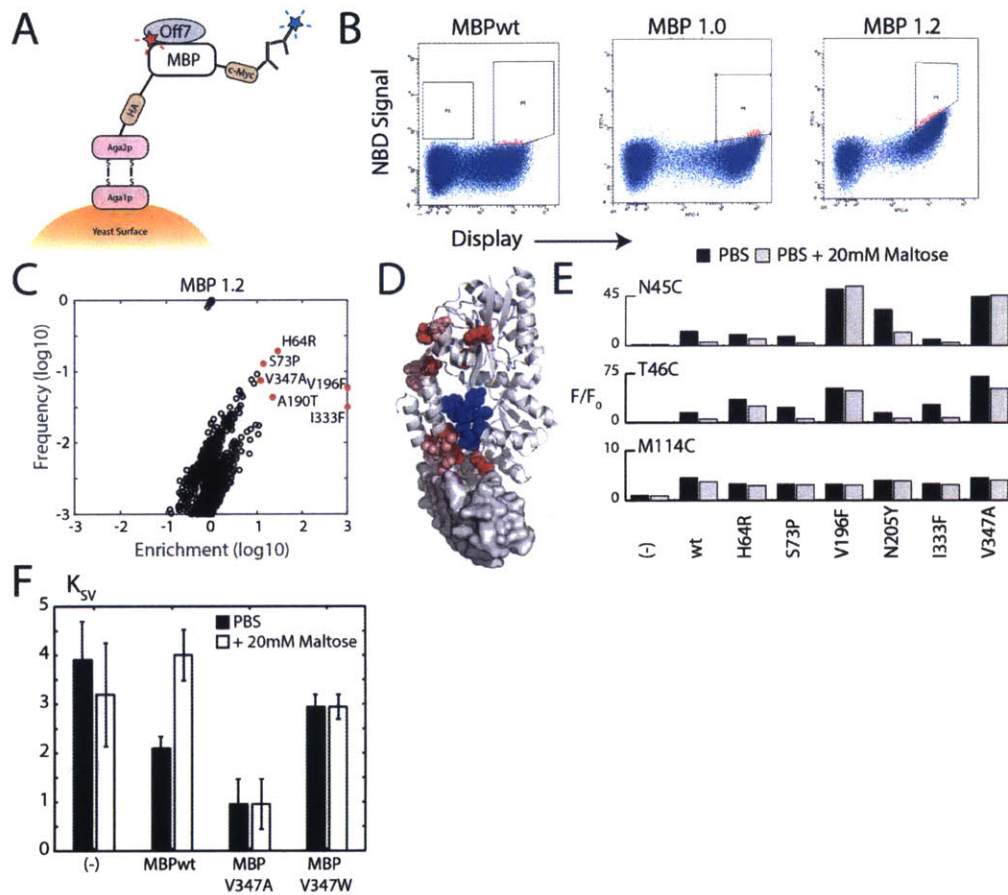


Figure 3-3: Off7(T46C)*NBD Sensor is regulated by MBP conformation. **A)** Graphical representation of experimental set-up. **B)** Example of library improvement. **C)** Enrichment versus frequency plot for the whole gene error prone PCR library. **D)** Graphical representation of enrichment onto Off7/MBP structure combined from all four libraries. Red and blue indicate positive and negative enrichment respectively. **E)** F/F_0 ratios for the YSD isolated mutants in the absence or presence of maltose. Mutations H64R, V196F and V347A greatly stabilized the activity of the sensor in the presence of maltose. **F)** Iodide quenching assay, reporting the Stern-Vollmer constant for wild-type MBP, the most strongly activating mutation V347A, and the most deteriorating mutation V347W.

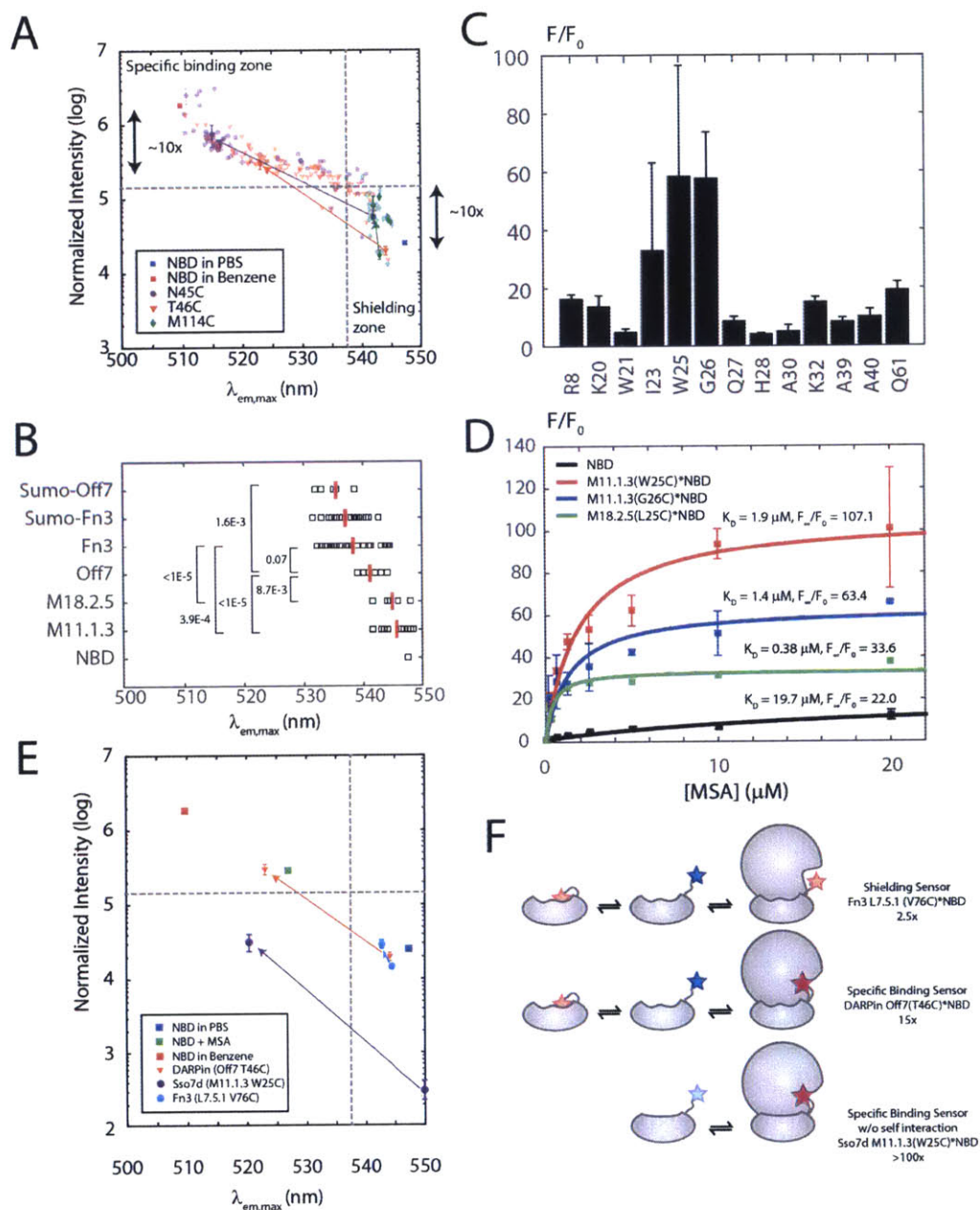


Figure 3-4: The balance between scaffold rigidity and specific binding pockets. **A)** Maximum emission wavelength versus signal intensity plotted for the three Off7 sensors in complex with all MBP mutants evaluated in this study. In addition, free NBD in PBS and benzene is depicted. **B)** Maximum emission wavelength of NBD labeled scaffolds in the absence of antigen. **C)** F/F_0 ratio for M11.1.3*NBD conjugates. **D)** Titrations of NBD and NBD labeled constructs with MSA. **E)** Maximum emission wavelength vs intensity for the best Fn3, DARPin, and Sso7d sensors. **F)** Mechanistic drawing of the three types of reagentless biosensors. The top one describes a sensors that suffers.

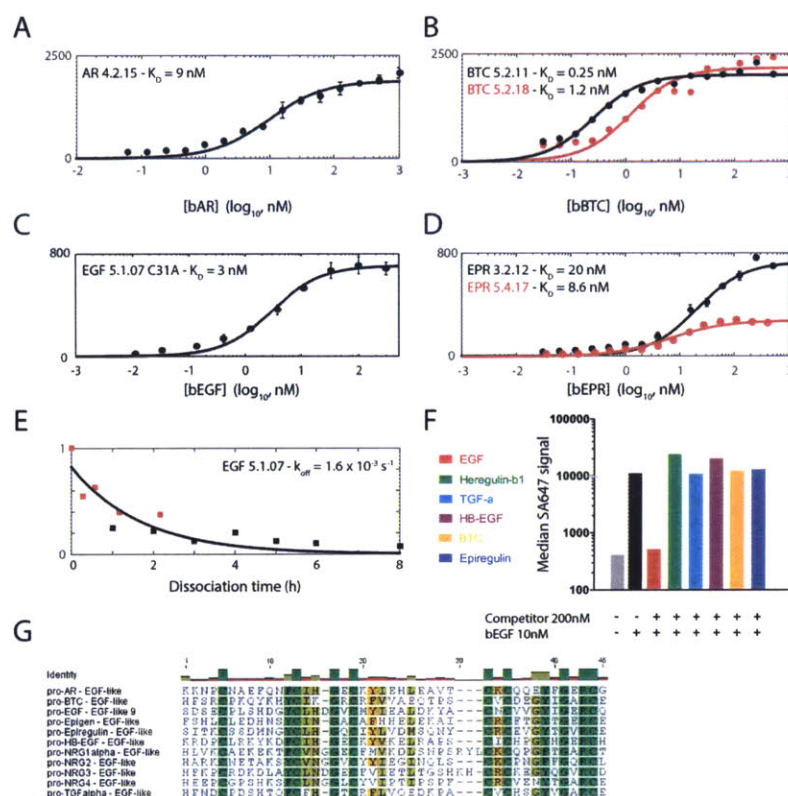


Figure 3-5: Characterization of ErbB ligand binders: **A-D)** Equilibrium titrations of yeast surface displayed fibronectins against their intended analyte. Equilibrium dissociation constant (K_D) are reported from a monovalent binding isotherm. **E** Dissociation constant measurement on yeast from two replicates (red and black squares). **F)** Specificity assay. Yeast cells displaying EGF 5.1.07 C31A were first incubated with 200 nM of ligand for 1h before adding 10 nM of biotinylated EGF. **G)** Sequence alignment of the EGF domain among the various ErbB ligands.

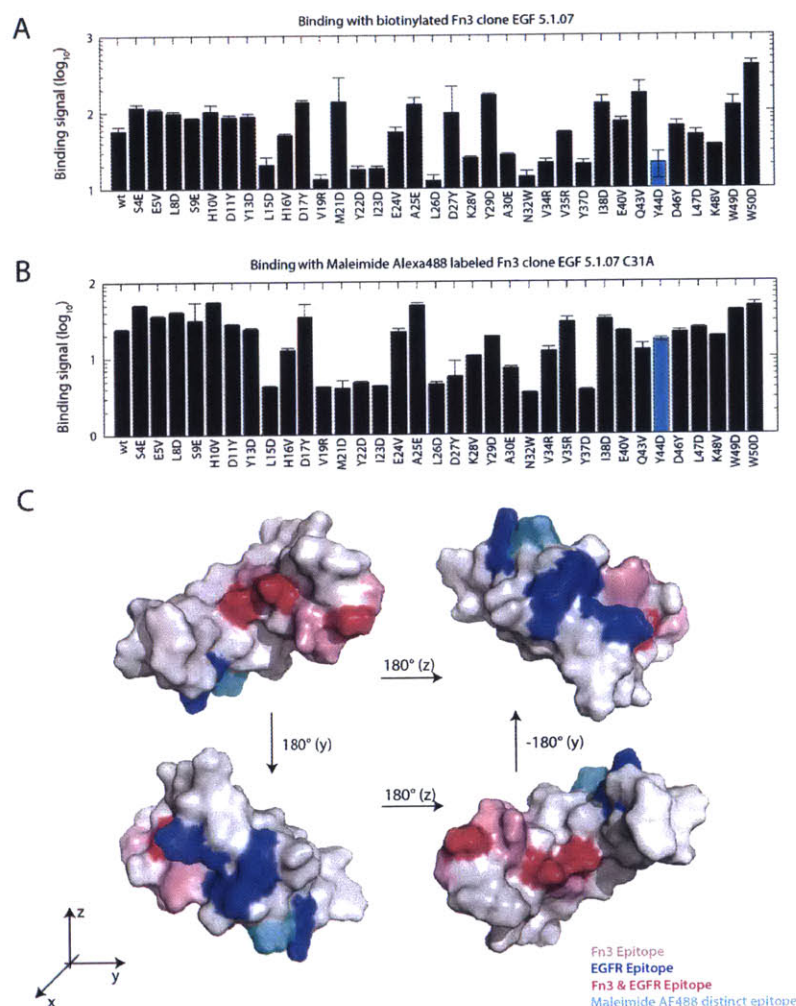


Figure 3-6: Epitope mapping of Fn3 EGF 5.1.07 on hEGF. A) Binding signal for EGF mutants displayed on yeasts detected with biotinylated Fn3 EGF 5.1.07. **B)** Binding signal for EGF mutants displayed on yeasts detected with maleimide-alexa488 labeled EGF 5.1.07. **C)** Representation of the residues important for binding, defined as a >80% loss in binding signal. In cyan is depicted a position that showed distinct effect when test with biotinylated Fn3 as opposed to maleimide-Alexa488 labeled at position C31S. This might suggest the orientation of the Fn3 in respect to EGF. Our EGF binder recognizes an epitope constituted by the two beta-strands in the protein. This region is likely more antigenic due to its structural order as opposed to the rest of the protein. Antibodies against EGF [93] have also been shown to bind to the same epitope [94], which is also shared with EGFR [95, 96].

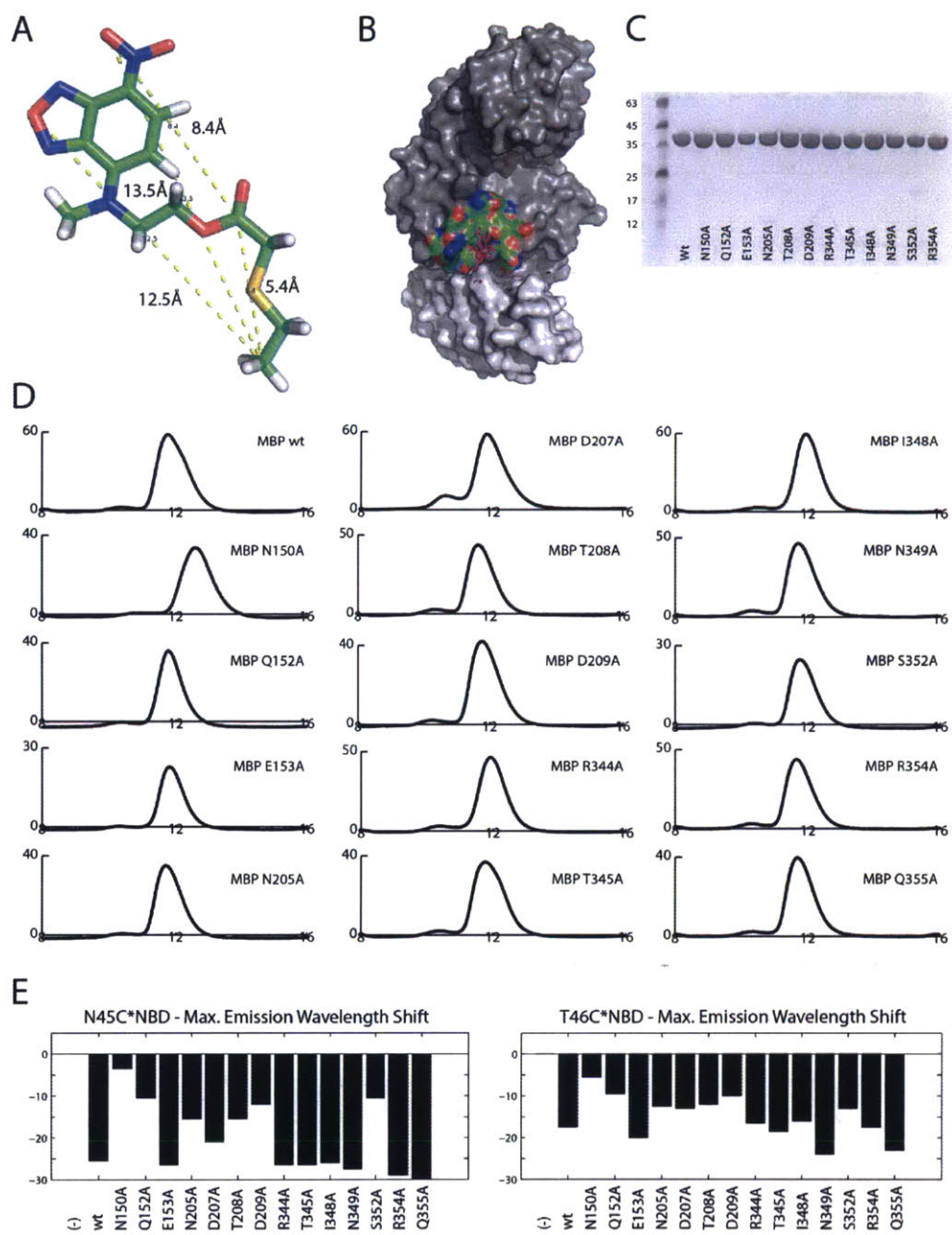


Figure 3-7: Chemical structures of the solvatochromic fluorophores used in this study.

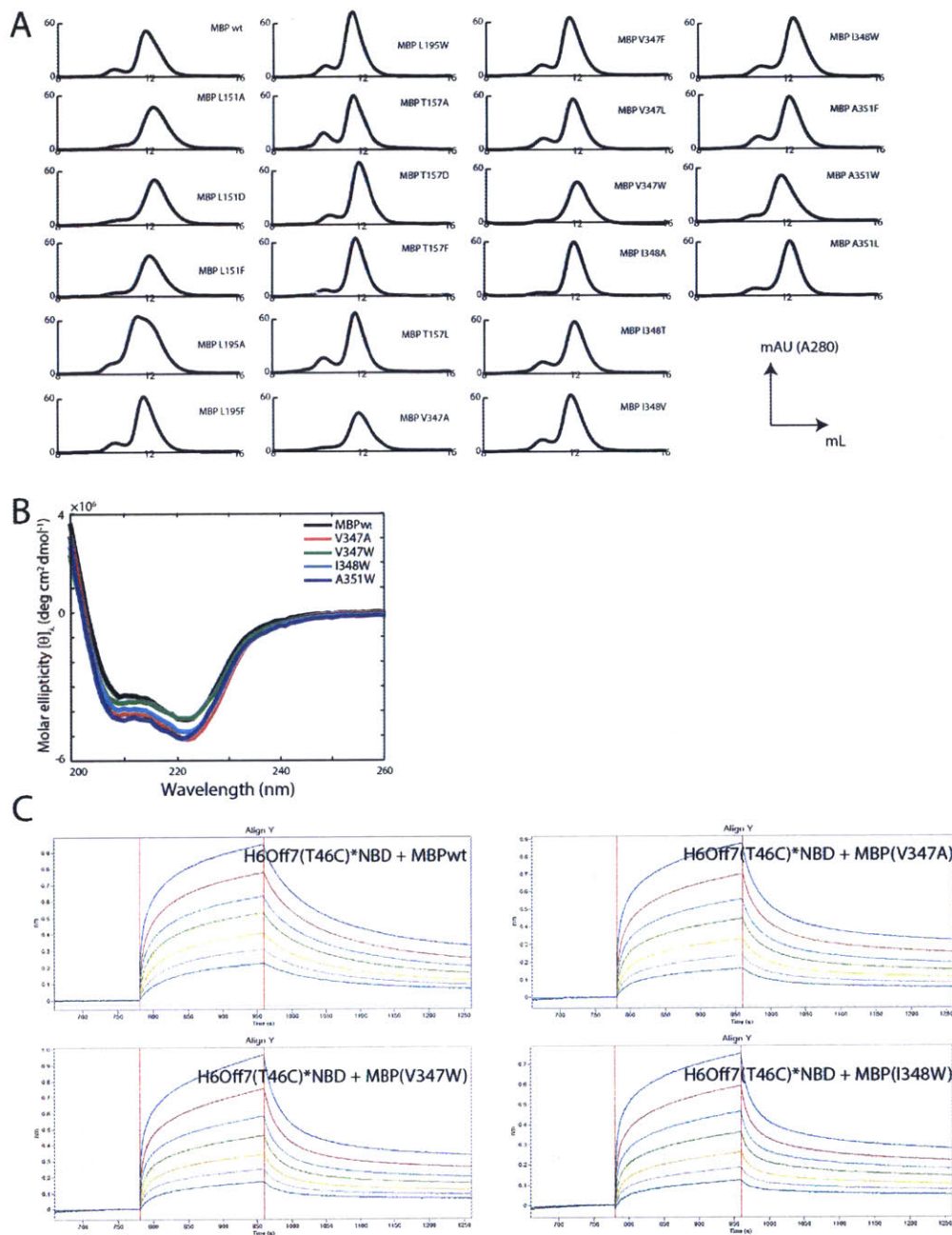


Figure 3-8: Epitope mapping of Off7(N45C)*NBD and Off7(T46C)*NBD on MBP. **A)** Structure of IANBD once labeled to a cysteine residue. Distances are measured from the alpha carbon. **B)** Residues within a 14 Å distance range from N45C/T46C sites. **C)** SDS-PAGE and **D)** size exclusion profiles for the expressed MBP mutants. **E)** Maximum wavelengths emission shift (nm) upon MBP addition.

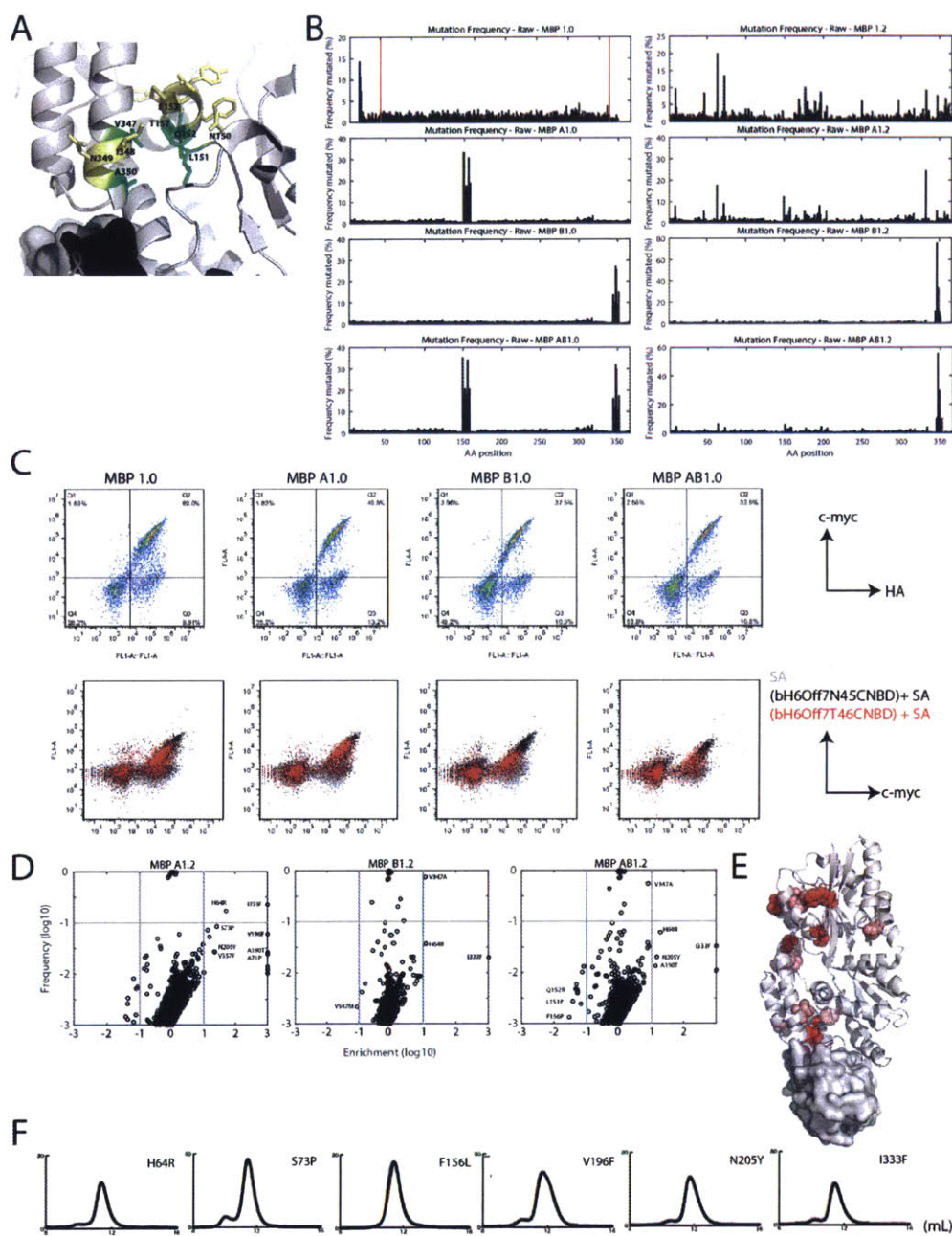


Figure 3-9: **Characterization of MBP cavity mutants.** **A)** Size exclusion profiles of MBP cavity mutants. **B)** Circular dichroism. **C)** Biolayer interferometry signals of biotinylated Off7(T46C)*NBD with 2, 1, 0.5, 0.25, 0.125, 0.61, 0.31 μ M MBP

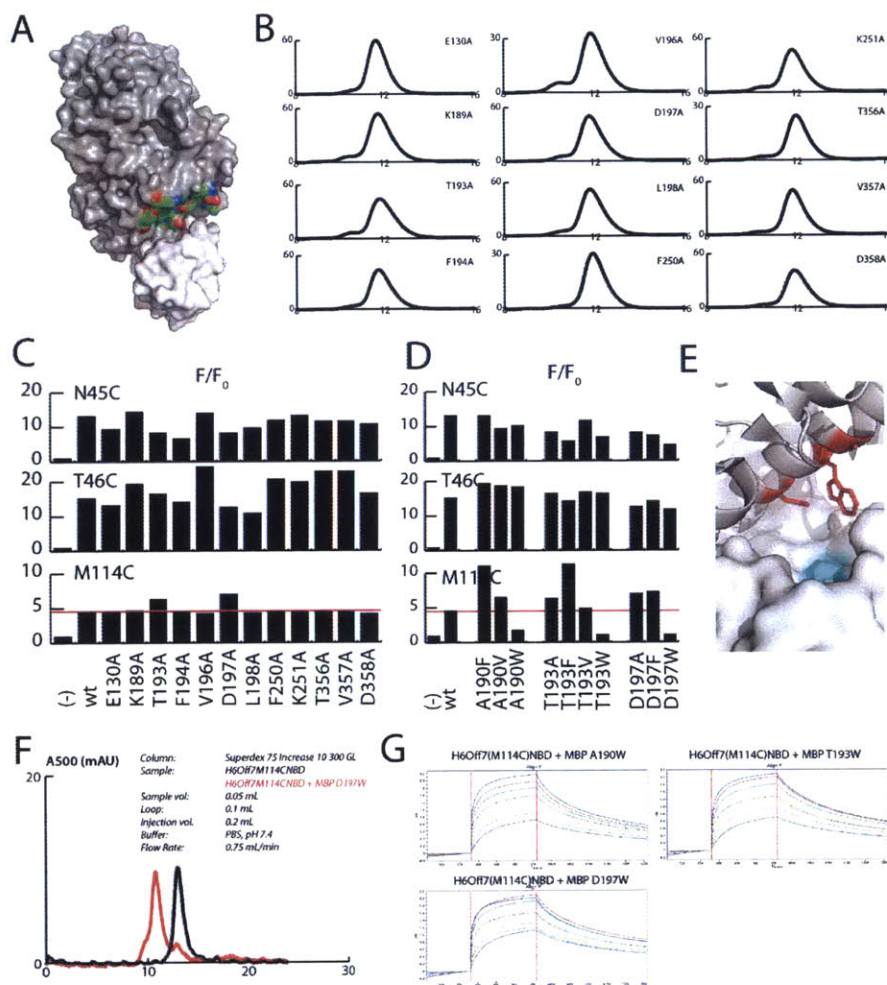


Figure 3-10: MBP maturation for sensor by YSD. **A)** Residues constituting the NBD binding pocket. **B)** Mutation frequencies at each position on MBP (amino acid 1-366) derived from MiSeq. Plots on the left are pre-sorting, plots on the right are post-sorting. Library 1.0 was designed by whole gene error-prone PCR. Library A1.0 was heavily mutated in the amino acid positions 150-158. Library B1.0 was heavily mutated in positions 345-352. Library AB1.0 was heavily mutated in both region A and B. **C)** Display (top) and binding (bottom) signal for the pre-sorted libraries. **D)** Enrichment versus frequency plots for the focused A, B and AB libraries. **E)** Most enriched sites plotted on the MBP structure. **F)** Size exclusion profiles of YSD isolated MBP mutants.

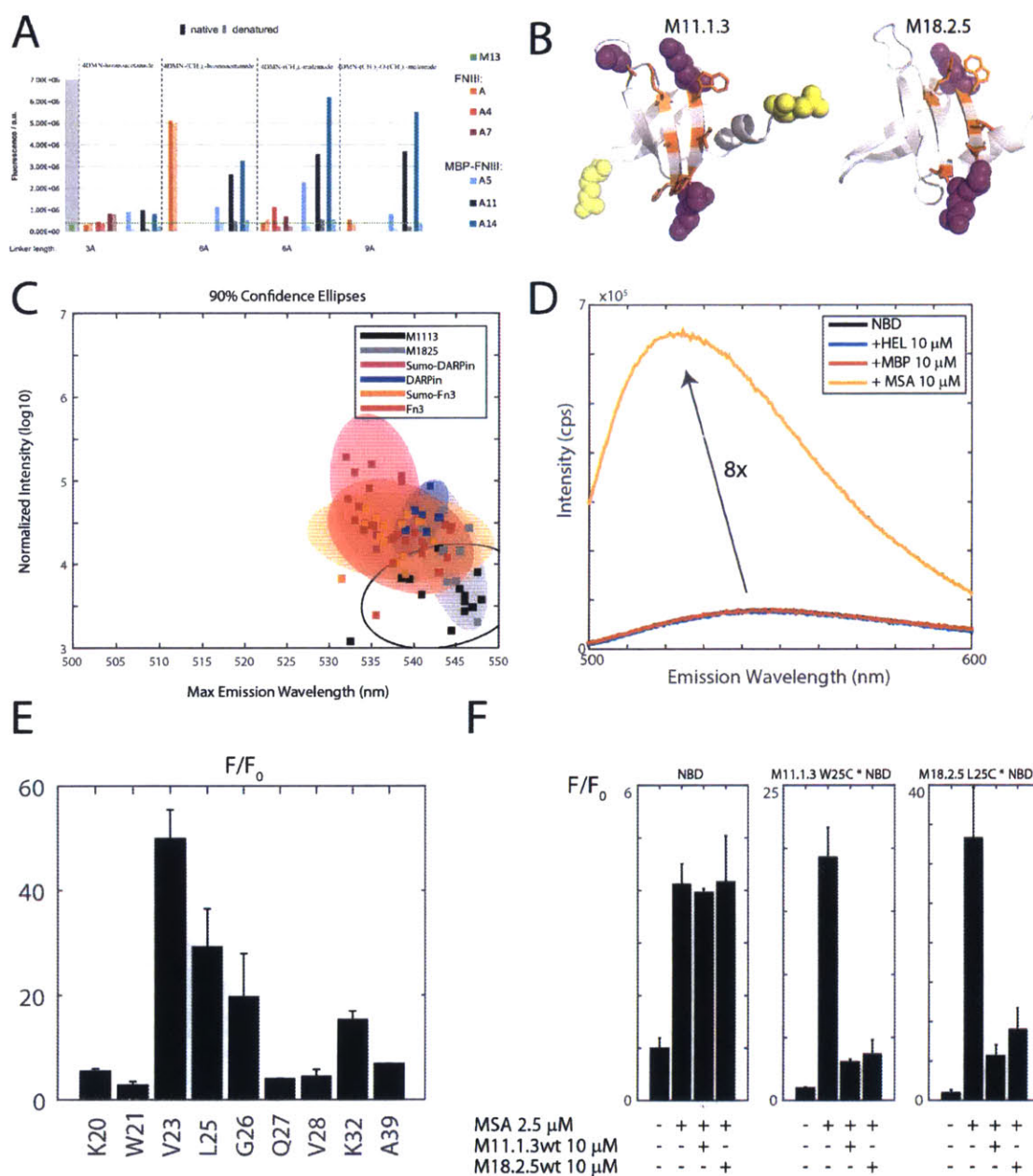


Figure 3-11: **Epitope mapping of Off7(M114C)*NBD on MBP.** **A)** Residues within 14A of M114C labeling site. **B)** Size exclusion profile of alanine mutants. **C)** F/F_0 ratio for alanine mutants. **D)** F/F_0 ratio for mutation along the alpha helix facing toward M114C labeling site. **E)** Representation of a possible T193W rotamer, suggesting sterical hindrance with the fluorophore NBD. **F)** Size exclusion profile of Off7(M114C)*NBD without and with MBP. The NBD signal can be seen to elute sooner in the presence of MBP, which indicates complex formation. **G)** Biolayer interferometry signals of biotinylated Off7(T46C)*NBD with 2, 1, 0.5, 0.25, 0.125, 0.61, 0.31 μ M MBP.

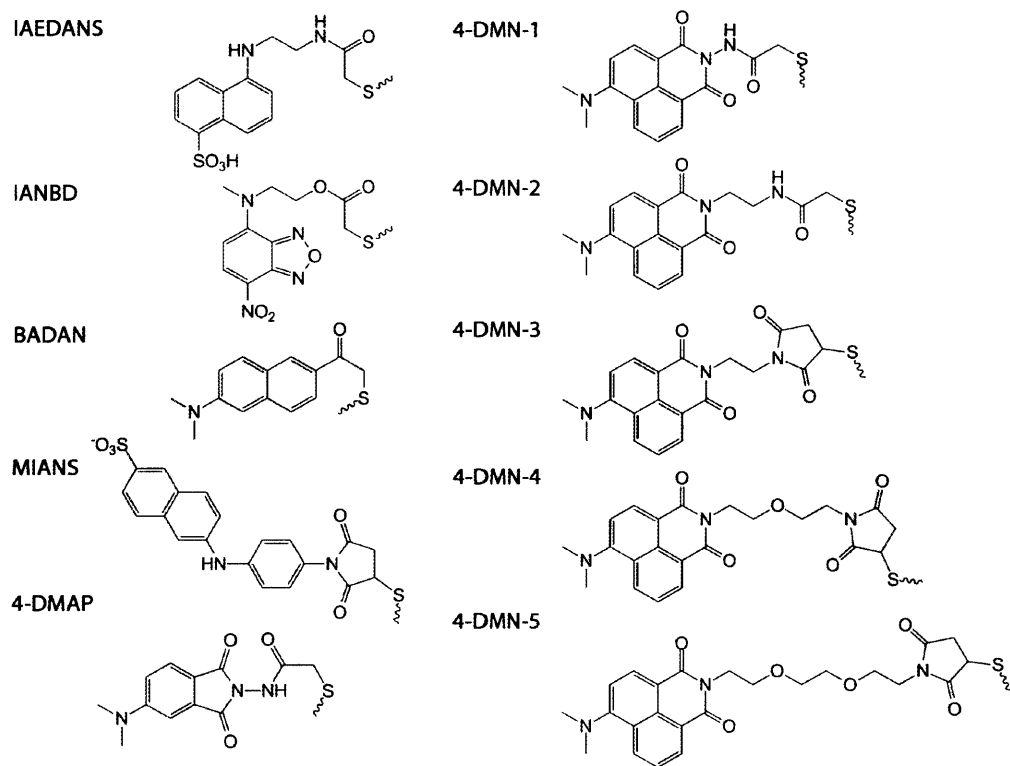


Figure 3-12: Balancing self and antigen interactions. **A)** Fluorescence of 4-DMN labeled EGF 5.1.07 based mutants with and without MBP as a fusion partner (Data from Elke Socher). **B)** Labeled site on M11.1.3 and M18.2.5. In yellow are depicted the near N and C-terminal residues, in orange the sites within the 9 randomized positions, in purple site within the loops connecting the beta-strands. **C)** Background fluorescence of NBD labeled constructs in the absence of ligands, 90% confidence interval represented as helices. **D)** Fluorescence of NBD in the presence of other antigens. **E)** F/F_0 ratio for M18.2.5 constructs. **F)** Blocking of MSA with wild-type binders M11.1.3 and M18.2.5 demonstrate these two clones share the same epitope.

Dye	Excitation (nm)	Emission (nm)	Integration Range (nm)
NBD	485	500-600	500-550
4-DMN	440	450-600	520-550
4-DMAP	421	480-600	500-560
Badan	389	420-650	475-525
EDANS	336	400-600	460-510
IANS	324	380-600	400-460

Table 3.1: **Fluorescence measurements:** excitation wavelength and integrated signal range.

Library	# Sequences
MBP 1.0	795,347
MBP A1.0	1,068,001
MBP B1.0	945,917
MBP AB1.0	946,408
MBP 1.2	1,656,318
MBP A1.2	1,245,396
MBP B1.2	940,962
MBP AB1.2	832,628
Total	8,430,977

Table 3.2: **MiSeq past filter reads**

Clone	Antigen	BC Loop	DE Loop	FG Loop	Framework	$K_{D,YSD}$ (nM)
WT	-	DAPAVTVR	GSKST	GRGDSPASSK	-	
AR 4.2.15	AR	YFPNKGLK	RSCS	RRRYYS		10
BTC 5.2.11	BTC	YDWSFAD	GSVST	DWYDYYSD	T75I	0.25
BTC 5.2.18	BTC	YDWSFAD	GSFSS	DYDDYSD	E46G	1.2
EGF 5.1.07	EGF	HHPYDSYSC	RSVS	THRWRYPF		5
EGF 5.1.07 (C31A)	EGF	HHPYDSYSA	RSVS	THRWRYPF		3.4
EPR 3.2.12	EPR	FSPVIYFVY	PGSAC	RKAHRSA		20
EPR 5.4.17	EPR	FSPRAADLF	RSRN	RSGLSA		8.6

Table 3.3: **Sequences of EGFR ligand Fn3 binders.** WT indicates the wild-type Fn3 scaffold. All antigens are human.

Table 3.4: List of selected binders, labeling sites and coupled fluorophores.

Scaffold	Clone	Target	Labeling Site	IANBD	Badan	IAEDANS	MIANS	4-DMN-1	4-DMN-2	4-DMN-3	4-DMN-4	4-DMN-5	4-DMAP	4-DMN-azide	4-DMN-II-azide
DARPin	Off7	MBP	D112C	x	x	x	x	x	x	x	x	x	x		
DARPin	Off7	MBP	L53C	x	x	x	x	x	x	x	x	x	x		
DARPin	Off7	MBP	M114C	x	x	x	x	x	x	x	x	x	x		
DARPin	Off7	MBP	N45C	x	x	x	x	x	x	x	x	x	x		
DARPin	Off7	MBP	R23C	x	x	x	x	x	x	x	x	x	x		
DARPin	Off7	MBP	S111C	x	x	x	x	x	x	x	x	x	x		
DARPin	Off7	MBP	T46C	x	x	x	x	x	x	x	x	x	x		
Fn3	1F11	c-Src SH3	A24C	x	x	x	x	x	x	x	x	x			
Fn3	1F11	c-Src SH3	52C					x	x	x	x	x			
Fn3	1F11	c-Src SH3	53C					x	x	x	x	x			
Fn3	BTC 5.2.18	BTC	F53C					x	x	x	x	x			
Fn3	BTC 5.2.18	BTC	V1C					x	x	x	x	x	x		
Fn3	BTC 5.2.18	BTC	Y23C					x	x	x	x	x	x		
Fn3	EGF 5.1.07 C31A	EGF	A20b										x		
Fn3	EGF 5.1.07 C31A	EGF	V1C					x	x	x	x	x	x		
Fn3	EGF 5.1.07 C31A	EGF	24C					x	x	x			x		
Fn3	EGF 5.1.07 C31A	EGF	26C					x	x	x			x		
Fn3	EGF 5.1.07 C31A	EGF	29C		x	x		x	x	x	x	x	x		
Fn3	EGF 5.1.07 C31A	EGF	2C					x	x	x	x				
Fn3	EGF 5.1.07 C31A	EGF	32C					x	x	x	x				
Fn3	EGF 5.1.07 C31A	EGF	34C					x	x	x	x				
Fn3	EGF 5.1.07 C31A	EGF	3C					x	x	x	x				
Fn3	EGF 5.1.07 C31D	EGF	48C					x	x	x	x			x	
Fn3	EGF 5.1.07 C31A	EGF	48C		x	x		x	x	x	x			x	x
Fn3	EGF 5.1.07 C31A	EGF	80C		x	x		x	x	x	x	x	x		
Fn3	EGF 5.1.07 C31A	EGF	81C					x	x	x	x				
Fn3	EGF 5.1.07 C31A	EGF	H24C					x	x	x		x			
Fn3	EGF 5.1.07 C31A	EGF	S54C					x	x	x	x	x			
Fn3	EGF 5.1.07 C31A	EGF	W80C									x		x	x
Fn3	EGF 5.1.07 C31A	EGF	Y26C					x	x	x		x			
Fn3	As15	Abl-SH2	C31					x	x	x	x	x			
Fn3	As15	Abl-SH2	C33					x	x	x	x	x			
Fn3	As15	Abl-SH2	C35					x	x	x	x	x			
Fn3	As15	Abl-SH2	C47					x	x	x	x	x			
Fn3	As15	Abl-SH2	C80					x	x	x	x	x			
Fn3	As15	Abl-SH2	C82					x	x	x	x	x			
Fn3	As15	Abl-SH2	C83					x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	CG	x											
Fn3	L7.5.1	Lysozyme	CG3	x											
Fn3	L7.5.1	Lysozyme	C-G4S	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	D81C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	E46C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	F80C	x	x					x	x	x			
Fn3	L7.5.1	Lysozyme	G32C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	G51C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	N54C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	R78C	x	x										
Fn3	L7.5.1	Lysozyme	S21C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	T29C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	T48C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	T53C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	V1C							x					
Fn3	L7.5.1	Lysozyme	V52C	x	x			x	x	x	x	x			

Continued on next page

Table 3.4 – Continued from previous page

Scaffold	Clone	Target	Labeling Site	IANBD	Badan	IAEDANS	MIANS	4-DMN-1	4-DMN-2	4-DMN-3	4-DMN-4	4-DMN-5	4-DMAP	4-DMN-azide	4-DMN-II-azide
Fn3	L7.5.1	Lysozyme	V73C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	V76C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	W27C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	Y25C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	Y30C	x	x			x	x	x	x	x			
Fn3	L7.5.1	Lysozyme	Y71C	x	x			x	x	x	x	x			
Fn3	YSX1	MBP	L81C	x	x		x	x							
Fn3	YSX1	MBP	S28C	x	x		x	x	x	x	x				
Fn3	YSX1	MBP	S83C	x	x		x	x	x	x	x	x			
Fn3	YSX1	MBP	Y28aC	x											
Fn3	YSX1	MBP	Y30C	x	x		x	x	x	x	x	x			
Fn3	YSX1	MBP	Y77C	x	x		x	x	x						
MBP-Fn3	EGF 5.1.07 C31A	EGF	C53					x	x	x	x				
MBP-Fn3	EGF 5.1.07 C31A	EGF	C81					x	x	x	x				
MBP-Fn3	EGF 5.1.07 C31A	EGF	C84					x	x	x	x				
Sumo-DARPin	Of7	MBP	D112C	x											
Sumo-DARPin	Of7	MBP	L53C	x											
Sumo-DARPin	Of7	MBP	M114C	x											
Sumo-DARPin	Of7	MBP	N45C	x											
Sumo-DARPin	Of7	MBP	R23C	x											
Sumo-DARPin	Of7	MBP	S111C	x											
Sumo-DARPin	Of7	MBP	T46C	x											
Sumo-Fn3	L7.5.1	Lysozyme	CG	x											
Sumo-Fn3	L7.5.1	Lysozyme	CG3	x											
Sumo-Fn3	L7.5.1	Lysozyme	C-G4S	x											
Sumo-Fn3	L7.5.1	Lysozyme	D81C	x											
Sumo-Fn3	L7.5.1	Lysozyme	E46C	x											
Sumo-Fn3	L7.5.1	Lysozyme	G32C	x											
Sumo-Fn3	L7.5.1	Lysozyme	G51C	x											
Sumo-Fn3	L7.5.1	Lysozyme	N54C	x											
Sumo-Fn3	L7.5.1	Lysozyme	R78C	x											
Sumo-Fn3	L7.5.1	Lysozyme	S21C	x											
Sumo-Fn3	L7.5.1	Lysozyme	T29C	x											
Sumo-Fn3	L7.5.1	Lysozyme	T48C	x											
Sumo-Fn3	L7.5.1	Lysozyme	T53C	x											
Sumo-Fn3	L7.5.1	Lysozyme	V52C	x											
Sumo-Fn3	L7.5.1	Lysozyme	V73C	x											
Sumo-Fn3	L7.5.1	Lysozyme	V76C	x											
Sumo-Fn3	L7.5.1	Lysozyme	W27C	x											
Sumo-Fn3	L7.5.1	Lysozyme	Y25C	x											
Sumo-Fn3	L7.5.1	Lysozyme	Y30C	x											
Sumo-Fn3	L7.5.1	Lysozyme	Y71C	x											

iz£

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Sortase A Catalyzed Site Specific Labeling of Yeast Surface Displayed Proteins

4.1 Abstract

Site-specific protein modification/conjugation is a powerful method to tune protein functionality and has wide-spread uses in biology and medicine. The evaluation of conjugated proteins has a low throughput due to the need for protein expression and purification. Here, we describe the use of Sortase A, a transpeptidase, for specific labeling of proteins displayed on the surface of yeast and report experimental conditions for fast and efficient labeling with sub-micromolar concentrations of substrate. This approach can be used for flow cytometry enabled screening of conjugated proteins or enzyme engineering by coupling substrates to the yeast cell wall.

4.2 Introduction

Site-specific conjugation of proteins can extend their functionality and has wide spread applications. Site-specific protein conjugation has been used to create antibody-drug conjugates [97], enhance drug pharmacokinetics [98], and introduce small molecule fluorophore for creating detection reagents [99] among many other applications [69]. For this purpose, two conjugation strategies have been used: chemical and enzymatic. Enzymatic modification is preferred over chemical modification as it allows a greater control of reaction, it requires less substrate and it is compatible with aqueous buffer under physiological conditions.

The conjugation process is hindered by the expression and purification of proteins, representing a major bottleneck. *In vivo* labeling on cell surface is a powerful solution to this problem as it circumvents the need for protein purification. Yeast surface display allows direct analysis of clones by flow cytometry and thus could be used to investigate post-translational modification of protein on the yeast surface in a high-

throughput manner. Shusta and colleagues have displayed intein fusions on the yeast surface and used this system to site specifically label proteins while detaching them off the yeast wall [100]. In a similar fashion, Sortase A could be used to covalently attach substrates of interest to proteins displayed on yeasts and thus expand the versatility of yeast surface display technology.

The *Staphylococcus aureus* Sortase A is an enzyme that covalently attaches protein to the bacterial cell wall. The free cysteine located in its catalytic site attacks specifically the $^N\text{LPXTG}^C$ motif and cleaves the peptide bond between the threonine and glycine. The N-terminal region to the threonine remains covalently linked to the enzyme via a thioester bond, while the C-terminal region is released. Subsequently, a cell wall precursor containing a N-terminal penta-glycine sequence ($^N\text{GGGGG}^C$) hydrolyses the intermediate complex forming an amide bond with the threonine. Sortase A has been used *in vitro* to conjugate fluorophores, oligonucleotide, polymers and more to proteins [101, 102] and to phage surface [103]. *In vivo*, it has also been used to conjugate peptides to antibodies on the surface of dendritic cells (Popp et al., 2007). Thus, Sortase A is an advantageous enzyme for site-specific conjugation of protein displayed on yeasts for several reasons. It is highly specific, the recognition motifs are small, and it allows the attachment of a very wide range of molecules easily synthesized as peptide conjugates. Moreover enzymatic reaction occurs under physiological conditions without affecting yeast viability.

Here we exploit the transpeptidase activity of Sortase A to specifically label yeast surface displayed proteins. We present the development of a plasmid vector for the facile insertion of a gene encoding the protein of interest for N-terminal labeling, as well as optimization of the labeling conditions. Using this system, we displayed a N-terminal tri-glycine sequence ($^N\text{GGG}^C$) protein fusion on yeasts and analyzed the extent of reaction with a biotin conjugated substrate peptide by varying the concentration of the enzyme, peptide, yeast cell density and the temperature of reaction. We identify reaction conditions that result in clear-cut gating for flow cytometry

analysis. Sortase A can be used to efficiently link any substrate to the yeast wall facilitating high-throughput screening of post-translational modifications. This method is cost-effective, rapid and highly versatile.

4.3 Results

We describe here conditions for the Sortase A catalyzed site-specific labeling of proteins displayed on the yeast surface. This procedure requires yeasts displaying a protein of interest bearing a N-terminal triglycine motif ($^N\text{GGG}^C$) and an exogenous peptide of sequence $^N\text{X-LPRTGG}^C$, where X can be any desired chemical moiety. The presence of the N-terminal $^N\text{GGG}^C$ motif did not affect the levels of protein display on yeast for a variety of constructs tested (data not shown). Furthermore, the pC-NSL vector designed in this study allows for the facile incorporation of any gene in a sortagging-ready system.

The extent of labeling on the cell wall should offer significant separation between labeled and non labeled cells when analyzed by flow cytometry. This is critical for downstream applications where sorting of protein library by flow cytometry would be desired. We sought to find reaction conditions that would fulfill this criterion in addition of being fast, inexpensive and maintain high specificity.

Several publications describe conditions for in vitro protein conjugation with Sortase A [104, 105]. In this study we sought to optimize the reaction conditions for yeast surface labeling. We systematically varied several parameters of the system such as temperature, concentration of yeast cells, exogenous peptide concentration and enzyme concentration. First, we used a low concentration of substrate, as it often is the most expensive reagent, and investigated the extent of reaction with various enzyme concentrations. We used a Biotin- $^N\text{YLPRTGG}^C$ peptide concentration

of 500 nM and varied the enzyme concentration ranging from 200 nM to 20 μ M. After reacting for only 15 minutes, already at low concentrations of enzyme (500 nM), a clear increase in biotin signal can be observed and augments with higher enzyme concentration (Fig. 4-2A and B). We also observed that yeasts that do not display the N GGG C -protein fusion sequence, negative for the c-myc antigen detection, maintain basal level signal intensity, demonstrating high specificity of the reaction. However, at high enzyme concentration (20 μ M), we started to observe a slight increase in signal of this population due to non specific labeling to the yeast wall. From this experiment we concluded that 10 μ M concentration of enzyme is required for quantitative and specific protein labeling in a short period of time. Next we sought to optimize the reaction temperature. We varied the temperature from 4 to 50°C and reacted cells in the presence of 500 nM of biotin substrate and 10 μ M of enzyme. The greatest extent of reaction was observed for temperatures between 30 to 37°C. However, the biotin signal was strong enough for clear flow cytometry gating even at low temperatures.

The sortase A enzyme first reacts with the N LPRTGG C motif, forming a covalent/thioester bond. Subsequently, this complex is released by the N GGG C motif acting as a nucleophile. We therefore sought to vary the yeast cell density and therefore displayed protein concentration, thus modulating nucleophile concentration. We varied the yeast density from 1 to 40 $\cdot 10^6$ yeasts/mL. Given that yeasts typically display about 50'000 molecules on their surface, this range would represent a protein molar concentration of 0.8-30 nM. We did not observe any changes in the extent of reaction (Fig. 4-2D).

Lastly, we varied the concentration of biotin- N YLPRTGG C peptide substrate from 100 nM to 50 μ M and incubated the reaction with 1 $\cdot 10^6$ induced yeasts and 10 μ M of Sortase A at 37°C for up to 4 hours (Fig. 4-2E). As expected, higher concentrations of peptide favored the labeling; however, it is also associated with increased non-specific binding to the yeast wall. Yeast cell wall non-specific binding is highly dependent on the nature of the substrate, similarly, we have observed with

a fluorescein-^NLPRTGG^C peptide non-specific signal starting at concentration of 1 μ M (data not shown). Beyond substrate concentrations of 3 μ M, increasing peptide concentration does not provide an improved labeling after 2 hours.

Finally, we sought to investigate the number of biotins attached on the cell wall. We reacted $1 \cdot 10^6$ yeasts in 100 μ L with 0.2 or 2 μ M of peptide and 10 or 20 μ M of enzyme for 30 minutes. Using calibrated beads, we were able to determine the number of biotins covalently attached to the yeast wall (Fig. 4-2F). For the lowest peptide concentration condition, there were about 2000 biotins per yeasts, whereas 20'000 biotins were measured for the 20 μ M enzyme, 2 μ M peptide condition.

4.4 Discussion

Here we described the procedure for specific and quantitative Sortase-mediated conjugation of protein substrates to the yeast surface. We identified reaction conditions that offer strong signal on flow cytometry allowing comfortable gating. We have shown that increased concentration of the enzyme or biotin-^NLPRTGG^C peptide offer greater labeling. Acylation of the ^NLPXTG^C peptide has been reported to be the rate limiting step for the transpeptidase activity of Sortase A. Accordingly; we have not observed any effect on the labeling of yeasts when varying the cell concentration. Temperature had a bell-shaped dependence on the reaction. The optimal temperature range is between 30-37°C, however labeling at 4°C yields already a strong signal by flow cytometry. This is of value for displayed proteins that are unstable and need to be maintained at low temperatures. We observed non-specific signal when the peptide concentration exceeded 1 μ M. It is not known whether this is the result of non-specific labeling or peptide adsorbed on the yeast wall in a non-covalent manner. We therefore recommend not using concentration above 1 μ M, although this limit depends on the nature of the peptide. Overall the reaction conditions are flexible and

allow the use of low concentration of peptide substrate which can be an expensive reagent depending on the functional groups attached to the ^NLPRTGG^C motif.

Site-specific enzymatic conjugation of the yeast cell wall is preferred over chemical labeling. It allows for more controlled extent of reaction and requires much less reagent. Chemical labeling results in homogenous mixtures and can damage the protein of interest by reacting on undesired epitopes rendering the protein non functional. Furthermore, diffusion through the membrane by certain permeable reactants can yield to cell death, which significantly hinders downstream processes.

We present here a fast and inexpensive strategy for site specific conjugation on the yeast surface. This method can be used for screening post-translational modification of protein without the need for tedious expression and purification. It can also be used for enzyme directed evolution.

4.5 Methods

4.5.1 Plasmid Construction

The pCHA vector (need citation) allows display of protein on the yeast surface fused at the N-terminus of the yeast mating protein Aga2p. It was linearized with forward primer NSL_for (see table 1) and pC_rev (IDT) with the high fidelity polymerase Pfu Ultra II (Agilent) according to manufacturer's directions. The oligonucleotide NSL_for introduces a sequence coding for a tri-glycine peptide and the restriction sites NheI, NdeI and BamHI. The reverse primer pC_rev anneals on the end of the alpha-pre pro secretion signal. The product of the amplification was run on a 1% agarose gel and the bands were extracted (Qiagen). The purified DNA was

transformed into XL1-Blue (Agilent). Product was confirmed by DNA sequencing, the resulting vector is shown in Fig. 4-1.

4.5.2 Sortase Expression and Purification

The plasmid for sortase expression in *E. coli* was transformed in the competent strain Rosetta 2 (DE3) (Agilent). Cells were expanded in LB media at 37°C and induced with 1 mM isopropyl- β -D-1-thiogalactopyranoside (Amresco) at optical density of 0.6 at 600nm overnight at 20°C. Cells were harvested by centrifugation and resuspended in 50 mM sodium phosphate, 300 mM sodium chloride (equilibration buffer), 1% Triton-X-100, 3% glycerol and 20 mM imidazole. The cell suspension was sonicated three times for one minute. The solution was centrifuged at 10'000 x g for 15 minutes and the supernatant filter sterilized with a 0.22 μ m bottle top filter (Nalgene). The supernatant was applied to 5 mL of Ni-NTA resin (Qiagen) and washed three times with 50 mL of equilibration buffer with 20 mM imidazole. Finally the protein was eluted with equilibration buffer with 250 mM imidazole. The fractions were run on a 4-12% Bis-Tris gel in MES buffer to confirm molecular weight and purity.

4.5.3 Yeast Transformation

The pC-NSL vector was transformed in yeast using the EZ-Transfo procedure following manufacturer's guidelines (ZymoResearch). For Fn3 insertion, the Fn3 gene was amplified using the primers W5_34 and W3_34 (IDT) with Taq polymerase (NEB) following the protocol described in [30]. The product was precipitated by ethanol precipitation and resuspended in 5 μ L of ddH₂O. The pC-NSL vector was digested with BamHI, NheI and NdeI and digested inserted was removed by spin column purification (QIAGEN). The yeast electrotransformation protocol was identical to described

previously [30].

4.5.4 Sortase Mediated Labeling

Induced yeasts were washed twice and resuspended in 100 μ L YSL buffer (300 mM Tris-HCl, 150 mM NaCl, 5 mM CaCl_2 , 1 mg/ml Bovine Serum Albumin, pH 8.0). Following addition of Sortase A and peptide, the cells were incubated in a shaker at 250 rpm. To quench the reaction, the cells were spun down and washed twice with 1 mL Phosphate Buffer Saline + 1 mg/ml Bovine Serum Albumin (PBSA).

4.5.5 Flow cytometry

10^5 cells were washed with 1mL PBSA then resuspended in 50 μ L PBSA, monoclonal mouse antibody against c-myc 9E10 (Covance) at 2 μ g/mL. After 30 minutes on ice, the cells were washed with 1mL PBSA and resuspended in 50 μ L PBSA, Goat anti Mouse Alexa488 2 μ g/mL (Invitrogen) and 2 μ g/ml Streptavidin Alexa647 (Invitrogen) for 15 minutes on ice. The cells were washed once with 1 mL PBSA and resuspended in 500 μ L PBSA just prior analysis on an Accuri C6 instrument (BD). For quantification of number of biotins on the yeast surface, 10^5 cells were washed with 1 mL PBSA then resuspended in 50 μ L PBSA, chicken antibody against c-myc (Gallus immunotech) at 2 μ g/mL. After 30 minutes on ice, the cells were washed with 1mL PBSA and resuspended in 50 μ L PBSA, Goat anti Chicken Alexa488 2 μ g/mL (Invitrogen) and anti biotin-APC antibody (Miltenyi Biotech) for 15 minutes on ice. The cells were washed once with 1 mL PBSA and resuspended in 500 μ L PBSA just prior analysis on an Accuri C6 instrument (BD). Quantum Simply Cellular anti mouse beads (Bangs Laboratories) were incubated with mouse anti biotin-APC antibody at 10-fold dilution.

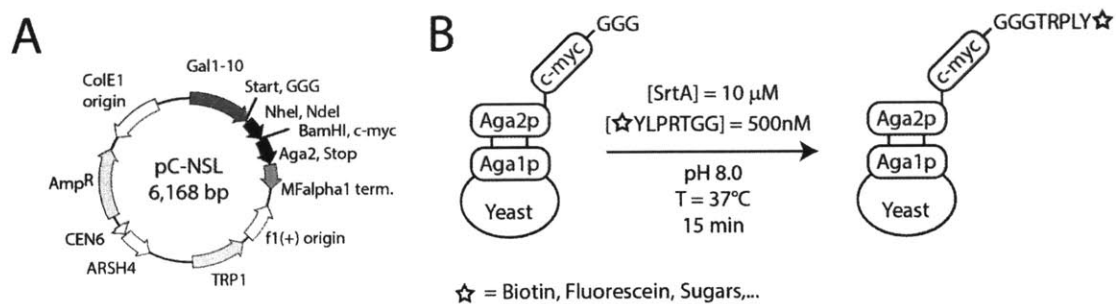


Figure 4-1: pC-NSL vector map and experimental overview. A) Vector map of the plasmid designed in this study. Gene coding for the protein of interest can be inserted using the NheI and BamHI restriction sites. The protein of interest is fused to the N-terminus of Aga2p. **B)** Schematic of the Sortase A mediated yeast surface labeling reaction.

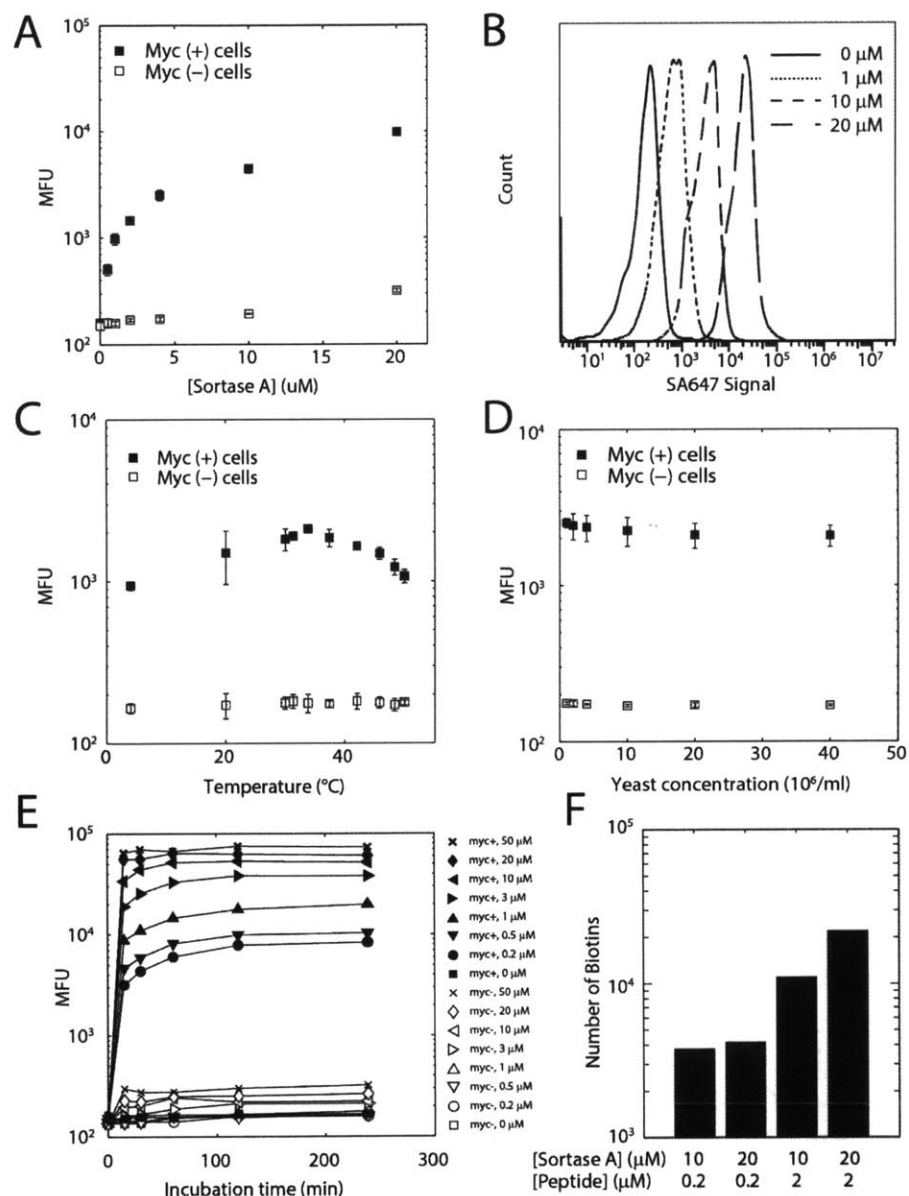


Figure 4-2: Labeling optimization. **A)** Effect of enzyme concentration on extent of labeling represented as median fluorescence signal. Representative displaying populations of **A)** were plotted as a histogram in **B)**. **C)** Temperature dependence. **D)** Nucleophile concentration dependence. **E)** Peptide substrate concentration dependence. **F)** Quantification of numbers of biotin per yeast.

Table 4.1: Oligonucleotide sequences

Primer Name	Sequence (5' - 3')
NSL_for	TCTTTGGACAAGAGAGGAGGTGGAGCTAGCCAGTCGCATATGACTGAGGGATCCACGCGTGCA
pC_rev	TCTCTTGTCCAAAGAGCC
W5_34	TGCAGAAGGCTCTTTGGACAAGAGAGGTGGTGGTGTCTGATGTTCCGAGGG
W3_3	CGGAGATAAGCTTTTGTCTGCACGCGTGGATCCTGTTCCGTAATTAATGAAATTGG

izf

Chapter 5

Competitive Inhibition of Fluorescence Quenching Based Sensors

5.1 Background

The development of technologies allowing the detection of analytes has been a major driver of our enhanced understanding of biological systems. Most of these assays are destructive and require extensive preparation. In the previous chapter, we have briefly discussed existing methods and established design principles for reagentless biosensors. Another strategy for protein detections consists of competitive inhibition of fluorescence quenching, as illustrated in Fig. 5-1. In this design, a fluorophore labeled binder can bind competitively either to its analyte or to an anti-fluorophore quencher. Simultaneous binding cannot be achieved due to steric hindrance. This concept has been demonstrated for the detection of small molecules [106], DNA [107] and proteins [108]. The underlying mechanism for this sensor requires steric hindrance, as opposed to more complex physical phenomena in the case of environment sensitive fluorophores. As a result, this system is intuitively more transposable to virtually any target of interest.

Here we present a general approach for the development of sensors based on inhibition of fluorescence quenching. We design sensors based on competitive inhibition of fluorescence quenching for Maltose Binding Protein (MBP) and Epidermal Growth Factor (EGF) using two distinct scaffolds. We develop a mathematical model to determine the optimal characteristic of the system to most adequately measure the concentration of an analyte. Based on our modeling results, we engineer a series of quenchers to expand the analyte detection range of a sensor system. Finally we lay the foundation for experimental validation of this mathematical model.

5.2 Results

We evaluated the affinity and effect on fluorescence of three different antibody/fluorophore pairs. The antibody 4-4-20 murine IgG2a recognizing fluorescein showed a quenching of 90% with an equilibrium dissociation constant (K_D) of 0.7nM in accordance with previously reported results [109, 110] (Fig. 5-2A). Fluorescein is self-quenched when reacted to protein surfaces; for this reason we also evaluated a commercial antibody against Alexa Fluor 488 (AF488) which does not exhibit this property. The anti AF488 antibody bound to AF488 with a K_D of 0.24 nM and quenched the fluorescence of AF488 up to 93% (Fig. 5-2A). As an additional reagent, we evaluated the single chain variable fragment HL4-MG recognizing the fluorophore malachite green ($K_D = 1.8 \mu\text{M}$) and drastically increasing fluorescence (610-fold) upon binding (Fig. 5-2A). Unfortunately, like fluorescein, malachite green (MG) is pre-activated upon labeling to protein surfaces, losing about a 10-fold dynamic range in the fluorescence enhancement potential (Fig. 5-2A).

To demonstrate the feasibility of sensors based on competitive inhibition of fluorescence quenching, we selected the DARPIn binder clone Off7, recognizing the analyte MBP [111]. We labeled Off7 at seven different sites near the binding interface [6], with three different derivatives of fluorescein (M5F, 5IAF, 5BrF) assuming that the different conjugation angle could be advantageous in identifying competitive epitopes. In the quenched state, the antibody 4-4-20 should be bound to the fluorescein derivative conjugated to Off7. In the active state, MBP is bound to Off7 and prevents simultaneous binding with 4-4-20. We found that the binding affinity for MBP of these constructs was strongly affected upon labeling with the bulky fluorescein dyes. Nonetheless, position D112C, when labeled with 5BrF, was responsive to MBP (Fig. 5-2B). The fluorescence was quenched by 40% upon addition of 4-4-20 and was partially recovered with 10 μM of MBP. High concentrations of MBP were necessary as the affinity of Off7(D112C)*5BrF was only 570 nM for MBP (measured by Bi-

olayer Interferometry, data not shown). When labeled with the fluorophore AF488 or MG, no changes in signal were observed upon addition of MBP, emphasizing the importance of the orientation for the competitive steric hindrance.

We further validated this proof of concept using an Fn3 scaffold engineered to bind to human EGF: clone EGF 5.1.07 C31A. We labeled the Fn3 at six different positions with AF488. The fluorophore could be quenched by the anti-AF488 antibody at all positions, but only for position Y29C did the addition of 15 μM of EGF affect the fluorescence (Fig. 5-2C). The recovery in fluorescence signal was only partial, suggesting the competition applied by 15 μM of EGF was insufficient to prevent the quencher from binding. These data suggest that sensors can be generated based on several scaffolds for distinct antigen using the competitive inhibition of quenching principle.

The key advantage of sensors based on competition is that the modulation of the competition stringency can allow for variation in the analyte detection range. To quantify in depth this unique property, the interactions between binder (R), analyte (L), and quencher (Q) were modeled as a set of ordinary differential equations and solved numerically. Variation of the kinetic rate constants for the R/L, and R/Q interactions, as well as the R and Q concentrations determine the detection range for L and the maximum fluorescence intensity change. To illustrate this, we used the literature reported kinetic rate constants for the fluorescein/4-4-20 interaction ($k_{\text{on,Q}} = 8 \cdot 10^7 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off,Q}} = 2.5 \cdot 10^{-2} \text{ s}^{-1}$) and a hypothetical R/L interaction with $K_D = 10 \text{ nM}$ ($k_{\text{on,R}} = 10^5 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off,R}} = 10^{-3} \text{ s}^{-1}$). We set arbitrarily the quencher and binder concentrations to 100 nM and 1 nM, respectively. Under these conditions the detectable range of analyte is 0.2 to 20 μM (Fig. 5-3), representing a shift of about two orders of magnitude from the Binder/Analyte equilibrium dissociation constant.

Several parameters can be modulated to shift the detectable analyte concentration range to lower concentrations. The simplest approach is to vary the binder or

quencher concentration. Indeed, lowering the concentration of quencher will result in less stringent competition and thus a lower detection range. In this particular example, a 100-fold reduction of the quencher concentration results in a 20-fold lower detectable range. For any sensors based on competitive inhibition of quenching, this approach has the advantage of requiring no further engineering. However, the ease of detectable analyte concentration range modulation comes at the cost of reduced relative maximum response magnitude. With lower quencher concentrations the equilibrium between free fluorescent binder and quenched binder is shifted toward the free form. As a consequence, the relative magnitude of the maximum response is reduced, a scenario that is not desirable. A second approach to lowering the analyte detection range is to engineer binders with stronger affinities for the analyte, resulting in lower concentration detection without affecting the fluorescence dynamic range. A 100-fold decrease in the dissociation rate constant ($k_{\text{off,R}}$) results in an equivalent 100-fold decrease in the detection range. Interestingly, further affinity improvement does not translate into a significantly reduced detectable concentration range, an effect observed due to the ultra-fast association rate of the quencher with the binder. This approach is beneficial as it maintains the relative magnitude of fluorescence increase, however, it requires engineering the binder which can be a difficult and resource intensive task. The third strategy for decreasing the detectable analyte concentration range is weakening the quencher/fluorophore interaction. Weakening this interaction by as little as a 4-fold decrease in the dissociation rate constant ($k_{\text{off,Q}}$) results in approximately a two-log decrease in the detectable concentration range without reducing the fluorescence dynamic range, in this example. This approach has the powerful advantage of being translatable to any sensor system as long as the same fluorophore is used.

While this system does not have an analytical solution, the magnitude of fluorescence response can be analytically determined as the fluorescence signal ratio with and without the analyte (Eqn. 5.8). Several important considerations can be observed from this equation. First, the response magnitude (S) is directly proportional to the

quenching factor (ϕ). Second, the tighter the affinity for the quencher to the fluorophore, the greater the response magnitude. Similarly, increased concentrations of quencher are favorable for achieving a greater response magnitude. Interestingly, the relationship between the response magnitude and the binder concentration is biphasic. Increased binder concentrations will favor interactions with the quencher, up to a stoichiometry of 1:1. Beyond this threshold, non-quenched binder will contribute to the signal and thus raise background levels, resulting in a lower response magnitude.

From the results of the mathematical model, it becomes clear that quencher affinity is the ideal parameter to vary, as it allows for modulation of the concentration detection range while maintaining the fluorescence signal range and can be transposed to different sensors. To verify experimentally these predictions we sought to engineer quenchers with lower affinity. The anti-fluorescein antibody has already been engineered as a single chain variable fragment, displayed on yeast (Fig. 5-3A), and further affinity matured in our lab to reach femtomolar affinities [8]. Our aim was the opposite: engineer an anti-fluorescein binder that maintains quenching while exhibiting a reduced equilibrium binding affinity. We performed an "affinity de-maturation" process: libraries of yeast displaying mutated single chain variable fragment of 4-4-20 were generated with low (FL library) and high (FH library) mutation rate. We performed three sequential selections by fluorescence-activated cell sorting, alternating between positive and negative selections (Fig. 5-4). We analyzed the sequences of the selected libraries and selected six clones for further characterization.

We expressed these mutants on a murine IgG2a backbone: clone FLB1 and FHC5 expressed with comparable yield as 4-4-20 (Table 1). On the other hand clone FLC1 showed poor expression and a strong propensity to aggregate (Fig. 5-5A). The expression levels correlated with the aggregation propensity as observed from the size exclusion chromatography profile; the best clones that produced the larger yields were found mostly as monomers. The mutants maintained the quenching potency and exhibited equilibrium dissociation constants between one and two orders of magnitude

greater than the wild-type antibody (Fig. 5-2B and Table 5.1). With the exception of clone FHA3, all mutants had a single mutation near the binding pocket but not affecting residues known to make contact with fluorescein4. Clone FHA3 has an I189T mutation in CDR-H2 facing inward of the beta-sandwich fold that might affect stability, as observed in the strong aggregation and poor expression of this clone. The additional mutation (E179D) is more difficult to rationalize as it is solvent exposed and distant from the fluorescein paratope. FLB1 has a K_D of 11.9 nM and the highest expression yield. It has one mutation, M242V, located in CDR-H3, making contacts with the light chain fragment. This mutation may alter the scFv conformation, resulting in a looser fit with fluorescein. Additionally, these engineered fluorescein binders maintained quenching of different thiol-reactive fluorescein derivatives, allowing their use to exploit different binding orientations (Fig. 5-6). The 4-4-20 antibody and its derivatives do not bind to 6IAF, as the attachment carbon C16 is buried in the 4-4-20 complex. Overall, yeast surface display affinity de-maturation enabled the identification of quenchers with weaker affinity for fluorescein while maintain the quenching potency.

The rationale for engineering a quencher with weaker affinity for fluorescein was that our mathematical model predicted it would allow the detection of lower MBP concentrations. Thus, we evaluated the MBP detection range for Off7(S112C)*5BrF by titrating MBP in the presence of 4-4-20 or FLB1. We observed that, as predicted by our model, increasing the concentrations of quencher resulted in greater potential magnitude response, with the quenched signal starting at 40% (2.5-fold potential) and 60% (1.6-fold potential) for the 50 nM and 5 nM 4-4-20 conditions respectively. The maximum response magnitude was lower for the FLB1 conditions, as the weakened binding for this clone lowers the fraction of quenched fluorophore in the absence of MBP. Unfortunately the use of the weakened fluorescein quencher FLB1 did not result in a strong shift in the concentration detection range, possibly due to an artifact of the experimental conditions used here.

5.3 Discussion

We demonstrated the design of two distinct sensors by competitive inhibition of quenching, using Fn3 and DARPin scaffolds for sensing of EGF and MBP respectively. Combined with literature evidence [106, 107, 112, 33], these results support the hypothesis of a broad applicability for this design. Furthermore, the Fn3 sensor for EGF was designed in the absence of structural information, an important advantage for future sensors developments. We established a strong mathematical framework for describing the interactions between the three components of this system: labeled binder, analyte, and quencher. The key advantage of this sensor design is its modularity: by means of binder and quencher concentration variations or the use of quenchers with altered affinities, one can modulate the analyte detection range. This feature is not available on other fluorescence based sensors. To support this evidence, we engineered fluorescein quenchers with a weakened affinity. Unfortunately, despite our successful quencher engineering, were unable to identify proper conditions to test the hypothesis of analyte concentration modularity. Nonetheless, our experimental results are the first stepping stone toward the corroboration of this mathematical model.

With further development needed, one questions stands out: how much potential does this sensor design have for as a tool for protein detection? There are two key criteria to be taken into consideration when designing sensors: the first is the detectable concentration range, and the second is the signal to noise ratio. The former is determined by the equilibrium binding constants between the labeled binder and the analyte as well as between the labeled binder and the quencher. The latter is determined in part by these affinities, the concentrations of the binder and quencher as well as by the intrinsic quenching capacity of the anti-fluorophore binder. Concerning the response magnitude, this system is well suited as it allows to use virtually any desired fluorophore, as long as a quenching antibody or other scaffold exists. Fluorescein is

quenched by a 20-fold factor upon binding with 4-4-20. Unfortunately, this potential response magnitude is strongly reduced as fluorescein is pre-quenched when labeled to protein surfaces. Some derivatives of fluorescein such as AF488 and Oregon Green 488 do not exhibit this undesirable property, making them more suitable for sensor applications, in addition to their greater photo-stability. Malachite green exhibits the same phenomenon: we observed a drastic 10-fold reduction in the response magnitude of this fluorophore when labeled to a protein. Twisted internal charge transfer and reduction of vibrational motion may be responsible for this behavior [23, 113].

The flexibility of the analyte concentration detection range by simply varying the concentration of binder, quencher, or the affinities is a great advantage for this design compared to other biosensor designs which are limited to a detection range between $0.1 \cdot K_D$ and $10 \cdot K_D$. Designing sensors can be a difficult task as this thesis has demonstrated, and the ability to vary the experimental conditions as opposed to engineering a new clone present a great advantage in terms of resources. However, this property is a double edged sword: in *in vitro* and *in vivo* applications, local changes in the binder or quencher concentrations will affect the read out and can lead to misinterpretation of the signal. To avoid this artifact, tight regulation of the binder and quencher concentration can be achieved for example by means of coupling to microspheres.

5.4 Conclusion & Future perspectives

Prior to our work, others had demonstrated the application of competitive inhibition of quenching based sensors to proteins [106], small molecules [112] and DNA [107] using antibodies or even DNA as a binder. Looking at this problem through the lens of an engineer, our contribution to this field has been the analytical characterization of interactions between the three components in this system. This modeling framework

has allowed us to quantify the detection range and response magnitude potential for this class of sensors. Furthermore, our validation of concept by developing two sensors based on structurally distinct scaffolds is critical for the widespread adoption of this concept. Nonetheless, a major unmet challenge in this sensor design is the control of the labeled-binder affinity for its analyte. We observed strong loss of binding when labeling with bulky fluorophore such as fluorescein. To address this issue, there are existing methods for selection from protein libraries with an expanded amino acid repertoire by phage display [114, 115], and yeast surface display (Chapter 4 and data from J. van Deventer), allowing the selection of fluorophore-labeled binders with desired affinity. While many challenges lie ahead, sensors based on competitive inhibition of quenching represent an attractive tool for the detection of proteins.

5.5 Material & Methods

5.5.1 4-4-20 de-maturation

The gene coding for the single chain variable fragment of 4-4-20 from the vector pCT302(4-4-20) was amplified and ligated into the vector pCTcon2 (AddGene 41483) in between the NheI and BamHI restriction sites. We introduced mutations by error-prone PCR over the entire gene using a protocol described previously [61], creating two libraries, FL and FH with low and high degree of mutation rate respectively. The generated DNA library was transformed in *S. cerevisiae* strain EBY100 (ATCC: MYA-4941) by electro-transformation as described previously [30]. After induction, the yeasts were incubated with 1 μ M of fluorescein-biotin (Anaspec) and chicken anti-c-myc antibody, followed by incubation with Streptavidin-PE (Life Technologies) and Goat anti Chicken Alexa 647 (Life Technologies). We first performed a positive selection collecting all clones that showed binding above background. We split the

subsequent population into three groups and performed sorting at antigen levels of 1 μ M, 100 and 10 nM (A, B, C libraries respectively). Cells showing a weaker binding signal than that of wild-type 4-4-20 were collected (negative sort). For the last sort we performed a positive sort at the same concentrations. This process yielded six distinct libraries; we sequenced 43 clones and discarded clones with a newly introduced cysteine.

5.5.2 Protein Expression & Purification

The light chain and heavy chain were incorporated into the gWiz mammalian expression vector (Genlantis) as murine IgG2a format. Constructs were verified by sequence analysis and plasmid DNA obtained with the DNA maxiprep kit (Invitrogen). Antibodies were secreted from Human Embryonic Kidney cells at the Protein Expression Core Facility (EPFL). The culture supernatants were filter sterilized through a 0.22 μ m filter (Nalgene) and purified on a protein A resin (Genscript). The samples were buffer exchanged into phosphate buffer saline pH 7.4 and stored at -80°C in aliquots of 1 mg/ml.

The DARPin proteins were cloned in a pE expression vector as a fusion protein His6-AviTag-Sumo-DARPin or GGG-His6-AviTag-Sumo-DARPin. These vectors were generated by modification of a pE-sumo vector (LifeSensors) received as a gift from Andrew Krueger in the Imperiali laboratory at MIT. The sequence confirmed plasmid were transformed into Rosetta 2 (DE3) cells (Millipore), grown in 1L LB media at 37°C and induced overnight with 500 μ M IPTG (AMRESCO) at 20°C when cells had reached an OD600 = 0.6. Cultures were centrifuged, pellet resuspended in 50 mM Sodium Phosphate (AMRESCO), 300 mM Sodium Chloride (AMRESCO), 1% Glycerol (AMRESCO), 3% Triton-X-100 (AMRESCO) pH 8.0. Cells were sonicated and centrifuged. The supernatant was filter sterilized and purified on a Ni-NTA resin (QIAGEN).

5.5.3 Protein Labeling

The protein sample was diluted to a final concentration of 100 μM in PBS and the reducing agent (tris(2-carboxyethyl)phosphine (TCEP) was added to a final concentration of 1 mM. After 60 minute incubation at room temperature for thiol reduction, the thiol reactive dye was added to a final concentration of 1mM and the sample was reacted overnight on a rotating wheel at room temperature. Free dye was removed on a G-25 desalting resin (GE Healthcare). Labeling efficiencies were in the range of 0.4-0.8 dye/protein molar ratio, as derived from the A280 and dye absorbance [6]. 5-iodoacetamidofluorescein (5IAF), 6-iodoacetamidofluorescein (6IAF), fluorescein 5-maleimide (M5F), 5-bromomethylfluorescein (5BrF), Oregon Green (OG), Rhodamine Green (RG) were purchased (Life Technologies). A thiol reactive derivative of malachite green was synthesized by reacting malachite green isothiocyanate (Life Technologies) with N-(2-Aminoethyl)maleimide trifluoroacetate salt and purified by high pressure liquid chromatography.

5.5.4 Model formulation

The system consists of five state variables: the concentrations of analyte (L), fluorophore labeled binder (R), quencher (Q), the complex between analyte and binder (C) and the complex between the binder and the quencher (D). The rate constants governing the complex C and D formations are the association ($k_{\text{on},\text{R}}$ and $k_{\text{on},\text{Q}}$) and the dissociation ($k_{\text{off},\text{R}}$ and $k_{\text{off},\text{Q}}$) rate constants. We assume that the system is homogenous and reaction limited. The described system is fully governed by the five differential equations:

$$\frac{d[L]}{dt} = -k_{on,R}[L][R] + k_{off,R}[C] \quad (5.1)$$

$$\frac{d[C]}{dt} = k_{on,R}[L][R] - k_{off,R}[C] \quad (5.2)$$

$$\frac{d[R]}{dt} = -k_{on,R}[L][R] + k_{off,R}[C] - k_{on,Q}[Q][R] + k_{off,Q}[D] \quad (5.3)$$

$$\frac{d[Q]}{dt} = -k_{on,Q}[Q][R] + k_{off,Q}[D] \quad (5.4)$$

$$\frac{d[D]}{dt} = k_{on,Q}[Q][R] - k_{off,Q}[D] \quad (5.5)$$

A convenient analytical solution to this system is not available. Therefore, we solved this equation system for a variety of parameter conditions by numerical integration (performed in MatLab). We report the values at 10,000 seconds, at which the system is assumed stable (not shown).

Response magnitude

The response magnitude of the system is defined by the maximum change in fluorescence observable. The maximum fluorescence signal (F_{max}) is observed when all fluorophore labeled fluorophore is unbound, it is therefore directly proportional (proportionality constant α) to the initial binder concentration (R_0).

$$F_{max} = \alpha \cdot R_0 \quad (5.6)$$

The minimum observed fluorescence (F_{min}) happens when the concentration of quenched binder ($[D]$) is maximal. This occurs in the absence of analyte (L). The concentration of quenched binder can be derived from the monovalent binding isotherm in presence of analyte depletion. The fluorescence from the quenched complex is proportional to its concentration with proportionality constant ($\phi \cdot \alpha$) where ϕ is a quenching factor.

The initial concentration of quencher is denoted (Q_0).

$$\begin{aligned}
 F_{min} &= \alpha[R] + \alpha\phi[D] \\
 &= \alpha R_0 (1 + (\phi - 1)) \frac{(\frac{k_{off,Q}}{k_{on,Q}} + R_0 + Q_0) \sqrt{(\frac{k_{off,Q}}{k_{on,Q}} + R_0 + Q_0)^2 - 4R_0 Q_0}}{2 \cdot R_0} \quad (5.7)
 \end{aligned}$$

Therefore, the response magnitude (S) of the sensor is defined as the ratio between Fmax and Fmin (Eqn. 8).

$$S^{-1} = 1 + (\phi - 1) \frac{(\frac{k_{off,Q}}{k_{on,Q}} + R_0 + Q_0) \sqrt{(\frac{k_{off,Q}}{k_{on,Q}} + R_0 + Q_0)^2 - 4R_0 Q_0}}{2 \cdot R_0} \quad (5.8)$$

Fluorescence Activating Protein

The equation system is identical as before and we keep the same nomenclature with the exception of Q which is replaced for A (activator). The fluorescence quenching factor (ϕ) is also substituted for the fluorescence activating factor (β). The response magnitude is still defined as the ratio between Fmax and Fmin, however, the conditions of occurrence for these are inversed and therefore we defined the response magnitude for a fluorescence activating system in equation (9).

$$S_A = 1 + (\beta - 1) \frac{(\frac{k_{off,Q}}{k_{on,Q}} + R_0 + Q_0) \sqrt{(\frac{k_{off,Q}}{k_{on,Q}} + R_0 + Q_0)^2 - 4 \cdot R_0 \cdot Q_0}}{2 \cdot R_0} \quad (5.9)$$

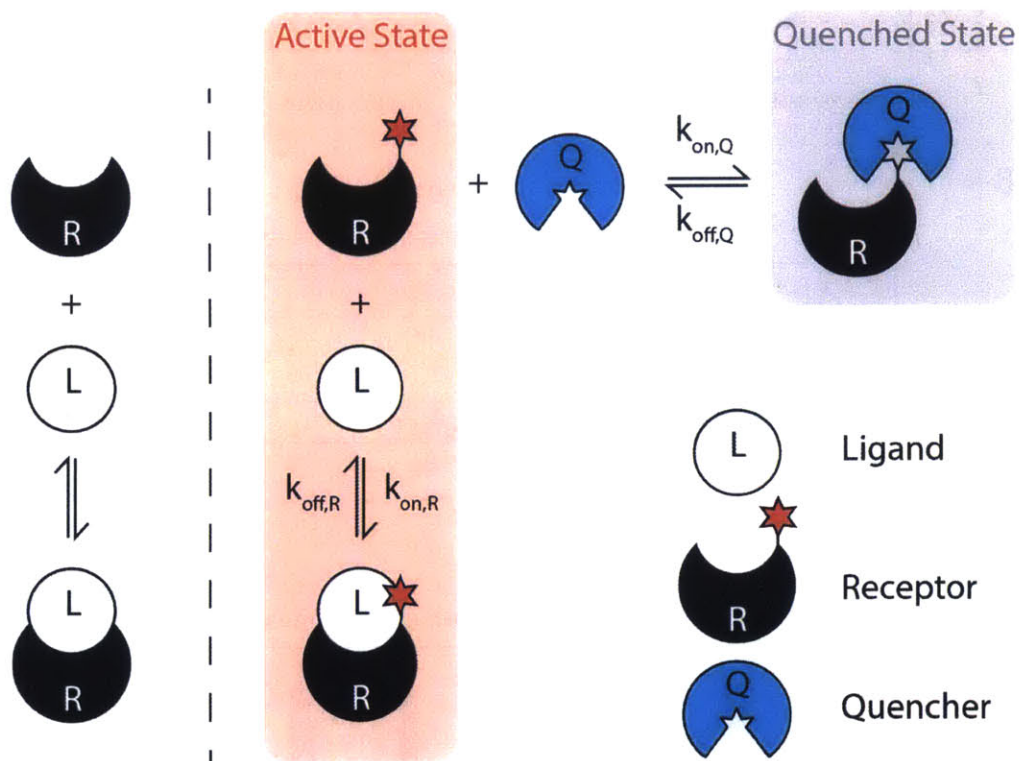


Figure 5-1: **Schematic representation of a sensor system based on competitive inhibition of quenching.** The binder (R) is engineered to recognize a particular analyte (L). The binder is then labeled with a fluorophore in a position such as the binder cannot be simultaneously bound by the quencher (Q) and the analyte. Presence of the analyte will shift the balance from the quenched state to the active state.

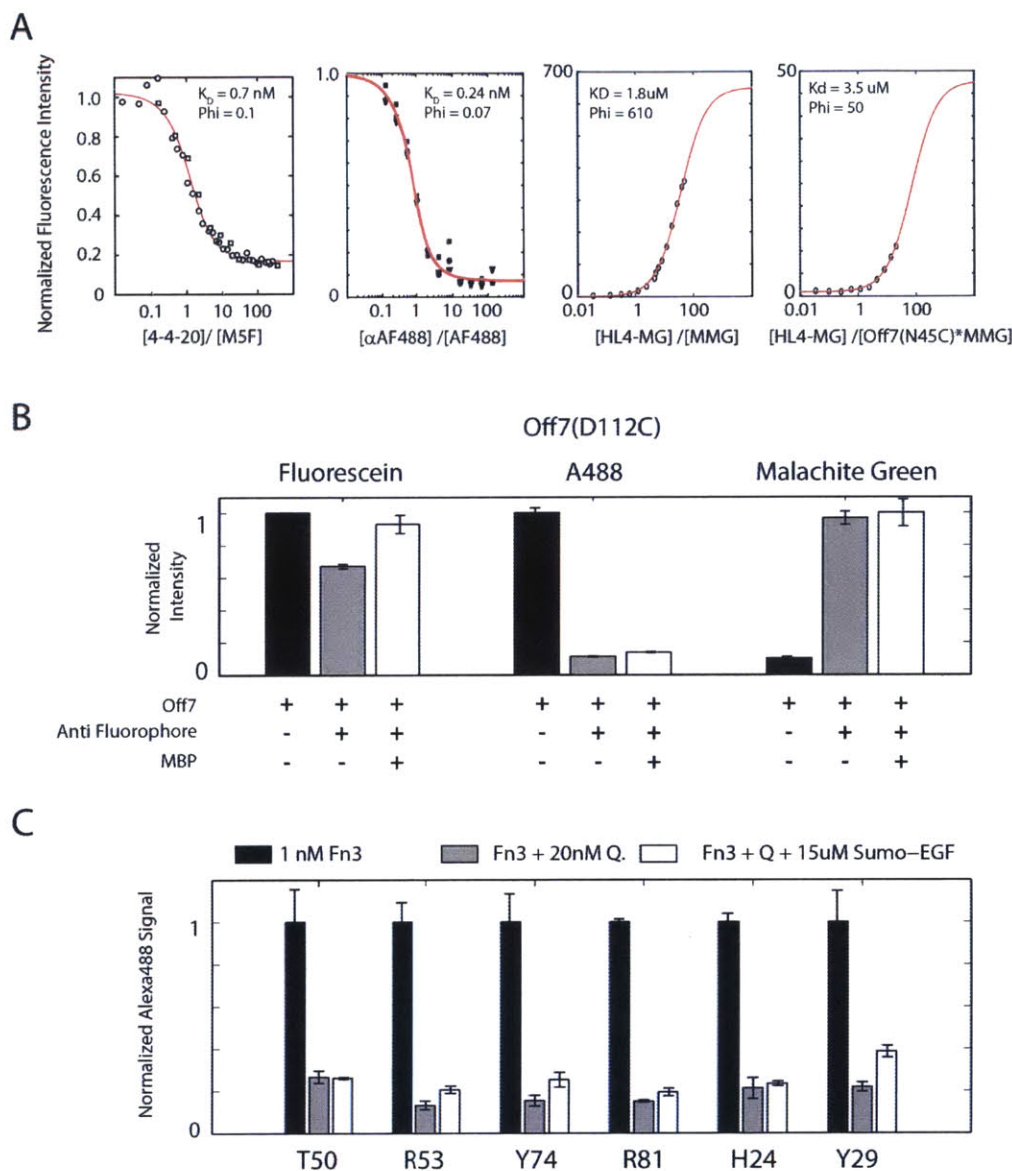


Figure 5-2: Proof of concept. **A)** Fluorescence titrations of antibodies and fluorophore pairs for fluorescein, AF488, maleimide Malachite Green (MMG) and MMG labeled DARPIn clone Off7(N45C) (left to right). The data points were fitted to a second order monovalent binding isotherm. **B)** Normalized Fluorescence intensity of clone Off7(D112C) labeled to fluorescein, AF488 or Malachite Green at 5 nM with or without 20 nM of 4-4-20 IgG2a and 10 μM of MBP. **C)** Normalized fluorescence intensity for clone EGF 5.1.07 C31A labeled at various positions with AF488 at 5 nM, with or without 20 nM of anti AF488 antibody and 15 μM of Sumo-EGF.

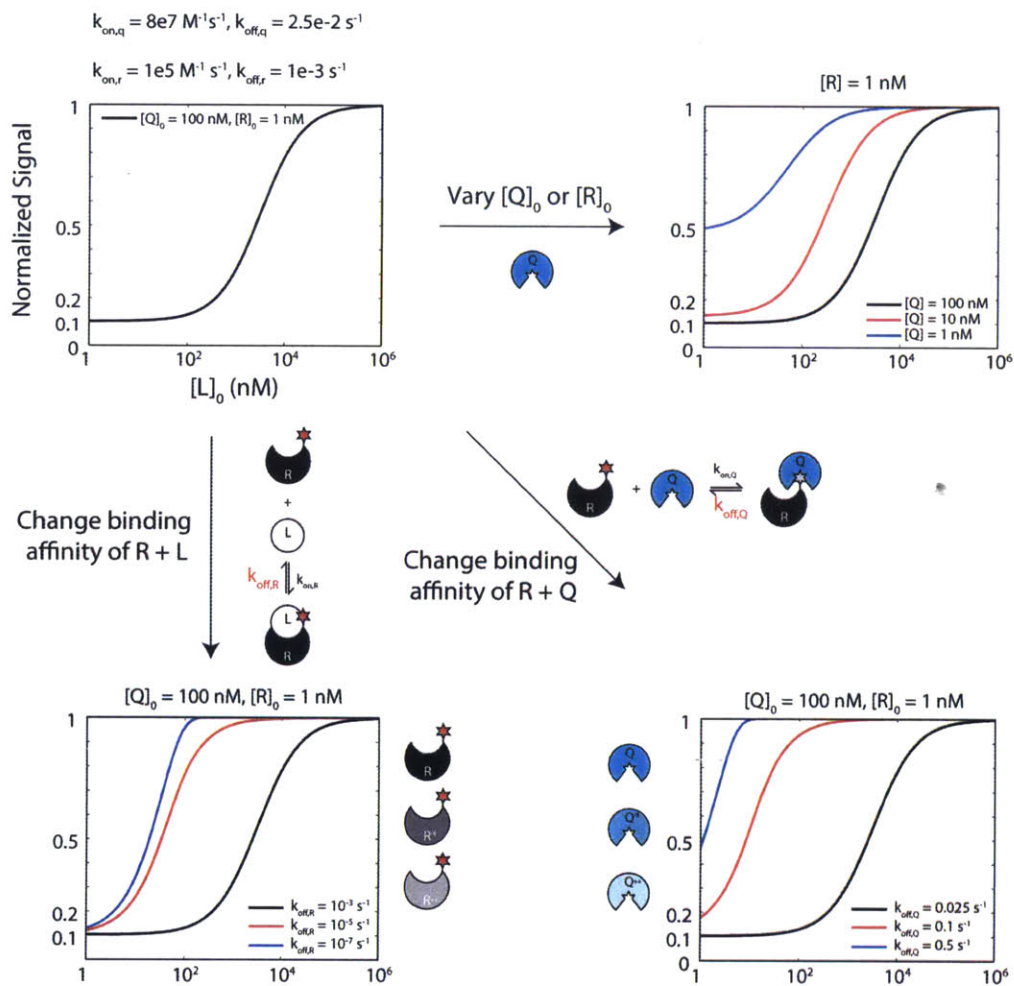


Figure 5-3: **Mathematical modeling:** fluorescence signal is shown as a function analyte concentration. The top left plot is the reference condition, all concentrations and kinetic rate constants are the same as this plot unless specified otherwise.

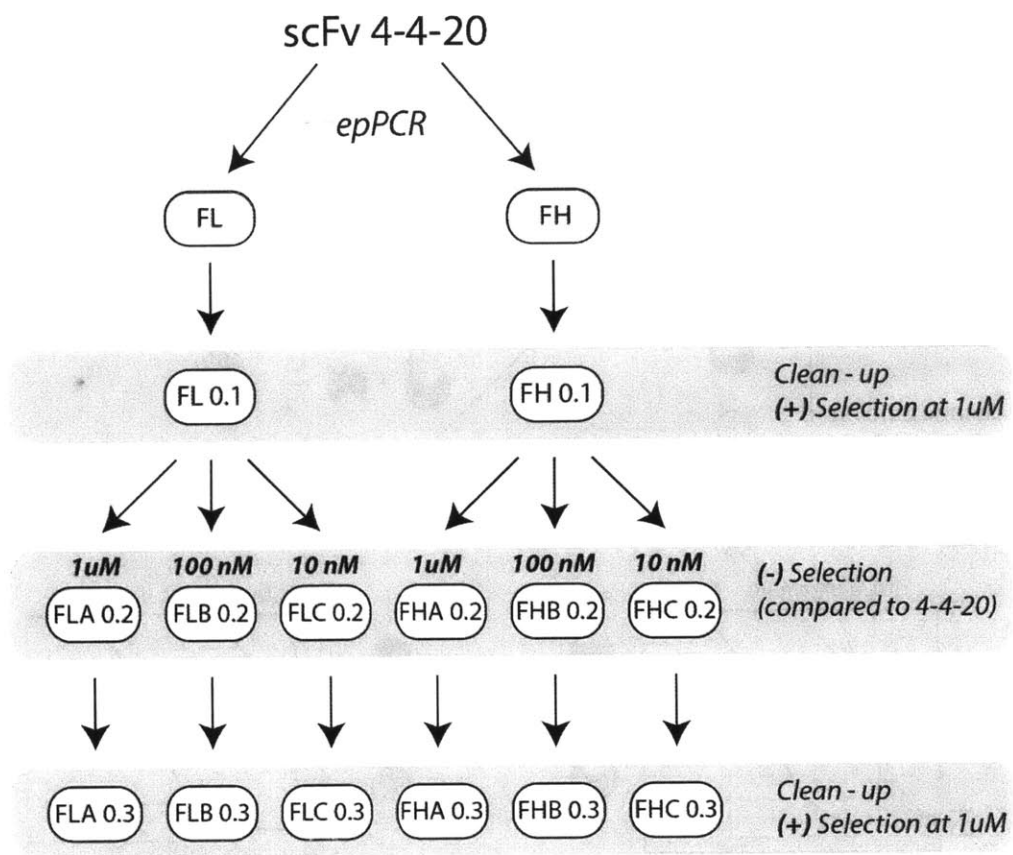


Figure 5-4: Overview of the 4-4-20 affinity de-maturation process.

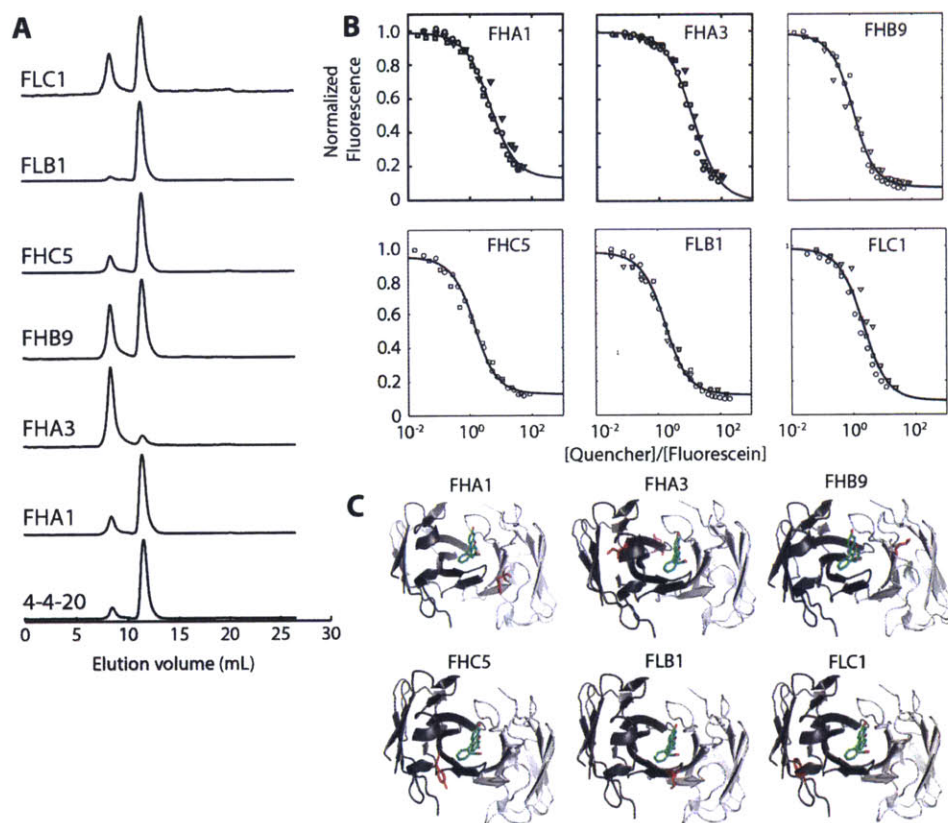


Figure 5-5: **Characterization of the anti-fluorescein quenchers.** **A)** Size exclusion profiles, where the first peak at an elution volume of 8 mL corresponds to aggregate and the second peak at around 11 mL corresponds to monomeric antibody. **B)** Fluorescence titrations. **C)** Mutated residues are highlighted in red on the scFv structure of the 4-4-20 and fluorescein complex (PDB ID: 1FLR).

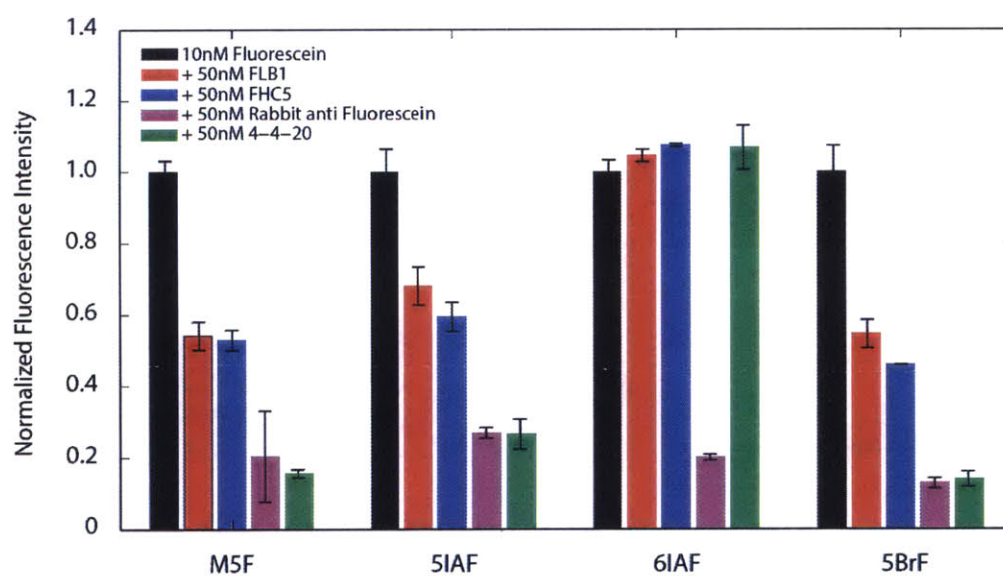


Figure 5-6: **Specificity of anti-fluorescein antibodies.** Rabbit anti fluorescein is a polyclonal preparation. All data points are normalized to the 10 nM fluorescein only condition.

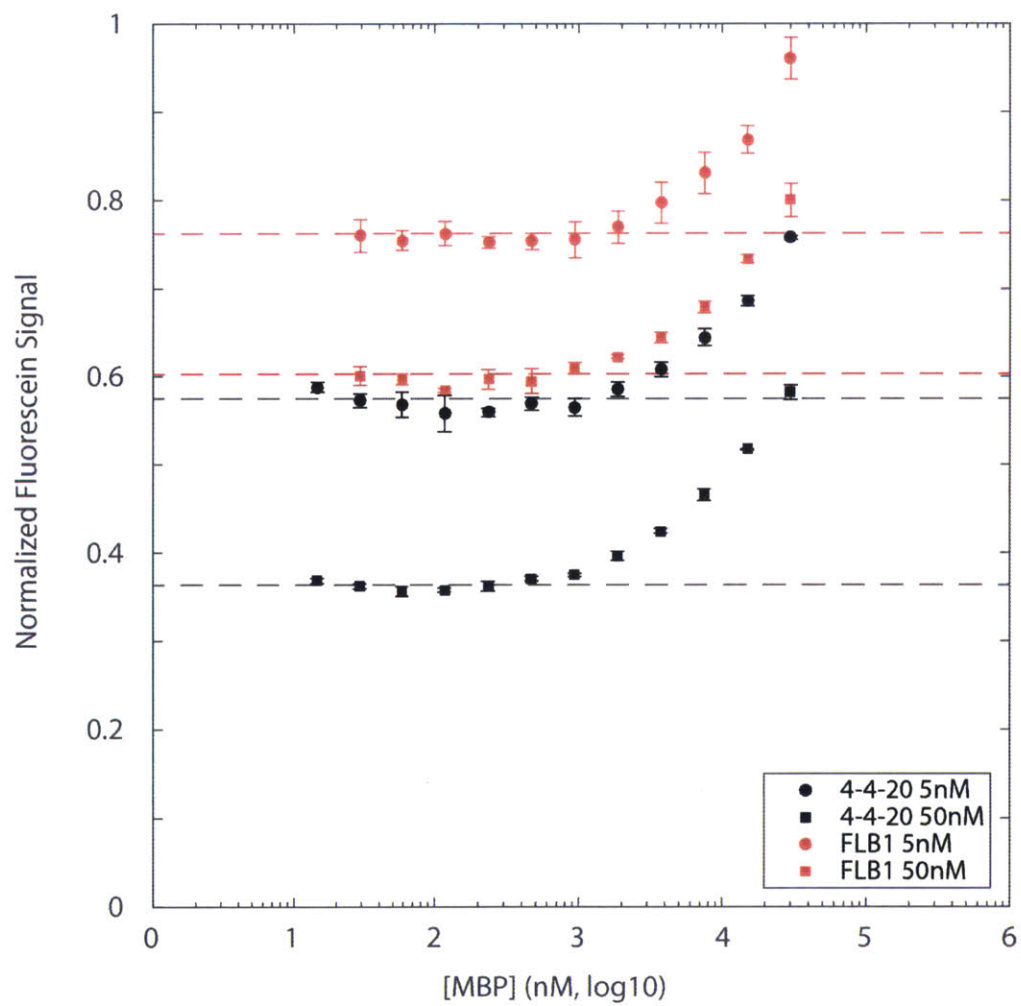


Figure 5-7: Titration of MBP to 5 nM of Off7(D112C)*5BrF.

Table 5.1: Anti-fluorescein antibody characterization

Clone ID	Mutation	Yield (mg/L)	Peak Elution (mL)	% Monomer	Quenching (%)	K_D (nM)
4420		18.4	11.48	86.9%	83.4 +/- 1.0	0.72 +/- 0.09
FHA1	Y54H	4.1	11.4	80.7%	87.2 +/- 2.3	41 +/- 4.7
FHA3	E179D, I189T	7	11.42	13.3%	100 +/- 1.2	118 +/- 13.4
FHB9	T36A	5.2	11.5	59.3%	92.5 +/- 1.5	8 +/- 1
FHC5	Y169S	16.7	11.48	82.3%	85.5 +/- 2.9	12.6 +/- 2.9
FLB1	M242V	59.3	11.4	94.2%	87.9 +/- 0.9	11.9 +/- 1.0
FLC1	F166L	0.8	11.48	63.2%	87.2 +/- 4.0	12.0 +/- 3.1