UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Engineering Molecular Modules Through Directed Evolution for Applications in Single-Cell Imaging and Immunotherapy

Permalink
https://escholarship.org/uc/item/7x01b05n

Author
Limsakul, Praopim

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Engineering Molecular Modules Through Directed Evolution for Applications in Single-Cell Imaging and Immunotherapy

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

By

Praopim Limsakul

Committee in charge:

Professor Yingxiao Wang, Chair
Professor Xiaohua Huang
Professor Shankar Subramaniam
Professor Wei Wang
Professor Jin Zhang

2019

Copyright

Praopim Limsakul, 2019

All rights reserved.

The Dissertation of Praopim Limsakul is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2019

DEDICATION

To my beloved family.

TABLE OF CONTENTS

Chapter 3 PEbody CAR: A Universal Chimeric Antigen Receptors for CAR T Cell Therapy…... 44

Chapter 4 Optimization of the Whole FRET Biosensors Through Directed Evolution………….. 61

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

During my journey through graduate school, I have received numerous support and encouragement from a great number of individuals. I would like to firstly acknowledge my advisor, Professor Yingxiao (Peter) Wang, who has encouraged, supported, and inspired me throughout my graduate career and has facilitated a collaborative and friendly laboratory environment that I have greatly enjoyed. It is with his unwavering confidence in me that I was able to accomplish this body of work. I would also like to thank my committee members Drs. Xiaohua Huang, Shankar Subramaniam, Wei Wang, and Jin Zhang for their suggestions and supports.

I am very fortunate to be surrounded by nice people. Without them, it would be very hard for me whose background is theoretical Physics to start working in experimental bioengineering laboratory. I would like to thank Dr. Shaoying (Kathy) Lu for her great support, suggestion, and kind help. I would like to thank Dr. Yijia Pan for introducing me to the Wang lab family and for her encouragement. I would like to thank Dr. Qin Peng for teaching me the fundamental techniques for laboratory and for always being a good friend. I would like to thank Yiqian Wu, Linshan Zhu, and Chi-Wei Man for being a great friend and for their role model as a PhD student which I tried to learn from. I would like to thank Dr. Ziliang Huang who always answers my questions. I also would like to thank Molly Allen who is going to graduate soon for her kind support. In addition, I would like to thank all lab members, i.e. Drs. Mingxing Ouyang, Lei Lei, Pengzhi Wang, Longwei Liu, Reed Harrison and many more, in the Wang lab and my mentees, i.e. Yanmin Ji, Fangchao He, Jiayi Li, Emma Zelus, Yuan Yuan, etc, for their help, advice, and friendship.

Last but not least, I would like to thank my family for their unconditionally love and support as well as their patience to wait for me. I would like to thank my best friend, Krit Charupanit, who is always beside me, teaches me, and helps me get through a difficult time. More importantly, I would like to thank The Royal Thai government for their financially support for my PhD study.

Since the portions of this dissertation have been published or will be published in various journals, I would like to acknowledge each of the authors for their contributions.

Chapter 1, in part, has been submitted for publication as a book chapter as it may appear in Wiley. P. Limsakul, C. Man, Q. Peng, S. Lu, and Y. Wang. “Development of Novel Cellular Imaging Tools using Protein Engineering”. Protein Engineering. Tools and Applications, UK. Wiley. The dissertation author was the first author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Cell Chemical Biology journal. P. Limsakul, Q. Peng, Y. Wu, M. Allen, J. Liang, A. G. Remacle, T. Lopez, X. Ge, B. Kay, H. Zhao, A. Y. Strongin, X.L. Yang, S. Lu, and Y. Wang. "Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface." Cell chemical biology 25.4 (2018). 370-379. The dissertation author was the first author of this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material. P. Limsakul, Y. Wu, Y. Pan, M. E. Allen, and Y. Wang. “PEbody CAR T. A Universal Chimeric Antigen Receptors for T Cell Immunotherapy”. The dissertation author was the first author of this paper.

Chapter 4, in full, is currently being prepared for submission for publication of the material. P. Limsakul, L. Lui, J. Huang, F. He, J. Li, R. E.S. Harrison, S. Lu, Y. Wang. “Optimization of the whole FRET biosensors for kinase activity through directed evolution”. The dissertation author was the first author of this paper.

Praopim Limsakul
La Jolla, CA
March 2019

VITA

EDUCATION

2010	Bachelor of Science in Physics Prince of Songkla University, Thailand
2012	Master of Science in Physics Mahidol University, Thailand
2015	Master of Science in Bioengineering University of California San Diego, USA
2019	Doctor of Philosophy in Bioengineering University of California San Diego, USA

JOURNAL PUBLICATION

P. Limsakul, Q. Peng, Y. Wu, M. Allen, J. Liang, A. G. Remacle, T. Lopez, X. Ge, B. Kay, H. Zhao, A. Y. Strongin, X.L. Yang, S. Lu, and Y. Wang. "Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface." Cell chemical biology 25.4 (2018). 370-379.

PATENT

P. Limsakul and Y. Wang. “Cell Surface Hybrid FRET Biosensor”. Pending

P. Limsakul and Y. Wang. “A Universal Chimeric Antigen Receptor for T Cell Therapy”. Submitted to UCSD Office of Innovation and Commercialization

ABSTRACT OF THE DISSERTATION

Engineering Molecular Modules Through Directed Evolution for Applications in Single-Cell Imaging and Immunotherapy

by

Praopim Limsakul

Doctor of Philosophy in Bioengineering

University of California San Diego, 2019

Professor Yingxiao Wang, Chair

Protein engineering through directed evolution has been extensively used to improve and modify the structure and function of proteins for wide range of applications. In this dissertation, I utilized the conventional directed evolution through yeast surface display library to engineer a single molecular module, i.e., monobody to have high binding affinity to R-Phycoerythrin. This engineered monobody was then applied to 1) a FRET biosensor with improved spatial resolution to monitor biological process of cancer cell invasion, and 2) a universal chimeric antigen receptor (CAR) for cancer immunotherapy. To further extend the power of directed evolution, I combined it with other technologies, including mammalian cell library, functional screening by FACS-based FRET, high-throughput DNA sequencing, and sequence-function analysis, to systematically optimize the sensitivity, specificity, and dynamic range of the FRET biosensors for monitoring kinase activities with high spatiotemporal resolution in living cells. These optimized biosensors can be used to study the cell signaling of CAR T cells upon engagement with cancer cells to improve the efficiency of CAR T cell immunotherapy. Overall, this dissertation illustrates the potential of directed evolution as a tool to engineer the molecular modules/sensors for biomedical diagnosis and immunotherapy.

Chapter 1

Introduction

1.1 Design by Directed Evolution

Over time, natural selection is the main mechanism to choose the beneficial and optimized proteins from the large pools of random genetic mutations within an organism. Mimicking this process, directed evolution, which involves interactive rounds of mutations at the genetic level followed by screening or selection for desired properties at the protein level, becomes a powerful tool in protein engineering to improve functionality, selectivity, and stability of proteins [1]. Although the concept of directed evolution is simple and straightforward, it requires efficient methods to search for the functional sequences in the vast space of possible sequences in which majority of them are deleterious or nonfunctional.

The size of protein sequence space is determined by the number of amino acids presented in a protein of interest. For example, the protein that contains N residues has a total number of possible sequences of 20^N (20 natural amino acids). Thus, finding an optimal sequence in a vast space of possible sequences through directed evolution can be described as climbing a hill on a fitness landscape where regions of high elevation represent proteins with desirable functions or properties [2]. Iterative cycles of mutation and functional-based screening/selection further direct the search to higher elevations on the fitness landscape. To find a sequence with a desired function, the following strategies are required: (1) using a good and functional starting sequence; (2) mutating this parent sequence to create a library of variants; (3) screening or selecting variants with improved function; (4) repeating the process until achieving variants with desired function [2].

Several procedures are used for mutagenizing the parent genes. Error prone PCR uses low fidelity polymerases and inappropriate PCR conditions to generate mutants; however, this method is not ideal if mutational space is limited [3]. Focused mutagenesis strategies make better use of mutational space and involve using oligonucleotides containing degenerate codons. However, this method requires prior knowledge (e.g. structure) of the protein being mutated. Paired with computational methods, focused mutagenesis is a powerful tool to generate mutant libraries [4]. After creating libraries, the next step is to screen/select for the functional variants in mutational space. Several display technologies and screening platforms have been used to identify desired mutants. Ribosome display uses ribosomes to translate the mutagenic library and has a large potential mutational space of up to 10^{15} library member [5]. Phage surface display involves fusion of library member proteins to the bacteriophage minor coat protein and then panning for high affinity library members using immobilized target proteins [6]. Phage display possesses a mutational space of 10^{10} library members. Yeast surface display involves expressing proteins of library members on the yeast surface, and then target protein is bound to the surface of yeast cells usually with a fluorescent protein to be used in high-throughput fluorescence activated cell sorting (FACS) [7]. Although a mutational space of yeast surface display contains around 10^8 library members, it is suitable for the eukaryotic expression system. During screening/selection, nonfunctional mutants are typically discarded, while improved functional mutants are enriched. This procedure is repeated until the mutant(s) with desired function or property is achieved.

1.2 Development of Protein-Based Imaging Tools Through Directed Evolution

A combination of directed evolution and rational design is a key method to improve functionality, stability, photophysical, and biochemical properties of proteins as imaging tools [8]. In particular, fluorescent proteins (FPs), which are derived from *Aequorea victoria* jellyfish GFP (avGFP) and other non-jellyfish proteins, have been utilized as fusion tags or markers for studying protein localization and interaction in living cells [9, 10]. In addition to fluorescent proteins, antibodies and protein scaffolds represent important tools for cellular imaging because of their ability to be genetically encoded and can recognize a wide range of antigens or molecules of interest.

Antibodies and protein scaffolds are often used as the specificity-determining region for imaging probes. The structure of antibodies is comprised of two types of peptide chains: a heavy chain and a light chain which forms the heterodimer through disulfide bonds. Although they provide specificity towards a target antigen, their large size (150 kDa), containing disulfide bonds, and triggering immunogenicity limit their applications as imaging tools [11]. As a result, a whole range of antibody fragments and other protein scaffolds, including Fabs [12], scFvs [13], nanobodies [14], and monobodies [15], have been engineered to retain the binding specificity. Fabs are the binding portion of full-length antibodies (~55 kDa), consisting of two single domains from each of the heavy and light chain. These retain the binding affinity of full-length antibodies but do not contain the signaling Fc domain. scFvs are a part of Fabs (~28 kDa) which consist of only variable domains from heavy and light chains connected via a peptide linker. scFvs are popular antibody derivatives because of their small size and high selectivity. Nanobody is a single domain antibody (~15 kDa), and monobody (~10 kDa) is an antibody single domain-like fragment derived from human fibronectin type III. They are easy to express, minimize steric hindrance, and retain the binding capabilities of their full-length parents, thus, nanobody and monobody are often used to target antigens buried within concave topologies.

With the advent of fluorescent proteins (FPs) and protein scaffolds for genetic labeling of biomolecules, genetically encoded biosensors have revolutionized research on cell biology, by allowing researchers to visualize and quantify a wide range of biological substances and events in living cells with high spatiotemporal resolutions [16-20]. Typically, biosensors consist of two main components: (1) a protein scaffold, which acts as a recognition module for the target analyte; and (2) a fluorescent transducer, which can convert the protein-analyte interaction into a fluorescent signal. The robustness of biosensors can be constructed toward targets of high diversity because protein scaffolds can be engineered to specifically recognize relevant biomolecules [21]. Thus, protein engineering through directed evolution has not only been applied to alter the properties of FPs [22] and protein scaffolds [23-25], but it has also been used to create novel fluorescent probes to reveal cellular processes in cell biology.

1.3 Motivation and Challenges

Fluorescent biosensors are increasingly providing insight to understand cellular processes and dynamics at a single-cell level. In the past decades, fluorescent biosensors have been extensively developed on several platforms, such as fluorescent proteins or combinations of protein scaffolds and fluorescent dyes, using fluorescence microscopy to examine cellular localization and cell signaling of important biomolecules. Moreover, these fluorescent biosensors can also be used in biomedicine for diagnostic applications. Therefore, the ultimate goal of this dissertation was to develop imaging tools, e.g. fluorescent biosensors, that can be used to elucidate molecular activities in living cells, and in biomedicine for diagnostic applications that lead to appropriate treatment.

Molecular activities typically involve protein-protein interaction and signaling cascades in cells that occur in proximity, thus Föster-resonance-energy-transfer based biosensors (FRET biosensors) are ideal for studying these systems (< 10 nm). The general construct of intramolecular FRET biosensors is simply based on modular design which consists of many variables, such as a pair of FPs, the protein sensing and binding domains, the length of linkers, and the order of each component; however, their developments still rely on trial-and-error based approach which is laborious, or structure-based design which requires further optimization. Adjusting one module affects the overall conformation of FRET biosensors. Moreover, currently available FRET biosensors usually have the limitation in sensitivity, specificity, and dynamic range that hinders their broader applications. For this reason, I would like to develop the method that can be used to systematically and universally optimize the whole FRET biosensors to report various enzymatic activities in living cells. Similar to searching for a desire mutant in fitness landscape, I utilized directed evolution approach to create biosensor variants with all possible conformations, screen for the functional ones, and identify the most optimal variant. In addition to the method development, I would like to apply directed evolution to engineer the protein with new function for applications in both cancer diagnostic and therapeutic.

1.4 Scope of Dissertation

In this doctoral dissertation, I developed the directed evolution technologies to engineer the molecular modules for FRET-based biosensors and cancer immunotherapy. In Chapter 2 (*Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface*), the directed evolution via yeast surface display library and sequence-function analysis technologies were integrated to engineer a monobody variant (PEbody) that binds to R-Phycoerythrin (R-PE) dye. R-PE binding to PEbody system was used as a FRET acceptor for a donor, enhanced cyan fluorescent protein (ECFP), to increase signal-to-noise ratio of the biosensor at a single cell surface. This hybrid FRET biosensor revealed new distinct subcellular features of cancer cell invasion with high spatiotemporal resolution. Then, I further applied the PEbody as a tool for therapeutics that I presented in Chapter 3 (*PEbody CAR: A Universal Chimeric Antigen Receptors for T-Cell Immunotherapy*). The PEbody, which was evolved through directed evolution, was utilized as a universal receptor of chimeric antigen receptors (CAR) system expressed on T cells for cancer therapy. Anti-PE–directed CAR T cell (PEbody CAR) is specific to PE-conjugated antibodies that target to diversities of antigens on both hematologic and solid tumors *in vitro*. The ability to control the interaction between PEbody CAR T cells and cancer cells via intermediate molecules, e.g. PE-conjugated antibodies, is crucial to enhance T cell response and specificity. In addition to engineer CAR T cells for cancer therapy, developing reporters for kinase activities in T cells is also needed because understanding the signaling pathway of CAR T cells during tumor engagement could improve the treatment. Thus, I chose to develop the FRET-based biosensor to monitor kinase activities, e.g. Fyn and Zap70 which are important for initiation of T cell receptor signaling pathways. Even though FRET biosensors provide high spatiotemporal resolution to monitor protein-protein interaction in living cells, their sensitivity and specificity are generally low which hinders their applications. Therefore, in Chapter 4 (*Optimization of The Whole FRET Biosensors for Kinase Activity Through Directed Evolution*), I have developed a systematic and universal method to optimize the whole FRET biosensors by combining the directed evolution via mammalian cell library, counter selection by flow cytometry-based FRET, and high-throughput sequencing technologies. As a proof of concept, a phosphorylation-sensing substrate of the Fyn FRET biosensor was engineered to be more specific and sensitive to the Fyn kinase activity. This biosensor screening template could also be easily modified and extended to improve any FRET biosensor to detect enzyme-based posttranslational modifications. Last, I will summarize the current work and possible future work in Chapter 5 (*Conclusion*). In conclusion, the combination of directed evolution, amplicon sequencing, and sequence-function analysis is a promising tool to optimize molecular binders and FRET biosensors for biomedical diagnostic and cancer immunotherapy.

Acknowledgements

Chapter 1, in part, has been submitted for publication as a book chapter as it may appear in Wiley. P. Limsakul, C. Man, Q. Peng, S. Lu, and Y. Wang. “Development of Novel Cellular Imaging Tools using Protein Engineering”. Protein Engineering. Tools and Applications, UK. Wiley. The dissertation author was the first author of this paper.

References

1. Arnold, F.H., *Design by directed evolution.* Accounts of chemical research, 1998. **31**(3): p. 125-131.
2. Romero, P.A. and F.H. Arnold, *Exploring protein fitness landscapes by directed evolution.* Nature reviews Molecular cell biology, 2009. **10**(12): p. 866.
3. McCullum, E.O., B.A.R. Williams, J.L. Zhang, and J.C. Chaput, *Random Mutagenesis by Error-Prone PCR.* In Vitro Mutagenesis Protocols, Third Edition, 2010. **634**: p. 103-109.
4. Rosenfeld, L., M. Heyne, J.M. Shifman, and N. Papo, *Protein Engineering by Combined Computational and In Vitro Evolution Approaches.* Trends in Biochemical Sciences, 2016. **41**(5): p. 421-433.
5. Hanes, J. and A. Pluckthun, *In vitro selection and evolution of functional proteins by using ribosome display.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(10): p. 4937-4942.
6. Mccafferty, J., A.D. Griffiths, G. Winter, and D.J. Chiswell, *Phage Antibodies - Filamentous Phage Displaying Antibody Variable Domains.* Nature, 1990. **348**(6301): p. 552-554.
7. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nature Biotechnology, 1997. **15**(6): p. 553-557.
8. Rodriguez, E.A., R.E. Campbell, J.Y. Lin, M.Z. Lin, A. Miyawaki, A.E. Palmer, X. Shu, J. Zhang, and R.Y. Tsien, *The growing and glowing toolbox of fluorescent and photoactive proteins.* Trends in biochemical sciences, 2017. **42**(2): p. 111-129.
9. Tsien, R.Y., *The green fluorescent protein*. 1998, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.
10. Shaner, N.C., R.E. Campbell, P.A. Steinbach, B.N. Giepmans, A.E. Palmer, and R.Y. Tsien, *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nature biotechnology, 2004. **22**(12): p. 1567-1572.
11. Levin, D., B. Golding, S.E. Strome, and Z.E. Sauna, *Fc fusion as a platform technology: potential for modulating immunogenicity.* Trends in Biotechnology, 2015. **33**(1): p. 27-34.

12. Dennis, M.S., H.K. Jin, D. Dugger, R.H. Yang, L. McFarland, A. Ogasawara, S. Williams, M.J. Cole, S. Ross, and R. Schwall, *Imaging tumors with an albumin-binding Fab, a novel tumor-targeting agent.* Cancer Research, 2007. **67**(1): p. 254-261.

13. Mazzocco, C., G. Fracasso, C. Germain-Genevois, N. Dugot-Senant, M. Figini, M. Colombatti, N. Grenier, and F. Couillaud, *In vivo imaging of prostate cancer using an anti-PSMA scFv fragment as a probe.* Scientific Reports, 2016. **6**.

14. Chakravarty, R., S. Goel, and W.B. Cai, *Nanobody: The "Magic Bullet" for Molecular Imaging?* Theranostics, 2014. **4**(4): p. 386-398.

15. Limsakul, P., Q. Peng, Y. Wu, M. Allen, J. Liang, A. Remacle, T. Lopez, X. Ge, B. Kay, H. Zhao, A. Strongin, X. Yang, S. Lu, and Y. Wang, *Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface.* Cell Chem Biol., 2018. **25**(3): p. 1-10.

16. Frommer, W.B., M.W. Davidson, and R.E. Campbell, *Genetically encoded biosensors based on engineered fluorescent proteins.* Chemical Society Reviews, 2009. **38**(10): p. 2833-2841.

17. Okumoto, S., A. Jones, and W.B. Frommer, *Quantitative Imaging with Fluorescent Biosensors.* Annual Review of Plant Biology, Vol 63, 2012. **63**: p. 663-706.

18. Tsien, R.Y., *Nobel lecture: constructing and exploiting the fluorescent protein paintbox.* Integrative Biology, 2010. **2**(2-3): p. 77-93.

19. Wang, H.X., E. Nakata, and I. Hamachi, *Recent Progress in Strategies for the Creation of Protein-Based Fluorescent Biosensors.* Chembiochem, 2009. **10**(16): p. 2560-2577.

20. Zhang, J., R.E. Campbell, A.Y. Ting, and R.Y. Tsien, *Creating new fluorescent probes for cell biology.* Nature Reviews Molecular Cell Biology, 2002. **3**(12): p. 906-918.

21. Tamura, T. and I. Hamachi, *Recent Progress in Design of Protein-Based Fluorescent Biosensors and Their Cellular Applications.* Acs Chemical Biology, 2014. **9**(12): p. 2708-2717.

22. Ai, H.W., M.A. Baird, Y. Shen, M.W. Davidson, and R.E. Campbell, *Engineering and characterizing monomeric fluorescent proteins for live-cell imaging applications.* Nature Protocols, 2014. **9**(4): p. 910-928.

23. Hattori, T., D. Lai, I.S. Dementieva, S.P. Montano, K. Kurosawa, Y.P. Zheng, L.R. Akin, K.M. Swist-Rosowska, A.T. Grzybowski, A. Koide, K. Krajewski, B.D. Strahl, N.L. Kelleher, A.J. Ruthenburg, and S. Koide, *Antigen clasping by two antigen-binding sites of an exceptionally specific antibody for histone methylation.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(8): p. 2092-2097.

24. Hattori, T., J.M. Taft, K.M. Swist, H. Luo, H. Witt, M. Slattery, A. Koide, A.J. Ruthenburg, K. Krajewski, B.D. Strahl, K.P. White, P.J. Farnham, Y.M. Zhao, and S. Koide, *Recombinant antibodies to histone post-translational modifications.* Nature Methods, 2013. **10**(10): p. 992-+.

25. Limsakul, P., Q. Peng, Y. Wu, M.E. Allen, J. Liang, A.G. Remacle, T. Lopez, X. Ge, B.K. Kay, and H. Zhao, *Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface.* Cell chemical biology, 2018.

Chapter 2

Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface

2.1 Introduction

Numerous genetically encoded intramolecular FRET biosensors have been developed to visualize various molecular events in live cells with high spatiotemporal resolution [1, 2]. However, improper protein folding frequently causes biosensor entrapment within cellular compartments [3], particularly when these biosensors are targeted through the secretory pathways to extracellular membrane to monitor local molecular events. The entrapped biosensors typically do not respond to stimuli as designed and often contribute significantly to unwanted signal noise during analysis. To overcome this deficiency, we sought to develop a hybrid FRET biosensor consisting of an enhanced CFP (ECFP; donor) and an R-Phycoerythrin (R-PE; acceptor) to be assembled *in situ* at the extracellular surface of plasma membrane. Since R-PE is a cell-impermeable fluorescent dye with a high extinction coefficient and large Stokes shift [4], the ECFP/R-PE pair is expected to provide strong FRET signals specifically at the plasma membrane with minimal intracellular background noise. However, R-PE cannot be genetically encoded [5]. Therefore, a protein scaffold fused to ECFP is needed to capture R-PE for FRET functionality.

Directed evolution technology is a powerful tool used to engineer protein domains and scaffolds, particularly when rational design alone is insufficient [6]. This technology has been used to develop numerous fluorescent proteins with improved properties including enhanced brightness, modified spectra, and increased photo-stability [7-9]. Directed evolution and rational design based on sequence and structure information have also been applied to optimize the sensing components or linker lengths for genetically encoded FRET biosensors [10-12].

Several protein scaffolds have been successfully optimized by directed evolution for different applications, including diagnostics [13], therapeutics [14], and imaging [15]. Among these, a short 94-residue monobody (Figure 2.1A), derived from the tenth type III domain of human fibronectin, is a versatile non-antibody protein scaffold with a structure similar to the immunoglobulin heavy chain domain [16]. The seven β-strands of the monobody can be randomized to create libraries of variants for protein binding sites [16, 17], with the BC and FG loops proximally positioned to form a binding interface for target biomolecules with high flexibility and affinity [16, 18].

Utilizing directed evolution and sequence-function analysis, we developed a monobody variant, PEbody, which serves as a specific binding partner for R-PE. The multivalent interaction between PEbody and R-PE significantly enhances signals at the cell-cell contact, allowing the precise monitoring of the dynamic formation and dissociation of cell-cell contacts. We have further applied PEbody for the assembly of a new ECFP/R-PE hybrid FRET biosensor at the extracellular surface of cancer cells to monitor the proteolytic activity of MT1-MMP, which is a key molecule regulating pericellular matrix degradation during cancer metastasis [19-24]. The results revealed that MT1-MMP is differentially regulated depending on the maturity of cell-cell contacts, with high and low proteolytic activities at loose and stable cell-cell contacts, respectively. Thus, our hybrid biosensors can serve as potent tools to study the dynamics of molecular activities at the cell surface with subcellular resolutions.

2.2 Materials and Methods

2.2.1 Yeast Strains and Growth Conditions

Saccharomyces cerevisiae EBY100 (a GAL1-AGA1::URA3 ura3-52 trp1 leu2Δ1 his3Δ200 pep4::HIS2 prb1Δ1.6R can1 GAL) was used throughout the work for yeast surface display library [25]. EBY100 was routinely cultured at 30 °C in rich media (YPD). Once transformed with the yeast display plasmid pYD1 (Thermo Fisher Scientific, Cat. No. V835-01), cells were grown in synthetic complete medium minus tryptophan (SC-Trp with 2% (w/v) glucose), and monobody library expression was induced in galactose media (SC-Trp with 2% (w/v) galactose). The pYD1 plasmid also carries ampicillin resistance for amplification in *E. coli* (DH5α; Thermo Fisher Scientific, Cat. No. 18258012), which was cultured in Luria Bertani (LB; Fisher Scientific, Cat. No. BP14262) medium at 37 °C.

2.2.2 Cell Lines

HEK293 cells (ATCC, Cat. No. CRL-1573, Gender: N/A) and HeLa cells (ATCC, Cat. No. CCL-2, Gender: female) were grown in DMEM (Thermo Fisher Scientific, Cat. No. 11995073) supplemented with 10% FBS (Thermo Fisher Scientific, Cat. No. 10438026) plus 100U/ml penicillin, 100 μg/ml streptomycin (Thermo Fisher Scientific, Cat. No. 15140122) at 37°C with 5% CO_2.

2.2.3 Library Construction

Two generations of R-PE-binding monobody libraries were created by site-saturation mutagenesis. For the first-generation libraries, five codons in the BC (residues 26-30) and FG loops (residues 77-81) of the G9 monobody underwent site-saturation mutagenesis by using NNK degenerate primers (IDT) with Q5 DNA polymerase (NEB, Cat. No. M0491), where N represents an equimolar distribution of A, T, G, and C; K denotes an equimolar distribution of T and G, to yield the BC and FG libraries. After generating the library by PCR, the fragments containing library were then extracted from agarose gel, inserted into pYD1 vector between the KpnI and XhoI restriction sites, transformed into *E. coli* (DH10B), and purified plasmid library with Qiagen HiSpeed Plasmid Maxi kit. These plasmids were then transformed into *Saccharomyces cerevisiae* (EBY100). The libraries were screened twice by FACS (BD FACSAria) using R-PE (Sigma-Aldrich, Cat. No. 52412-1MG-F) as a ligand (Figures 2.2B-C). The clones with the best binding capability and fluorescence intensity were combined to generate the R-PE-binding monobody library.

To construct the combined library, the selected DNA libraries of BC and FG loops were amplified by the primer P1 (forward) and P2 (reverse) to produce the BC library, and the primer P3 (forward) and P4 (reverse) to generate the FG library. P1 anneals to a region in the G9 sequence upstream of the randomized BC loop; whereas P2 anneals to a region between the BC and FG loop. P3 was complementary to the 20 nucleotides at the 5' end of P2; P4 anneals to a region downstream of the FG loop. These two DNA fragments were assembled through the overlap extension PCR with a 20-nucleotide sequence complementary. The combined library was then transformed into *S. cerevisiae* jointly with the original pYD1 vector which has a complementary sequence of G9 upstream of the BC loop and downstream of the FG loop. The size of the resulting R-PE-binding monobody library was approximately 1×10^6 individual clones.

2.2.4 Selection of the R-PE-Binding Monobody Using FACS

The surface expression of the variant proteins was induced by culturing yeast in 2% galactose-containing medium (synthetic complete medium minus tryptophan, SC-Trp) at 20°C for 24–48 hours during which the OD600 increased 2-4-fold. In order to identify and select the yeast cells that displayed the G9 monobody variants with the high R-PE-binding capability, the yeast cells were incubated with 10 µg/mL R-PE in wash buffer containing the phosphate buffered saline (PBS, Sigma-Aldrich) with 0.5% bovine serum albumin (BSA) (Sigma-Aldrich) for 1 hour at room temperature. The level of the expressed proteins on the yeast surface was also measured using the anti-V5 epitope antibody (Thermo Fisher Scientific, Cat. No. R960-25). In each step of staining, the unbound reagents were washed away with wash buffer.

Each of the BC and FG libraries were sorted twice by FACS using R-PE as a ligand. In the first round of sorting, the top 10% of the R-PE positive cells were collected for further expansion. In the second round of sorting, 0.1% of the R-PE positive cells were selected (Figures 2.2B and 2.2D). Later, the R-PE-binding monobody library was constructed by combining both selected BC and FG libraries. For the combined library, the top 10%, 0.5%, and 0.1% of the R-PE binders were consecutively screened in three rounds of FACS. After the third round of sorting, the cells, which were most efficient in binding to R-PE, were selected and seeded into wells of 96-well plates for DNA sequencing (Genewiz).

2.2.5 Improvement of the R-PE-Binding Monobody

After three rounds of FACS, the S4 monobody variant, which efficiently binds to R-PE, was identified. To further improve the R-PE-binding monobody, the second generation of the S4-based monobody library was generated by site-saturation mutagenesis. The sequence in the BC loop of S4 was fixed; while the sequence of the FG loop was subjected to site-saturation mutagenesis using the NNK degenerate primers.

To increase the selection stringency of the R-PE binders, only the top 0.1% of the brightest R-PE positive cells were selected in the first and second round of FACS. In the second round of FACS, cells were further seeded into wells of 96-well plates. 20 single colonies were then randomly selected to measure their R-PE-binding capability by flow cytometer (BD Accuri C6). Later, ten individual clones, including five clones of the highest and lowest R-PE-binding capability, were subjected to the DNA sequence analysis. The R-PE binding capability was defined as the ratio between the % of R-PE binding yeasts (gated by R-PE) and the % of monobody expressing yeast (gated by a labeling tag V5 on the monobody).

The DNA sequences of S4 and those 10 representative clones, were used to predict the mutants with the improved R-PE-binding capability (Figure 2.2G). The amino acids in the FG loop of each clone were evaluated based on the results from flow-cytometry analysis. The contribution score of the amino acid residue A at position j in the FG loop was calculated by averaging the relative R-PE binding capability among all mutants of A at j. If the contribution score of A at position j was higher than 10% of the maximum contribution score at that position, amino acid residues with the highest ranking for this position were then selected for further examinations (Figures 2.2H-I). As a result, amino acids W, R/P, W/F, W, and Y/N for positions 77-81 of the FG loop sequence were selected for experimental examination. A combination of these residues led us to the eight individual mutant sequences (Figure 2.1D). The R-PE binding capability of these mutants were later examined by flow cytometry (BD Accuri C6). Data from flow cytometry were collected from about 100,000 events per sample with three independent experiments. Our approach greatly reduced the searching space and provided an efficient path to reach an optimized R-PE binder.

2.2.6 Measuring K_D by Flow Cytometry

The protein-protein dissociation constant K_D of different monobody variants was measured using the yeast surface display as described [26]. Antigen concentration ranged from 10-fold lower to 10-fold higher than the expected K_D; concentrations from 0.5 nM to 50 nM were used to label 2×10^6 induced yeast cells which were incubated at room temperature for 1 hour. The unbound R-PE was then washed away, and the samples were analyzed by flow cytometry. The mean fluorescence intensity (MFI) of the R-PE bound to the yeast cell surface was plotted against the increasing R-PE concentration. The K_D was calculated from nonlinear least-squares regression of the relationship between the MFI and the R-PE concentration (Figure 2.2J).

The dissociation constant (K_D) of the PEbody and R-PE using a yeast display method by flow-cytometry analysis [26] was calculated to be 9.6 ± 3.0 nM (mean±SEM). Similarly, the K_D of anti-V5 antibody toward V5 epitope tag, which is located at C-terminus of PEbody as an indicator of protein expression on the yeast surface (Figure 2.2C), was measured to be about 2 nM (Figure 2.2J), which is consistent with previously reported values [27].

2.2.7 Measuring K_D by Bio-Layer Interferometry

PEbody proteins were expressed with N-terminal 6x His tag in *Escherichia coli* and purified by nickel chelation chromatography as previously described [1]. Binding kinetics of PEbody and R-PE were measured by bio-layer interferometry (ForteBio BLItz system). PEbody with His tag was loaded onto an anti-HIS biosensor for 15 sec to establish baselines. A variety of concentrations of R-PE (40, 80, 160, and 320 nM) were introduced and their association to immobilized PEbody was monitored for 60 sec, after which it was allowed to dissociate into PBS with 0.5% BSA (pH 7.4) for 60 s. K_D values were calculated from the 1:1 binding model to be 5.7 ± 3.5 nM (mean ± SEM) by determined k_{on} and k_{off} parameters.

2.2.8 Construction of the MT1-MMP FRET Biosensor

The MT1-MMP substrate peptide sequence (CRPA<u>HL</u>RDSG, the scissile bond is underlined) flanked by the GGSGGT linker peptides was inserted between the C-terminus of ECFP and the N-terminus of PEbody (GGSGGTCRPAHLRDSGGGTGGS) [28]. The flexible linker (GGSGGT) was added to extend the cleavable peptide and to allow its access by MT1-MMP. The biosensor sequence was then inserted to the pDisplay vector (Thermo Fisher Scientific, Cat. No. V66020) between the BglII and SalI restriction sites. The pDisplay vector contains an N-terminal murine Ig κ-chain leader sequence, which directs the protein cargo to the secretory pathway, and a C-terminal transmembrane (TM) domain of the platelet-derived growth factor receptor β (PDGFR-β), which directs target proteins to the extracellular surface of the plasma membrane [29]. The mCherry-conjugated MT1-MMP was constructed by fusing mCherry sequence at the C-terminal of MT1-MMP with GGS and inserted to pcDNA3.1 between HindIII and XhoI restriction sites [29].

2.2.9 Cell Culture and Transfection

Cell culture reagents were purchased from Thermo Fisher Scientific. The human embryonic kidney cells (HEK293) and the cervical cancer cells (HeLa) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM of L-glutamine, 100 units/mL of penicillin, 100 µg/mL of streptomycin, and 1mM of sodium pyruvate at 37°C with 5% CO_2. The DNA plasmids were transfected into cells with Lipofectamine 3000 (Thermo Fisher Scientific, Cat. No. L3000015). To visualize MT1-MMP activity in HeLa cells, the ECFP/PEbody MT1-MMP FRET biosensor was co-transfected with the full-length MT1-MMP gene. 36-48 hours after transfection, R-PE (10 µg/mL) was added into the culture medium and incubated with the cells for 30 minutes. The unbound R-PE was then washed out and the samples were analyzed by fluorescence microscopy.

2.2.10 Microscopy, Image Acquisition, and Analysis

Cells expressing the exogenous biosensor proteins were cultured in a glass bottom dish (Cell E&G) coated with 20 µg/mL fibronectin for 16-20 hours before imaging. Cells were then incubated with R-PE in serum-free DMEM without phenol red at 37°C for 30 min, washed twice using PBS, and maintained in the starvation medium (DMEM with 0.5% FBS). The images were taken with a Nikon Eclipse Ti inverted microscope with a cooled charge-coupled device camera and analyzed using MetaFluor 7.8 or MetaMorph 7.8 software (Molecular Devices). The FRET ratio images were obtained by dividing the fluorescence image pixel-by-pixel of FRET wavelength (420/40 nm excitation, 575/20 emission, with 560 LP dichroic mirror) by the R-PE wavelength (495/10 nm excitation, 575/20 emission, with 560 LP dichroic mirror) after the background subtraction. The images were analyzed and displayed in the intensity modified display mode by using MetaFluor and Fluocell software [30, 31].

For data presentation, the normalized values were shown to clearly compare the differences among the experimental groups and to minimize the cell-cell heterogeneity. The pre-stimulation baseline for each cell was established by averaging the FRET or other fluorescent signals of each individual cell before stimulation. This provides an internal normalization reference for establishing the stimulation-dependent FRET changes of an individual cell which reduces the cell-cell heterogeneity and noise. This normalization procedure had been used for Figures 2.6, 2.8, and 2.10. In Figure 2.5, the normalization was calculated by dividing the intensity at each time point of an individual cell with the intensity when that cell contact was completely dissociated (minimum R-PE intensity was observed during time = 50-60 minutes after addition of EGTA).

2.2.11 Dissociation and Formation of Cell-Cell Contacts

For R-PE-label experiment, R-PE (10 µg/mL) was directly added onto HEK293 cells expressing ECFP-PEbody under microscope. For cell-cell contact dissociation experiments, HEK293 cells expressing ECFP-PEbody were stained with R-PE (10 µg/mL) and cell-cell junctions were disrupted by the addition of 2 mM EGTA (Sigma-Aldrich) to monitor the change of R-PE and ECFP signal at cell-cell contacts. Once EGTA was washed away, cells started to reform the junctions. During this phase, R-PE signal was monitored together with ECFP signal at reforming cell-cell contacts.

2.2.12 Inhibitor Washout FRAP Assay

For inhibitor washout assays, HeLa cells were pre-incubated with 20 µM GM6001 (EMD Millipore, Cat. No. 364205), a broad-spectrum matrix metalloproteinase inhibitor, in the starvation medium for 16-20 hours. The inhibitor was then washed out with the Hank's balanced salt solution (HBSS, with Ca^{2+} and Mg^{2+}, Thermo Fisher Scientific). For FRAP experiments, cells expressing the MT1-MMP-mCherry fusion protein were imaged using a 100x objective as a reference for setting the pinhole position which allows strong light to pass through. Photobleaching was performed by exciting mCherry in the chosen pinhole region with the full power of excitation light for 1 minute. Cells were then monitored for the recovery of the fluorescence intensity. To estimate the diffusion coefficients, our previous published finite element mathematical model and algorithm were used [31, 32]. This method is flexible to handle different imaging and photobleach protocols, as well as variable cell geometry. Briefly, two images from the recovery time course were used to estimate a diffusion coefficient, based on a linear correction between the temporal and spatial variations of fluorescence signal. The estimated diffusion coefficient was considered to be sufficiently accurate when the correlation calculated from a pair of images is larger than 0.7 [31].

2.2.13 Quantification and Statistical Analysis

All the experiments were replicated at least three times and represented biological replicates. Data are presented as mean values ± standard deviation (SD) or mean values ± standard error of the mean (SEM) as indicated in the figure legend. Statistical analyses were carried out using MATLAB (version R2015b) and were defined by P values calculated from the two-tailed Student's t tests or Wilcoxon rank sum test by comparison with the relevant control as specified in the figures or figure legends. Levels of significance were $P<0.05$ (*), $P<0.01$ (**), $P<0.001$ (***). When the sample size $n>=4$, the Wilcoxon rank sum test was used since it is applicable to random samples of all distributions. For some flow cytometry experiments, when $n=3$, each data point is an average reading from more than 100,000 single cells. According to the central limit theorem, these data follow the Gaussian distribution. Therefore, we used the two-tailed Student's t-tests.

2.3 Results

2.3.1 Directed Evolution of an R-PE Binder

We first applied directed evolution technology to engineer a monobody variant for the capture of R-PE and subsequent *in situ* assembly of a FRET biosensor at the extracellular surface. The G9 monobody, initially engineered to bind to the Src Homology 3 (SH3) of Fyn [33], was used as a template to generate mutant libraries by applying site-saturation mutagenesis [34] in the BC (residues 26-30) and FG (residues 77-81) loops of G9 (Figures 2.1A and 2.2A). These libraries were displayed on the yeast surface and used to screen for R-PE binders (Figure 2.1B and Figure 2.2B). After three rounds of selection by fluorescence-activated cell sorting (FACS), a varaint named Single4 (S4) containing the FINFK and WRWWY sequences in the BC and FG loops, respectively, was identified as an efficient R-PE binder (Figures 2.1C and 2.2F). The R-PE binding capability of S4, which is defined as the ratio between the % of R-PE binding yeast (gated by R-PE) and the % of monobody expressing yeast (gated by a labeling tag V5 on the monobody, Figure 2.2C), was 500-fold greater than the original G9 (Figure 2.1C), but was not further improved by additional rounds of cell sorting. It is therefore possible that the S4 sequence represents a locally optimal solution around the neighboring mutation space. Replacing a single BC or FG loop of G9 with that of S4, however, was not sufficient to increase the binding capability, suggesting that mutations on both loops are necessary for the overall enhanced R-PE-binding of S4 (Figure 2.1C).

Based on the structure and sequence of S4, we found that the FG variants consistently showed larger improvement on target binding capability than the BC variants (Figures 2.2D-E) [17]. Consequently, it is possible to improve R-PE binding capability of S4 by further optimizing sequences in the FG loop of S4 while holding the BC loop constant. Additional mutagenesis was focused on the FG loop of S4 to create a new library (Figure 2.2F). However, no variant showed significant improvement of R-PE binding capability comparing to S4 after two more rounds of cell sorting (Figure 2.2G). Therefore, a rational approach was explored to further engineer and improve S4 for R-PE binding.

We surveyed the mutation space around the S4 monobody sequence by choosing 10 variants for sequence-function analysis, including 5 clones each with the highest and lowest binding affinities toward R-PE (Figure 2.2G). The high-affinity sequences (along with S4) were used to identify the residues contributing positively to R-PE-binding, while the low-affinity sequences allowed us to identify and exclude residues with negative contribution. Amino acids (AAs) at each position were ranked by their contribution scores. For example, the contribution score of amino acid A at the position *j* was calculated by averaging the R-PE binding capability among all mutants containing A at position *j* (see Method, Figures 2.2H-I). We then selected the highest ranked AAs for each position. This selection resulted in eight variant sequences, including S4 (Figure 2.1D). The flow cytometry results showed that two mutants, M4 and M5, have significantly improved binding capability toward R-PE compared to S4 (Figure 2.1D). The mutant M4 was named as PEbody, having an R-PE-binding capability of 96%, and was used for all of the following experiments. The mutant M5 had similar R-PE-binding capability but less efficiency in protein production and hence was not further tested. Comparing the amino acid sequences of PEbody and S4 revealed that replacing arginine with proline at position 78 in S4 caused the increase in R-PE binding for PEbody (Figures 2.1D).

To further characterize the specificity and binding affinity of PEbody in recognizing R-PE, we stained the PEbody-expressing yeast cells with various fluorescent dyes, such as PerCP-Cy5.5, FITC, Alexa Fluor 488, and R-PE with streptavidin, that have been extensively used in flow cytometry applications. Only R-PE-based dyes could bind to PEbody, indicating that it is specific to R-PE (Figure 2.1E). The binding affinity between PEbody and R-PE was further characterized by using biolayer interferometry (BLI), with the K_D determined to be 5.7 ± 3.5 nM (mean±SEM) (Figure 2.1F) which is consistent with the K_D calculated using the yeast surface display [26]. The mean fluorescence intensity (MFI) of the R-PE bound to the yeast cell surface was plotted against increasing R-PE concentrations. We estimated the K_D to be 9.6 ± 3.0 nM (mean±SEM) using nonlinear least-squares regression of the relationship between the MFI and the R-PE concentration (Figure 2.2J). These results suggest that PEbody suffices as an efficient and specific R-PE binder.

2.3.2 Characterization of PEbody in Mammalian Cells

PEbody was then displayed and characterized for R-PE-binding at the surface of human embryonic kidney cells (HEK293, Figure 2.3A). R-PE staining was found to be specific in cells expressing PEbody at the surface, especially efficient at the cell-cell contacts, but not in those expressing the non-specific G9 monobody or lacking the construct altogether (Figures 2.3B and 2.4A). These results further demonstrate that PEbody can be expressed in mammalian cells to serve as a specific binder of R-PE.

Because R-PE effectively binds to PEbody expressed on the cell surface, the R-PE/PEbody system can be utilized as a new acceptor for FRET biosensor applications. ECFP was hence fused with PEbody to create a FRET-based biosensor and allow energy transfer from ECFP to R-PE upon its capture by PEbody (Figures 2.4B-C), with the length of the linker between ECFP and PEbody affecting the FRET efficiency (Figure 2.4D). Indeed, upon the addition of R-PE to the cell culture medium, the R-PE signal dramatically increased at cell surfaces particularly at the cell-cell contacts, while exhibiting a corresponding ECFP signal decreased due to its energy transfer to R-PE (Figures 2.4E-F). The fluorescence images of HEK293 cells also revealed the expression of ECFP/PEbody and clear R-PE staining at the cell surface, with relatively strong signals at the cell-cell contacts compared to cell edges without contacts (Figure 2.3C, middle panel). We noticed that the R-PE signal at the cell-cell contact was 3.6-fold brighter than that at the free end, while the ECFP signal only increased 1.4-fold, which reflects the accumulated copies of local PEbody (Figure 2.3D). These results reveal high enrichment of R-PE at local sites when multiple PEbodies are present. We hypothesize that this phenomenon is due to the multivalent interaction between R-PE and PEbody. In contrast to the efficient binding between PEbody and R-PE, which resulted in a strong FRET between R-PE and ECFP, we observed no binding of R-PE in the cells expressing the control ECFP-G9 construct (Figure 2.3C, top panel). Consistently, cells without PEbody did not gain R-PE signals at the cell surface upon R-PE addition in the cell culture medium (Figure 2.4F, red broken line), verifying the specificity of PEbody in R-PE binding.

Interestingly, fusing PEbody at the N-terminus of ECFP showed weaker R-PE staining, indicating that the position and orientation of ECFP is important for the accessibility of PEbody to R-PE (Figure 2.3E). Taken together, these results indicate that the fusion of PEbody at the C-terminus of ECFP provides a platform for the development of biosensors based on FRET between ECFP and R-PE to visualize active molecular events at extracellular surface.

2.3.3 Visualization of Cell-Cell Contact Dissociation and Formation

Due to the multivalent interaction between R-PE and PEbody at cell-cell contacts, our hybridized assembly of ECFP/PEbody and R-PE could be ideally suited for monitoring cell-cell contact dissociation and formation. We performed an experiment in which cell-cell contacts between HEK293 cells were first disrupted with EGTA for 1 hour while the change of fluorescence signals was monitored. As shown in Figures 2.5A-B, cells started to contract upon the addition of EGTA, causing the local morphological thickening of cell-cell contacts and a slight, transient increase in R-PE signal, followed by sharp decrease when cells started to dissociate physically. Once EGTA was washed away, cells started to reform the junctions. During this phase, R-PE signal significantly increased while ECFP intensity was less changed (Figures 2.5C-D). Overall, the results showed a rapid increase in the junctional R-PE signal when cells form cell-cell contacts, and a fast decrease of the signal when cells dissociate. As such, the multivalent binding property between R-PE and PEbody at the cell-cell contacts can be used as a sensitive marker for monitoring the maturity of cell-cell junctions during the dynamic formation and dissociation processes, which is difficult to be monitored using ECFP or other genetically encoded fluorescent proteins alone.

2.3.4 Visualization of the Surface MT1-MMP Activity

To demonstrate the functionality of the ECFP-PEbody construct in engineering hybrid FRET biosensors, we inserted an MT1-MMP substrate peptide (CRPAHLRDSG) flanked by linkers in between ECFP and PEbody. R-PE can then be applied to produce an ECFP/PEbody/R-PE hybrid FRET biosensor to visualize the *in situ* substrate cleavage by MT1-MMP at the extracellular surface (Figure 2.6A) [28, 35]. The functionality of the FRET biosensor was verified in MT1-MMP deficient HeLa cells transfected with full-length MT1-MMP. The cells were pretreated with MMP inhibitor, GM6001, and monitored before and after inhibitor washout [35]. After inhibitor washout, MT1-MMP proteolysis of the substrate in the biosensor allows ECFP to diffuse away from R-PE, which leads to a reduction in energy transfer. We observed a decrease in FRET/R-PE ratio of approximately 50%, while R-PE intensity remained relatively constant (Figures 2.6B-C and 2.7A). In contrast, the FRET/R-PE ratio was not affected by inhibitor washout in the control HeLa cells deficient in MT1-MMP (Figures 2.6B(ii) and 2.6C (black line)), confirming the specificity of this hybrid FRET biosensor. We have also compared this hybrid ECFP/PEbody/R-PE biosensor with a previously developed ECFP/YPet MT1-MMP FRET biosensor [28]. After we quantified the biosensor signals at the plasma membrane where MT1-MMP localizes, it is clear that the overall kinetics of MT1-MMP activities monitored by both biosensors had similar characteristics, while our hybrid biosensor produced significantly cleaner images (Figures 2.7B-D). These results indicate that our new ECFP/PEbody/R-PE hybrid biosensor provides high sensitivity and clean membrane signals with low intracellular noise.

2.3.5 Asymmetric Dynamics of the Cell-Surface MT1-MMP Activity

FRET/R-PE ratio images clearly revealed distinct kinetics of MT1-MMP activation at different subcellular locations upon inhibitor washout (Figure 2.8A). The quantified time courses of FRET/R-PE ratio showed similar basal values before GM6001 inhibitor washout (Figure 2.8B). After washout, however, the ratio values decreased at different rates in different cell-cell contacts (Figures 2.8C). To explored these differences, actin-GFP was used as an indicator to investigate different cell-cell contact types [36, 37]. We found that the rapid decrease of FRET/R-PE ratio, representing fast MT1-MMP activation, was predominantly observed at loose cell-cell contacts where actin fibers formed wide and diffusive branches (Figures 2.9A-C) [37, 38]. Conversely, stable cell-cell contacts where actin filaments accumulate in packed bundles at the cell borders [37] showed a slow decrease of FRET/R-PE ratio, representing slow MT1-MMP activation (Figures 2.9A-C). Interestingly, images of cells with mCherry-tagged MT1-MMP showed that the MT1-MMP protein localized at cell-cell contacts without particular spatial preferences (Figures 2.9D-E) [39]. Taken together, these results indicate that although the total amount of MT1-MMP is similar at both stable and loose contacts, the proteolytic activity of MT1-MMP may be inhibited at the stable cell-cell contacts by the local microenvironment.

The cytoskeleton plays a crucial role in controlling cell-cell adhesions and MT1-MMP dynamics [20-22]. Actin filaments and microtubules regulate the localization and trafficking of MT1-MMP [29, 40, 41], so we reasoned that they may also mediate the observed asymmetric distribution of MT1-MMP activity. To investigate the relationship between the MT1-MMP activity and the cytoskeletal components, cytochalasin D and nocodazole were applied to cells to disrupt actin filaments and microtubules, respectively. The application of either inhibitor resulted in a uniformly rapid decrease of the FRET/R-PE ratio at different cell-cell contacts, similar to the loose contacts in MT1-MMP-expressing HeLa cells after inhibitor washout (Figures 2.8D-E). These results indicate that actin filaments and microtubules play crucial roles in the regulation of MT1-MMP activity at different cell-cell contacts.

To further explore the spatially asymmetric regulation of MT1-MMP activity, we measured the apparent diffusion rate of MT1-MMP using the fluorescence recovery after photobleaching (FRAP) assay. The full-length MT1-MMP-mCherry fusion protein was monitored before and after photobleaching (Figure 2.10). A mathematical model was used to quantify the diffusion rate of MT1-MMP-mCherry based on the recovery images [31]. The analysis showed that the average diffusion rate of MT1-MMP-mCherry at the membrane regions lacking cell-cell contacts was 2.2-fold greater than at regions with stable cell-cell contacts (4.40±2.17 vs. 1.98±0.82 μm^2/sec) (Figures 2.8F and 2.10). These results indicate that the MT1-MMP molecules are less mobile at stable contacts, which is consistent with the results from our biosensor that showed less proteolytic activity of MT1-MMP at stable contacts. This possibly reflects its inhibited forms in the presence of inhibitory interacting molecules at these local microenvironments.

2.4 Discussion

FRET biosensors are valuable molecular imaging tools that allow for real-time visualization of specific molecular activities during important cellular processes. Through an integration of iterative directed evolution and rational design (Figure 2.1), we developed a hybrid FRET biosensor composed of ECFP and an engineered PEbody that functions as a selective and effective binder of the R-PE fluorescent dye. Because R-PE is cell-impermeable, this hybrid biosensor can be assembled *in situ* at the live cell surface to monitor extracellular molecular activities with high spatiotemporal resolution. Indeed, the binding between R-PE and PEbody can be clearly detected on the cell surface with minimal intracellular background noise, particularly at cell-cell contacts (Figure 2.3). Our design provides an alternative approach to the SNAP-tag and Halo-tag technologies [42, 43], with the advantage of allowing for the direct and reversible assembly of biosensors without additional chemical modifications. Integrating PEbody into a FRET biosensor, we successfully applied our hybrid biosensor to monitor cell-cell junction maturity during the dynamic formation and dissociation of cell-cell contacts (Figure 2.5), utilizing the multivalent binding property of R-PE toward PEbody. Therefore, the integration of the directed evolution technology and rational sequence-function analysis can be used as powerful tools to engineer protein binding scaffolds for the assembly of efficient hybrid FRET biosensors capable of monitoring subcellular molecular events at live cell surfaces (Figures 2.6 and 2.8).

Monobody is widely used as a protein scaffold because of its small size, flexible loops, and proper folding within living cells [16, 44]. Accordingly, with the G9 monobody serving as a starting template for directed evolution [33], FACS screenings directly allowed the identification of the S4 variant with improved, specific R-PE binding capability. However, additional mutagenesis and screenings failed to identify a better binder to R-PE than S4. This may be due to the nonlinear relation between the library sequence and the functional readout, as well as our limited library size [6, 45, 46]. Further improvement was only achieved by the integration of screening and rational design. Optimizing amino acid residues in both BC and FG loops resulted in the higher affinity of PEbody toward R-PE. As such, the integration of directed evolution and rational analysis provided powerful tools for the molecular engineering of novel proteins.

PEbody was integrated into a FRET-based biosensor for monitoring the proteolytic activity of MT1-MMP. We found that although the amount of MT1-MMP protein is relatively uniform at cell-cell contacts [20, 39], the proteolytic activity of MT1-MMP is surprisingly spatially heterogeneous. Perhaps the E-cadherin complexes of stable cell-cell contacts can recruit antagonistic molecules to create an inhibitory microenvironment, suppressing MT1-MMP-mediated cadherin cleavage and thereby maintaining stable junctions [47]. Consistent with this hypothesis, FRAP results indicated that the MT1-MMP molecules are less mobile at stable cell-cell contacts which could account for the lower activity of MT1-MMP. This supports the notion that MT1-MMP molecules may be more restricted by inhibitory partners at these local regions. The results from FRAP experiment also agree well with the results of cytoskeletal disruption. Cytochalasin D or nocodazole treatment eliminated this differential distribution of MT1-MMP activity at different cell-cell contacts. It is hence possible that the distinct characteristics of the cytoskeleton at different cell-cell contacts play a role in determining the local microenvironment and regulating MT1-MMP distinctively at different cell-cell contacts.

2.5 Figures

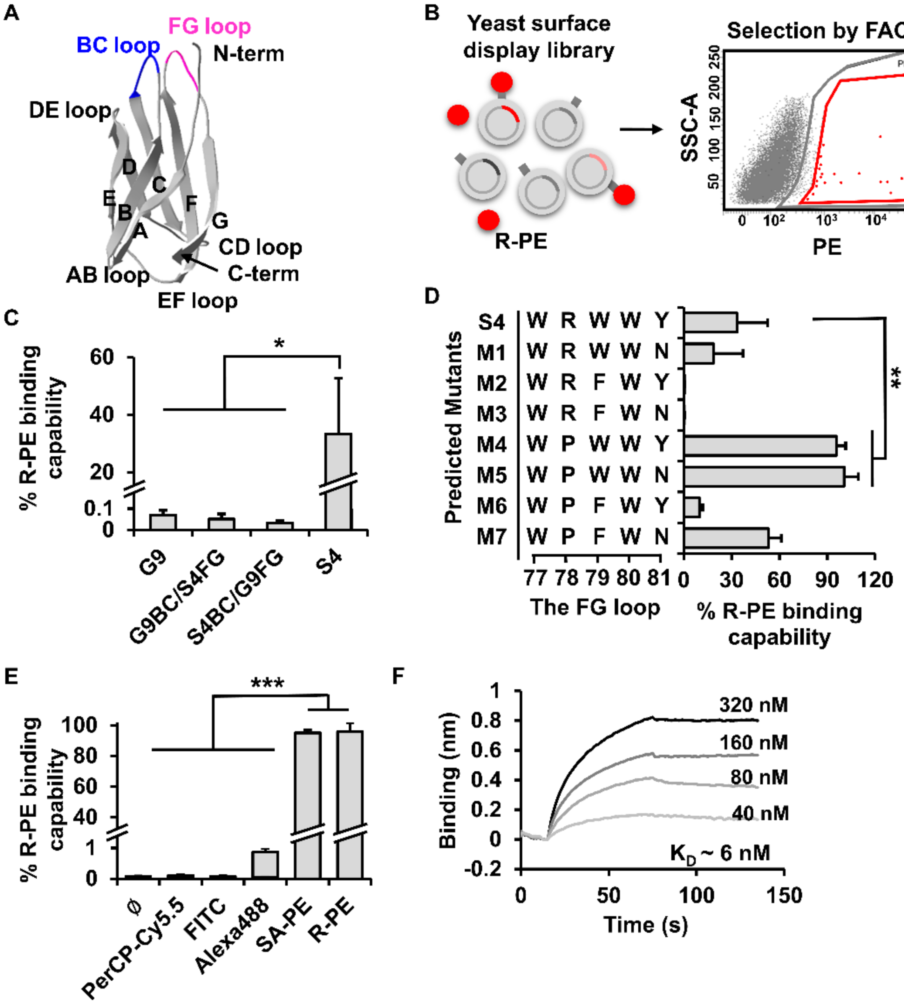
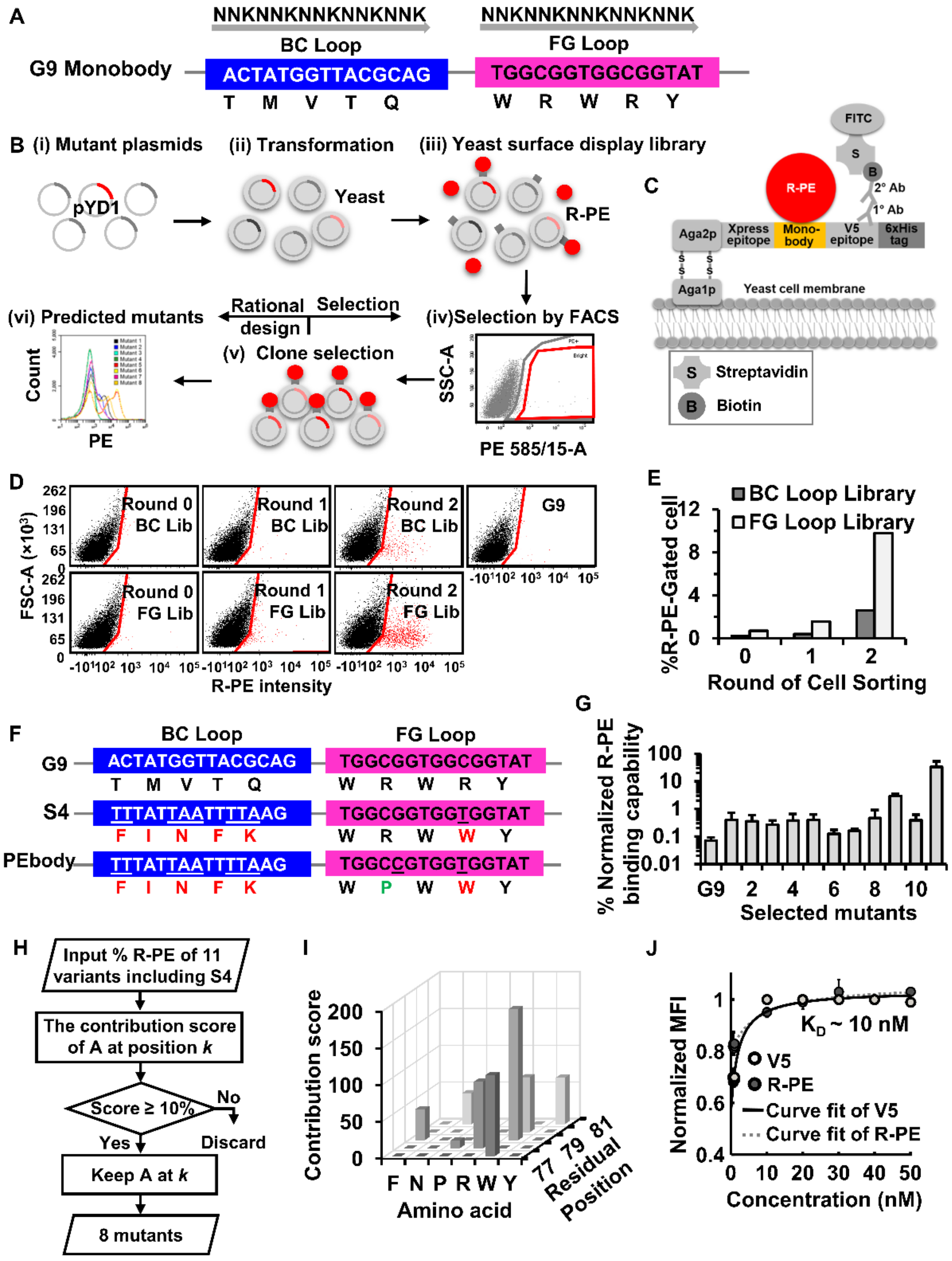


Figure 2.1: The development of PEbody

(A) The structure of the G9 monobody (modified from PDB ID: 1TTG). (B) The schematic diagram of the yeast display monobody library and the selection of the R-PE-binding monobody clones via FACS. (C) The R-PE binding capability of different monobody mutants as indicated: G9, a mutant with the FG loop of S4 (G9BC/S4FG), a mutant with the BC loop of S4 (S4BC/G9FG), and S4. The R-PE binding capability is defined as the ratio of the % of R-PE-positive yeast to the % of V5-positive yeast. The V5 epitope tag fused at C-terminus of PEbody was used as the indicator of protein expression on the yeast surface, see Figure 2.2C. (D) The improvement of R-PE-binding monobodies after further rounds of mutagenesis and sequence-function analysis. Eight mutants with different amino acid sequences in the FG loop were predicted and their R-PE binding capabilities were analyzed through flow cytometry. (E) Testing the specificity of R-PE-binding monobody. The binding capability of different dyes, including PerCP-Cy5.5, FITC, Alexa488, streptavidin-PE (SA-PE), and R-PE, to PEbodies displayed on the yeast surface was measured by flow cytometry. (F) The determination of binding affinity between R-PE and PEbody by bio-layer interferometry. Different concentrations of R-PE were used to determine k_{on} and k_{off} parameters which were used to calculate K_D values. Data in (C-E) are represented as mean ± SD. The asterisk indicates a significant difference (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$ with the two-tailed Student's t test). See also Figure 2.2.

Figure 2.2: The directed evolution strategy for identifying the R-PE-binding monobody

(A) Sequence information of the site-saturation mutagenesis by degenerate primers (NNK) on the BC and FG loops of the G9 monobody. The blue and pink boxes represent the DNA sequences in the BC and FG loops, respectively. The encoded amino acid sequence is shown under the boxes. (B) The diagram showing the directed evolution of the R-PE-binding monobody. Mutagenesis in the BC and FG loops of the G9 monobody generated a library of the mutant DNA fragments. (i-ii) These DNA libraries were inserted into the yeast expression vector (pYD1), followed by transformation of the recombinant constructs into *S. cerevisiae* (EBY100) to generate a yeast surface display library. (iii-v) The yeasts that expressed the monobody mutants were labeled by R-PE and selected by FACS. The process from (iii) to (v) was repeated until the yeasts exhibiting the enhanced R-PE-binding were identified. (vi) The DNA sequence was then analyzed to identify monobody variants with high-efficient R-PE binding. (C) Schematic representation of yeast surface display and detection strategies. The mutants of monobody are encoded between two epitope tags, including Xpress and V5 epitopes, which are in-frame with Aga2p. This construct is displayed on the yeast surface by forming disulfide bonds between Aga1p and Aga2p. The expression of protein display is detected by using antibodies, streptavidin reagents onto V5 epitope. (D) An improvement in the R-PE-binding affinity in the different rounds of cell sorting of the BC and FG libraries. (E) The comparison of the % R-PE-gated cells among the first and second cell sorting of the BC and FG libraries. (F) Nucleotide and amino acid sequences of the BC and FG loops of the G9 monobody, S4 monobody, and PEbody. The second-generation of the R-PE-binding monobody library was generated in the FG loop of S4. (G) The comparison of the normalized % R-PE-binding capability in a log scale among G9, the selected mutants from the library shown in (F), and S4. The mutants were selected from the second round of FACS screening. (H) The flow chart to predict the R-PE-binding monobody. The normalized percentage of the R-PE-binding yeasts was used to predict the sequence of the optimized R-PE binder. (I) The residues at each position in the FG loop and their contribution score calculated from (H). (J) The determination of binding affinity between R-PE and PEbody by yeast surface display. R-PE or anti-V5 antibody were used to stain yeast cells expressing PEbody. The normalized mean fluorescence intensity (MFI) was plotted against the R-PE concentration along with the least-squares regression fits of the 1:1 binding model. The errors are the standard deviations.


A
NNKNNKNNKNNKNNK
BC Loop
NNKNNKNNKNNKNNK
FG Loop
G9 Monobody
ACTATGGTTACGCAG
TGGCGGTGGCGGTAT
T M V T Q
W R W R Y
B (i) Mutant plasmids
(ii) Transformation
(iii) Yeast surface display library
pYD1
Yeast
R-PE
(iv)Selection by FACS
SSC-A
PE 585/15-A
(v) Clone selection
Rational design
Selection
(vi) Predicted mutants
Count
PE
C
FITC
S
B
2° Ab
1° Ab
R-PE
Aga2p
Xpress epitope
Mono-body
V5 epitope
6xHis tag
Aga1p
Yeast cell membrane
S Streptavidin
B Biotin
D
Round 0 BC Lib
Round 1 BC Lib
Round 2 BC Lib
G9
Round 0 FG Lib
Round 1 FG Lib
Round 2 FG Lib
FSC-A (×10³)
262
196
131
65
0
-10¹10² 10³ 10⁴ 10⁵
R-PE intensity
E
BC Loop Library
FG Loop Library
%R-PE-Gated cell
12
8
4
0
0
1
2
Round of Cell Sorting
F
BC Loop
FG Loop
G9
ACTATGGTTACGCAG
T M V T Q
TGGCGGTGGCGGTAT
W R W R Y
S4
TTTATTAATTTTAAG
F I N F K
TGGCGGTGGTGGTAT
W R W W Y
PEbody
TTTATTAATTTTAAG
F I N F K
TGGCCGTGGTGGTAT
W P W W Y
G
% Normalized R-PE binding capability
100
10
1
0.1
0.01
G9
2
4
6
8
10
Selected mutants
H
Input % R-PE of 11 variants including S4
The contribution score of A at position k
Score ≥ 10%
No
Discard
Yes
Keep A at k
8 mutants
I
Contribution score
200
150
100
50
0
F N P R W Y
Amino acid
77 79 81
Residual Position
J
Normalized MFI
1
0.8
0.6
0.4
K_D ~ 10 nM
V5
R-PE
Curve fit of V5
Curve fit of R-PE
0 10 20 30 40 50
Concentration (nM)

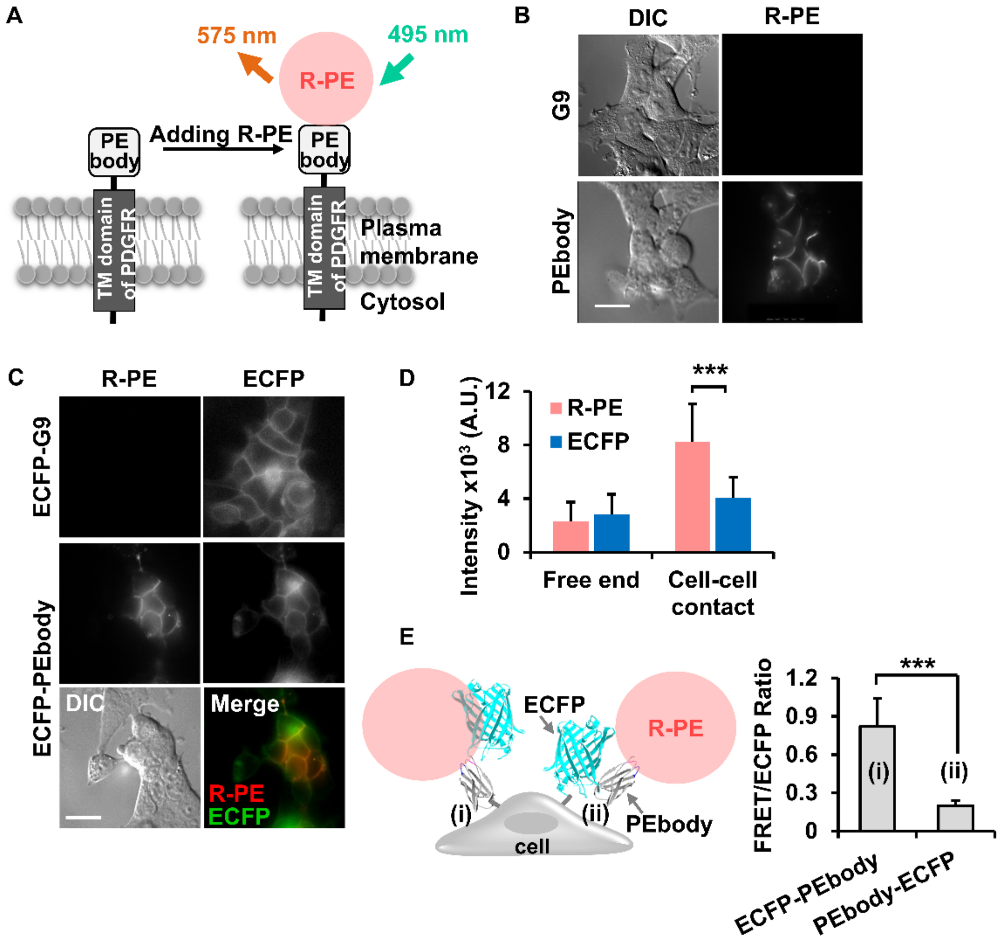


Figure 2.3: Characterization of PEbody on the surface of live mammalian cells

(A) A schematic diagram of R-PE binding to PEbody displayed on the surface of a live mammalian cell. The green arrow represents the excitation wavelength (495 nm) and the orange arrow represents the emission wavelength (575 nm) of R-PE. (B) The differential interference contrast (DIC) and the R-PE intensity images of HEK293 cells expressing the G9 monobody or PEbody. (C) The R-PE staining of HEK293 cells expressing ECFP-G9 monobody or ECFP-PEbody. The respective DIC and overlay images of HEK293 cells expressing ECFP-PEbody are shown in the bottom panel. (D) The average R-PE or ECFP intensity at free end and cell-cell contact regions. Bar graphs represent mean ± SD (n = 33 and 69, where n is the number of analyzed cell free ends and cell-cell contacts, respectively). (E) The comparison of FRET/ECFP ratio between HEK293 cells expressing (i) ECFP-PEbody and (ii) PEbody-ECFP. Bar graphs are represented as mean ± SD (n = 36 and 42, where n is the number of analyzed cell contacts). The asterisk indicates a significant difference (****P* < 0.001 with Wilcoxon rank sum test). Scale bar: 20 μm. See also Figure 2.4.

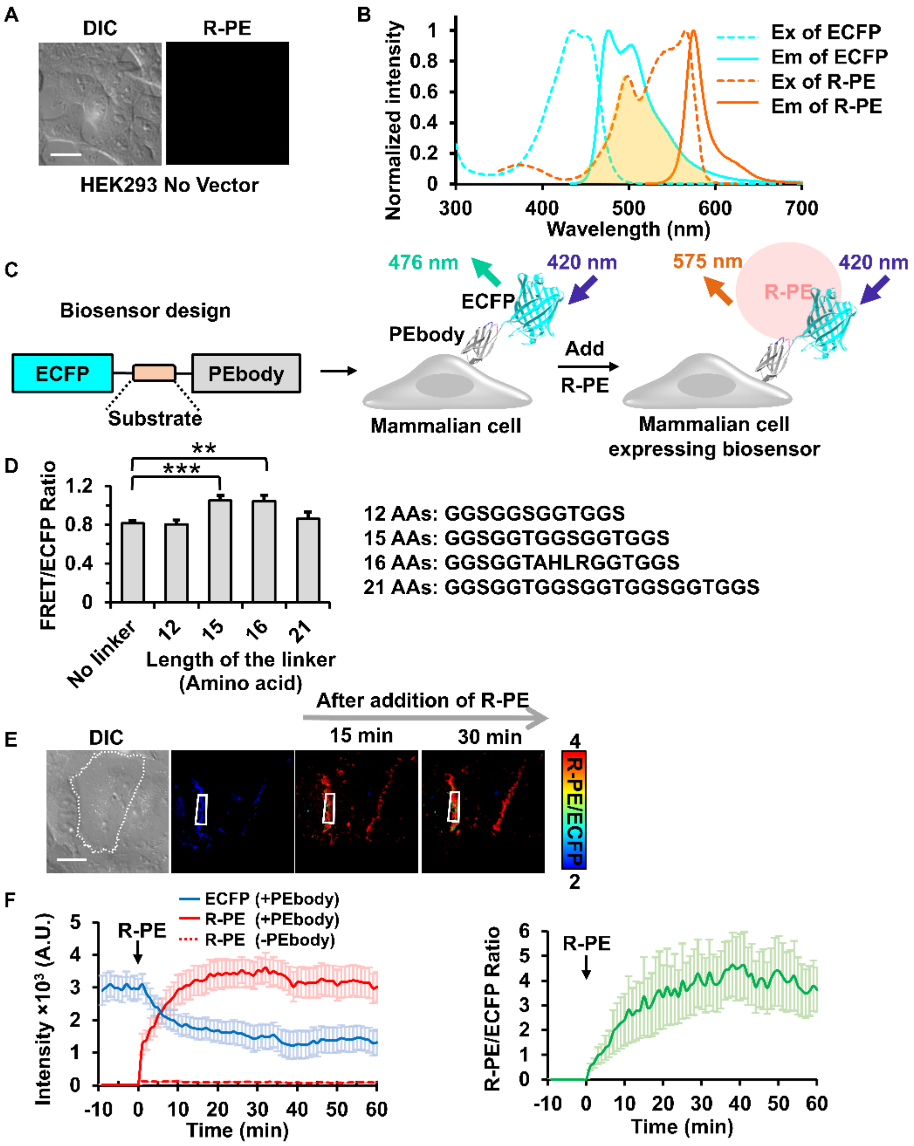
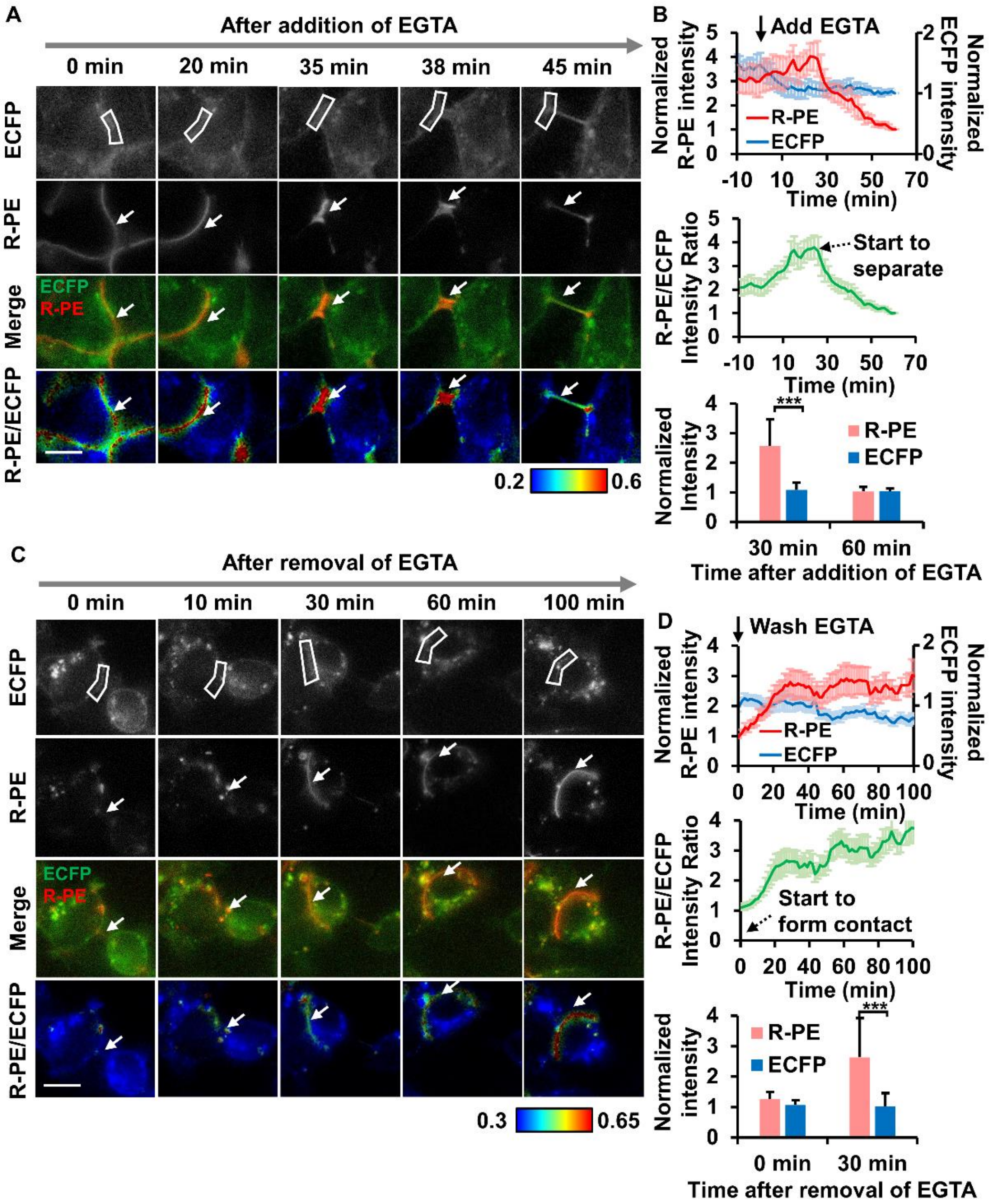


Figure 2.4: Characterization of PEbody and R-PE binding for FRET biosensors

(A) The DIC and R-PE intensity images of HEK293 cells (no vector) stained with R-PE. (B) The excitation and emission spectra of the ECFP and R-PE (data from http://spectra.arizona.edu/). The shaded area represents the overlap spectra of ECFP excitation and R-PE absorption. (C) The diagram of the biosensor design. The PEbody (the R-PE binder) was engineered into the FRET biosensor with ECFP as a donor. By utilizing the pDisplay, the biosensor can be expressed at the extracellular surface of the plasma membrane in mammalian cells. (D) Left: The effect of the linker length on the FRET/ECFP ratio. Data are represented as mean ± SEM (n = 48, 48, 33, 39, 45, where n represents the number of cell contact; **P < 0.01, **P < 0.001 with Wilcoxon rank sum test). Right: The amino acid sequence of the substrate linkers. (E) The DIC and R-PE/ECFP intensity ratio images of the representative HeLa cells expressing ECFP-PEbody before and after adding R-PE. The color bar indicates the R-PE/ECFP intensity ratio, with hot and cold colors representing the high and low ratios, respectively. The white square indicates the region where fluorescent signals were quantified for (F). Scale bar: 20 µm. (F) The average quantified time courses of Left: The R-PE and ECFP intensity and Right: The R-PE/ECFP intensity ratio from multiple HeLa cells in the R-PE-label assay. Data are represented as mean ± SEM (n = 10; n represents the number of analyzed cell contact).

Figure 2.5: The change of the R-PE and ECFP signals upon cell-cell contact dissociation and formation

(A and C) The ECFP, R-PE, overlay, and R-PE/ECFP ratio images in (A) the EGTA-induced cell-cell contact dissociation and (C) the reformation of cell-cell contacts after washing out EGTA. The color bar indicates the R-PE/ECFP intensity ratio, with hot and cold colors representing the high and low ratios, respectively. The white boxes and arrows indicate the regions of interest where an R-PE/ECFP intensity ratio was quantified for (B) and (D). (B and D) Upper panel: The normalized time courses of the average R-PE (left *y*-axis) and ECFP intensity (right *y*-axis); Middle panel: The average R-PE/ECFP ratio during the periods of cell-cell contact dissociation (B) and reformation (D). The removal of EGTA was performed at time = 60 min after the addition of EGTA in (B), and time = 0 min in (D). The black arrow with broken line represents the time point where multiple cells start to separate (B) or form the contact (D). Bottom panel: The bar graphs represent the normalized average intensities of R-PE and ECFP. The level of R-PE dye was maintained in the medium at all times during imaging. The normalized intensity value for each cell was established by dividing the fluorescent signals of each individual cell with that when cell contact was completely dissociated (dimmest R-PE intensity). Data in (B) and (D) are represented as mean ± SEM (n = 8 and 9 where n is the number of analyzed cell contacts in (B) and (D), respectively; $***P < 0.001$ with Wilcoxon rank sum test). Scale bar: 10 µm.


A
After addition of EGTA
0 min
20 min
35 min
38 min
45 min
ECFP
R-PE
Merge
ECFP
R-PE
R-PE/ECFP
0.2
0.6
B
Add EGTA
Normalized R-PE intensity
Normalized ECFP intensity
R-PE
ECFP
Time (min)
R-PE/ECFP Intensity Ratio
Start to separate
Time (min)
Normalized Intensity
***
R-PE
ECFP
30 min
60 min
Time after addition of EGTA
C
After removal of EGTA
0 min
10 min
30 min
60 min
100 min
ECFP
R-PE
Merge
ECFP
R-PE
R-PE/ECFP
0.3
0.65
D
Wash EGTA
Normalized R-PE intensity
Normalized ECFP intensity
R-PE
ECFP
Time (min)
R-PE/ECFP Intensity Ratio
Start to form contact
Time (min)
Normalized intensity
***
R-PE
ECFP
0 min
30 min
Time after removal of EGTA

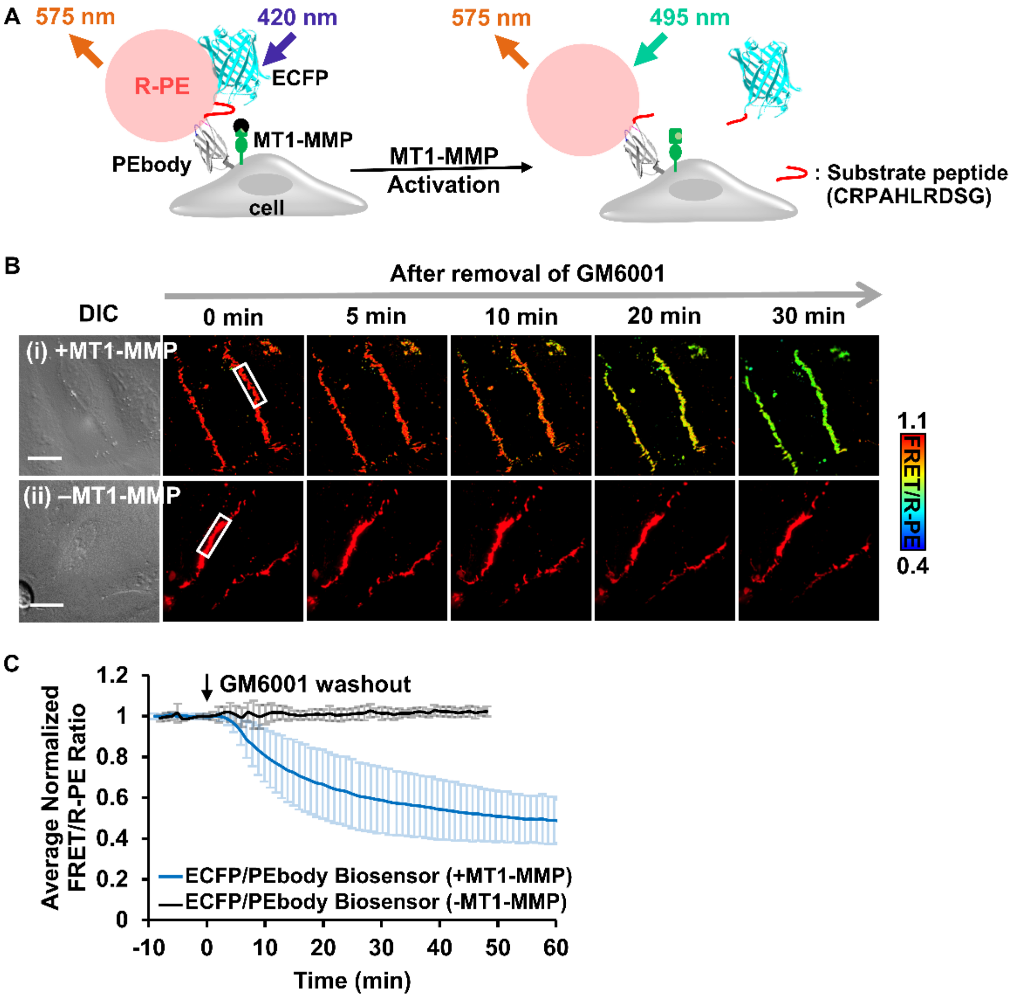


Figure 2.6: The ECFP/PEbody hybrid FRET biosensor for visualizing MT1-MMP activity on the surface of live cancer cells

(A) The design strategy and the sensing mechanism of the biosensor. The biosensor was anchored on the extracellular surface using the transmembrane domain of PDGFR. The N-terminus of the transmembrane domain was linked to PEbody to allow for the R-PE binding. A MT1-MMP substrate sequence flanked by flexible linkers was inserted between ECFP and PEbody. Left: R-PE-staining of the intact biosensor allowed the energy transfer from ECFP to R-PE; Right: Following activation, MT1-MMP cleaved the biosensor substrate sequence, disrupted FRET, and reduced the FRET/R-PE ratio. (B) The DIC and FRET/R-PE ratio images before and after GM6001 inhibitor washout of the representative HeLa cells expressing the MT1-MMP ECFP/PEbody biosensor (i) with or (ii) without the full-length MT1-MMP gene. The color bar indicates the FRET/R-PE ratio, with hot and cold colors representing the high and low ratios, respectively. The white boxes indicate the region where the ratio was quantified for (C). Scale bar: 20 μm. (C) The average quantified time course of the normalized FRET/R-PE ratio from multiple HeLa cells in the GM6001 washout assay was compared in cells with and without the full-length MT1-MMP gene. Data are normalized with an individual basal FRET/R-PE ratio before GM6001 washout and represented as mean ± SD (n = 64 and 29; n is the number of analyzed cell contacts with and without the full-length MT1-MMP, respectively). See also Figure 2.7**.**

Figure 2.7: The comparison of localizations and FRET dynamic ranges between ECFP/YPet and ECFP/PEbody MT1-MMP FRET biosensors

(A) The average quantified time courses of the R-PE intensity in GM6001 washout assay of HeLa cells expressing ECFP/PEbody MT1-MMP biosensors with no treatment, cytochalasin D, or nocodazole treatment. Data are represented as mean ±SEM (n = 46, 22, or 16, where n is the number of analyzed cell contact) (B) The DIC and FRET ratio images of representative HeLa cells before and after GM6001 inhibitor washout. The cells co-expressed full-length MT1-MMP with (i) the genetically encoded ECFP/YPet MT1-MMP biosensor, or (ii) ECFP/PEbody MT1-MMP biosensor. The color bar indicates (i) the FRET/ECFP ratio and (ii) FRET/R-PE ratio, with hot and cold colors representing high and low ratios, respectively. Scale bar: 20 µm. (C) Line-scan analyses of the FRET intensity at the indicated junctions of cells expressing ECFP/YPet (green line in B(i)) or ECFP/PEbody MT1-MMP biosensors (blue line in B(ii)). (D) The average quantified time courses of the FRET/ECFP or FRET/R-PE ratio from multiple HeLa cells expressing the full-length MT1-MMP in the GM6001 washout assay. Data are represented as mean ±SD (n = 64 and 55, where n is the number of analyzed cell contact).

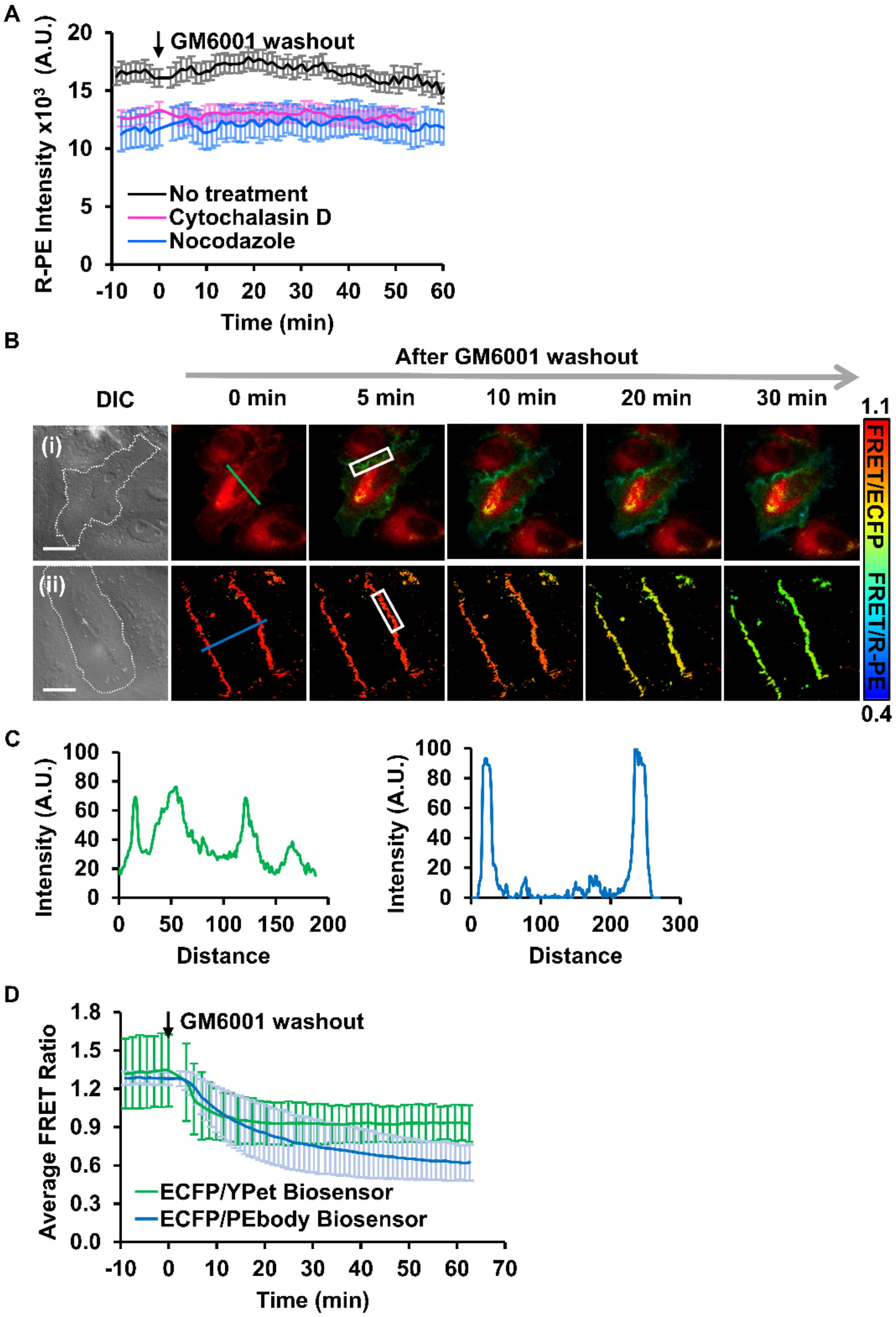

A
R-PE Intensity x10³ (A.U.)
20
15
10
5
0
GM6001 washout
No treatment
Cytochalasin D
Nocodazole
-10
0
10
20
30
40
50
60
Time (min)
B
After GM6001 washout
DIC
0 min
5 min
10 min
20 min
30 min
(i)
(ii)
1.1
FRET/ECFP
FRET/R-PE
0.4
C
Intensity (A.U.)
100
80
60
40
20
0
0
50
100
150
200
Distance
Intensity (A.U.)
100
80
60
40
20
0
0
100
200
300
Distance
D
Average FRET Ratio
1.8
1.5
1.2
0.9
0.6
0.3
0.0
GM6001 washout
ECFP/YPet Biosensor
ECFP/PEbody Biosensor
-10
0
10
20
30
40
50
60
70
Time (min)

Figure 2.8: A distinct distribution of MT1-MMP activity at different subcellular cell-cell contacts

(A) The representative FRET/R-PE ratio images of HeLa cells co-expressing the ECFP/PEbody MT1-MMP FRET biosensor and the full-length MT1-MMP. After GM6001 inhibitor washout, the discrete regions of the membrane showed distinctive FRET/R-PE ratios. Hot and cold colors indicate high and low ratios, respectively. The position 1 and 2 with white boxes are used to quantify the FRET/R-PE ratio in (B). Scale bar: 20 µm. (B) The quantified time course of the FRET/R-PE ratio was compared between positions 1 and 2 in (A). (C) The quantified time course of the normalized FRET/R-PE ratio from multiple cells. The normalized FRET/R-PE ratios illustrated in green or red represent results predominantly from the loose or stable cell-cell contacts, respectively. (D) The average time course of the normalized FRET/R-PE ratio quantified from the loose or stable cell-cell contacts with or without Cytochalasin D or nocodazole treatment. (E) The bar graphs represent the average half-time of signal reduction in (D) with n = 34, 30, 18, and 38, respectively, where n is the number of analyzed cell contacts. (F) The comparison of the MT1-MMP diffusion speed between free edge regions (denoted as loose cell-cell contact) and cell-cell contacts (denoted as stable cell-cell contact) (n = 7 for each group, where n is the number of cells from three independent experiments. Data in (C-D) are normalized with an individual basal FRET/R-PE ratio before GM6001 washout and data in (D-F) are represented as mean ± SD. The asterisk indicates a significant difference (**$P < 0.01$ and ***$P < 0.001$ with Wilcoxon rank sum test). See also Figures 2.9 and 2.10.

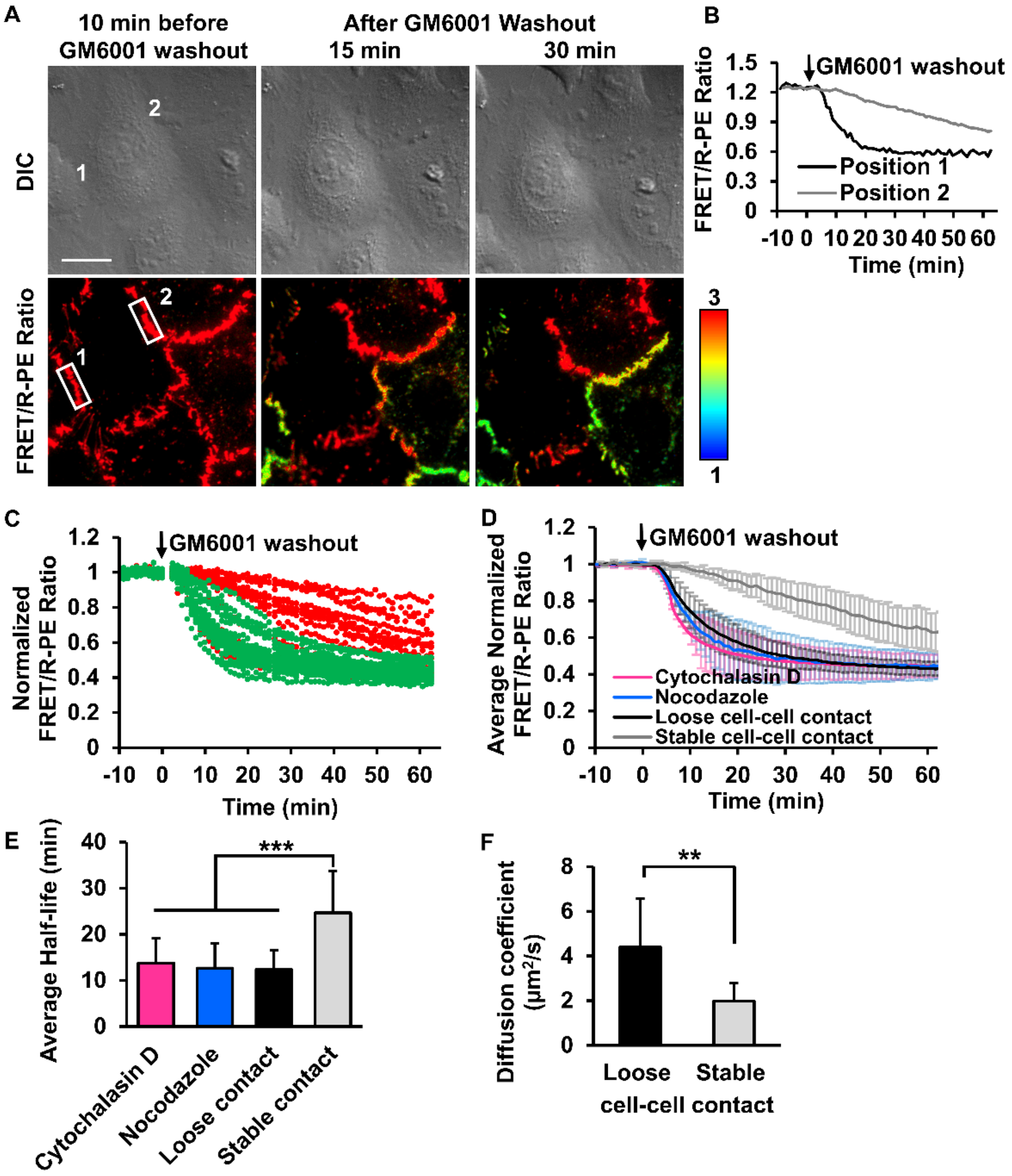
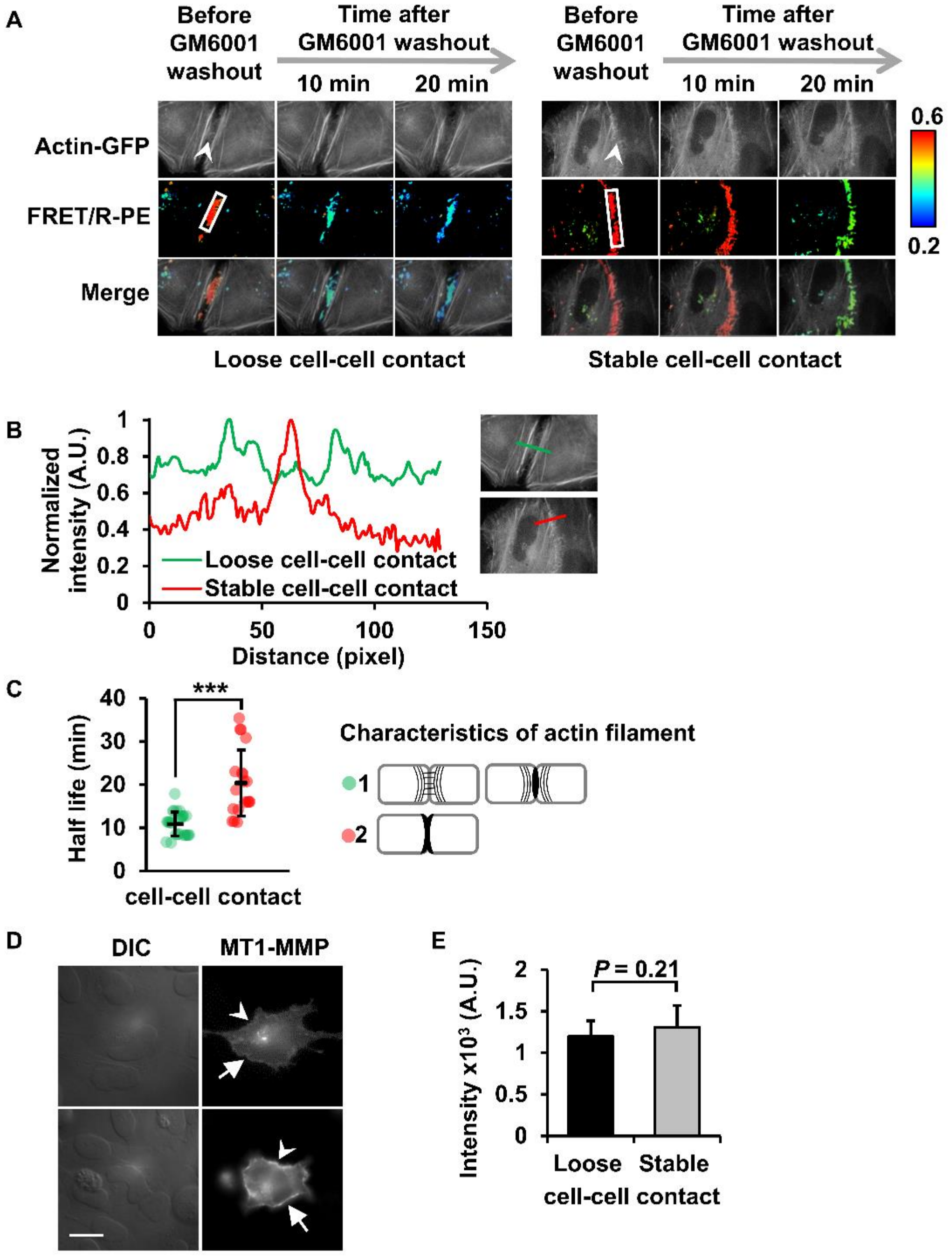

A
10 min before GM6001 washout
After GM6001 Washout
15 min
30 min
DIC
FRET/R-PE Ratio
1
2
3
1
B
FRET/R-PE Ratio
GM6001 washout
Position 1
Position 2
1.5
1.2
0.9
0.6
0.3
0
-10 0 10 20 30 40 50 60
Time (min)
C
Normalized FRET/R-PE Ratio
GM6001 washout
1.2
1
0.8
0.6
0.4
0.2
0
-10 0 10 20 30 40 50 60
Time (min)
D
Average Normalized FRET/R-PE Ratio
GM6001 washout
Cytochalasin D
Nocodazole
Loose cell-cell contact
Stable cell-cell contact
-10 0 10 20 30 40 50 60
Time (min)
E
Average Half-life (min)
***
40
30
20
10
0
Cytochalasin D
Nocodazole
Loose contact
Stable contact
F
Diffusion coefficient (μm^2/s)
**
8
6
4
2
0
Loose
Stable
cell-cell contact

Figure 2.9: Live-cell images showing the relationship among actin patterns, asymmetric FRET ratio reduction, and the localization of MT1-MMP

(A) Top: The actin-GFP; Middle: FRET/R-PE ratio; and Bottom: overlay images of Actin-GFP and FRET/R-PE ratio of representative HeLa cells expressing the ECFP/PEbody MT1-MMP FRET biosensor, full-length MT1-MMP, and GFP-tagged actin. After GM6001 inhibitor washout, the regions with different actin organizations showed different reduction rates of FRET ratios. Hot and cold colors indicate the high and low FRET ratios, respectively. Scale bar: 20 µm. (B) Line-scan analysis of actin-GFP across the loose (green line) or stable (red line) cell-cell contacts. The intensity was normalized by the maximum intensity from each data. (C) The relationship between the half-life of FRET signal reduction (the time that the FRET signal reduces to half of the basal signal) and the characteristics of actin bundles at loose and stable cell-cell contact. The number of cells classified into two different categories according to the spatial organization of actin bundles: (1) loose bands with diffusive branches of actin filaments localized close to cell-cell contacts; (2) actin filaments form tight actin bundles localized at the stable cell-cell contacts. Date are represented as mean ± SD where the number of analyzed loose and stable cell-cell contacts are 25 and 19, respectively. The asterisk indicates a significant difference (***$P < 0.001$ with Wilcoxon rank sum test). (D) HeLa cells expressing the full-length MT1-MMP gene fused with mCherry. Left: The DIC image of the cell; Right: the MT1-MMP-mCherry image. The white arrow head and arrow represent loose and stable cell-cell contacts, respectively. Scale bar: 20 µm. (E) The average intensity of mCherry at loose and stable cell-cell contacts from (D). Date are represented as mean ± SEM where the number of analyzed loose and stable cell-cell contacts are 58 and 44, respectively.


A
Before GM6001 washout
Time after GM6001 washout
10 min
20 min
Actin-GFP
FRET/R-PE
Merge
Loose cell-cell contact
Stable cell-cell contact
0.6
0.2
B
Normalized intensity (A.U.)
1
0.8
0.6
0.4
0.2
0
0
50
100
150
Distance (pixel)
Loose cell-cell contact
Stable cell-cell contact
C
***
Half life (min)
40
30
20
10
0
cell-cell contact
Characteristics of actin filament
1
2
D
DIC
MT1-MMP
E
P = 0.21
Intensity x10^3 (A.U.)
2
1.5
1
0.5
0
Loose
Stable
cell-cell contact

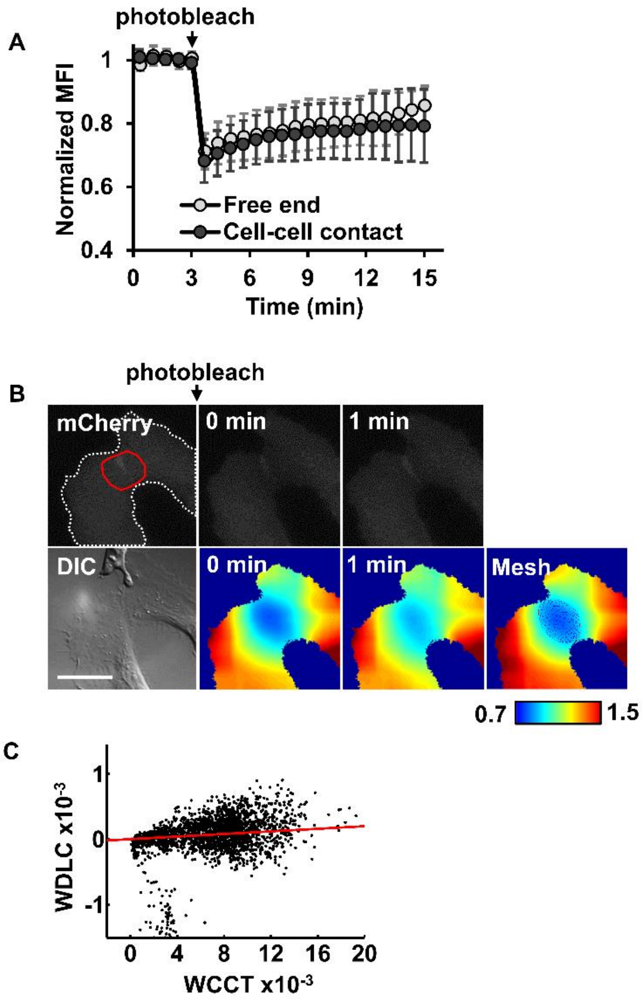


Figure 2.10: The FRAP images of the MT1-MMP-mCherry fusion protein

(A) The recovery intensity curve of mCherry at free end and cell-cell contact after photobleaching. (B) The FRAP images taken from free end and cell-cell contact regions. Top left: the fluorescence intensity image of a cell before photobleaching. Top middle and right: the fluorescence intensity images at 0 and 1 min after photobleaching, respectively. Bottom left: DIC image. Bottom middle and right: the concentration maps at 0 and 1 min after photobleaching, analyzed by normalizing the fluorescence intensity with the image before photobleaching. Bottom far right: the concentration map with a triangular mesh generated for the finite element discretization. The broken white-colored outline defining the cell edge and red-colored outline defining the region of interest monitored for fluorescence recovery. Scale bar: 20 µm. (C) The scattered plot of the weighted discrete Laplacian of concentration (WDLC) and the weighted change of concentration in time (WCCT) on each mesh node, with the linear fitting indicated by the solid line.

Acknowledgements

Chapter 2, in full, is a reprint of the material as it appears in Cell Chemical Biology journal. P. Limsakul, Q. Peng, Y. Wu, M. Allen, J. Liang, A. G. Remacle, T. Lopez, X. Ge, B. Kay, H. Zhao, A. Y. Strongin, X.L. Yang, S. Lu, and Y. Wang. "Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface." Cell chemical biology 25.4 (2018). 370-379. The dissertation author was the first author of this paper.

This work is supported by grants from NIH HL098472, HL109142, HL121365 (Y. Wang), NSF CBET1360341, DMS1361421 (Y. Wang and S.L.), UC San Diego, and Beckman Laser Institute, Inc. (Y. Wang) and NIH NS 085092 (X.L. Yang). This research was also supported by NSF China NSFC 11428207 (Y. Wang). P.L. acknowledges fellowship support from the Thai Ministry of Science and Technology. Q.P. acknowledges fellowship support from China Scholarship Council (CSC). J.L. acknowledges fellowship support from the Agency for Science, Technology and Research of Singapore. The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Wang, Y., E.L. Botvinick, Y. Zhao, M.W. Berns, S. Usami, R.Y. Tsien, and S. Chien, *Visualizing the mechanical activation of Src.* Nature, 2005. **434**(7036): p. 1040-1045.
2. Wang, Y., J.Y.-J. Shyy, and S. Chien, *Fluorescence proteins, live-cell imaging, and mechanobiology: seeing is believing.* Annu. Rev. Biomed. Eng., 2008. **10**: p. 1-38.
3. Shaner, N.C., G.H. Patterson, and M.W. Davidson, *Advances in fluorescent protein technology.* Journal of Cell Science, 2007. **120**(24): p. 4247-4260.
4. Glazer, A., *Light harvesting by phycobilisomes.* Annual review of biophysics and biophysical chemistry, 1985. **14**(1): p. 47-77.
5. Isailovic, D., I. Sultana, G.J. Phillips, and E.S. Yeung, *Formation of fluorescent proteins by the attachment of phycoerythrobilin to R-phycoerythrin alpha and beta apo-subunits.* Analytical biochemistry, 2006. **358**(1): p. 38-50.
6. Arnold, F.H., *Design by directed evolution.* Accounts of chemical research, 1998. **31**(3): p. 125-131.
7. Shaner, N.C., R.E. Campbell, P.A. Steinbach, B.N. Giepmans, A.E. Palmer, and R.Y. Tsien, *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nature biotechnology, 2004. **22**(12): p. 1567-1572.

8. Shaner, N.C., M.Z. Lin, M.R. McKeown, P.A. Steinbach, K.L. Hazelwood, M.W. Davidson, and R.Y. Tsien, *Improving the photostability of bright monomeric orange and red fluorescent proteins.* Nature methods, 2008. **5**(6): p. 545-551.

9. Shaner, N.C., G.G. Lambert, A. Chammas, Y. Ni, P.J. Cranfill, M.A. Baird, B.R. Sell, J.R. Allen, R.N. Day, and M. Israelsson, *A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum.* Nature methods, 2013. **10**(5): p. 407-409.

10. Hires, S.A., Y. Zhu, and R.Y. Tsien, *Optical measurement of synaptic glutamate spillover and reuptake by linker optimized glutamate-sensitive fluorescent reporters.* Proc Natl Acad Sci U S A, 2008. **105**(11): p. 4411-6.

11. Komatsu, N., K. Aoki, M. Yamada, H. Yukinaga, Y. Fujita, Y. Kamioka, and M. Matsuda, *Development of an optimized backbone of FRET biosensors for kinases and GTPases.* Mol Biol Cell, 2011. **22**(23): p. 4647-56.

12. Ibraheem, A., H. Yap, Y. Ding, and R.E. Campbell, *A bacteria colony-based screen for optimal linker combinations in genetically encoded biosensors.* BMC Biotechnol, 2011. **11**: p. 105.

13. Binz, H.K., P. Amstutz, and A. Plückthun, *Engineering novel binding proteins from nonimmunoglobulin domains.* Nature biotechnology, 2005. **23**(10): p. 1257-1268.

14. Wittrup, K.D., G.M. Thurber, M.M. Schmidt, and J.J. Rhoden, *Practical theoretic guidance for the design of tumor-targeting agents.* Methods in enzymology, 2012. **503**: p. 255.

15. Gulyani, A., E. Vitriol, R. Allen, J. Wu, D. Gremyachinskiy, S. Lewis, B. Dewar, L.M. Graves, B.K. Kay, and B. Kuhlman, *A biosensor generated via high-throughput screening quantifies cell edge Src dynamics.* Nature chemical biology, 2011. **7**(7): p. 437-444.

16. Koide, A., C.W. Bailey, X. Huang, and S. Koide, *The fibronectin type III domain as a scaffold for novel binding proteins.* Journal of molecular biology, 1998. **284**(4): p. 1141-1151.

17. Batori, V., A. Koide, and S. Koide, *Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain.* Protein engineering, 2002. **15**(12): p. 1015-1020.

18. Carr, P.A., H.P. Erickson, and A.G. Palmer, *Backbone dynamics of homologous fibronectin type III cell adhesion domains from fibronectin and tenascin.* Structure, 1997. **5**(7): p. 949-959.

19. Hotary, K.B., E.D. Allen, P.C. Brooks, N.S. Datta, M.W. Long, and S.J. Weiss, *Membrane type I matrix metalloproteinase usurps tumor growth control imposed by the three-dimensional extracellular matrix.* Cell, 2003. **114**(1): p. 33-45.

20. Gálvez, B.G., S. Matías-Román, M. Yáñez-Mó, F. Sánchez-Madrid, and A.G. Arroyo, *ECM regulates MT1-MMP localization with β1 or αvβ3 integrins at distinct cell compartments modulating its internalization and activity on human endothelial cells.* The Journal of cell biology, 2002. **159**(3): p. 509-521.

21. Woskowicz, A.M., S.A. Weaver, Y. Shitomi, N. Ito, and Y. Itoh, *MT-LOOP-dependent localization of membrane type I matrix metalloproteinase (MT1-MMP) to the cell adhesion complexes promotes cancer cell invasion.* Journal of Biological Chemistry, 2013. **288**(49): p. 35126-35137.

22. Nawrocki-Raby, B., C. Gilles, M. Polette, C. Martinella-Catusse, N. Bonnet, E. Puchelle, J.-M. Foidart, F. Van Roy, and P. Birembaut, *E-Cadherin mediates MMP down-regulation in highly invasive bronchial tumor cells.* The American journal of pathology, 2003. **163**(2): p. 653-661.

23. Covington, M.D., R.C. Burghardt, and A.R. Parrish, *Ischemia-induced cleavage of cadherins in NRK cells requires MT1-MMP (MMP-14).* Am J Physiol Renal Physiol, 2006. **290**(1): p. F43-51.

24. Rozanov, D.V., E.I. Deryugina, E.Z. Monosov, N.D. Marchenko, and A.Y. Strongin, *Aberrant, persistent inclusion into lipid rafts limits the tumorigenic function of membrane type-1 matrix metalloproteinase in malignant cells.* Exp Cell Res, 2004. **293**(1): p. 81-95.

25. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nature biotechnology, 1997. **15**(6): p. 553-557.

26. Feldhaus, M.J., R.W. Siegel, L.K. Opresko, J.R. Coleman, J.M.W. Feldhaus, Y.A. Yeung, J.R. Cochran, P. Heinzelman, D. Colby, and J. Swers, *Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library.* Nature biotechnology, 2003. **21**(2): p. 163-170.

27. Kastritis, P.L., I.H. Moal, H. Hwang, Z. Weng, P.A. Bates, A.M. Bonvin, and J. Janin, *A structure-based benchmark for protein–protein binding affinity.* Protein Science, 2011. **20**(3): p. 482-491.

28. Ouyang, M., H. Huang, N.C. Shaner, A.G. Remacle, S.A. Shiryaev, A.Y. Strongin, R.Y. Tsien, and Y. Wang, *Simultaneous visualization of protumorigenic Src and MT1-MMP activities with fluorescence resonance energy transfer.* Cancer research, 2010. **70**(6): p. 2204-2212.

29. Ouyang, M., S. Lu, X.-Y. Li, J. Xu, J. Seong, B.N. Giepmans, J.Y.-J. Shyy, S.J. Weiss, and Y. Wang, *Visualization of polarized membrane type 1 matrix metalloproteinase activity in live cells by fluorescence resonance energy transfer imaging.* Journal of Biological Chemistry, 2008. **283**(25): p. 17740-17748.

30. Lu, S., T.-j. Kim, C.-E. Chen, M. Ouyang, J. Seong, X. Liao, and Y. Wang, *Computational analysis of the spatiotemporal coordination of polarized PI3K and Rac1 activities in micro-patterned live cells.* PloS one, 2011. **6**(6): p. e21293-e21293.

31. Lu, S., M. Ouyang, J. Seong, J. Zhang, S. Chien, and Y. Wang, *The spatiotemporal pattern of Src activation at lipid rafts revealed by diffusion-corrected FRET imaging.* PLoS Comput Biol, 2008. **4**(7): p. e1000127-e1000127.

32. Vandenberg, W. and P. Dedecker, *Effect of probe diffusion on the SOFI imaging accuracy.* Scientific Reports, 2017. **7**: p. 44665.

33. Huang, R., P. Fang, and B.K. Kay, *Isolation of monobodies that bind specifically to the SH3 domain of the Fyn tyrosine protein kinase.* New biotechnology, 2012. **29**(5): p. 526-533.

34. Miyazaki, K. and F.H. Arnold, *Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function.* Journal of molecular evolution, 1999. **49**(6): p. 716-720.

35. Lu, S., Y. Wang, H. Huang, Y. Pan, E.J. Chaney, S.A. Boppart, H. Ozer, A.Y. Strongin, and Y. Wang, *Quantitative FRET imaging to visualize the invasiveness of live breast cancer cells.* PloS one, 2013. **8**(3): p. e58569.

36. Drees, F., S. Pokutta, S. Yamada, W.J. Nelson, and W.I. Weis, *α-catenin is a molecular switch that binds E-cadherin-β-catenin and regulates actin-filament assembly.* Cell, 2005. **123**(5): p. 903-915.

37. Zhang, J., M. Betson, J. Erasmus, K. Zeikos, M. Bailly, L.P. Cramer, and V.M. Braga, *Actin at cell-cell junctions is composed of two dynamic and functional populations.* Journal of cell science, 2005. **118**(23): p. 5549-5562.

38. Burridge, K. and E.S. Wittchen, *The tension mounts: stress fibers as force-generating mechanotransducers.* J Cell Biol, 2013. **200**(1): p. 9-19.

39. Remacle, A.G., S.A. Shiryaev, V.S. Golubkov, J.N. Freskos, M.A. Brown, A.S. Karwa, A.D. Naik, C.P. Howard, C.J. Sympson, and A.Y. Strongin, *Non-destructive and selective imaging of the functionally active, pro-invasive membrane type-1 matrix metalloproteinase (MT1-MMP) enzyme in cancer cells.* Journal of Biological Chemistry, 2013. **288**(28): p. 20568-20580.

40. Zucker, S., M. Hymowitz, C.E. Conner, E.A. DiYanni, and J. Cao, *Rapid trafficking of membrane type 1-matrix metalloproteinase to the cell surface regulates progelatinase a activation.* Laboratory investigation, 2002. **82**(12): p. 1673-1684.

41. Remacle, A., G. Murphy, and C. Roghi, *Membrane type I-matrix metalloproteinase (MT1-MMP) is internalised by two different pathways and is recycled to the cell surface.* J Cell Sci, 2003. **116**(19): p. 3905-3916.

42. Brun, M.A., K.T. Tan, E. Nakata, M.J. Hinner, and K. Johnsson, *Semisynthetic fluorescent sensor proteins based on self-labeling protein tags.* J Am Chem Soc, 2009. **131**(16): p. 5873-84.

43. Maurel, D., L. Comps-Agrar, C. Brock, M.L. Rives, E. Bourrier, M.A. Ayoub, H. Bazin, N. Tinel, T. Durroux, L. Prezeau, E. Trinquet, and J.P. Pin, *Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization.* Nat Methods, 2008. **5**(6): p. 561-7.

44. Batard, P., J. Szollosi, I. Luescher, J.C. Cerottini, R. MacDonald, and P. Romero, *Use of phycoerythrin and allophycocyanin for fluorescence resonance energy transfer analyzed by flow cytometry: advantages and limitations.* Cytometry, 2002. **48**(2): p. 97-105.

45. Lutz, S., *Beyond directed evolution--semi-rational protein engineering and design.* Curr Opin Biotechnol, 2010. **21**(6): p. 734-43.

46. Benatuil, L., J.M. Perez, J. Belk, and C.-M. Hsieh, *An improved yeast transformation method for the generation of very large human antibody libraries.* Protein Engineering Design and Selection, 2010: p. gzq002.

47. Liu, P., J. Yang, J. Pei, D. Pei, and M.J. Wilson, *Regulation of MT1-MMP activity by beta-catenin in MDCK non-cancer and HT1080 cancer cells.* J Cell Physiol, 2010. **225**(3): p. 810-21.

Chapter 3

PEbody CAR:

A Universal Chimeric Antigen Receptor for CAR T Cell Therapy

3.1 Introduction

Cancer is one of the leading causes of death worldwide. Traditional methods to cure cancer such as surgery, chemotherapy and radiation therapy suffer from low efficiency and high relapse rates due to nonspecific targeting and the heterogeneity of cancer cells. Also, these treatments often cause side effects. In recent years, immunotherapy including checkpoint inhibitors and chimeric antigen receptor (CAR) T cells, gives promising results in clinical trials [1, 2]. Particularly, CAR T cell therapy, which employs and enhances a patient's own immune cells to attack cancer cells and trigger long-lived immune memories by generating memory T cells to consistently recognize and attack cancer cells [3], has been approved by the US Food and Drug Administration (FDA), and is currently utilized for the treatment of leukemia and lymphoma [1].

Mimicking T cell receptors (TCRs), CARs are synthetic receptors that redirect the specificity, function, and metabolism of T cells upon engaging to cancer cells. CAR design has evolved over years to enhance efficacy and safety immunologic settings. In general, genetically engineered CAR T cells consist of an extracellular domain, typically derived from an antibody single chain variable fragment (scFv) to direct specificity, and intracellular signaling domains, typically derived from endogenous TCRs and costimulatory signals, to provide activation signals (Figure 3.1A) [4]. This modular design of CARs allows more flexible to target antigen which is independent of the major histocompatibility complex (MHC). The use of CAR T cell therapies has been applied to target CD19 on B cell acute lymphoblastic leukemia (B-ALL) with the greatest promise in the field to date [4]. The extracellular glycoprotein CD19 was chosen because it is highly expressed in B-cell leukemias and lymphomas relative to other targets, such as CD20 or CD22, and it has minimal expression in normal tissues. [1]. However, due to evolution of tumor cells, depletion of CD19 antigen on leukemia cells causes CAR T cell dysfunction and relapse in patients treated with anti-CD19 CAR T cells. Moreover, CAR T cell therapies cause significant toxicities, such as cytokine-release syndrome (CRS) and neurotoxicity [5]. The "on-target, off-tumor" activity of anti-CD19 CAR T cells also affects long-term depletion of B-cells in patients [6].

To mitigate treatment-related toxicities but maintain potent efficacy of CAR T cells, several strategies have been applied. Inhibitory CARs, which contains the intracellular domain from inhibitory molecules, can turn off activation of CAR T cells when binding to antigens presenting on normal cells [7]. CARs containing suicide system can reduce the lifespan of CAR T cells [8]. In addition, multi-input CARs that utilize the Boolean logic gate, such as AND gate, OR gate, and AND-NOT gate, can improve selectivity of CAR T cells toward tumor cells [9]. For example, dual targeting CARs require the engagement of two tumor-associated antigens (TAAs) to induce full activation of CAR- T cells [10]. Moreover, switchable CARs, which comprise of two components (1) an adaptor molecule that is tumor-targeting antibody or scFv or small molecule, and (2) a second molecule that selectively binds the CAR but not an endogenous receptor, allow for flexibility and controllability [11-13]. One of the switchable CARs that have been recently reported is the split, universal, and programmable (SUPRA) CAR system, which comprises of zipCAR and zipFv fragments. The zipCAR is the CAR of which extracellular portion is a leucine zipper, and the zipFv contains an antigen-targeting scFv fused to a second leucine zipper. Because SUPRA CAR requires binding of leucine zippers, its universal designs could be used to target various antigens, combat relapse, control activation, and improve specificity [12].

Universal CARs could become "off-the-shelf" CAR T cell therapy because it requires only one engineered T cells while antigen-targeting molecules can be many. This could make manufacturing process becomes faster, easier, and more economic. Although the SUPRA CAR or other universal CARs showed promising results in vivo, the design can be further improved. scFv which was derived from murine was used as an extracellular receptor of CAR or an adaptor molecule. This could cause immunogenic reaction in human. Furthermore, the adaptor molecules, such as scFv or Fab, have shorter half-life in vivo than the full-length antibody. This could be problematic because it requires frequently injection of adaptors with large quantity. To overcome these problems, here, we presented a simple and universal receptor for CAR. Our CAR is also a two-component receptor system: (1) a PE-targeting receptor (PEbody) expressed on T cells; (2) a tumor-targeting antibody conjugated with PE (Ab-PE) which acts as an adaptor molecule (Figure 3.1A). Instead of using scFv, we used the engineered monobody variant, which was derived from human fibronectin type III [14], with nM range affinity (K_D ~ 6 nM) toward R-Phycoerythrin (PE) [15]. For an adaptor molecule, we used a full-length antibody conjugated with PE which is commercially available. Even though PE was extracted from red algae [16], it was reported to be biocompatible and it was also applied to food coloring industry [17]. Combining these new features, our PEbody CAR system could be an alternative to improve the safety and efficacy of current cellular cancer immunotherapy.

3.2 Materials and Methods

3.2.1 Constructs and Plasmids

For anti-CD19 CAR, the nucleotide sequence encoding the CD19-CAR, including the anti-human CD19 scFv, the human CD8α hinge, the CD28 transmembrane and costimulation domain, the human 4–1BB costimulation domain, and the CD3ζ ITAM-signaling domain were synthesized (Integrated DNA Technologies), and amplified by PCR following the order from [18]. All cDNA fragments were fused together and inserted into the lentiviral transfer vector, pSIN, between SpeI and EcoRI by Gibson assembly (New England Biolabs). For PEbody CAR, the cDNA encoding PEbody [15] was used to replace the anti-CD19 scFv region while the rest remained the same. All fragments were amplified and assembled following the standard molecular cloning protocols. After verifying the plasmid by Sanger sequencing, all plasmids were amplified by Maxiprep (Qiagen).

3.2.2 Cell Culture and Reagents

Cell culture reagents were purchased from Thermo Fisher Scientific and cell lines were purchased from the American Type Culture Collection (ATCC) unless otherwise stated. The human embryonic kidney cell (HEK293T), Lenti-X 293T cells (#632180; Clontech Laboratories) the prostate cancer cell (PC-3), the breast cancer cells e.g. MCF-7 and MDA-MB-231 were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM of L-glutamine, 100 units/mL of penicillin, 100 µg/mL of streptomycin, and 1mM of sodium pyruvate. Jurkat T cell, Toledo cell, and NAML6 cell were maintained in RPMI-1640 medium supplemented with 10% FBS, penicillin, and streptomycin. Peripheral blood mononuclear cells or PBMCs (Primary human T cells) were isolated from an anonymous healthy donor's blood after apheresis (San Diego Blood Bank). Purified PBMCs were cryopreserved in culture RPMI-1640 medium supplemented with 10% DMSO until use. During experiments, PBMCs were maintained in RPMI-1640 medium supplemented with 10% FBS, penicillin/streptomycin, and 100 units/mL IL-2 (PeproTech). All cells were cultured at 37 °C with 5% CO_2.

3.2.3 Lentiviral Infection of human T Cells

Pantropic VSV-G pseudotyped lentivirus was produced from Lenti-X 293T cells co-transfected with a pSIN transgene expression vector and the viral packaging plasmids, including pCMVdR8.91 and pCMV-VSV-G, using ProFection® Mammalian Transfection System (Promega, Cat #E1200). Viral supernatant was collected 48 h after transfection and concentrated using PEG-it (System Biosciences, Cat# LV825A-1) following the supplier's protocol. 48-72 h before viral infection, PBMCs were thawed and activated using 2 µg/mL phytohemagglutinin (PHA) (ThermoFisher Scientific) in RPMI-1640 medium supplemented with 10% FBS, penicillin/streptomycin, and 100 units/mL IL-2. PHA was then removed before cells were infected. Rectronectin (Clontech Cat#T100B) was used to transduce cells. Briefly, a non-TC treated 24-well plate was coated with rectronectin following the supplier's protocol. Concentrated viral supernatant was mixed with $0.3x10^6$ PMBCs and then this mixture was added into a well and spun for 60 min at 1800xg. After centrifugation, cells were move to an incubator at 37˚C. Infected PBMCs were maintained at ~$0.5x10^6$/mL in culture RPMI medium for 5 d before enriching infected cells.

3.2.4 Verification of CAR Expression in T Cells

Jurkat or primary human T cells were resuspended in FACS wash buffer (PBS + 0.5% BSA) and stained with Myc-tag antibody conjugated with Alexa Fluor 647 (Cell Signaling Technology, Cat# 2233). Stained cells were washed three times with wash buffer and processed with a BD Accuri C6 cytometer

(BD). FlowJo software (TreeStar) was used to quantify fluorescence intensities of Alexa dyes and percentages.

3.2.5 Enrichment of CAR T cells

Our CAR construct has c-myc peptide on the extracellular domain at N-terminal of PEbody or anti-CD19 scFv fragment. Thus, to enrich T cells expressing CARs, cells were stained with anti-c-myc antibody (Milteny, Cat# 130-092-471), and then bound to anti-biotin microbeads (Milteny, Cat# 130-090-485) before separated by MACS columns (Milteny, Cat# 130-042-401) following the supplier's protocols. Degassed and cold MACS buffer (PBS (pH 7.2), 0.5% BSA, and 2mM EDTA) was used for staining and washing cells. Purified CAR-T cells were confirmed by flow cytometry using anti-c-myc Alexa Fluor 647 (following the protocol in 3.2.4) after three days of enrichment or before preforming luciferase killing assay.

3.2.6 Quantitation of CD69 Surface Expression

Jurkat or primary human T cells after co-cultured with target Toledo or NAML6 cells at a different T-cell/target cell ratio (E:T ratio) in a 24-well plate or 96-well plate. After 24 h of incubation, cells were pelleted by centrifugation at 300 × *g* for 5 min. Cells then were resuspended in FACS wash buffer (PBS + 0.5% BSA) and stained with APC anti-human CD69 antibody (BioLegend, Cat #310910). Stained cells were washed three times in FACS wash buffer and processed with a BD Accuri C6 cytometer. FlowJo software (TreeStar) was applied to quantify the fluorescent intensity and percentage of activated cells.

3.2.7 Luciferase-Based Cell Killing Assay

Cytotoxicity assays were carried out using Dual-Luciferase reporter assay system (Promega, Cat#E1910). Briefly, CAR T cells were incubated with target cells (e.g. NALM-6, MDA-MB-231, PC3, and MCF-7) that were engineered to express firefly luciferase at varying E:T ratio for 24h at 37˚C unless otherwise stated. Target cells were stained with different antibodies conjugated with PE and unbound antibodies were washed away. Stained target cells were seeded at 10,000 or 50,000 cells per well of a 96-well plate, and then CAR T cells were added. After 24h of incubation, cells were pelleted by centrifugation at 300 × *g* for 5 min and resuspended in 50 μL of lysis buffer (Promega, Cat#E1910). The cell lysates were transferred to each well of the 96-well plate with black wall. Luciferase reagent was added into each well and luciferase signals were detected with a plate reader (Tecan Infinite M1000 Pro). Target cell cytotoxicity was calculated using the following formula: Cytotoxicity = 100 x [(Total target cell luminescence – luminescence of remaining cells after lysis) / (Total target cell luminescence)].

3.2.8 Quantification and statistical analysis

All the experiments were replicated at least three times. Data are presented as mean values ± standard deviation (SD) as indicated in the figure legend. Statistical analyses were carried out using MATLAB (version R2018a) and were defined by P values calculated from the two-tailed Student's t tests by comparison with the relevant control as specified in the figures or figure legends. Levels of significance were $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***). For flow cytometry experiments, when n=3, each data point is an average reading from more than 30,000 single cells. According to the central limit theorem, these data follow the Gaussian distribution. Therefore, we used the two-tailed Student's t-tests.

3.3 Results

3.3.1 Design and Characterization of the PEbody CAR System

Instead of expressing CARs as full-length fusion proteins that include antigen-binding, transmembrane, and signaling domains, the PEbody CAR is a two-component system containing a PE-targeting receptor (PEbody) expressed on T cells and a tumor-targeting antibody conjugated with PE (Ab-PE) (Figures 3.1A-B). Unlike other CARs of which scFvs are derived from murine monoclonal antibodies [19, 20], the PEbody was evolved from the human fibronectin type III domain to bind to PE with the affinity of 6 nM (Figure 3.2) [15]. The PEbody was inserted into an extracellular domain of the 3rd generation CAR of which intracellular domains contain CD28, 4-1BB co-stimulatory and a CD3ζ signaling domain. A functional CAR is reconstituted when Ab-PE is added to engineered T cells that express PEbody CARs.

The PEbody has been examined to be specific binding to R-Phycoerythrin in yeast and mammalian cell systems (Figure 3.2). To test specificity of the PEbody in the CAR system, we engineered human primary T cells to express the PEbody CAR or the G9 CAR. The G9 monobody, which has a similar structure as PEbody but cannot bind to PE (Figures 3.2B-C), was replaced the extracellular portion of CAR.

T cells expressing these CAR constructs were co-cultured with NALM6 cells, which were stained with anti-CD19 antibody conjugated with PE. We varied E:T ratio between T cells (E; effector) and tumor cells (T, target) to validate the killing ability of CAR T cells by measuring the luciferase signal of remaining NALM6 cells. Our result showed that the PEbody CAR T cell selectively targeted the Ab-PE but not for the G9 CAR and plain T cells (Figure 3.1C).

Although the PEbody on the extracellular domain of CAR can bind to PE, we observed the slight killing ability of the PEbody CAR T cells when non-stained NAML6 cells were co-cultured with PEbody CAR T cells. This is possible because purified peripheral blood mononuclear cells (PMBC) from human blood may contain natural killing cells that causes higher cytotoxicity. To reduce non-specific killing, different E:T ratios of PEbody CAR T cells to NALM6 cells were varied. We observed that the differences of cytotoxicity between stained and non-stained NALM 6 cells were larger when the number of NALM 6 cells increased (Figure 3.3).

3.3.2 In Vitro Comparison of Anti-CD19 CAR vs. PEbody CAR T Cells

To determine the relative activity of the PEbody CAR T cells, we compared killing ability of the PEbody CAR with the anti-CD19 CAR in the presence of NALM6 cells. The PEbody CAR or anti-CD19 CAR lentiviral particles were generated and used to transduced T cells (purified from the same donor). CAR T cells were further enriched by magnetic-activated cell sorting (MACS) before co-culturing with cancer cells. For the PEbody CAR, we firstly stained NALM6 cells with anti-CD19 Ab-PE and co-cultured them with PEbody CAR-T cells. For anti-CD19 CAR, T cells were directly co-cultured with Nalm6 cells without any staining. The results from the luciferase-based killing assay illustrated that the PEbody CAR T cells can kill NALM6 cells to a similar extent as anti-CD19 CAR T cells in 12 h and 24 h of co-culturing (Figure 3.4A). Overall, our in vitro findings suggest that PEbody CAR T cells together with the anti-CD19 Ab-PE switch, are comparable to conventional anti-CD19 CAR T cells in targeting tumor antigen.

3.3.3 The PEbody CAR System Targeting Different Tumor Antigens

Despite the conventional fixed CAR design, the PEbody CAR design allows for targeting of different antigens without further genetic manipulations of a patient's immune cells. To test the versatile ability of the PEbody CAR system in targeting different antigens, we introduced the PEbody CARs into human primary T cells and co-cultured them *in vitro* with either hematologic cancers, e.g. NALM6 leukemia cells that express CD19, CD20, and CD38, or solid tumors e.g. MCF-7, MDA-MB-231, and PC-3 that express HER2, MUC1, and PSMA tumor antigens, respectively.

For hematologic cancers, multiple antibodies (conjugated with PE) including anti-CD19, anti-CD20, and anti-CD38 were used to target NALM6 cells. After staining, the unbounded antibodies were washed away and stained NALM6 cells were co-cultured with the PEbody CAR T cells. The results consistently showed that the PEbody CAR T cells killed the corresponding tumor cells when Ab-PE binds to tumor antigens whereas slight killing ability was observed in non-stained NALM6 cells (Figure 3.4B). In addition to hematologic cancers, similar experiment has been conducted for solid cancer cells (Figure 3.5A). Different cancer cell lines were co-cultured with PEbody CAR T cells such as the breast cancer cell lines, e.g. MCF-7 and MDA-MB-231, and the prostate cancer cell line, e.g. PC-3. Each cell line endogenously expresses multiple antigens such as (1) MCF-7 express Mucin-1 (MUC-1) and human epidermal growth factor receptor 2 (HER-2); (2) MDA-MB-231 express MUC-1 and programmed death-ligand 1 (PD-L1). However, PC-3 cells were over expressed the prostate-specific membrane antigen (PSMA) on the cell surface. All cancer cell lines were also genetically encoded with fire-fly luciferase for the luciferase killing assay. To verify the killing ability of PEbody CAR system to target different antigens on the solid cancers, PEbody CAR T cells were co-cultured with different cancer cells in the presence and absence of appropriate PE-conjugated antibodies (Figure 3.5B). The results showed that the killing efficiency of PEbody CAR T cells in the presence of PE-conjugated antibodies was significantly better than those in the absence of PE-conjugated antibodies. This indicated that PE-conjugated antibodies are selective to PEbody CAR and can increase the killing ability of PEbody CAR T cells. In addition, this PEbody CAR system can modulate the T cell response because the number of antigens is limited on the surface of cancer cells such as MDA-MB-231 (Figure 3.6), adding multiple antibodies to target different antigens on one cancer cell could enhance the killing ability of PEbody CAR T cells.

3.4 Discussion

We reported here the development of a simple, modular, and universal CAR system. A unique feature of this CAR design is that it has multiple tunable variables, such as (1) the affinity between tumor antigen and antibody, (2) the concentration of PE-conjugated antibody, and (3) the expression level of the PEbody CAR, that can be used to modulate the T cell response. Our results consistently showed the killing efficacy of PEbody CAR T cells to both hematologic and solid tumor cells *in vitro*, but there is a need to further validate this system *in vivo*. Although there are many reported universal CARs which mostly based on scFv from murine [11, 12], the PEbody (10 kDa) has its own advantages that it is derived from human protein which could not cause immunogenicity in human. Furthermore, we previously observed that the interaction between the PEbody and PE was multivalent which means that a certain amount of the PEbody was required to interact with PE [15]. This could be a beneficial property for CAR T cells to target solid tumor cells. Because some antigens are present in the normal tissues with low expression level, using high-affinity CARs can cause on-target but off-tumor effect [21]. Thus, multivalent interaction between the PEbody and PE could allow for antigen-induced receptor clustering on only highly expressed antigen-presenting tumor cells which enhances the killing efficacy of PEbody CAR T cells. Furthermore, the PEbody CARs can only recognize PE whereas antibodies targeting antigens can be changed. This one CAR construct could avoid antigen loss on tumor cells which usually occurs when engineered T cells targeting a single antigen were applied [22, 23]. In addition to the PEbody receptor expressed on T cells, the tumor-targeting PE-conjugated full-length antibodies are more stable and have longer half-life in circular system than scFv fragments or other small molecules [24]. This could reduce the frequency of adaptors injection. However, producing full-length antibodies requires more sophisticated methods and higher cost than producing antibody fragments. In conclusion, our PEbody CAR system, whose extracellular portion is engineered through directed evolution, provides an alternative design of CAR that can target different antigens without having to re-engineer the T cells and it could also manage major safety issues associated with fixed antigen CAR.

3.5 Figures

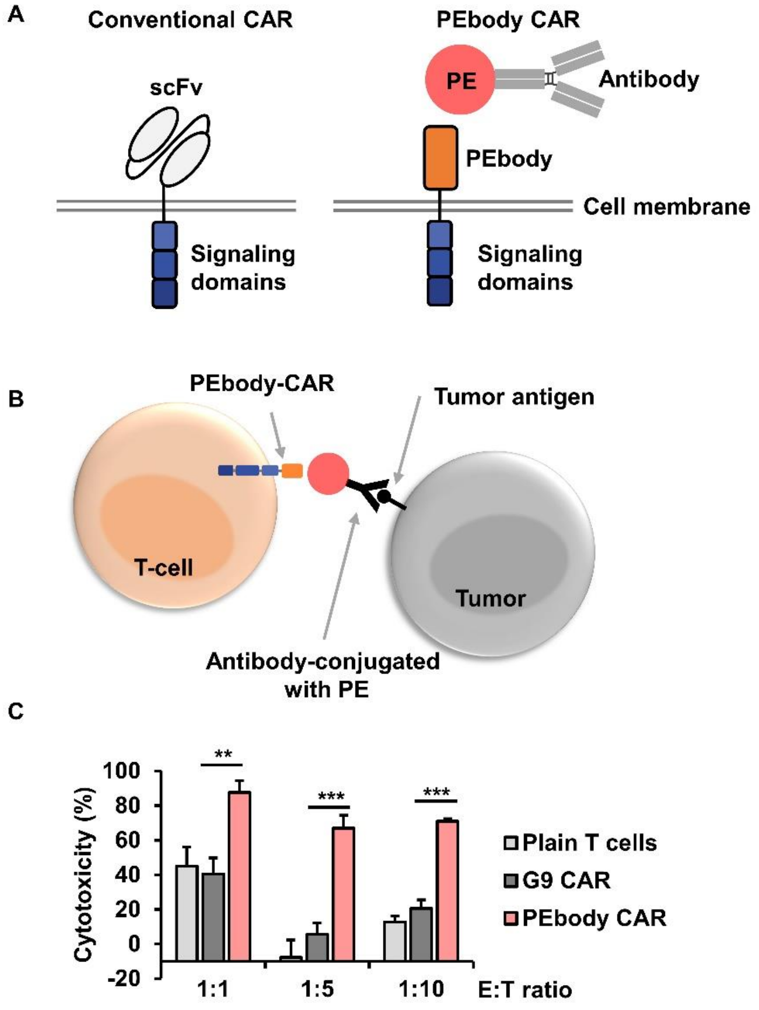
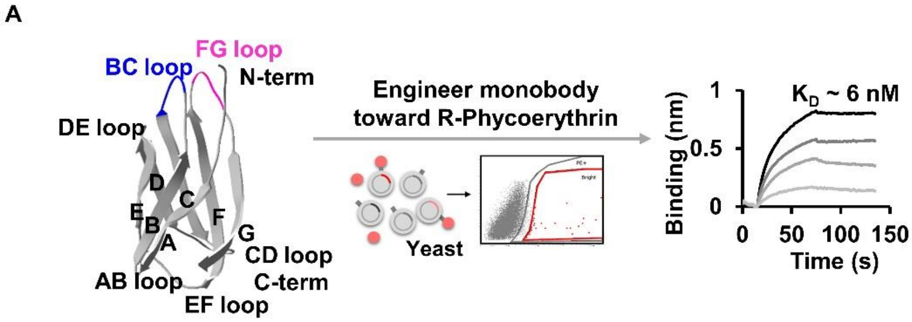


Figure 3.1: Design and characterization of the PEbody CAR system

(A) Comparison between the conventional CAR and PEbody CAR design. The conventional CAR structure includes an extracellular antigen recognition domain (e.g. scFv) fused to intracellular TCR signaling domains (CD3ζ) and co-stimulatory domains (e.g. CD28 and 4-1BB). While the intracellular of the PEbody CAR system is similar to the conventional CAR, the extracellular recognition domain is composed of two components: a PEbody, an engineered human fibronectin type III domain, and an antibody conjugated with PE dye. (B) The schematic diagram of how PEbody CAR-T cell targets tumor cells. The PEbody on the extracellular domain of the PEbody CAR binds to PE-conjugated antibody which targets specific antigen on the surface of tumor cells. (C) The PEbody CAR specifically targeting PE-conjugated anti-CD19 antibody on tumor cell surface. Human primary T cells without (Plain) and with CAR constructs (G9 CAR or PEbody CAR) were co-cultured *in vitro* with NALM6 cancer cells which were prior stained with the anti-CD19 antibody conjugated with PE. The luciferase signal for cytotoxicity calculation was measured by a plate reader. Data in (C) are represented as mean ± SD (n = 4). The asterisk indicates a significant difference (**$p < 0.01$, ***$p < 0.001$; t-test).



B

Monobody	BC loop	FG loop
G9	TMVTQ	WRWRY
PEbody	FINFK	WPWWY

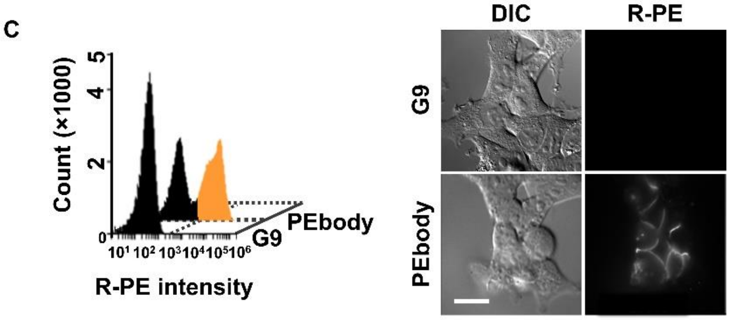


Figure 3.2: Engineering of the PEbody, an R-Phycoerythrin (PE) binder

(A) The schematic diagram showing directed evolution of PE-binding monobody (PEbody). Left: The structure of the G9 monobody (modified from PDB ID: 1TTG). Middle: The G9 monobody was used as a template for directed evolution to evolve itself to a PE binder. The amino acids in the BC and FG loops were randomly mutated These monobody variants were then introduced into yeast. After induction, protein variants were expressed on yeast surface and screened against PE dye. Only cells with high intensity of PE were selected, amplified, and sequenced. Right: The determination of binding affinity between PE and the PEbody by bio-layer interferometry. (B) Nucleotide and amino acid sequences in the BC and FG loops of the G9 and PEbody. (C) PE binding to PEbody but not G9. Left: The flow cytometry histogram results showing the distribution of PE staining intensity on the yeast cell surface displaying the G9 or PEbody. The population with black color refers to cells that cannot bind to PE and orange color represents those that bind to PE. Right: The imaging results from fluorescence microscope showing PE intensity and the differential interference contrast (DIC) images of HEK293 cells expressing the G9 or PEbody. Scale bar: 20 µm.

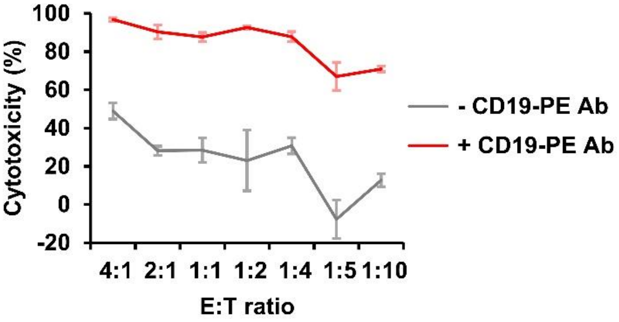


Figure 3.3: The luciferase-based killing assay between the PEbody CAR T cells and tumor cells at different E:T ratios

T cells expressing the PEbody CAR (E, effector cells) were co-cultured with NAML6 cells (T, target cells) with different ratios. The red line represents NAML6 cells with the anti-CD19-PE antibody staining and the grey line represents cells without antibody staining. The luciferase signal for cytotoxicity calculation was measured by a plate reader. Data are represented as mean ± SD (n = 3).

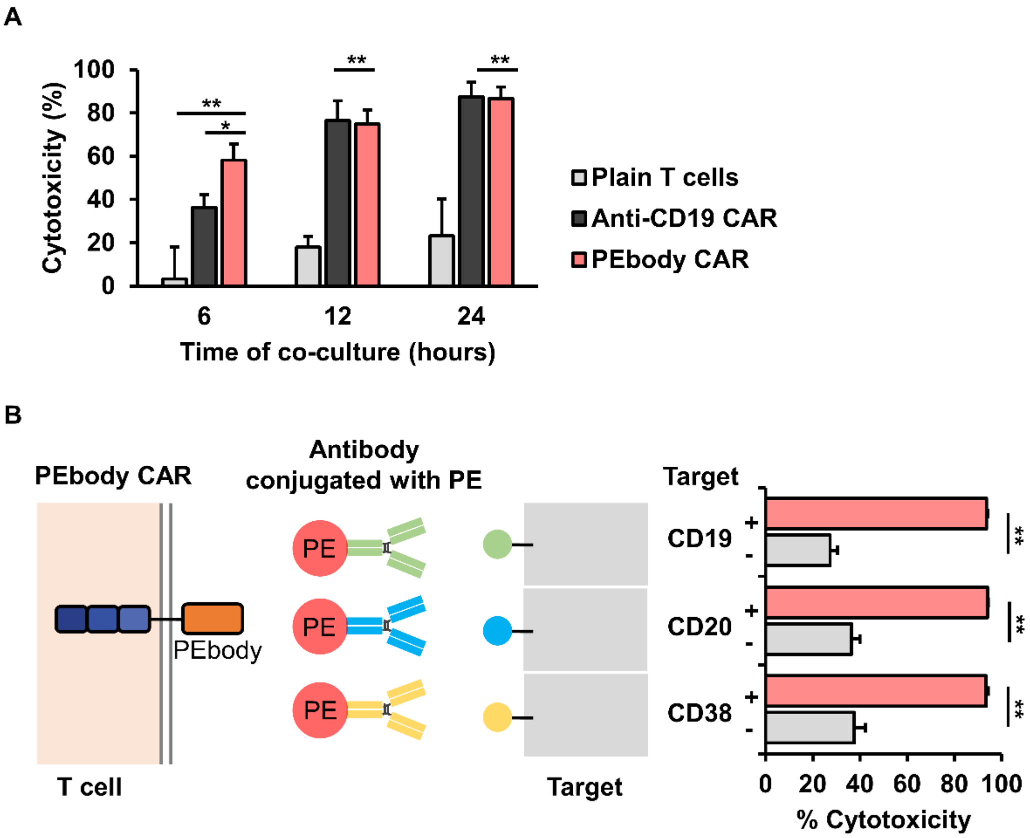


Figure 3.4: The PEbody CAR system targeting different antigens on hematologic cancers

(A) The comparison of cytotoxicity between the anti-CD19 CAR and PEbody CAR at different length of co-culturing. Human primary T cells without (Plain) and with anti-CD19 CAR constructs were co-cultured *in vitro* with NALM6 cells for 6, 12, and 24 h. In addition, T cells expressing the PEbody CAR were also co-cultured with NALM6 which were prior stained with the anti-CD19 antibody conjugated with PE for 6, 12, and 24 h. After co-culturing, cells were harvested for the luciferase-based killing assay. (B) The PEbody CAR targeting different antigens on the NALM6 cells. Different PE-conjugated antibodies such as anti-CD19, CD20, and CD38 antibodies were used to stained NALM6 cells. PEbody CAR T cells recognized PE on the surface of tumor cells and killed them. The luciferase signal of remaining cells was measured for cytotoxicity calculation. % of cytotoxicity of PEbody CAR T cells in the presence of PE-conjugated antibodies (+) is significantly higher than those in the absence of PE-conjugated antibodies (-). Data in (A-B) are represented as mean ± SD (n = 4). The asterisk indicates a significant difference ($*p < 0.05$, $**p < 0.01$; t-test).

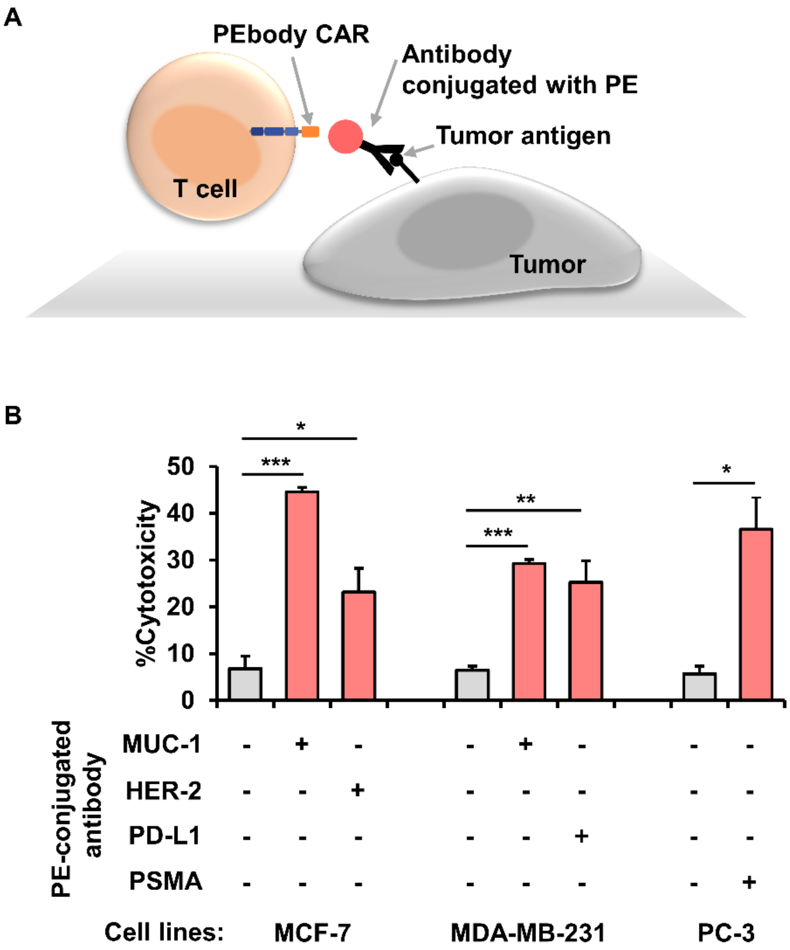


Figure 3.5: The PEbody CAR system targeting different antigens on solid cancers

(A) The schematic diagram of PEbody CAR-T cells directing tumor cells. A PEbody CAR has a PEbody as the extracellular domain of the CAR which binds to a PE-conjugated antibody, while this antibody targets specific antigen on the surface of solid tumor cells. (B) Killing efficiency of PEbody CAR T cells on different solid cancer cell lines such as breast cancer cell lines, e.g. MCF-7 and MDA-MB-23, and the prostate cancer cell line, e.g. PC-3. Different PE-conjugated antibodies were added to PEbody CAR T cells co-cultured with different cancer cell lines such as (1) anti-MUC-1 or anti-HER-2 for MCF-7, (2) anti-MUC-1 or anti-PD-L1 for MDA-MB-231, and (3) anti-PSMA for PC-3. The (+) sign indicates the addition of the PE-conjugated antibody. The (-) sign indicates the absence of the PE-conjugated antibody. The killing efficiency was calculated as % of cytotoxicity. Data in (B) are represented as mean ± SD (n = 3). The asterisk indicates a significant difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; t-test).

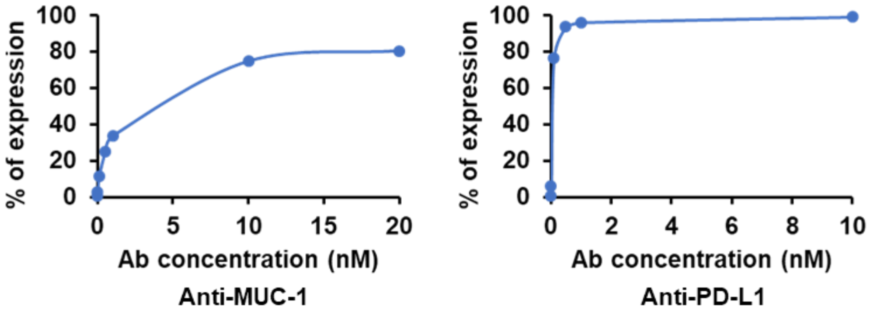


Figure 3.6: Antigen presenting on solid tumor cells

The % expression of different antigens (Left: MUC-1 and Right: PD-L1) on the surface of MDA-MB-231. The expression was measured by flow cytometry.

Acknowledgements

Chapter 3, in full, is currently being prepared for submission for publication of the material. P. Limsakul, Y. Wu, Y. Pan, M. E. Allen, and Y. Wang. “PEbody CAR T. A Universal Chimeric Antigen Receptors for T Cell Immunotherapy”. The dissertation author was the first author of this paper.

This work is supported by grants from NIH HL098472, HL109142, HL121365 (Y. Wang), NSF CBET1360341, DMS1361421 (Y. Wang and S.L.), UC San Diego, and Beckman Laser Institute, Inc. (Y. Wang). P.L. acknowledges fellowship support from the Thai Ministry of Science and Technology. The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. June, C.H. and M. Sadelain, *Chimeric antigen receptor therapy.* New England Journal of Medicine, 2018. **379**(1): p. 64-73.
2. Ribas, A. and J.D. Wolchok, *Cancer immunotherapy using checkpoint blockade.* Science, 2018. **359**(6382): p. 1350-1355.
3. Kalos, M., B.L. Levine, D.L. Porter, S. Katz, S.A. Grupp, A. Bagg, and C.H. June, *T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia.* Science translational medicine, 2011. **3**(95): p. 95ra73-95ra73.
4. Fesnak, A.D., C.H. June, and B.L. Levine, *Engineered T cells: the promise and challenges of cancer immunotherapy.* Nature Reviews Cancer, 2016. **16**(9): p. 566.
5. Neelapu, S.S., S. Tummala, P. Kebriaei, W. Wierda, C. Gutierrez, F.L. Locke, K.V. Komanduri, Y. Lin, N. Jain, and N. Daver, *Chimeric antigen receptor T-cell therapy—assessment and management of toxicities.* Nature reviews Clinical oncology, 2018. **15**(1): p. 47.
6. Davila, M.L., I. Riviere, X. Wang, S. Bartido, J. Park, K. Curran, S.S. Chung, J. Stefanski, O. Borquez-Ojeda, and M. Olszewska, *Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia.* Science translational medicine, 2014. **6**(224): p. 224ra25-224ra25.
7. Fedorov, V.D., M. Themeli, and M. Sadelain, *PD-1–and CTLA-4–based inhibitory chimeric antigen receptors (iCARs) divert off-target immunotherapy responses.* Science translational medicine, 2013. **5**(215): p. 215ra172-215ra172.
8. Kenderian, S.S., M. Ruella, O. Shestova, M. Klichinsky, V. Aikawa, J.J. Morrissette, J. Scholler, D. Song, D.L. Porter, and M. Carroll, *CD33-specific chimeric antigen receptor T cells exhibit potent preclinical activity against human acute myeloid leukemia.* Leukemia, 2015. **29**(8): p. 1637.
9. Lim, W.A. and C.H. June, *The principles of engineering immune cells to treat cancer.* Cell, 2017. **168**(4): p. 724-740.

10. Zah, E., M.-Y. Lin, A. Silva-Benedict, M.C. Jensen, and Y.Y. Chen, *T cells expressing CD19/CD20 bi-specific chimeric antigen receptors prevent antigen escape by malignant B cells.* Cancer immunology research, 2016: p. canimm. 0231.2015.

11. Ma, J.S., J.Y. Kim, S.A. Kazane, S.-h. Choi, H.Y. Yun, M.S. Kim, D.T. Rodgers, H.M. Pugh, O. Singer, and S.B. Sun, *Versatile strategy for controlling the specificity and activity of engineered T cells.* Proceedings of the National Academy of Sciences, 2016. **113**(4): p. E450-E458.

12. Cho, J.H., J.J. Collins, and W.W. Wong, *Universal Chimeric Antigen Receptors for Multiplexed and Logical Control of T Cell Responses.* Cell, 2018.

13. Rodgers, D.T., M. Mazagova, E.N. Hampton, Y. Cao, N.S. Ramadoss, I.R. Hardy, A. Schulman, J. Du, F. Wang, and O. Singer, *Switch-mediated activation and retargeting of CAR-T cells for B-cell malignancies.* Proceedings of the National Academy of Sciences, 2016. **113**(4): p. E459-E468.

14. Koide, A., C.W. Bailey, X. Huang, and S. Koide, *The fibronectin type III domain as a scaffold for novel binding proteins1.* Journal of molecular biology, 1998. **284**(4): p. 1141-1151.

15. Limsakul, P., Q. Peng, Y. Wu, M.E. Allen, J. Liang, A.G. Remacle, T. Lopez, X. Ge, B.K. Kay, and H. Zhao, *Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface.* Cell chemical biology, 2018. **25**(4): p. 370-379. e4.

16. MacColl, R., L.E. Eisele, E.C. Williams, and S.S. Bowser, *The discovery of a novel R-phycoerythrin from an Antarctic red alga.* Journal of Biological Chemistry, 1996. **271**(29): p. 17157-17160.

17. Kubitscheck, U., *Fluorescence Microscopy: From Principles to Biological Applications*. 2nd ed. John Wiley & Sons. 2017.

18. Grupp, S.A., M. Kalos, D. Barrett, R. Aplenc, D.L. Porter, S.R. Rheingold, D.T. Teachey, A. Chew, B. Hauck, and J.F. Wright, *Chimeric antigen receptor–modified T cells for acute lymphoid leukemia.* New England Journal of Medicine, 2013. **368**(16): p. 1509-1518.

19. Porter, D.L., B.L. Levine, M. Kalos, A. Bagg, and C.H. June, *Chimeric antigen receptor–modified T cells in chronic lymphoid leukemia.* New England Journal of Medicine, 2011. **365**(8): p. 725-733.

20. Alabanza, L., M. Pegues, C. Geldres, V. Shi, J.J. Wiltzius, S.A. Sievers, S. Yang, and J.N. Kochenderfer, *Function of novel anti-CD19 chimeric antigen receptors with human variable regions is affected by hinge and transmembrane domains.* Molecular Therapy, 2017. **25**(11): p. 2452-2465.

21. Liu, X., S. Jiang, C. Fang, S. Yang, D. Olalere, E.C. Pequignot, A.P. Cogdill, N. Li, M. Ramones, and B. Granda, *Affinity-tuned ErbB2 or EGFR chimeric antigen receptor T cells exhibit an increased therapeutic index against tumors in mice.* Cancer research, 2015. **75**(17): p. 3596-3607.

22. Lee, D.W., J.N. Kochenderfer, M. Stetler-Stevenson, Y.K. Cui, C. Delbrook, S.A. Feldman, T.J. Fry, R. Orentas, M. Sabatino, and N.N. Shah, *T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial.* The Lancet, 2015. **385**(9967): p. 517-528.

23. June, C.H., R.S. O'Connor, O.U. Kawalekar, S. Ghassemi, and M.C. Milone, *CAR T cell immunotherapy for human cancer.* Science, 2018. **359**(6382): p. 1361-1365.

24. Nelson, A.L., *Antibody fragments: hope and hype.* mAbs, 2010. **2**(1): p. 77-83.

Chapter 4

Optimization of the Whole FRET Biosensors Through Directed Evolution

4.1 Introduction

Chimeric antigen receptor (CAR) T cell therapy, which is based on the genetic reprogramming of T cells with a synthetic CAR to directly destroy malignant cells, have become a paradigm-shifting therapeutic approach for cancer treatment [1, 2]. The CAR T cell activation is mediated through the engagement of the CAR with a tumor-associated antigen, resulting in intricate series of signaling events that involves the activation of protein tyrosine kinases (PTKs) such as Src and Syk families [3]. For the T cell receptor (TCR), upon engagement of the TCR and a co-receptor CD4 or CD8 by a peptide-bound MHC molecule, the Src family kinase Lck phosphorylates ITAMs present in the cytoplasmic regions of the CD3 chains and TCR ζ-chains [4, 5]. The Syk family kinase ZAP70 then binds to doubly phosphorylated ITAMs resulting in a conformational change that allows for its phosphorylation by Lck. Activated ZAP70 phosphorylates the adaptor molecules LAT and SLP76, which have a central role in nucleating the recruitment of downstream effector molecules [4, 5]. In addition to activation of Lck, the Src family kinase Fyn is also involved in the initiation of TCR signaling pathways [4]. These events further activate downstream signaling pathways leading to T cell activation. However, the regulation mechanisms on how the PTKs control the CAR signals is poorly understood. Therefore, it is important to unravel the spatiotemporal dynamics of these kinase activities in single live cells which can advance our understanding in CAR T cell regulation and pave the way for appropriate therapeutic designs in the future.

Developing imaging tools that allow for studying dynamics of PTKs signal transduction in living cell is necessary to understand kinase activities. Genetically-encoded biosensors based on FRET (FRET biosensors) have been widely used to dynamically track activity of PTKs in living cells because of their sensitive ratiometric fluorescence readout, high signal-to-noise ratio, precise molecular targeting, and high spatiotemporal resolution [6, 7]. A general modular design of FRET biosensors based on a phosphorylation-dependent conformational switch has been successfully applied to many tyrosine kinases [7-9]. These reporters share a common mechanism that is a substrate peptide becomes phosphorylated and then is recognized by and becomes bound to a binding domain. However, the limitation in sensitivity, specificity, and dynamic range of these biosensors have hindered their broader applications [10, 11]. Several research groups have developed FRET biosensors by changing the FP pairs, the distance between them [7], the relative orientation between them [12], the order of each domain of FRET biosensors [13], or the selectivity/affinity of phosphobinding domain and a phosphorylated substrate peptide [14, 15]. Although the general construct of intramolecular FRET biosensors is simply based on modular design, their development involves trial and error. Thus, a more efficient method which allows for simultaneously screening a large number of biosensor variants for ratiometric response to PTK activity is required.

With a goal of establishing a rapidly and systematic approach for FRET biosensor optimization, directed evolution which involves generating genetic diversity and screening for protein variants with desired properties has been applied [16-19]. Schulz al et. created a set of 36 vectors carrying a variety of FPs and different length linkers [17]. The activation of each variant upon stimulation was monitored by low-throughput time-lapse microscopy and the variant with largest dynamic range was manually selected for further identification. Apart from a small library size of mammalian cell-based screening, Campbell at el. developed the method to optimize the linkers between the peptide substrate and the phosphobinding domain of FRET biosensors using a bacterial colony-based screening under two expression regimes, e.g. expression of sensor variant alone or in the presence of a co-expressed enzyme [18]. However, throughput was limited because each colony had to be manually selected. Later, the same group created a larger library, in which the linker sequence was varied and different variants of CFP and YFP were used, and screened functional colonies based on large FRET efficiency before and after inducible kinase expression [20]. In addition to FPs pair and linkers, proper affinity between the peptide substrate and the phosphobinding domain also affects FRET efficiency. Recently, we utilized a yeast surface display technology to optimize binding affinity between phosphobinding domain (SH2 domain) and substrate containing tyrosine [15]. We were able to improve the dynamic range of existing FRET biosensors to more than 100% change from the basal level upon activation, but additional selection and optimization steps were required because higher affinity *in yeast* does not directly translate into better FRET sensitivity *in mammalian cell*. Overall, these methods were generally aimed for optimizing the conformation of FRET biosensor to gain larger FRET dynamic range, but still lacked of optimizing their specificity toward a certain target. Besides, more adjustment is required when screening is carried out in other display systems before able to use in mammalian cells. Therefore, optimizing the FRET biosensors directly in mammalian cell through function-based screening is desirable.

In this chapter, we introduced a systematic and universal method, which combines several technologies together such as site-saturation mutagenesis, mammalian cell-based and function-based high-throughput screening by FACS, and high-throughput DNA sequencing, to directly optimize the whole FRET biosensors for kinase activity associated with T cell activation (Figure 4.1). In general, the component of these FRET biosensors consists of a pair of FPs (for FRET action), a sensor region (a substrate peptide containing a tyrosine phosphorylation site), a ligand region (SH2 domain), and a linker peptide to connect each domain. Because the FRET efficiency of intramolecular sensors is primarily influenced by the distance and the relative orientation of the two fluorophores but not highly dominated by FP pairs [7], we chose to optimize the sensor region of FRET biosensors (ECFP/YPet pair) to increase sensitivity, specificity, and FRET dynamic range in reporting kinase activities. To find an optimal sequence for the sensor region, directed evolution was applied to create all possible sequences of this substrate peptide in FRET biosensors. Biosensor variants were introduced into mammalian cells and screened by FACS based on an individual FRET ratio (ECFP emission divided by YPet emission). The sequence information of selected biosensor variants was revealed by high-throughput sequencing (amplicon sequencing) which allows us to explore the protein sequence space of selected biosensors compared with the input library. Based on the information of protein sequence space, indeed, we predicted functional substrate peptide sequences and verified them in FRET biosensors by using live-cell imaging. Therefore, our newly developed method can be used not only to improve the kinase-substrate specificity, but it can also be used to optimize the conformational changes of the whole FRET biosensors.

4.2 Materials and Methods

4.2.1 Plasmid Construction

The gene template for the protein tyrosine kinase biosensor (ELYK) was constructed by polymerase chain reaction (PCR) amplification of the complementary DNA of an enhanced CFP, LacZ, YPet, EV linker (116 amino acids), and either active or mutated kinase domains. All cDNA fragments were fused together (Figure 4.3) and inserted into the lentiviral transfer vector, pSIN, between SpeI and EcoRI with T4 ligation (New England Biolabs) where ECFP is at N-terminal and the kinase domain is at C-terminal. Several restriction sites were introduced such as two of Esp3I sites at each end of LacZ for replacing to the sensing domain (including SH2 domain and substrate peptide) and XbaI/EcoRI for replacing to different kinase domains (Figure 4.3). To construct the biosensor, the cDNA of the sensing domain, which was amplified by PCR from the mutated c-Src SH2 domain (C185A) with a sense primer containing an Esp3I and a reverse primer containing the cDNA of a flexible linker (15 amino acids), a substrate peptide, and an Esp3I site, replaced the LacZ domain via the Golden Gate assembly (New England Biolabs). The substrate peptide sequence can be changed by using different reverse primers.

4.2.2 Library Construction

Two biosensor libraries were created by site-saturation mutagenesis [21]. For Lib1 and Lib2, three residues upstream of the tyrosine residue (EKIXXXYGVV) and downstream of the tyrosine residue (EKIEGTYXXX), respectively, of the wild-type substrate underwent site-saturation mutagenesis by using NNK degenerate primers (IDT), where N represents an equimolar distribution of A, T, G, and C; K represents an equimolar distribution of T and G; X represents any amino acid. Briefly, the cDNA of the substrate variants was generated by PCR with Q5 DNA polymerase (NEB, Cat#M0491) from the c-Src SH2 domain (C185A) with a sense primer containing an Esp3I and a reverse primer containing a flexible linker (15 amino acids), NNK codons on a substrate peptide (an antisense primer for Lib1: 5’-

GGATCGTCTCTTAGAGAGCTCCACGACGCCGTAMNNMNNMNNGATCTTCTCAGAACCCTCACCAGAACCC -3’, and that for Lib2: 5’-GGATCGTCTCTTAGAGAGCTCMNNMNNMNNGTACGTGCCCTCGATCTTCTCAGAACCCTCACCAGAACCC -3’), and an Esp3I site. The template for PCR was the wild-type substrate. The PCR condition of annealing temperature was varied from 55-70˚C and thermocycling condition was 20 cycles. After generating the substrate library with NNK degenerated reverse primers, the fragments containing cDNA library were then extracted from agarose gel, inserted into pSIN-ELYK template vector between Esp3I restriction sites through the Golden Gate assembly (NEB), transformed into *E. coli* (DH10B, Thermo Fisher Scientific), purified plasmid library with Qiagen HiSpeed Plasmid Maxi kit (Qiagen), and sequenced to confirm the mutation regions

4.2.3 Generation of Mammalian Cell Library

The plasmids of biosensor libraries were introduced into mammalian cells (HEK293T cells from ATCC) through virus infection with low MOI to allow low copy number of plasmids per a single cell. Pantropic lentiviruses were produced from Lenti-X 293T cells (Clontech Laboratories, Cat# 632180) co-transfected with a pSIN containing biosensor variants and the viral packaging plasmids pCMVdR8.91 and pCMV-VSV-G using the ProFection Mammalian Transfection System (Promega, Cat#E1200). Viral medium/supernatant was collected 48 h after transfection, filtered with 0.45µm filter (Sigma-Millipore), and concentrated using PEG-*it* virus precipitation solution (System Biosciences, Cat. # LV825A-1). The virus titer was measured by flow cytometry. To generate the mammalian cells library, the concentrated virus with MOI between 0.1-0.3 was directly added into HEK293T cells which were seeded $2x10^6$ cells in 10-cm dish a day before transfection. Cells were then cultured in DMEM medium containing puromycin (2 µg/mL) after 48 h of transfection until screening by fluorescent activated cell sorting (FACS).

4.2.4 Library Screening by FACS

HEK293T cells containing biosensor variants were screened by FACS (BD FACS Aria II Cell Sorter)-based FRET ratio (ECFP/FRET ratio) which was calculated from emission of ECFP (Ex 405-nm laser; Em 450/50 nm) divided by that of FRET (Ex 405-nm laser; Em 454/35 nm). Several controls were used to gate the cell population, for example, plain HEK 293T cells were used to gate for the live cells;

cells with pcDNA3.1-ECFP were used to gate for the cells that expressed ECFP only (Ex 405 nm, Em 450/50 nm); cells with pcDNA3.1-YPet were used to gate for the cells that expressed Ypet only (Ex 488 nm, Em 545/35 nm); the mixture of cells that express either ECFP or YPet was used as a negative gating of FRET signal (Ex 405 nm, Em 545/35 nm); cells co-transfected with both pcDNA3.1-ECFP and pcDNA3.1-YPet were also used to gate for the intermolecular FRET signal. The conformations of intramolecular FRET biosensors can be measured based on the FRET ratio. Thus, cells expressing biosensors fused with active kinase domain (KA) were used to gate for the active conformation of FRET biosensor (high FRET ratio), while, those expressing biosensors with kinase dead domain (KM) were used to gate for the inactive conformation of FRET biosensor (low FRET ratio). In addition to all gating, only cells expressing medium intensity of biosensor (by YPet filter; Ex 488 nm, Em 545/35 nm) were selected for sorting to avoid the abnormal expression of biosensors in cells. After gate setting, cells containing biosensor libraries were sorted into high and low FRET ratio where the median FRET ratio was used to as a threshold for sorting. For example, cells expressing biosensors that had higher FRET ratio than median FRET ratio will be selected for the high FRET ratio and vice versa. Sorted cells were then cultured and further selected in the second rounds of FACS.

4.2.5 Illumina Sequencing of and Analysis

Cells containing biosensor variants were sorted based on the FRET ratio and their sequences in the substrate peptide were sequenced by Illumina HiSeq 4000 sequencing system. The total RNA of each pool of sorted cells were extracted by RNeasy Mini Kit (Qiagen, Cat# 74104). During column purification, the genomic DNA was removed by RQ1 RNase-Free DNase (Promega, Cat# M6101). This allows only RNA that can be encoded to the biosensor proteins to be purified. The RNA was quantified by Nanodrop and gel electrophoresis. The purified total RNA (~500 ng) were used as a template for cDNA synthesis via the SuperScript IV reverse transcriptase (ThermoFisher Scientific, Cat# 18090010) with gene-specific primer. Adaptor sequences with different indexes for Illumina sequencing were added into cDNA by PCR using Q5 DNA polymerase (NEB, Cat# M0491S) with low PCR cycles (< 16 cycles). The amplicon containing all adaptors were confirmed by gel electrophoresis (2% agarose gel) and purified by Zymoclean gel DNA recovery kit (Zymo Research, Cat# D4008). The purified amplicon libraries were sequenced by Sanger sequencing (Genewiz) to verify the success of library preparation and quantified by Qubit prior being sequenced by Illumina HiSeq4000 with 50-bp single-end sequencing (for the entire libraries).

4.2.6 Cell Culture and Transfection

Cell culture reagents were purchased from Thermo Fisher Scientific. The human embryonic kidney cells (HEK293T) was cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM of L-glutamine, 100 units/mL of penicillin, 100 µg/mL of streptomycin, and 1mM of sodium pyruvate at 37°C with 5% CO_2. The DNA plasmids were transfected into cells with Lipofectamine 3000 (Thermo Fisher Scientific, Cat#L3000015) unless otherwise stated.

4.2.7 Microscopy, Image Acquisition, and Analysis

Cells expressing the exogenous biosensor proteins were starved with 0.5% FBS DMEM for 12 h before being subjected to PP1 (10 µg/mL) stimulation. Images were taken with a Nikon Eclipse Ti inverted microscope with a cooled charge-coupled device (CCD) camera with a 420DF20 excitation filter, a 450DRLP dichroic mirror, and two emission filters controlled by a filter changer (480DF30 for ECFP and 535DF35 for YPet). The time-lapse fluorescence images were acquired at 90 s interval by MetaMorph 7.8 software (Molecular Devices). The pixelwise ECFP/FRET ratio images were calculated and visualized with the intensity modified display (IMD) method by Fluocell software [22] (Github http://github.com/lu6007/fluocell) and further analyzed by MATLAB or Excel.

For data presentation, the normalized values were shown to clearly compare the differences among the experimental groups and to minimize the cell-cell heterogeneity. The pre-stimulation baseline for each cell was established by averaging the FRET ratio of each individual cell before stimulation. This provides an internal normalization reference for establishing the stimulation-dependent FRET changes of an individual cell which reduces the cell-cell heterogeneity and noise.

4.2.8 Immunoblotting Analysis.

The immunoblotting analysis was performed using standard procedures and visualized using SuperSignalTM West Pico PLUS Chemiluminescent Substrated (ThermoFisher Scientific). The antibodies that recognize phosphotyrosine (PY20 from Abcam, Cat# ab10321) and GFP (anti-GFP from Abcam, Cat# ab290) were purchased from Abcam.

4.2.9 Quantification and Statistical Analysis

All the experiments were replicated at least three times and represented biological replicates. Data are presented as mean values ± standard deviation (SD) as indicated in the figure legend. Statistical analyses were carried out using MATLAB (version R2018a) and were defined by P values calculated from the two-tailed Student's t tests or Wilcoxon rank sum test by comparison with the relevant control as specified in the figures or figure legends. Levels of significance were $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***). When the sample size $n >= 4$, the Wilcoxon rank sum test was used since it is applicable to random samples of all distributions. For some flow cytometry experiments, when $n=3$, each data point is an average reading from more than 100,000 single cells. According to the central limit theorem, these data follow the Gaussian distribution. Therefore, we used the two-tailed Student's t-tests.

4.3 Results

4.3.1 The Direct Optimization of the Whole FRET Biosensor in Mammalian Cells

Starting with a functional protein is one of the key steps for directed evolution to search for higher elevations on the fitness landscape. [23, 24]. The rationally designed Fyn FRET biosensor [14] was used as a starting template to generate biosensor variants. This biosensor, which has similar structure as Src kinase biosensor [8], consists of an ECFP, a SH2 domain, a flexible linker, a peptide sequence (from $p34^{cdc2}$) that is specific to Fyn kinase, and a YPet (Figure 4.2A). To optimize the specificity, sensitivity, and FRET dynamic range of the Fyn biosensor, the wild-type substrate peptide EKIEGT**Y**GVV was subjected to site-saturation mutagenesis on the neighboring residues of the consensus tyrosine site (e.g. either Lib1: -1, -2, -3 or Lib2: +1, +2, +3, see Figures 4.3A-B) to create a biosensor library consisting of 8,000 variants by using degenerate primers (NNK) [25, 26]. The randomness of the mutation region was verified by the Sanger sequencing that showed equimolar ratios of A, T, C, and G (Figure 4.3C) suggesting the unbiased library generation.

Ratiometric fluorescence readout of FRET biosensors is used to dynamically track kinase activities in living cells. To identify the ECFP/YPet emission ratio (FRET ratio) change of biosensor variants in living cells, the Fyn kinase-active (KA) domain was fused with biosensor variants to allow for intramolecular interaction between the kinase domain and the substrate peptide. Because the linker between the kinase domains and biosensors affects the accessibility of the kinase domain to phosphorylate the substrate peptide, we tested several flexible linkers with different lengths and found that the EV linker consisting of 116 amino acids [7] showed highest phosphorylation level and largest FRET dynamic range (Figures 4.2B and 4.2D). Furthermore, the regulation of phosphorylation level also depends on the specificity between the substrate and the kinase domain. Varying the length of the kinase domain, we found that the domain with 266 amino acids gave the optimum activity (Figure 4.2B-C). In addition to the positive regulation of the Fyn KA, the Fyn kinase-dead (KD) domain was fused with the biosensor variants to allow for the negative selection of the library

To systematically optimize substrate peptides of the Fyn biosensor, four libraries including two mutation regions (e.g. Lib1: EKIXXXYGVV or Lib2: EKIEGTYXXX where X can be any amino acid) and two kinase domains (e.g. KA or KD) were created (Figure 4.4A). Utilizing iterative cycles of directed evolution, we introduced each library into HEK293T cells using lentiviral transduction with low MOI of 0.3 to allow ~30% of cells to express one or a few copies of biosensor variants, whereas those without biosensors were removed by purocymin selection medium. The individual library was then screened by FACS based on the conformation of each variant which can be represented by the FRET ratio.

4.3.2 Counter Sorting by Using the Flow Cytometry-Based FRET

A high-throughput screening method by FACS was used to screen functional mutants from large pool of FRET biosensor variants. Two conformations of biosensors e.g. active (high FRET ratio) and inactive (low FRET ratio) which were regulated by attached kinase domains can be identified by flow cytometry-based FRET (Figure 4.4B). To test this hypothesis, we attached the Fyn KA and KD to the wild-type Fyn biosensor (substrate of EKIEGTYGVV) and tested its activity using western blot, fluorescent imaging, and flow cytometry (Figures 4.3B and 4.4C-D). The results from the western blot showed that KA can phosphorylate the substrate of Fyn biosensor but not for KD (Figure 4.3B). The phosphorylation event following by the conformational change of the biosensor can be also visualized by fluorescence microscope. At the resting state, the FRET ratio of the biosensor with KA was high and reduced upon treatment with the PP1, a selective inhibitor of Fyn kinase, while no change was observed in the biosensor with KD (Figure 4.4C). Interestingly, the fold difference of basal FRET ratio between biosensors with either KA or KD measured by microscope was similar to that measured by flow cytometry regardless of dissimilar imaging setups (Figures 4.4C-D). Moreover, the FRET emission intensity profile of biosensors containing KA or KD were clearly separated in the scatter plot between FRET intensity (*y*-axis) and ECFP intensity (*x*-axis) (Figure 4.5A) in both measurements. This distinct profile was important to create a region of interest for flow cytometry. Overall, the characteristics of basal FRET ratio detected from either flow cytometry or fluorescence microscope were consistent and could be exchangeable (from our experiment and others [27]). Despite monitoring FRET dynamic change via microscope as a single-cell fashion, the FRET signal detected from flow cytometry can be used to screen for functional biosensor variants in a high-throughput fashion.

Because different conformations of biosensors can be identified by FRET ratios, the flow cytometry-based FRET was used to screen for functional biosensor variants (Figure 4.6A). To precisely screen those desired biosensor variants, seven controls were used to identify populations of cells with different characteristics, including (1) plain HEK293T cells, cells transfected with (2) ECFP, (3) YPet, (4) the mixture of cells transfected with either ECFP or YPet, (5) both ECFP and YPet, and the wild-type biosensor with (6) KA, or with (7) KD (Figure 4.5B). The first four controls were used for background and bleed through subtraction, the fifth control was used to discard the effect of the intermolecular FRET signal, and the last two controls were used to distinguish low and high FRET ratio. Moreover, to avoid the false positive signals due to the abnormally low or high biosensor expressions, we only selected cells expressing the medium level of biosensors which was represented by the medium intensity of YPet. Using these gating systems, biosensor variants from each library whose signals appeared in either low or high FRET ratio gates were selected (Figures 4.5C and 4.6B). For example, variants from Lib1 (with either KA or KD) were sorted into two sub-population e.g. low and high FRET ratios. Therefore, there were four sub-population sorted from Lib1 (Figure 4.6C). Because functional variants should have high FRET ratio when attaching with KA and low FRET ratio when attaching with KD, only these two groups were sequentially enriched for the next round of screening. To verify our sorting strategy, the phosphorylation level of sorted cells was detected by the western blot and the FRET ratio changes of sorted cells were monitored by fluorescence microscope before and after PP1 treatment (Figures 4.6D-F). We found that more than 60% of cells showed FRET ratio changes upon stimulation. This indicated that our library screening strategy can indeed enrich functional biosensors.

4.3.3 Assessing Sequence-Function Relationship in the Substrate Peptide

The information of protein sequence space from different libraries were revealed by high-throughput DNA sequencing. After cells were sorted based on their individual FRET ratio into low and high ratios, their RNA was extracted and synthesized for cDNA. The Illumina’s adaptors were added to cDNA by PCR with low cycle numbers to create unbiased amplification of amplicon libraries which were later sequenced by the Illumina HiSeq4000. Individual pool of library was labeled with different barcode. We acquired sequencing data for the input libraries as well as after one and two rounds of screening, with an average of 21.6 million raw reads and the mean quality score more than 39 per library. The average of 16 million reads per library passed quality filtering and were converted to unique amino acid sequences for further analysis.

Because the screening by FACS enriched cells containing functional protein variants and depleted nonfunctional variants, the change in frequency of each variant from input to selection served as a measure of its function [28]. In other word, functional variants should increase in abundance after selection and be sequenced more often and vice versa for nonfunctional variants [29]. Thus, we compared the population of protein variants presented in the input library to the populations found in the round-one and round-two libraries. To assess sequence-function relationships of the substrate peptide, several parameters were computed and filters were applied (Figure 4.7A) [29]. The frequency of unique sequences (variants) was computed by normalizing the variant count in each library to the total number of sequencing reads for that library. This calculation could avoid the bias due to sequencing depths because different libraries had different total sequencing reads. Thus, that frequency of a unique sequence can be compared across different libraries. The frequency for a given variant, v, in the X^{th} round ($F_{v,X}$) was

$$F_{v,X} = \frac{count_{v,X}}{\sum_{all\ v} count_{v,X}}.$$

The frequency data was later used to compute variant enrichment ratios between successive rounds which allowed us to find the fold enrichment of that variant before and after sorting. Moreover, this calculation could avoid the bias due to library preparation i.e. different starting number of variants from input library. The enrichment ratio for a given variant, v, the X^{th} round was

$$E_{v,X} = \frac{F_{v,X}}{F_{v,input}}.$$

Like the enrichment ratio, the fitness score was used to compare the enrichment ratio of each variant to that of the wild-type in the same generation. Because the wild-type was generally expected to be enriched after selection, the variant of which the fitness score was close to or more than one was likely to behave similar to the wild-type. Fitness scores were calculated for each variant, v, to enable comparison to wild type in the X^{th} round was

$$W_{v,X} = \frac{E_{v,X}}{E_{wide\ type,X}}.$$

The enrichment ratio is commonly used to construct the protein sequence space [28]. Due to different conformational changes of FRET biosensors depending on kinase activities, the functional biosensor variants should have high FRET ratio (KAH) when the kinase was active and low FRET ratio (KDL) when the kinase was inactive (Figures 4.4 and 4.8A). We called this condition as “a positive selection”. While the nonfunctional biosensor variants did not change their conformation, they stayed in the regions outside of a positive selection. To increase the stringent condition for screening of “a negative selection”, we discarded variants that had low FRET ratio conformation (KAL) when the kinase was active and high FRET ratio (KDH) when the kinase was inactive. Following our counter selection, the number of sequencing reads of functional variants should be higher in a positive selection group (i.e. KAH and KDL) and lower or none in a negative selection group (i.e. KAL and KDH) than that of input variants and vice versa for nonfunctional variants. In other word, optimal variants tended to have high enrichment ratio in a positive selection but low ratio in a negative selection group (Figures 4.7B and 4.8B). This counter relation can be clearly illustrated by the sequence-function heat map of the enrichment ratio of all existing variants (Figure 4.8C). The results showed that the sequences of Lib2 were more diverse and had higher enrichment ratio than those of Lib1. Besides, the amino acid sequences that were highly enriched in the positive selection group contained the non-polar side chains. The trend was opposite for the negative selection group that the amino acid sequences preferred the positively charged side chains. However, the enrichment ratio of KDH had minimal change. It is likely that most of variants in this group were deleterious and cannot provide any information for sequence-function analysis. In addition to the heat map, the relationship of the positive selection group can be illustrated in a scatter plot between the enrichment ratios of KAH/KA and KDL/KD (Figure 4.7C). We selected the set of variants whose enrichment ratios were on the boundary of this scatter plot because these dots represented variants with the optimum tradeoff of enrichment ratio between KAH/KA and KDL/KD. The selected set of variants was then filtered by the following conditions: (KAH/KA > 1 and KDL/KD > 1) and (KAL/KA < 0.5 and KDH/KD < 0.5) (Figures 4.7B). The variants which were satisfied all conditions were selected, and their FRET ratio change upon PP1 treatment were measured by fluorescence microscope.

To verify the FRET ratio change of selected substrates, the sequences of these substrates were inserted into the ELYK biosensor template attached with Fyn KA. The FRET response was monitored before and after the treatment of PP1 which allows for the biosensor to change the confirmation from high to low FRET ratio. More than 86% or 94% of selected substrates from either Lib1 (XXXYGVV) or Lib2 (EGTYXXX), repectively, have higher FRET ratio changes than the wild-type substrates. For example, the time-lapse imaging of selected variants (with KA or KD) from Lib1 (#19) and Lib2 (#4) demonstrated that these variants had larger FRET dynamic range and FRET ratio change than the wild-type (Figure 4.7D). We also combined the selected sequences which have large FRET ratio changes (>40%) from Lib1 and Lib2 to generate the combined library (LibC: XXXYXXX). More than 86% of selected combined library have greater FRET ratio changes compared with that of the wild-type one. After verifying the FRET ratio changes of > 100 variants that appeared on or inside the boundary of the KAL/KA vs KDH/KD plots, we used this information to further predict the relation between sequence and function (% of FRET ratio change). The score on the heat map in Figure 4.7E was calculated from the FRET ratio changes of the mutant normalized with that of the wide type. The amino acid sequence of which score was high (>0) represented a preferable amino acid that could involve in “active” or “inactive” conformation of biosensor. This result was similar to the heat map results of all variants (Figure 4.8C). The amino acid sequences in position +1, +2, +3, which mostly contained the non-polar side chains, had higher score than position -3, -2, -1, which likely contained polar uncharged side chains. Based on this information, we found that the hypothetical sequence-function heat map of the enrichment ratios of all variants behaved similar to the experimental results. Thus, combining counter sorting and amplicon sequencing is a powerful high-throughput approach to search for the optimal variants.

4.3.4 ZAP70 Kinase Biosensor

In addition to Fyn FRET biosensor optimization, we further utilized our systematic and universal approach to improve the ZAP70 FRET biosensor. The ZAP70 kinase is another important kinase that works along with Fyn and Lck kinase to activate T cells. Using rational design, other researcher groups and our group had developed the ZAP70biosensors. However, their FRET dynamic ranges upon physiological stimulation in T cells were around 10-20%, and they were reported to have cross-interaction with other kinases in Syk family kinase [9, 30, 31]. Thus, the optimized ZAP70 biosensor with large dynamic range and high specificity is needed.

Similar to the Fyn FRET biosensor, we first started with optimization of the biosensor construct (Figure 4.9A). Different length of ZAP70 kinase domains and substrate peptides were inserted into pSIN-ELYK template (Figures 4.3A). The western blot results revealed the basal phosphorylation levels of the biosensor with kinase#3 were higher than other kinase domains with the same substrate peptide: GDDIYEDIIKVE (wild-type) (Figure 4.9B) [31]. Using this kinase domain for activation, we later changed the substrate domains to the ROZA [9] (SCEDYVNVPES) and ROZA-XL [30] (SREYVNVSGEL) and verified the phosphorylation level by the western blot and monitored the ECFP/FRET ratio (FRET ratio) ratio change in living cells upon stimulation by fluorescence microscope (Figures 4.9C-D). The results consistently showed that ZAP70 biosensor with ROZA-XL substrate and kinase#3 performed better than others in many aspects such as high phosphorylation level large difference of FRET dynamic range, and large FRET ratio change upon the treatment of TAK-659, which is a selective inhibitor for ZAP70 [32]. Consequently, we selected the ZAP70 FRET biosensor which consisted of ROZA-XL substrate and kinase#3 (both KA and KD) for generating the ZAP70 biosensor substrate libraries including zLib1 (XXXYVNVSGEL) and zLib2 (SREYXXXSGEL). Each of these four libraries was introduced to HEK293T cells using lentiviral transduction with low MOI to generate the mammalian cell library containing ZAP70 FRET biosensor variants. Following counter sorting strategy, all libraries will be screened based on their FRET ratio and sequenced. Finally, protein sequence spaces of each sorted group will be explored, and the functional biosensors will be identified.

4.4 Discussion

Here we presented the method which allows for high-throughput screening a large number of FRET biosensor variants to systematically optimize the substrate peptide of the FRET biosensor. We illustrated that our method, which combines directed evolution, the counter sorting via the flow cytometry-based FRET, and amplicon sequencing, is generalized for optimizing Fyn biosensor and could be used to optimize ZAP70 biosensors. Based on the simplicity of library generation, screening, and identification, we can apply this approach to improve any genetically encoded FRET biosensors for post-translational modification.

Developing the universal method to optimize FRET biosensors is challenging because FRET efficiency depends on several factors such as the intrinsic properties of fluorescent proteins (FPs), distance, and orientation between them (our previous designs and their drawbacks can be found in Table 4.1). For simplicity, we based our work on previous biosensors (ECFP/YPet pair) [14, 30] as a starting point for mutation and chose to optimize only the substrate peptide to be sensitive and specific to kinase activities. Three amino acids surrounding the tyrosine (Y) residue (-1, -2, -3, Y, +1, +2, +3) on the substrate peptide were reported to be important for kinase phosphorylation and SH2 interaction [33, 34] of which interactions affected the switching between ON and OFF conformations of FRET biosensors. Thus, nucleotide sequences encoding those six amino acids were mutated with degenerate primers (NNK) which yielded $6.4x10^7$ peptide variants (~10^{10} nucleotide sequences). Because genetically encoded FRET biosensors are generally used to study cell signaling in mammalian cells, we carried out our screening in the cell line, e.g. HEK293T, in which the biosensor is to be used. However, due to the small library size of mammalian cell-based screening (~10^4 variants for genome integrated libraries [35]), we had to reduce the library size by separately mutating three amino acids at position -3, -2, and -1 N-terminal (Lib1) or position +1, +2, and +3 C-terminal (Lib2) to the Y residue. This scaled down the library size to 8,000 peptide variants (~33,000 nucleotide sequences) which can be efficiently handled in the mammalian cells.

The counter sorting that we introduced in this work is one of the key components for library screening. We used FACS to screen a large number of FRET biosensor variants based on their ECFP/FRET ratio but as a drawback of high-throughput screening, undesirable variants can be also selected. To avoid noises, we included multiples controls and used two libraries with different activation, e.g. KA and KD, to identify the optimum substrates for one mutation region. However, our amplicon sequencing results revealed more than 80% of variants were found in round-one of sorting which possibly indicated that there were multiple copies of variants in a single cell. This can cause false positive selection because sorted cells may include both functional and nonfunctional variants. There are several ways to minimize this problem. First, in the library generation step, using a combination of degenerate primers that allow only one codon to be encoded to one amino acid could avoid the bias due to codon degeneracy [36]. Second, when introducing cDNA of biosensor variants to mammalian cells, using lower MOI (< 0.1) for virus infection can decrease the copy number of variants to be introduced into a single cell [37]. Third, multiple rounds of FACS screening may be needed because cells with functional variants could be highly enriched than those with nonfunctional ones [29].

While the high-throughput mammalian-based screening and sequencing did allow us to identify a population of functional variants from a large library, it was not sufficient for identifying the single best variant. Because the changes in the frequency of each variant from input to selection can serve as a measure of its function, we found that most of functional variants had higher enrichment ratio (selection/input of that variant) in the positive selection group (e.g. high and low FRET ratio from KA and KD, respectively) than the negative controls (e.g. low and high FRET ratio from KA and KD, respectively). This indicated that our counter sorting method and sequence-function analysis can be efficiently used to identify functional substrates. However, a small population (< 14%) of identified variants had lower FRET ratio change after PP1 treatment than the wild-type biosensor. It is reasonable that our screening and sequence-function analysis cannot reach 100% efficiency because of the low overall density of functional sequences in the large library in which the vast majority do not code for any functional protein [38]. Another possibility is the uneven distribution of functional sequences. Although representing a very small fraction of all possible sequences, functional sequences are often next to other functional sequences [24, 39]. Our results showed that some of biosensors with similar sequences usually had similar FRET ratio change. Therefore, it is still challenging to identify the best variant with the largest FRET ratio change from the library unless all variants were functionally verified.

Protein-tyrosine kinases (PTKs) and SH2 domains provide a double selection to specific substrate peptides which maintain the fidelity of signaling events [34]. For optimizing substrate peptides of Fyn biosensors, our results showed that the FRET biosensor whose substrate peptide has mutation at position +1, +2, and +3 C-terminal of Y residue showed larger FRET ratio change than those at N-terminal. This result agreed well with previous reports that SH2 domains in general have specific binding sites that include the amino acids at these positions while there is no apparent preference for amino acids at positions N-terminal to Y. [33, 34] Moreover, selectivity of cytosolic PTKs predominantly preferred substrates with hydrophobic amino acids at certain positions, such as at the -1, +1, and +3 positions surrounding Y. This trend was also consistent with our selected substrates that had good performance. Therefore, our hypothetical sequence-function heat map (showing the enrichment ratio, see Figure 4.8C) can be viewed as an all-residue scan [28] which provides insightful information of the relation between protein structure and function.

Upon success, the optimized Fyn, and ZAP70 FRET biosensors can allow for spatiotemporal visualization of kinase responses in single natural and engineered immune cell when they are engaged with target cancer cells. While kinases crucial for immune responses are chosen as the first targets to provide the proof-of-concept for this systematic approach of optimizing biosensors, this biosensor screening template can be easily modified and extended to any FP-based biosensor to detect enzyme-based posttranslational modifications. We believe that this will revolutionize the development of genetically-encoded biosensors and have transformative impact on live cell imaging and cancer therapeutics in general.

4.5 Figures

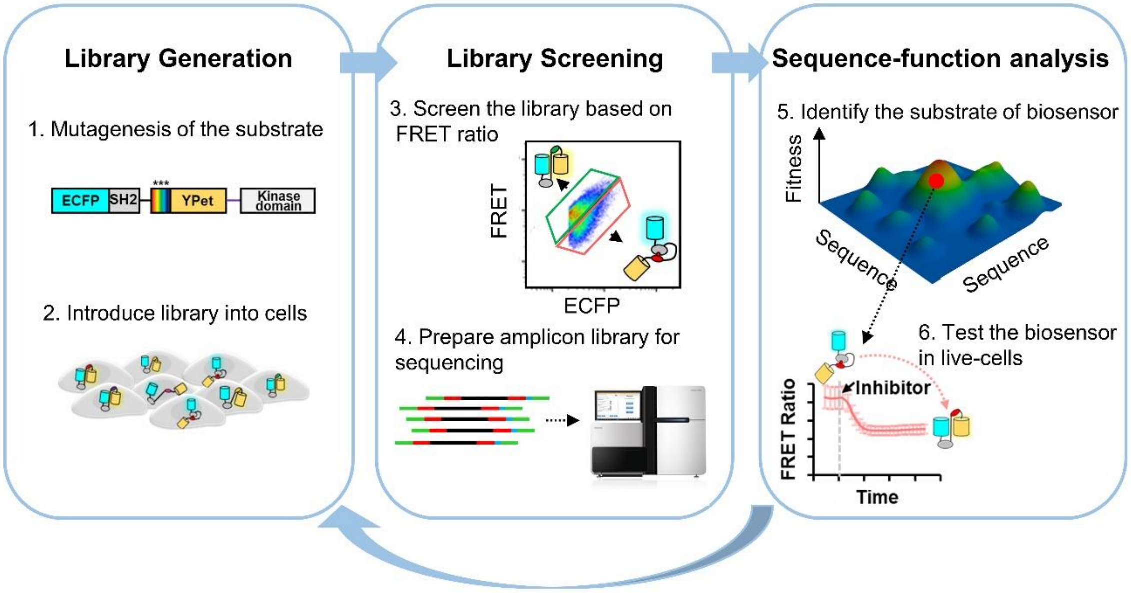
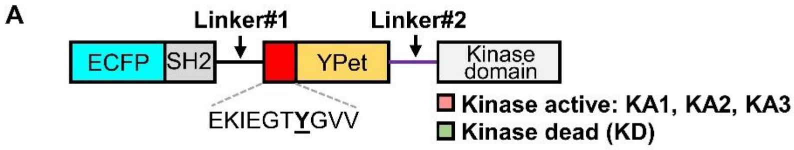
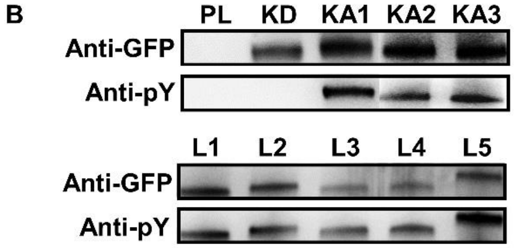


Figure 4.1: The schematic diagram for optimizing FRET biosensors through directed evolution

Several steps were conducted to achieve the desired FRET biosensor variants that have large FRET dynamic range, high specificity and sensitivity. Library generation allows for (1) creating large diversity of biosensor variants and (2) their functions can be expressed via mammalian cells. Library screening was used to (3) screen functional variants based on FRET ratio then (4) these selected variants were revealed by high-throughput DNA sequencing. Finally, the functional scores of each variant from sequencing data were calculated and their sequence space were explored. The variants with good functional scores were selected and further validated their FRET ratio changes by using microscope.





C

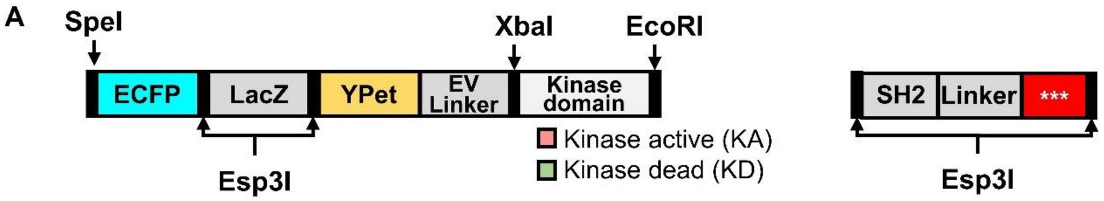
Kinase	Length (aa)	Position
KA1	262	265-526
KA2	266	261-526
KA3	277	261-537

D

Linker	Length (aa)	Amino acid sequence
L1	34	GSTSGSGKPGSGEGSTGAGGAGSTSGSGKPSGEG
L2	52	SAGGSAGGSAGGSAGGSAGGSGSAGGSAGGSTSAGGSAGGSAGGSAGGSAGG
L3	18	GGSGGTGGSGGTGGSGGT
L4	30	GGSGGTGGSGGTGGSGGTGGSGGTGGSGGT
L5	116	SAGGSAGGSAGGSAGGSAGGSGSAGGSAGGSTSAGGSAGGSAGGSAGGSAGGSGSAGGSAGGSTSAGGSAGGSAGGSAGGSAGGSGSAGGSAGGSTSAGGSAGGSAGGSAGGSAGG

Figure 4.2: Characterization of the DNA template for optimizing the Fyn FRET biosensor

(A) The construct of biosensor template designed for library screening. A FRET biosensor consists of ECFP, SH2, linker#1 which is fixed, substrate peptide (EKIEGTYGVV for the wild-type Fyn biosensor), YPet, linker#2 which is varied, and kinase domain which is varied. KA and KD represent the kinase active dead, respectively (B) The western blot results showing the expression of biosensor (using anti-GFP) and phosphotyrosine on the substrate peptide (using anti-pY). Top: Different kinase domain are used e.g. KA1, KA2, KA3 of which information are found in (C). PL denotes for plain HEK293T cells without transfection, and KD denotes for the biosensor with the KD. Bottom: Different linker#2 are varied e.g. L1, L2, L3, L4, and L5. The information of linker#2 are illustrated in (D).



B

Name	Nucleotide sequence	Amino acid sequence
Lib1	GAGAAGATCNNKNNKNNK**TAC**GGCGTCGTG	EKIXXX**Y**GVV
Lib2	GAGAAGATCGAGGGCACG**TAC**NNKNNKNNK	EKIEGT**Y**XXX

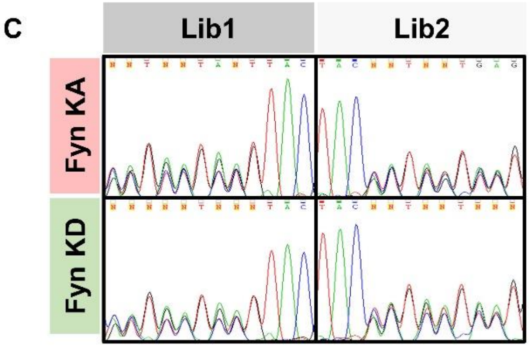


Figure 4.3: Site-saturation mutagenesis of substrate peptides

(A) Schematic representation of the domain structures of a FRET biosensor template (ELYK template). Left: An ELYK consists of ECFP, LacZ, YPet, EV linker, and kinase domain e.g. kinase active (KA) or kinase dead (KD). Right: An insert consists of SH2, linker, and the variants of the substrate peptide. The insert can replace the LacZ domain in ELYK template through the Golden Gate assembly. (B) Mutagenesis of substrate peptide for Lib1 and Lib2. The insert in (A, Right) is randomly mutated by NNK degenerated primers at the C-terminus of the insert. (C) The DNA sequencing chromatogram by Sanger sequencing confirming random mutation on the substrates of both Lib1 and Lib2.

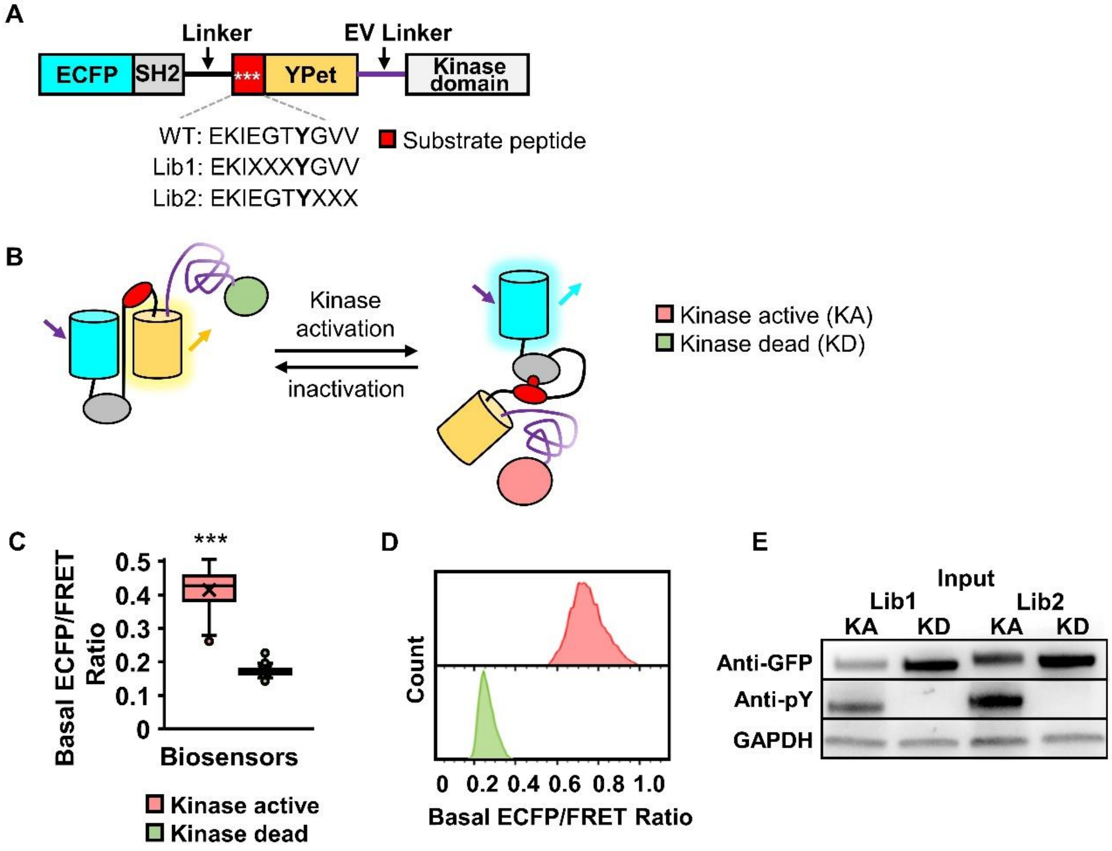


Figure 4.4: Library generation of the Fyn FRET biosensor

(A) Schematic representation of the domain structures of a FRET biosensor. A FRET biosensor consists of ECFP, SH2, linker, substrate peptide (EKIEGTYGVV for WT, EKIXXXYGVV for Lib1, and EKIEGTYXXX for Lib2, where X can be any amino acid), YPet, EV linker, and kinase domain e.g. kinase active (KA) or kinase dead (KD). (B) The cartoon illustrates the FRET effect of the Fyn reporter in the presence of KD (represented in green) or KA (represented in red). The KA phosphorylated the WT substrate leading to the intramolecular association between phosphotyrosine and SH2. This brings YPet away from ECFP, causing no or low FRET. While high FRET signal is observed when the biosensor is attached with the KD. (C) The basal ECFP/FRET ratio of biosensors attached with KA or KD measured by microscope. (D) The basal ECFP/FRET ratio of biosensors attached with KA or KD measured by flow cytometry. (E) The western blot results showing the expression of biosensor (using anti-GFP), phosphotyrosine on the substrate peptide variants (using anti-pY), and GAPDH. Biosensors which are attached with either KA or KD and contain variants of substrate peptides of either Lib1 or Lib2 showed different phosphorylation level in HEK293T cells.

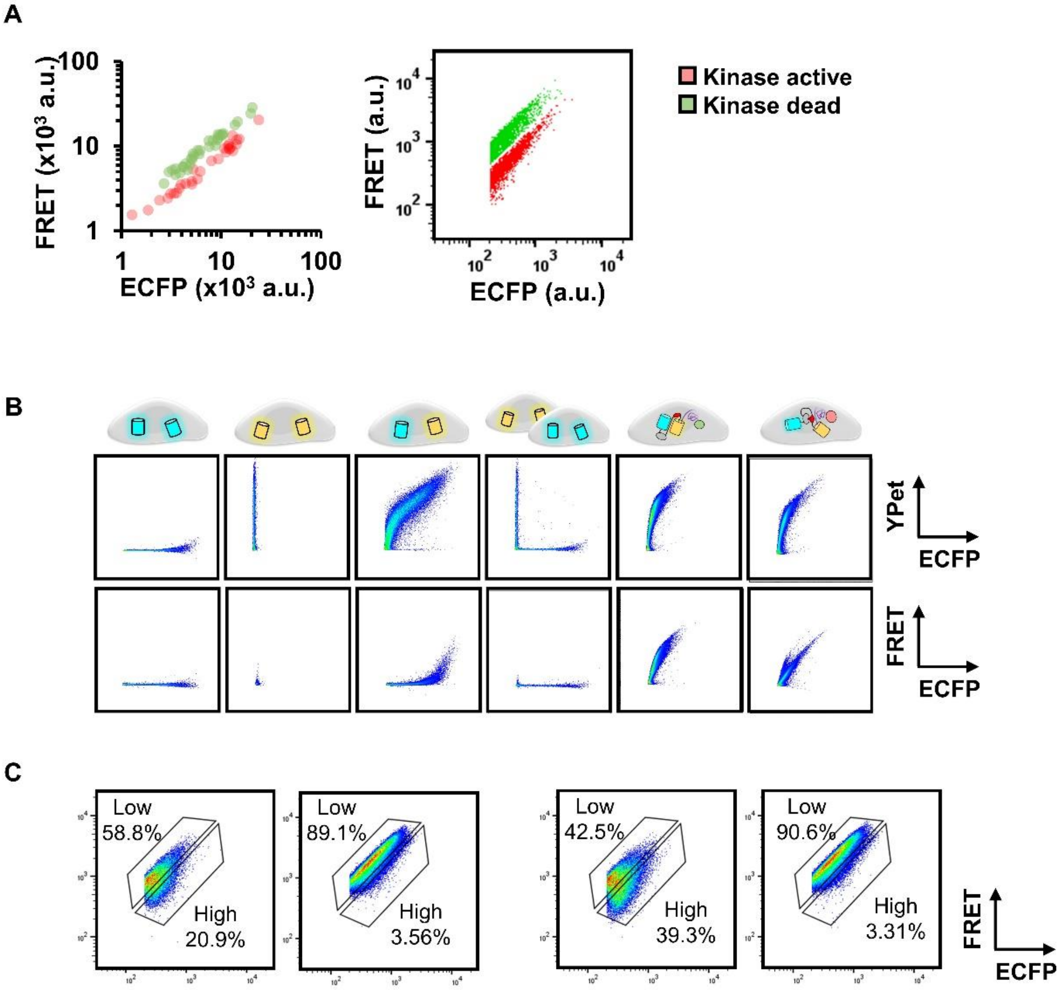


Figure 4.5: The counter soring

(A) The scatter plots showing the relationship between FRET and ECFP. The data were measured by microscope (Left) and flow cytometry (Right). Colors represent biosensors with different kinase domains e.g. KA (red) or KD (green). (B) Controls for counter sorting. HEK293T cells were transfected with only ECFP, only YPet, both ECFP and YPet, mixture of only ECFP and only Ypet, biosensor with KD, and biosensor with KA. The fluorescent signals were measured from flow cytometry and plotted as the relation of ECFP vs YPet, and ECFP vs FRET. (C) The scatter plot showing relationship between ECFP and FRET of each library. The irregular squares represent the gates for low or high FRET ratio which are created from controls in (B). The % represent the fraction of cells whose signals were in these gates.

Figure 4.6: Library screening of Fyn FRET biosensor

(A) Schematic presentation of FACS-based ECFP/FRET ratio selection. The individual mammalian cell containing biosensor variants is sorted based on their ECFP/FRET ratio into high and low ratio. (B) The basal ECFP/FRET ratio of WT and biosensor libraries fused with KA or KD using flow cytometry. The light-blue region, which is defined from FRET ratio = 0 to the mode of WT with KD, is used to select variants with low FRET ratio. The light-yellow region, which is defined from FRET ratio starting at the mode of WT with KA to maximum ratio, is used to select variants with high FRET ratio. (C) The counter screening strategy. Four input libraries (KA-Lib1, KD-Lib1, KA-Lib2, and KD-Lib2) are sorted based on FRET ratio into high and low FRET. (D) The western blot results showing the expression of biosensor (using anti-GFP), phosphotyrosine on the substrate peptide variants (using anti-pY), and GAPDH after the first round of sorting (Sort 1). The data represent sorted variants which are attached with KA from Lib1 or Lib2. (E) The imaging results of variants from Sort 1 before and after adding PP1. (F) The total number of cells from (E) that showed FRET ratio change (black) out of the imaged cells (grey).

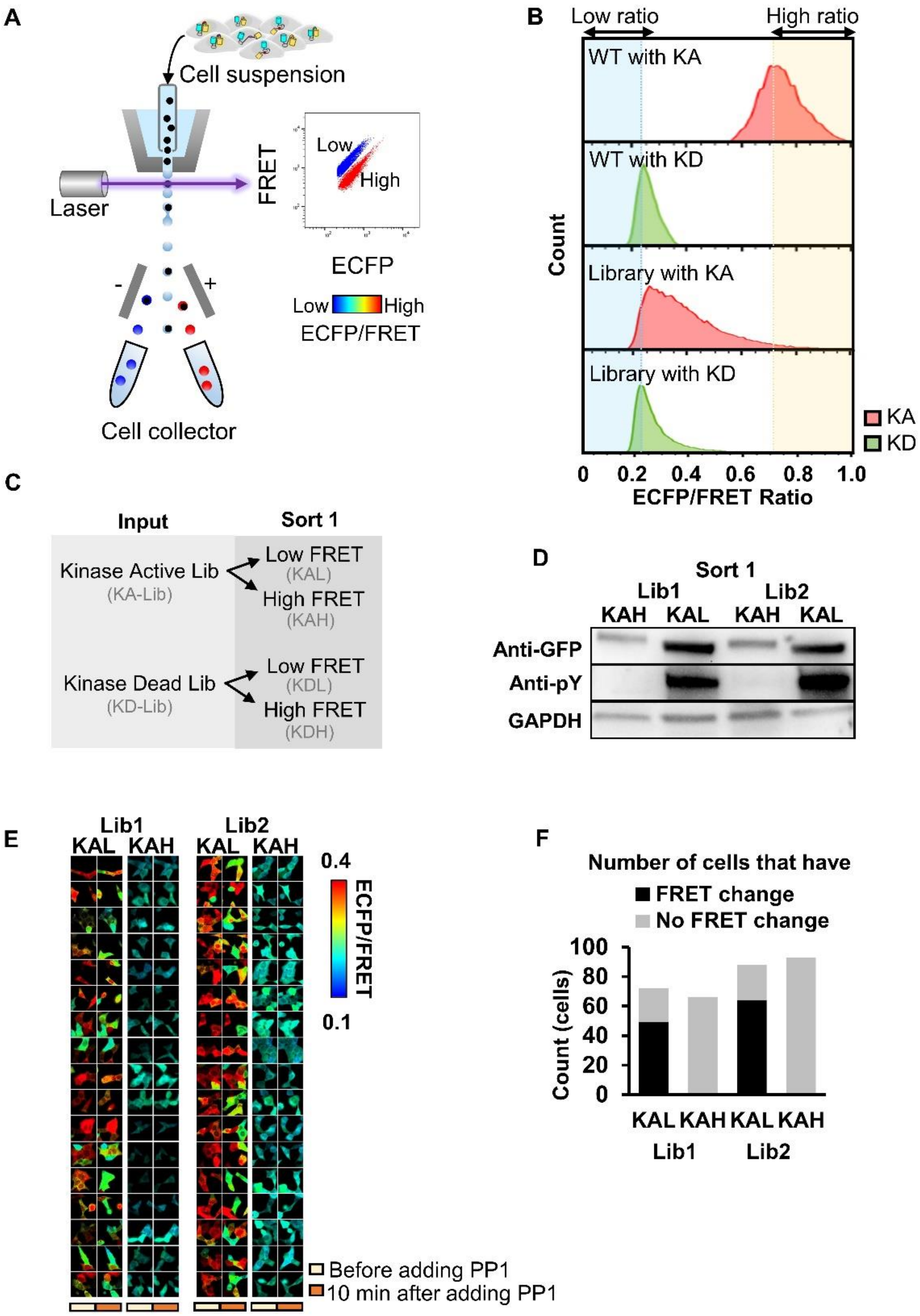

A
Cell suspension
Laser
FRET
Low
High
ECFP
Low
High
ECFP/FRET
-
+
Cell collector
B
Low ratio
High ratio
WT with KA
WT with KD
Library with KA
Library with KD
Count
KA
KD
0
0.2
0.4
0.6
0.8
1.0
ECFP/FRET Ratio
C
Input
Sort 1
Kinase Active Lib
(KA-Lib)
Low FRET
(KAL)
High FRET
(KAH)
Kinase Dead Lib
(KD-Lib)
Low FRET
(KDL)
High FRET
(KDH)
D
Sort 1
Lib1
Lib2
KAH
KAL
KAH
KAL
Anti-GFP
Anti-pY
GAPDH
E
Lib1
Lib2
KAL
KAH
KAL
KAH
0.4
ECFP/FRET
0.1
Before adding PP1
10 min after adding PP1
F
Number of cells that have
FRET change
No FRET change
100
80
60
40
20
0
Count (cells)
KAL
KAH
KAL
KAH
Lib1
Lib2

Figure 4.7: The sequence-function analysis of Fyn biosensor libraries

(A) The flow chart describing data manipulation. The sequencing data in FASTQ format from HiSeq400 sequencer are analyzed and their frequencies are calculated to obtain the functional scores. (B) Schematic presentation of the method that is used for selecting functional variants. The dots, which appear on the boundary of the KDL/KD vs KAH/KA scatter plot, are selected because these dots represent variants with the optimum tradeoff of enrichment ratio within a positive selection (e.g. KAH and KDL). Then these variants are further screened by filters that take the enrichment ratio of all sorted groups into account. (C) The scatter plot between the enrichment ratio of KDL and KAH from each variant in Sort 1 comparing with input library. The selected variants of substrate peptides were cloned to biosensor attached with KA and monitored their FRET ratio change upon PP1 stimulation by using microscope. Gray dots represent the enrichment ratio of each substrate from sequencing results and the dots labeled with different color represent substrates that have been selected and tested FRET ratio change by microscope. The >50%, 40%-50%, 30-40%, and <30% of FRET ratio changes before and after adding PP1 are illustrated in red, yellow, green, and blue dots, respectively. (D) The time-lapse imaging of the optimum substrate variants from both libraries (L1-19 is from Lib1, and L2-4 is from Lib2) compared with the WT. The solid line indicates the biosensors attached with KA, and the broken line indicates the biosensors attached with KD. (E) The sequence-function heat map of selected variants for a 7-residue of a substrate peptide. Positions are indicated numerically (*x*-axis), and each mutation is indicated by its single-letter amino acid abbreviation (*y*-axis) and the properties of each amino acid side chain are labeled in different color. The color of each element in the heat map illustrates the score, which is calculated from the $\log_2$ of the FRET ratio change of that variant normalized with that of WT ($\log_2$(FRET ratio changes of that mutant/WT)).

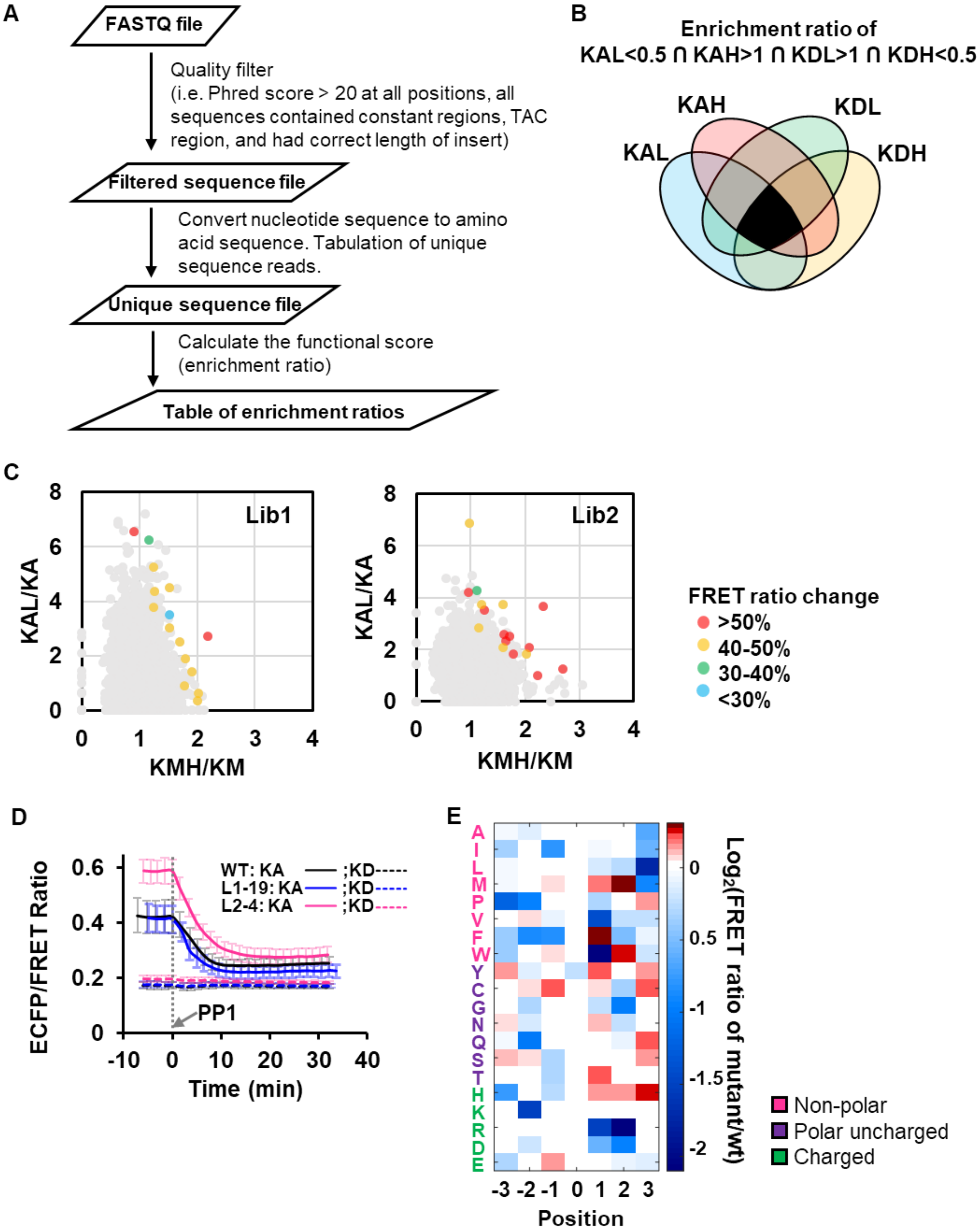
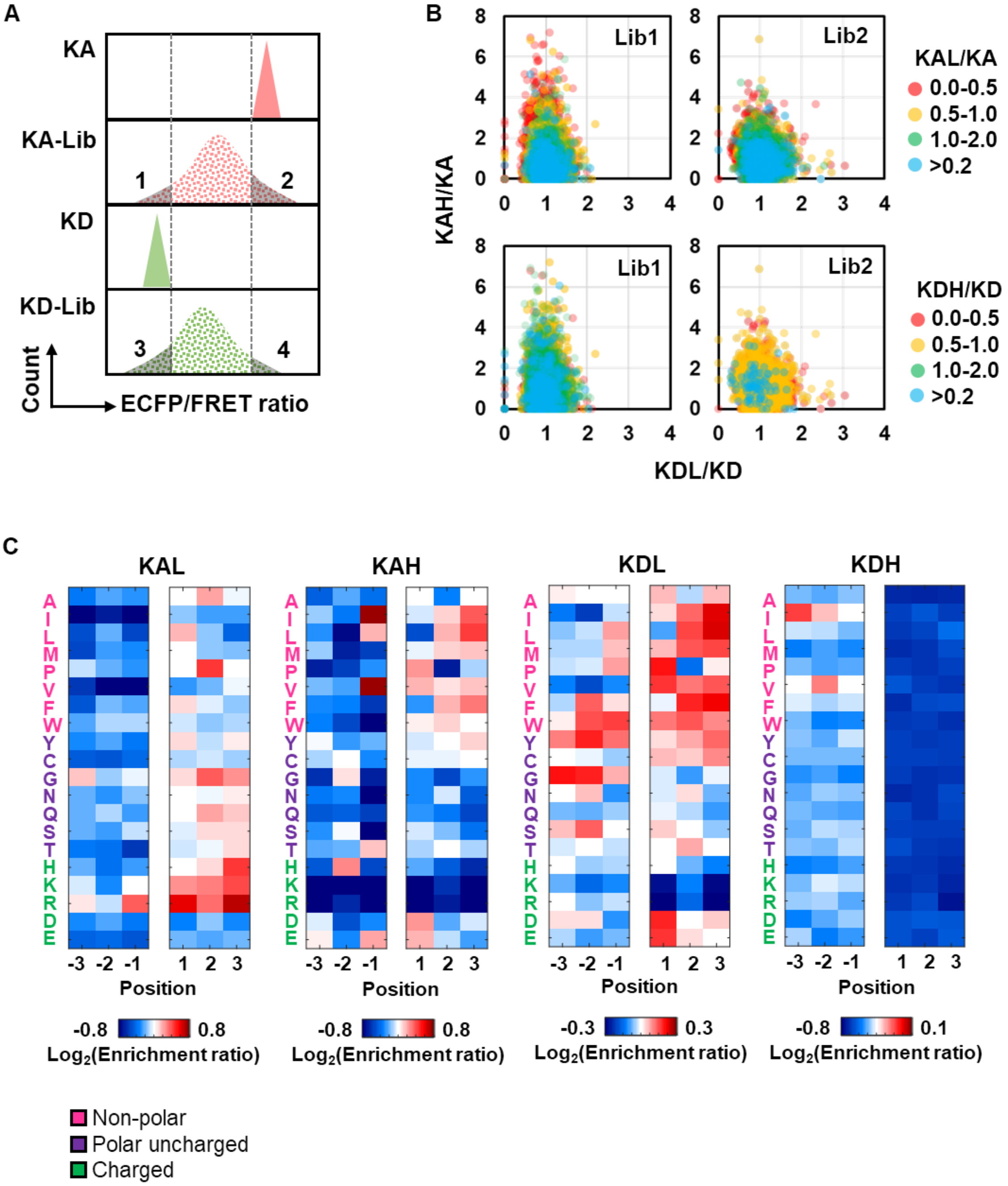
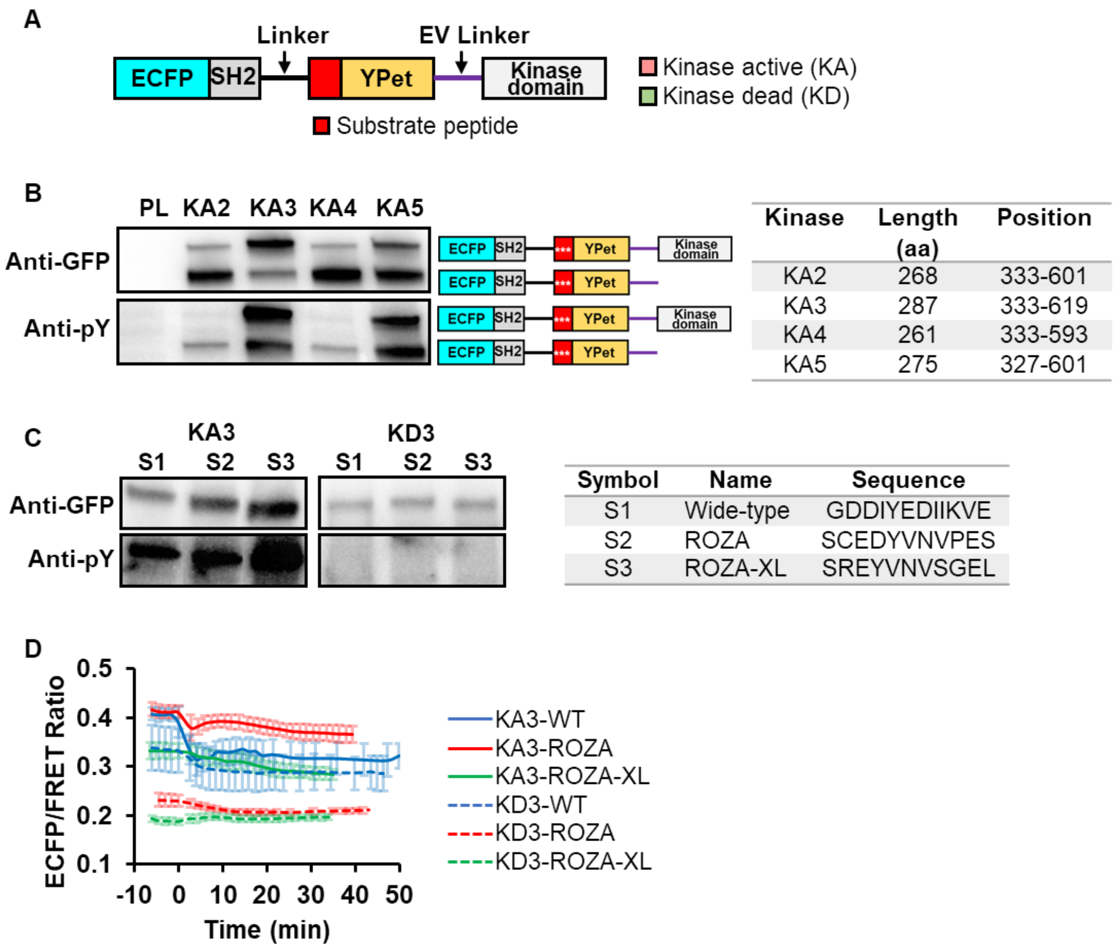

A
FASTQ file
Quality filter
(i.e. Phred score > 20 at all positions, all sequences contained constant regions, TAC region, and had correct length of insert)
Filtered sequence file
Convert nucleotide sequence to amino acid sequence. Tabulation of unique sequence reads.
Unique sequence file
Calculate the functional score (enrichment ratio)
Table of enrichment ratios
B
Enrichment ratio of
KAL<0.5 ∩ KAH>1 ∩ KDL>1 ∩ KDH<0.5
KAH
KDL
KAL
KDH
C
Lib1
Lib2
KAL/KA
KMH/KM
0
1
2
3
4
6
8
FRET ratio change
>50%
40-50%
30-40%
<30%
D
ECFP/FRET Ratio
0.6
0.4
0.2
0
-10
10
20
30
40
Time (min)
WT: KA ;KD
L1-19: KA ;KD
L2-4: KA ;KD
PP1
E
A
I
L
M
P
V
F
W
Y
C
G
N
Q
S
T
H
K
R
D
E
-3 -2 -1 0 1 2 3
Position
0
0.5
-1
-1.5
-2
Log2(FRET ratio of mutant/wt)
Non-polar
Polar uncharged
Charged

Figure 4.8: Calculation of functional scores for each variant

(A) Schematic presentation of counter sorting strategy. The KA and KD represent the WT biosensor attached with KA and KD, respectively. The KA-Lib and KD-Lib represent the biosensor variants attached with KA and KD, respectively. The broken lines are used to confine the regions of low and high FRET ratio which are determined from KD and KA, respectively. Variants from KA-Lib are sorted into two groups e.g. low (region 1) and high (region 2) FRET ratio. Similarly, variants from KD-Lib are sorted into two groups e.g. low (region 3) and high (region 4) FRET ratio. Following results from Figures 4.4D and 4.6B, the functional variants should appear in region 2 and 3, while the non-functional ones tend to be highly enriched in region 1 and 4. (B) The scatter plot of the enrichment ratio of the positive selection group (e.g. KDL and KAH) where the color of each dot represents the enrichment ration of the negative selection group (e.g. KDH and KAL). (C) The sequence-function map of all variants from each selected library (Sort 1). Positions are indicated numerically (*x*-axis), and each mutation is indicated by its single-letter amino acid abbreviation (*y*-axis) and the properties of each amino acid side chain are labeled in different color. The color of each element of the heat map illustrates the enrichment score, which is calculated from the $\log_2$ of the frequency of that variant from Sort 1 divided by that from input library.


A
KA
KA-Lib
1
2
KD
KD-Lib
3
4
Count
ECFP/FRET ratio
B
KAH/KA
KDL/KD
Lib1
Lib2
KAL/KA
0.0-0.5
0.5-1.0
1.0-2.0
>0.2
KDH/KD
0.0-0.5
0.5-1.0
1.0-2.0
>0.2
C
KAL
KAH
KDL
KDH
A I L M P V F W Y C G N Q S T H K R D E
-3 -2 -1 1 2 3
Position
-0.8 0.8
Log₂(Enrichment ratio)
-0.3 0.3
-0.8 0.1
Non-polar
Polar uncharged
Charged



Kinase	Length (aa)	Position
KA2	268	333-601
KA3	287	333-619
KA4	261	333-593
KA5	275	327-601

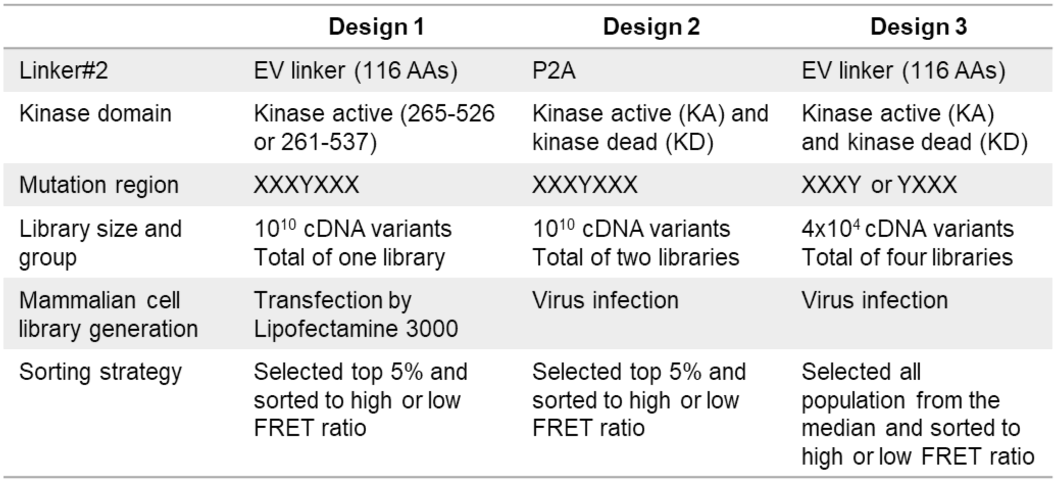
Symbol	Name	Sequence
S1	Wide-type	GDDIYEDIIKVE
S2	ROZA	SCEDYVNVPES
S3	ROZA-XL	SREYVNVSGEL

Figure 4.9: Characterization of the DNA template for optimizing the ZAP70 FRET biosensor

(A) The construct of biosensor template designed for library screening. Similar to Fyn FRET biosensor, a ZAP70 FRET biosensor consists of ECFP, SH2, flexible linker, substrate peptide, which is varied, YPet, EV linker, and kinase domain which is varied. KA and KD represent the kinase active dead, respectively. (B) The western blot results showing the activity of biosensor with different kinase domains. The expression of biosensor is detected using anti-GFP antibody and phosphotyrosine on the substrate peptide is detected using anti-pY antibody. Left: Different kinase domain are used e.g. KA2, KA3, KA4, and KA5 of which information are found in (B, Right). PL denotes for plain HEK293T cells without transfection. (C) The western blot results showing the activity of biosensor with different substrate peptides. Left: Different substrate peptides are used e.g. S1, S2, and S3 of which information are found in (C, Right). KD denotes for the biosensor with the KD. (D) The time-lapse imaging of different substrate peptides with fixed kinase#3 (KA3). The solid line indicates the biosensors attached with KA, and the broken line indicates the biosensors attached with KD.

4.6 Tables

Table 4.1: The detailed description of each library design



	Design 1	**Design 2**	**Design 3**
Linker#2	EV linker (116 AAs)	P2A	EV linker (116 AAs)
Kinase domain	Kinase active (265-526 or 261-537)	Kinase active (KA) and kinase dead (KD)	Kinase active (KA) and kinase dead (KD)
Mutation region	XXXYXXX	XXXYXXX	XXXY or YXXX
Library size and group	10^{10} cDNA variants Total of one library	10^{10} cDNA variants Total of two libraries	$4x10^{4}$ cDNA variants Total of four libraries
Mammalian cell library generation	Transfection by Lipofectamine 3000	Virus infection	Virus infection
Sorting strategy	Selected top 5% and sorted to high or low FRET ratio	Selected top 5% and sorted to high or low FRET ratio	Selected all population from the median and sorted to high or low FRET ratio

Acknowledgements

Chapter 4, in full, is currently being prepared for submission for publication. P. Limsakul, L. Lui, J. Huang, F. He, J. Li, R. E.S. Harrison, S. Lu, Y. Wang. "Optimization of the whole FRET biosensors for kinase activity through directed evolution". The dissertation author was the first author of this paper.

This work is supported by grants from NIH HL098472, HL109142, HL121365 (Y. Wang), NSF CBET1360341, DMS1361421 (Y. Wang and S.L.). P.L. acknowledges fellowship support from the Thai Ministry of Science and Technology. The funding agencies had no role in preparation of the manuscript.

References

1. Davila, M.L., I. Riviere, X. Wang, S. Bartido, J. Park, K. Curran, S.S. Chung, J. Stefanski, O. Borquez-Ojeda, M. Olszewska, J. Qu, T. Wasielewska, Q. He, M. Fink, H. Shinglot, M. Youssif, M. Satter, Y. Wang, J. Hosey, H. Quintanilla, E. Halton, Y. Bernal, D.C. Bouhassira, M.E. Arcila, M. Gonen, G.J. Roboz, P. Maslak, D. Douer, M.G. Frattini, S. Giralt, M. Sadelain, and R. Brentjens, *Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia.* Sci Transl Med, 2014. **6**(224): p. 224ra25.

2. Grupp, S.A., M. Kalos, D. Barrett, R. Aplenc, D.L. Porter, S.R. Rheingold, D.T. Teachey, A. Chew, B. Hauck, J.F. Wright, M.C. Milone, B.L. Levine, and C.H. June, *Chimeric antigen receptor-modified T cells for acute lymphoid leukemia.* N Engl J Med, 2013. **368**(16): p. 1509-18.

3. Walunas, T.L., D.J. Lenschow, C.Y. Bakker, P.S. Linsley, G.J. Freeman, J.M. Green, C.B. Thompson, and J.A. Bluestone, *CTLA-4 can function as a negative regulator of T cell activation.* Immunity, 1994. **1**(5): p. 405-413.

4. Salmond, R.J., A. Filby, I. Qureshi, S. Caserta, and R. Zamoyska, *T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance.* Immunological reviews, 2009. **228**(1): p. 9-22.

5. Smith-Garvin, J.E., G.A. Koretzky, and M.S. Jordan, *T cell activation.* Annual review of immunology, 2009. **27**: p. 591-619.

6. Zhang, J., R.E. Campbell, A.Y. Ting, and R.Y. Tsien, *Creating new fluorescent probes for cell biology.* Nature reviews Molecular cell biology, 2002. **3**(12): p. 906.

7. Komatsu, N., K. Aoki, M. Yamada, H. Yukinaga, Y. Fujita, Y. Kamioka, and M. Matsuda, *Development of an optimized backbone of FRET biosensors for kinases and GTPases.* Molecular biology of the cell, 2011. **22**(23): p. 4647-4656.

8. Wang, Y., E.L. Botvinick, Y. Zhao, M.W. Berns, S. Usami, R.Y. Tsien, and S. Chien, *Visualizing the mechanical activation of Src.* Nature, 2005. **434**(7036): p. 1040.

9. Randriamampita, C., P. Mouchacca, B. Malissen, D. Marguet, A. Trautmann, and A.C. Lellouch, *A novel ZAP-70 dependent FRET based biosensor reveals kinase activity at both the immunological synapse and the antisynapse.* PLoS One, 2008. **3**(1): p. e1521.

10. Seong, J., A. Tajik, J. Sun, J.L. Guan, M.J. Humphries, S.E. Craig, A. Shekaran, A.J. Garcia, S. Lu, M.Z. Lin, N. Wang, and Y. Wang, *Distinct biophysical mechanisms of focal adhesion kinase mechanoactivation by different extracellular matrix proteins.* Proc Natl Acad Sci U S A, 2013. **110**(48): p. 19372-7.

11. Paster, W., C. Paar, P. Eckerstorfer, A. Jakober, K. Drbal, G.J. Schutz, A. Sonnleitner, and H. Stockinger, *Genetically encoded Forster resonance energy transfer sensors for the conformation of the Src family kinase Lck.* J Immunol, 2009. **182**(4): p. 2160-7.

12. Miyawaki, A., *Visualization of the spatial and temporal dynamics of intracellular signaling.* Developmental cell, 2003. **4**(3): p. 295-305.

13. Lindenburg, L. and M. Merkx, *Engineering genetically encoded FRET sensors.* Sensors, 2014. **14**(7): p. 11691-11713.

14. Ouyang, M., R. Wan, Q. Qin, Q. Peng, P. Wang, J. Wu, M. Allen, Y. Shi, S. Laub, and L. Deng, *A sensitive FRET biosensor reveals Fyn kinase regulation by sub-membrane localization.* ACS sensors, 2018.

15. Wang, P., J. Liang, L.Z. Shi, Y. Wang, P. Zhang, M. Ouyang, D. Preece, Q. Peng, L. Shao, and J. Fan, *Visualizing spatiotemporal dynamics of intercellular mechanotransmission upon wounding.* ACS Photonics, 2018. **5**(9): p. 3565-3574.

16. Fritz, R.D., M. Letzelter, A. Reimann, K. Martin, L. Fusco, L. Ritsma, B. Ponsioen, E. Fluri, S. Schulte-Merker, and J. van Rheenen, *A versatile toolkit to produce sensitive FRET biosensors to visualize signaling in time and space.* Sci. Signal., 2013. **6**(285): p. rs12-rs12.

17. Piljić, A., I. De Diego, M. Wilmanns, and C. Schultz, *Rapid development of genetically encoded FRET reporters.* ACS chemical biology, 2011. **6**(7): p. 685-691.

18. Ibraheem, A., H. Yap, Y. Ding, and R.E. Campbell, *A bacteria colony-based screen for optimal linker combinations in genetically encoded biosensors.* BMC biotechnology, 2011. **11**(1): p. 105.

19. Thestrup, T., J. Litzlbauer, I. Bartholomäus, M. Mues, L. Russo, H. Dana, Y. Kovalchuk, Y. Liang, G. Kalamakis, and Y. Laukat, *Optimized ratiometric calcium sensors for functional in vivo imaging of neurons and T lymphocytes.* Nature methods, 2014. **11**(2): p. 175.

20. Belal, A.S.F., B.R. Sell, H. Hoi, M.W. Davidson, and R.E. Campbell, *Optimization of a genetically encoded biosensor for cyclin B1-cyclin dependent kinase 1.* Molecular BioSystems, 2014. **10**(2): p. 191-195.

21. Zheng, L., U. Baumann, and J.-L. Reymond, *An efficient one-step site-directed and site-saturation mutagenesis protocol.* Nucleic acids research, 2004. **32**(14): p. e115-e115.

22. Lu, S., T.-j. Kim, C.-E. Chen, M. Ouyang, J. Seong, X. Liao, and Y. Wang, *Computational analysis of the spatiotemporal coordination of polarized PI3K and Rac1 activities in micro-patterned live cells.* PloS one, 2011. **6**(6): p. e21293-e21293.

23. Arnold, F.H., *Design by directed evolution.* Accounts of chemical research, 1998. **31**(3): p. 125-131.

24. Romero, P.A. and F.H. Arnold, *Exploring protein fitness landscapes by directed evolution.* Nature reviews Molecular cell biology, 2009. **10**(12): p. 866.

25. Nair, S.A., M.H. Kim, S.D. Warren, S. Choi, Z. Songyang, L.C. Cantley, and D.G. Hangauer, *Identification of efficient pentapeptide substrates for the tyrosine kinase pp60c-src.* J Med Chem, 1995. **38**(21): p. 4276-83.

26. Songyang, Z. and L.C. Cantley, *SH2 domain specificity determination using oriented phosphopeptide library*. Methods Enzymol, 1995. **254**: p. 523-35.

27. Banning, C., J. Votteler, D. Hoffmann, H. Koppensteiner, M. Warmer, R. Reimer, F. Kirchhoff, U. Schubert, J. Hauber, and M. Schindler, *A flow cytometry-based FRET assay to identify and analyse protein-protein interactions in living cells.* PloS one, 2010. **5**(2): p. e9344.

28. Fowler, D.M. and S. Fields, *Deep mutational scanning: a new style of protein science.* Nature methods, 2014. **11**(8): p. 801.

29. Fowler, D.M., C.L. Araya, S.J. Fleishman, E.H. Kellogg, J.J. Stephany, D. Baker, and S. Fields, *High-resolution mapping of protein sequence-function relationships.* Nature methods, 2010. **7**(9): p. 741.

30. Cadra, S., A. Gucciardi, M.-P. Valignat, O. Theodoly, A. Vacaflores, J.C. Houtman, and A.C. Lellouch, *ROZA-XL, an improved FRET based biosensor with an increased dynamic range for visualizing zeta associated protein 70 kD (ZAP-70) tyrosine kinase activity in live T cells.* Biochemical and biophysical research communications, 2015. **459**(3): p. 405-410.

31. Li, K., X. Xiang, J. Sun, H.-T. He, J. Wu, Y. Wang, and C. Zhu, *Imaging spatiotemporal activities of ZAP-70 in live T cells using a FRET-based biosensor.* Annals of biomedical engineering, 2016. **44**(12): p. 3510-3521.

32. Purroy, N., P. Abrisqueta, J. Carabia, E. Calpe, C. Carpio, C. Palacio, M. Crespo, and F. Bosch, *Syk Inhibition By TAK-659 Overcomes Proliferative, Survival and Migratory Signals from the Microenvironment in Chronic Lymphocytic Leukemia Cells*. 2014, Am Soc Hematology.

33. Zhou, S., S.E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W.G. Haser, F. King, T. Roberts, S. Ratnofsky, and R.J. Lechleider, *SH2 domains recognize specific phosphopeptide sequences.* Cell, 1993. **72**(5): p. 767-778.

34. Songyang, Z., K.L. Carraway III, M.J. Eck, S.C. Harrison, R.A. Feldman, M. Mohammadi, J. Schlessinger, S.R. Hubbard, D.P. Smith, and C. Eng, *Catalytic specificity of protein-tyrosine kinases is critical for selective signalling.* Nature, 1995. **373**(6514): p. 536.

35. Mason, D.M., C.R. Weber, C. Parola, S.M. Meng, V. Greiff, W.J. Kelton, and S.T. Reddy, *High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis.* Nucleic acids research, 2018.

36. Currin, A., N. Swainston, P.J. Day, and D.B. Kell, *Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently.* Chemical Society Reviews, 2015. **44**(5): p. 1172-1239.

37. Klasse, P., *Molecular determinants of the ratio of inert to infectious virus particles*, in *Progress in molecular biology and translational science*. 2015, Elsevier. p. 285-326.

38. Keefe, A.D. and J.W. Szostak, *Functional proteins from a random-sequence library*. Nature, 2001. **410**(6829): p. 715.

39. Taverna, D.M. and R.A. Goldstein, *Why are proteins so robust to site mutations? 1.* Journal of molecular biology, 2002. **315**(3): p. 479-484.

Chapter 5

Conclusion

5.1 Summary of Dissertation

The main goal of this dissertation was to utilize the directed evolution technology to improve the sensors for detecting molecular activities in living cells. The story of this dissertation can be separated into two parts. First, I employed directed evolution following a conventional approach through yeast surface display library to engineer the molecular module to have a special function. Subsequently, I applied this engineered molecule as a part of FRET biosensor for cancer diagnostic application (Chapter 2) as well as a universal receptor of CARs for cancer therapeutic application (Chapter 3). Second, I developed a universal and systematic approach to optimize FRET biosensors (Chapter 4). These optimized biosensors will be used to study the signaling cascade of protein kinases that are involved in T cell activation. Although all chapters were united by directed evolution, the experimental concepts were different depending on the scientific problems. As a result, the summary of each chapter was provided below.

Monitoring enzymatic activities at the cell surface is challenging due to the poor efficiency of transport and membrane integration of FRET biosensors. Thus, in Chapter 2, I developed a hybrid biosensor with separate donor and acceptor that assemble *in situ*. The directed evolution and sequence-function analysis were integrated to engineer a monobody variant (PEbody) that binds to R-PE dye. PEbody was used for visualizing the dynamic formation/separation of intercellular junctions. I further fused PEbody with ECFP and an enzyme-specific peptide at the extracellular surface to create a hybrid FRET biosensor upon R-PE capture for monitoring MT1-MMP activities. This biosensor revealed asymmetric distribution of MT1-MMP activities, which were high and low at loose and stable cell-cell contacts, respectively.

T cells expressing CARs are promising cancer therapeutic agents. However, due to evolution of tumor cells, depletion of a certain antigen causes CAR T cell dysfunction and relapse in patients treated with CAR T cells targeting a single antigen. To avoid antigen loss and other toxicities, in Chapter 3, I developed a simple and universal receptor for CAR that increases the ability to switch targets without re-engineering the T cells. The design of this CAR is a two-component receptor system: (1) a PE-targeting receptor (PEbody presented in Chapter 2) expressed on T cells; (2) a tumor-targeting PE-conjugated antibody which acts as an adaptor molecule. This feature is useful to reduce relapse, tune activation, and enhance specificity. I tested the PEbody CAR system *in vitro* to demonstrate its broad utility. Multiple PE-conjugated antibodies were used as adaptors to target various antigens on tumor cell surfaces. With only one construct, PEbody CAR T cells showed killing efficacy in both hematologic and solid tumor cells.

Understanding the spatiotemporal patterns of kinase activities underlying T cell activation is crucial for our capability to apply cell-based engineering for therapy. The currently available biosensors to study this mechanism are still limited in their sensitivity and specificity. To overcome the limitation, in Chapter 4, I developed a systematic technology platform for optimizing FRET biosensors. I combined the directed evolution technology with function-based screening in mammalian cells, high-throughput DNA sequencing, and sequence-function analysis to improve the performance of FRET biosensors. The results showed that the optimized Fyn FRET biosensors perform better than the wide-type biosensor in both sensitivity and specificity.

5.2 Future work

Directed evolution is a powerful optimization and engineering strategy for biosensor development, but success requires a good starting material that leads to the desired function and a good screening or selection strategy to identify it. In the past few years, I spent most of the time optimizing directed evolution protocol to engineer and characterize the molecular modules with special functions. Thus, the step afterward is to apply these tools for biological applications.

In Chapter 2, I introduced a new design for FRET biosensors which combines genetically encoded fluorescent protein and PEbody as the binding scaffold for a cell-impermeable fluorescent dye, R-PE, which provides high spatial resolution on the cell surface. In addition to the applications that I mentioned in the chapter, this hybrid FRET biosensor, could be used to detect the invasiveness of cancer cells from clinical samples, and to screen for anti-cancer drugs. Moreover, it could be used to monitor other molecular activities at the extracellular surface as well as the endocytosis of surface molecules, due to the low-pH stability of both R-PE and ECFP. As such, the method developed in this chapter can be employed to systematically design hybrid FRET biosensors capable of monitoring important molecular dynamics at the cell surface and other subcellular regions of living cells.

In Chapter 3, I demonstrated another application of the PEbody as a universal receptor for CAR T cell therapy. I have tested the killing efficacy of PEbody CAR T cells using several PE-conjugated antibodies to target different antigens on either hematologic or solid tumor cells. The results were promising *in vitro*. The next step is to apply this system *in vivo*. Thus, many optimizations will be needed. Furthermore, due to the size of PEbody is small (10 kDa) and easy to manipulate, the PEbody CAR can be integrated with other systems such as the Boolean logic gates, including AND gate, OR gate, and AND-NOT gate, to improve selectivity of CAR T cells toward tumor cells.

In Chapter 4, I developed the systematic approach to optimize the whole FRET biosensors. For example, the optimized Fyn and ZAP70 FRET biosensors can be applied to provide feedback readings of these kinase activities for the calibration and optimization of the CAR T cell functionality when engaging with target cancer cells. The understanding of these kinases could be crucial guidance for the design of therapeutic approaches. Besides, this systematic approach could allow for the direct development of sensitive and specific FRET biosensors detecting a broad range of enzymatic activities in living cells. The screening template for optimizing the biosensor sensing unit is modularized and generally applicable to the development of FRET biosensors for, in principle, any post-translational modification such as kinases, methyltransferase, and acetyltransferase. This technology is expected to have transformative impact in biomedical applications, ranging from the in-depth understanding of spatiotemporal characteristics of post-translational regulations to the screening of regulators and drugs of crucial post-translational events directly in mammalian cells.