Synthetic Protein Switches

Methods and Protocols

Edited by

Viktor Stein

Fachbereich Biologie, Technische Universität Darmstadt, Darmstadt, Germany

Humana Press
Editor
Viktor Stein
Fachbereich Biologie
Technische Universität Darmstadt
Darmstadt, Germany
Preface

Synthetic protein switches with custom response functions have become invaluable tools in basic research and biotechnology for monitoring biomolecular analytes or actuating cellular functions in a rapid, specific, integrated, and autonomous fashion. This book provides a comprehensive summary of state-of-the-art protocols to facilitate the construction of synthetic protein switches for a variety of applications in biotechnology and basic research. Protocols are applicable to life scientists from diverse research fields that range from traditional, discovery-centered disciplines such as cancer research to newly emerging disciplines such as synthetic biology.

Chapters are grouped into separate sections focusing on different types of switches, sensors, and actuators. Starting with a general view, I first discuss the experimental challenges and theoretical considerations that underlie the construction of synthetic protein switches, also highlighting an increasing number of computational approaches which aim to render the design cycle more rational and therefore more efficient. In the second chapter, Ha and Loh provide an overview on the construction of synthetic protein switches by means of alternative frame folding and intermolecular fragment exchange which promises a generic route to convert any conventional binding receptor or enzyme into an allosterically regulated protein switch. This is followed up by a detailed protocol by Ribeiro, Ostermeier, et al. on the construction of synthetic protein switches by means of domain insertion describing the underlying non-homology-dependent DNA recombination process to build DNA libraries.

Subsequent chapters become increasingly specific, providing case studies on how to engineer synthetic protein switches for different types of applications. Starting with protocol chapters that describe the construction of fluorescent and bioluminescent sensors, Mitchell, Jackson, et al. and Clifton, Jackson, et al. demonstrate how computational strategies based on molecular modeling and statistical sequence analysis can be applied to engineer small molecule FRET sensors with enhanced biophysical properties. Farrants, Johnsson, et al. then describe a general route toward small molecule sensors based on semisynthetic fluorescent and bioluminescent sensors that are built with the SNAP-tag protein conjugation system. Finally, Nyati et al. and Matysuma, Ueda, et al. illustrate the construction of bioluminescent sensors based on proximity-dependent and allosterically regulated firefly luciferases.

Beyond fluorescent and bioluminescent sensors, three chapters by Iwai et al., Wouters et al., and Nirantar et al. focus on the construction of synthetic protein switches based on β-lactamase, which has served as a model enzyme for pioneering a number of design strategies, for instance, by means of domain insertion and competitive autoinhibition. This is followed up by two chapters that describe the construction of protease-based switches as Wintgens, Wehr, et al. and Stein and Alexandrov illustrate how viral proteases can be reengineered into synthetic protease sensors with custom input-output functions based on split- and competitively autoinhibited architectures.

The book concludes with chapters focusing on the construction of protein switches that can actuate biological signaling functions in live cells. To this end, Muehlhaeuser,
Radzwilli, et al.; Stabel, Moeglich, et al.; Cosentino, Moroni, et al.; and Taxis provide protocols on how to regulate protein kinase function, ion channel permeability, and protein degradation by means of light-regulated protein switches. This is followed up with protocol chapters by Castillo, Ghosh, et al. and DiRoberto, Peisajovich, et al. who devise strategies for regulating cellular signal transduction systems through biologically inert ligands and rewiring key nodes of intracellular signaling systems.

Darmstadt, Germany

Viktor Stein
Contents

Preface .......................................................... v
Contributors ................................................................ ix

PART I GENERAL STRATEGIES AND CONSIDERATIONS

1 Synthetic Protein Switches: Theoretical and Experimental Considerations . . . 3
   Viktor Stein

2 Construction of Allosteric Protein Switches by Alternate Frame Folding and
   Intermolecular Fragment Exchange .................................................. 27
   Jeung-Hoi Ha and Stewart N. Loh

3 Construction of Protein Switches by Domain Insertion and Directed Evolution . . . 43
   Lucas F. Ribeiro, Tiana D. Warren, and Marc Ostermeier

PART II PEPTIDE SWITCHES

4 Catalytic Amyloid Fibrils That Bind Copper to Activate Oxygen .................. 59
   Alex Sternisha and Olga Makhlynets

PART III FLUORESCENT AND BIOLUMINESCENT SENSORS

5 Ancestral Protein Reconstruction and Circular Permutation for Improving the
   Stability and Dynamic Range of FRET Sensors .......................... 71
   Ben E. Clifton, Jason H. Whitfield, Inmaculada Sanchez-Romero,
   Michel K. Herde, Christian Henneberger, Harald Janovjak,
   and Colin J. Jackson

6 Method for Developing Optical Sensors Using a Synthetic Dye-Fluorescent
   Protein FRET Pair and Computational Modeling and Assessment .......... 89
   Joshua A. Mitchell, William H. Zhang, Michel K. Herde,
   Christian Henneberger, Harald Janovjak, Megan L. O’Mara,
   and Colin J. Jackson

7 Rational Design and Applications of Semisynthetic Modular
   Biosensors: SNIFITs and LUCIDs ............................................ 101
   Helen Farrants, Julien Hiblot, Rudolf Griss, and Kai Johnsson

8 Ultrasensitive Firefly Luminescent Intermediate-Based Protein-Protein
   Interaction Assay (FlimPIA) Based on the Functional Complementation
   of Mutant Firefly Luciferases ...................................................... 119
   Yuki Ohmuro-Matsuyama and Hiroshi Ueda

9 Quantitative and Dynamic Imaging of ATM Kinase Activity ...................... 131
   Shyam Nyati, Grant Young, Brian Dale Ross, and Alnawaz Rebentulla
Part IV  β-lactamase Sensors

10  Creation of Antigen-Dependent β-Lactamase Fusion Protein Tethered by Circularly Permuted Antibody Variable Domains .......................... 149
   Hiroto Iwai, Miki Kojima-Misaizu, Jinhua Dong, and Hiroshi Ueda

11  Protein and Protease Sensing by Allosteric Derepression ......................... 167
   Hui Chin Goh, Farid J. Ghadessy, and Saurabh Nirantar

12  DNA-Specific Biosensors Based on Intramolecular β-Lactamase-Inhibitor Complex Formation ............................................ 179
   Wouter Engelen and Maarten Merkx

Part V  Proteolytic Sensors

13  Engineering and Characterizing Synthetic Protease Sensors and Switches .... 197
   Viktor Stein and Kirill Alexandrov

14  Characterizing Dynamic Protein–Protein Interactions Using the Genetically Encoded Split Biosensor Assay Technique Split TEV ...................... 219
   Jan P. Wintgens, Moritz J. Rosner, and Michael C. Wehr

Part VI  Optogenetic Switches

15  Development of a Synthetic Switch to Control Protein Stability
    in Eukaryotic Cells with Light ........................................... 241
    Christof Taxis

16  Light-Regulated Protein Kinases Based on the CRY2-CIB1 System .......... 257
    Wignand W.D. Mühlhäuser, Maximilian Hörner, Wilfried Weber,
    and Gerald Radziwill

17  Yeast-Based Screening System for the Selection of Functional
    Light-Driven K⁺ Channels .................................................. 271
    Cristian Cosentino, Laura Alberio, Gerhard Thiel, and Anna Moroni

18  Primer-Aided Truncation for the Creation of Hybrid Proteins ................. 287
    Robert Stabel, Birthe Stüven, Robert Obelendorf, and Andreas Möglich

Part VII  Cellular Signaling Switches

19  Engineering Small Molecule Responsive Split Protein Kinases ............... 307
    Javier Castillo-Montoya and Indraneel Ghosh

20  Directed Evolution Methods to Rewire Signaling Networks .................... 321
    Raphaël B. Di Roberto, Benjamin M. Scott, and Sergio G. Peisajovich

Index ................................................................. 339
Contributors

Laura Alberio • Department of Biosciences, University of Milan and Biophysics Institute, National Research Council (CNR), Milan, Italy

Kirill Alexandrov • Institute for Molecular Biosciences, The University of Queensland, St. Lucia, QLD, Australia

Javier Castillo-Montoya • Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA

Hui Chin Goh • p53 Laboratory, A*STAR Agency for Science, Technology and Research, Singapore, Singapore

Ben E. Clifton • Research School of Chemistry, The Australian National University, Canberra, ACT, Australia

Cristian Cosentino • Department of Biosciences, University of Milan and Biophysics Institute, National Research Council (CNR), Milan, Italy

Jinhua Dong • Laboratory for Chemistry and Life Science, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan; College of Chemistry and Chemical Engineering, Linyi University, Shandong, China

Wouter Engelen • Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

Helen Farrants • National Centre of Competence in Research (NCCR) Chemical Biology, Institute of Chemical Sciences and Engineering (ISIC), Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Farid J. Ghadessy • p53 Laboratory, A*STAR Agency for Science, Technology and Research, Singapore, Singapore

Indraneel Ghosh • Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA

Rudolf Griss • National Centre of Competence in Research (NCCR) Chemical Biology, Institute of Chemical Sciences and Engineering (ISIC), Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Jeung-Hoi Ha • Department of Biochemistry and Molecular Biology, State University of New York Upstate Medical University, Syracuse, NY, USA

Christian Henneberger • Institute of Cellular Neurosciences, University of Bonn, Bonn, Germany; German Centre for Neurodegenerative Diseases, Bonn, Germany; University College of London, London, UK

Michel K. Herde • Institute of Cellular Neurosciences, University of Bonn, Bonn, Germany

Julien Hiblot • National Centre of Competence in Research (NCCR) Chemical Biology, Institute of Chemical Sciences and Engineering (ISIC), Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Maximilian Höner • Faculty of Biology and BIOSS – Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany
Contributors

HIROTO IWAI • Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Tokyo, Japan

COLIN J. JACKSON • Research School of Chemistry, The Australian National University, Canberra, ACT, Australia

HARALD JANOVJAK • Institute of Science and Technology Austria (IST Austria), Klosterneuburg, Austria

KAI JOHNSON • National Centre of Competence in Research (NCCR) Chemical Biology, Institute of Chemical Sciences and Engineering (ISIC), Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; Max-Planck Institute for Medical Research, Department of Chemical Biology, Heidelberg, Germany

MIKI KOJIMA-MISAIZU • Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Tokyo, Japan

STEWART N. LOH • Department of Biochemistry and Molecular Biology, State University of New York Upstate Medical University, Syracuse, NY, USA

OLGA MAKHLYNETS • Department of Chemistry, Syracuse University, Syracuse, NY, USA

MAARTEN MERKX • Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

JOSHUA A. MITCHELL • Research School of Chemistry, The Australian National University, Canberra, ACT, Australia

ANDREAS MOGLICH • Lehrstuhl für Biochemie, Universität Bayreuth, Bayreuth, Germany; Institut für Biologie, Biophysikalische Chemie, Humboldt-Universität zu Berlin, Berlin, Germany

ANNA MORONI • Department of Biosciences, University of Milan and Biophysics Institute, National Research Council (CNR), Milan, Italy

WIGNAND W.D. MÜHLHÄUSER • Faculty of Biology and BIOSS – Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany

SAURABH NIRANTAR • p53 Laboratory, A*STAR Agency for Science, Technology and Research, Singapore, Singapore

SHYAM NYATI • Center for Molecular Imaging, University of Michigan, Ann Arbor, MI, USA; Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA

MEGAN L. O’MARA • Research School of Chemistry, The Australian National University, Canberra, ACT, Australia

ROBERT OHLENDORF • Institut für Biologie, Biophysikalische Chemie, Humboldt-Universität zu Berlin, Berlin, Germany; Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

YUKI OHMUNO-MATSUYAMA • Laboratory for Chemistry and Life Science, Institute for Innovative Research, Tokyo Institute of Technology, Yokohama, Japan

MARC OSTERMEIER • Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA

SERGIO G. PESAIJOVICH • Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

GERALD RADZIWILL • Faculty of Biology and BIOSS – Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany

ALNAWAZ REHMTULLA • Center for Molecular Imaging, University of Michigan, Ann Arbor, MI, USA; Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA
LUCAS F. RIBEIRO • Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA

RAPHAËL B. DI ROBERTO • Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

BRIAN DALE ROSS • Center for Molecular Imaging, University of Michigan, Ann Arbor, MI, USA; Department of Radiology, University of Michigan, Ann Arbor, MI, USA

MORITZ J. ROSSNER • Department of Psychiatry, Ludwig Maximilian University of Munich, Munich, Germany

INMACULADA SANCHEZ-ROMERO • Institute of Science and Technology Austria (IST Austria), Klosterneuburg, Austria

BENJAMIN M. SCOTT • Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

ROBERT STAUBL • Lehrstuhl für Biochemie, Universität Bayreuth, Bayreuth, Germany

VIKTOR STEIN • Fachbereich Biologie, Technische Universität Darmstadt, Darmstadt, Germany

ALEX STERNISHA • Department of Chemistry, Syracuse University, Syracuse, NY, USA

BIRTHE STUVEN • Lehrstuhl für Biochemie, Universität Bayreuth, Bayreuth, Germany

CHRISTOF TAXIS • Department of Biology/Genetics, Philipps-Universität Marburg, Marburg, Germany

GERHARD THIEL • Plant Membrane Biophysics, Technical University Darmstadt, Darmstadt, Germany

HIROSHI UEDA • Laboratory for Chemistry and Life Science, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan

TIANA D. WARREN • Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA

WILFRIED WEBER • Faculty of Biology and BIOSS – Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany

MICHAEL C. WEHR • Department of Psychiatry, Ludwig Maximilian University of Munich, Munich, Germany

JASON H. WHITFIELD • Research School of Chemistry, The Australian National University, Canberra, ACT, Australia

JAN P. WINTGENS • Department of Psychiatry, Ludwig Maximilian University of Munich, Munich, Germany

GRANT YOUNG • Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA

WILLIAM H. ZHANG • Research School of Chemistry, The Australian National University, Canberra, ACT, Australia
Part I

General Strategies and Considerations
Chapter 1

Synthetic Protein Switches: Theoretical and Experimental Considerations

Viktor Stein

Abstract

Synthetic protein switches with tailored response functions are finding increasing applications as tools in basic research helping dissect the molecular mechanisms that underlie the function of a cell, or in biotechnology as diagnostic reagents reporting in an autonomous fashion on distinct molecular biomarkers that are specific for a disease process [1, 2]. Common to all synthetic protein switches is a receptor that recognizes a distinct molecular queue (such as ligand binding or a posttranslational modification) and an actuator that is functionally coupled to the receptor and thus able to translate the primary molecular recognition event into a change in biophysical, chemical, or enzymatic signal depending on the preferred readout.

At the molecular level, a number of architectures have been successfully devised to construct synthetic protein switches with tailored response functions: These range from integrated designs...
featuring allosteric-binding receptors that are inserted into the tertiary structure of an actuator such as a fluorescent protein (FP) or an autoinhibited enzyme module, to modularly organized binding receptors and actuators where independently folding functional domains are organized along a linear polypeptide chain. For integrated designs, a binding event is typically transduced from the receptor to the actuator through a complex network of conformational transitions in the tertiary structure of a protein. In contrast, modularly organized synthetic protein switches are typically regulated through mutually exclusive binding interactions where conformational transitions are limited to the linkers connecting independently folding functional domains. Beyond single-component protein switches, synthetic protein switches can also be composed of multiple molecularly distinct components. These are typically regulated through the induced proximity of a transducer with an actuator or two split protein halves [3, 4].

In terms of specific applications, synthetic protein switches are increasingly employed as intracellular sensors that monitor molecular functions in an integrated and autonomous fashion in real time, e.g., reporting on the presence or absence of key metabolites, protein-protein interactions, or posttranslational modifications based on fluorescence or bioluminescence readouts [5–8]. In comparison, conventional techniques that have traditionally been employed to analyze protein-associated functions by means of antibodies or mass spectrometry only provide snapshots of molecular states as cells and tissues need to be broken up and/or fixed for analysis. In this case, monitoring time courses of biological processes based on successive time points quickly becomes laborious and also introduces variability from repeated sampling. Beyond applications in basic research, synthetic protein switches are increasingly developed as diagnostic reagents to detect clinically important biomarkers in an integrated fashion with no need for laborious work-up steps such as the successive binding and washing steps necessitated by immunological techniques based on antibodies.

Beyond applications as molecular sensors, synthetic protein switches can also be employed to actuate biological functions [9–12]. Traditionally, this has been realized through small molecular weight ligands that can bind and thus control the function of key signaling proteins inside the cell. In the majority of cases, small molecular weight ligands primarily inhibit protein-associated functions. In contrast, synthetic protein switches can regulate cellular functions in both positive and negative ways, for instance, by introducing artificial control elements into key regulatory proteins of intracellular signal transduction pathways.

With a number of applications emerging in basic research and biotechnology, a key bottleneck has been to devise generally applicable strategies to engineer synthetic protein switches with tailored response functions [1, 2, 13, 14]. Notably, current
design strategies extensively rely on iterative cycles of designing, building, and testing synthetic protein switches (Fig. 1) with the emphasis on empirical testing that is costly and time-consuming. The following chapter thus provides a summary of the key experimental techniques and theoretical considerations that apply to the construction of synthetic protein switches.

2 Designing Synthetic Protein Switches

A key goal in synthetic biology is to engineer biological functions a priori [15, 16]. This is to accelerate the design-build-test cycle and reduce the need for costly empirical optimization. In addition, a capacity to engineer biological functions a priori reflects on our fundamental understanding of the underlying biological processes and phenomena. In the context of proteins, significant progress
has been made in the computational design of protein structures, protein assemblies, protein-protein interactions, ligand and substrate specificities, as well as catalytic mechanisms [17–21]-yet, progress in the computational design of synthetic protein switches with tailored response functions has been limited. Notably, synthetic protein switches are dynamic entities and undergo conformational transitions that are critically important for their function, yet challenging to analyze and even more challenging to predict, control, and engineer in a systematic fashion. The majority of synthetic protein switches have thus been designed based on an intuitive molecular understanding of protein structure and function while computational strategies increasingly assist in the rational optimization of key functional or biophysical properties.

The protein database (PDB) features over 120,000 solved protein structures that can be exploited for the structure-guided engineering of protein switches by (semi-)rationally recombining binding receptors with enzymes, fluorescent, or bioluminescent proteins. Protein structures are readily accessible through structural visualization programs such as PyMol (DeLano WL, 2002 The PyMOL Molecular Graphics System) that provide an indispensable design aid. For instance, in domain insertion strategies, an allosteric receptor is typically inserted into surface exposed loop regions such that ligand-induced conformational changes are efficiently transmitted to the actuator modulating its function. In this way, synthetic protein switches and sensors have been engineered based on GFP [22–24], β-lactamase [25–27], tyrosine protein kinases [28–30], xylanase [31], and PQQ-dependent glucose dehydrogenase (GDH) [32]. Similarly, alternative frame folding relies on a thorough structural analysis to identify, duplicate, and modify structural elements that are important for the binding or catalytic function of a synthetic protein switch [33–35]. Structurally related to synthetic protein switches engineered by domain insertion are split protein complementation sensors that reassemble into a functional protein upon induced localization of the two protein halves. Here, structural intuition frequently guides the choice of the split sites that separate a protein into two structurally well-defined subdomains. In this way, it has been possible to reengineer a number of split luciferases to report on intracellular signaling events [36–39], split tobacco etch virus (TEV) proteases to sense and actuate intracellular signaling function [40–43], split tyrosine protein kinases to actuate cellular signaling functions and screen for drugs [44, 45], and split PQQ-dependent GDH as a universal biosensor platform [46]. In addition, the construction of modularly organized protein switches based on structurally distinct allosteric receptors and actuators benefits from high-resolution structural information as it provides clues about the position and relative orientation of the N- and C-termini that assist in the
construction of the connecting linkers and facilitate rapid diversification of input functions. Notably, many intracellular signal transducers are organized in a modular fashion that facilitates rewiring the response functions of bacterial [47–49] and eukaryotic [50–55] signal transducers or the construction of genetically encoded [56–61] and semisynthetic protein sensors [62–67]. Visual inspections of protein structures are, however, relatively crude design strategies that are nonquantitative, rely on manual assessment, and frequently need to be optimized empirically through experimental screening. Ideally, the function of a synthetic protein switch can be engineered computationally in an automated fashion based on quantitative parameters, which also reflects on our fundamental understanding how protein sequence relates to protein structure and function.

2.2 Design by Molecular Modeling

Toward this goal, a number of computational strategies have been pursued to analyze and engineer structural and functional properties of a protein a priori by means of computational design [68–70]. In its most elementary form, molecular dynamic simulations compute the behavior of an ensemble of molecules based on the physical forces that every single atom is subject to. Such high-resolution models are however computationally expensive, and in practice take prolonged periods of time to model the structure or the conformational dynamics of proteins. As a result, molecular dynamics simulations are primarily restricted to analytical studies and thus not suited to iterate through large numbers of protein mutants as necessitated in rational protein design.

Instead, increasing grades of abstraction and simplification are introduced aiming to limit the conformational search space and accelerate computation times [68–70]. This usually requires identifying, approximating, and weighing the key parameters that underlie a structural, biophysical, or functional property. Specific simplifications include restricting the dihedral angles of the polypeptide backbone and amino acid sidechains to the most frequently occurring rotamers (in the same way structural biologists match the tertiary structure of a protein to its electron density map) or approximating secondary structure propensities, solvation terms, electrostatic energies, and hydrogen bond potentials. This is increasingly complemented by bioinformatic approaches mining protein structures for functional motifs that can be grafted onto a desired binding or enzyme catalyzed reaction.

In this way, a number of new protein structures and functions could be computationally engineered including new folds [71, 72], new ligand and substrate specificities [73, 74], as well as new catalytic functions [75, 76]. In contrast, predictably engineering the conformational transitions that underlie the switch-like behavior of synthetic protein switches has proven more difficult and primarily relied on redesigning individual properties. In one recent example,
an allosterically regulated Ca\(^{2+}\)-sensitive Kemp Eliminase was engineered by introducing a binding site and reactive groups for a Kemp Eliminase reaction into the EF hand of calmodulin, while preserving its natural propensity to undergo a conformational transition from compact to extended upon binding Ca\(^{2+}\) [77]. Similarly, the ligand specificity of the bacterial transcription factor LacI was computationally reengineered to recognize fucose, gentiobiose, lactitiol, and sucralose [78], while preserving the natural propensity of LacI to bind DNA in a ligand-dependent fashion. However, preserving natural allosteric transitions while introducing new ligand specificities is nontrivial, and in case of bacterial transcription factors additionally involved experimental screening and selection of a large library of mutant protein switches [78].

In contrast, predictably engineering the conformational transitions that underlie synthetic protein switches have so far met with limited success. This particularly applies to integrated designs, where allosteric changes are regulated through complex networks of amino acids in the tertiary structure of a protein that are difficult to recapitulate in a rational manner. In contrast, for modularly organized protein switches with structurally distinct receptor, actuator, and AI-domains, the behavior of the connecting linkers can be described with synthetic polymer models to assist balancing steric strain in ligand-bound and unbound conformational states. In one example, the worm-like chain (WLC) model was successfully applied to quantify the behavior of Gly-Ser-rich linkers connecting two FPs undergoing resonance energy transfer in a Zn\(^{2+}\)-specific protein sensor [79]. Yet, these models have so far primarily been used to rationalize the behavior of a linker post-experimentally, but not engineer linkers a priori.

Beyond structure-guided protein engineering, the evolutionary history of proteins provides a rich source of information that can be computationally analyzed to derive useful functional and biophysical properties of proteins. Notably, next-generation sequencing technologies have generated an unprecedented wealth of sequence data that provides a detailed snapshot on the evolution of proteins and protein families. This data is increasingly mined and analyzed using sophisticated computational algorithms to extract valuable information on how the primary structure of a protein correlates with key biophysical and functional properties.

In the simplest case, the consensus sequence of a protein can highlight functionally and structurally important residues that are conserved within a protein family [80]. Enriching proteins with conserved consensus motifs has previously been shown to improve their thermal and conformational stability that constitutes a critical parameter in the development of recombinant proteins for many biotechnological applications including therapeutic binding agents
or enzymes for large-scale, industrial biosynthesis [84, 85]. It is worth noting that the consensus sequence of a protein does not yield a true protein sequence, but an averaged one which neglects that individual mutations are subject to epistatic effects [86]. This means, depending on their context, combination of mutations can have synergistic, neutral, or detrimental effects on a specific structural, biophysical, or function property. Considering this correlation is lost in the consensus sequence, the resulting proteins are not necessarily functional and, thus, frequently have to be correlated with additional sequence, biochemical, biophysical, or structural information to yield proteins with the desired properties.

In contrast to the consensus sequence approach, ancestral gene resurrection (AGR) aims to identify the true sequence of a primordial protein [87, 88]. This approach is unique in that it allows to resurrect and experimentally study extinct proteins. Notably, from a protein engineer’s perspective, ancestrally resurrected proteins display a number of superior properties over their contemporary counterparts. This includes superior folding, improved thermodynamic stability [89–91], and greater levels of substrate promiscuity [89, 92], which, in the context of engineering synthetic proteins, has already been exploited to reengineer the ligand specificity of allosteric binding receptors [92]. Similar to the consensus sequence approach, the evolutionary tree of a protein family is retraced based on multiple sequence alignments and different statistical methods. These include maximum likelihood, maximum parsimony, or Bayesian reconstruction to calculate the posterior probability of a protein sequence at every evolutionary branch point. While the specific evolutionary ancestral resurrection algorithm is frequently of debate—especially, if the true ancestral sequence of a protein is to be determined in the context of evolutionary studies—this is a lesser concern in protein engineering as long as the resurrected protein sequences yield improved functional or biophysical properties. For instance, AGR has been employed to improve the thermodynamic and folding efficiency of L-arginine-specific periplasmic-binding proteins (PBPs). This turned out critical for their efficient recombination with FPs to engineer L-arginine-specific FRET sensors [90] and also facilitated their subsequent reengineering into L-glutamine-specific FRET sensors [92].

Finally, statistical coupling analysis (SCA) has been successfully applied to identify co-evolving networks of residues that are distant in primary, but continuous in tertiary structure highlighting 3D hotspots that are functionally coupled in a protein [93, 94]. Notably, recombining AsLov2 and PDZ receptor domains with dihydrofolate reductase (DHFR) via computationally predicted allosteric hotspots yielded a regulated enzyme that transduces light- and ligand-induced conformational transitions from the receptor to the actuator [95].
3 Building Synthetic Protein Switches

Considering computational approaches can only optimize a limited number of biophysical and functional properties in a protein; this means the construction of synthetic protein switches relies, to a significant extent, on empirical optimization based on medium-to high-throughput screening assays. As a general rule of thumb, synthetic protein switches generated by means of random domain insertion rely on higher throughput screening approaches due to the less predictable effect of recombining two structurally well-defined protein domains on fold, structure, and function. In contrast, modularly organized synthetic protein switches can be engineered in a more rational manner solely focusing on the length and structure of the linkers connecting individual domains. Consequently, each of the individual design strategies imposes different challenges on the underlying DNA assembly process.

3.1 Non-Homology-Dependent Recombination Strategies

Before the advent of highly affordable synthetic DNA, random domain insertions were created following a limited endonuclease digest of a circular DNA construct coding for an actuator and subsequent fusion with a linear DNA construct coding for an allosteric receptor. The latter may also be circularly permutated resulting in a set of new N- and C-termini which potentially enhances the transmission of conformational changes between the receptor and the actuator; these are not necessarily confined to the original N- and C-termini, but most pronounced at internal sites [96]. The resulting libraries are then empirically screened for domain insertion mutants that are functionally recombined in allosteric hotspots (c.f. SCA that aims to predict allosteric hotspots as opposed to experimentally screen for them). This strategy has, for instance, been successfully applied to engineer a number of allosterically regulated enzymes, including maltose regulated β-lactamase [25, 97, 98], xylose regulated xylanase [99], and HIF1-binding domain cytosine deaminase [100]. Considering only 1 in 6 constructs are in frame and the unpredictable effect of domain insertion on protein structure and function, a large number of domain insertion mutants need to be screened using a suitable high-throughput screening assay. These can either be directly screened for functional protein switches, e.g., based on antibiotic resistance conferring β-lactamases [97] or in case of more technically challenging enzyme assays fused with GFP to identify in frame, non-homologously recombined genes before assaying for the relevant enzyme function in a multiwell plate assay format [99].

3.2 Homology-Dependent Recombination Strategies

Alternatively, more focused DNA insertion libraries can be created by means of homology-dependent DNA cloning methods overcoming the limitations associated with out-of-frame insertions: e.g., overlap extension PCR (OE-PCR) constitutes one of the
earliest homology-dependent recombination methods [101, 102] and has recently been applied to engineer defined linker libraries for light-activated histidine protein kinase switches [103]. Here, a small number of DNA templates with overlapping homologous sequences prime each other during every reannealing step to recombine two DNA fragments. Recombination by means of OE-PCR can, however, prove technically challenging considering the relatively low efficiency of recombination between two larger single-stranded DNA fragments. This is further aggravated by the exponential nature of PCR amplification, which potentially renders OE-PCR susceptible to nonspecific DNA amplification products and limits the number of DNA fragments that can be simultaneously recombined.

More recently, Gibson assembly has originated as a powerful, homology-dependent cloning strategy relying on the combined action of a dsDNA 5’ to 3’ exonuclease, a thermostable DNA ligase and a thermostable DNA polymerase [104]. Reactions are typically conducted at 50 °C and initiated by the exonuclease-dependent chew back of the 5’ end. This results in the formation of single-stranded 3’ DNA extensions that guide the reannealing of homologous DNA sequences that are subsequently extended and filled by the DNA polymerase and eventually sealed by the DNA ligase. Unlike OE-PCR, Gibson Assembly occurs at a constant temperature without the need for thermal cycling coordinating successive reannealing and amplification steps. This significantly increases the efficiency of recombination, enables the simultaneous assembly of multiple DNA fragments, and prevents any bias that may arise through successive reannealing and amplification cycles. While technically easy, the efficiency of Gibson assembly can be reduced by secondary structures, repeat regions and GC-rich regions as they frequently occur in the glycine- and serine-rich polypeptide linkers as is applicable in the construction of synthetic protein switches.

Beyond Gibson assembly, a number of alternative methods have been devised that rely on similar principles such as sequence and ligase-independent cloning (SliC) [105], circular polymerase extension cloning (CPEC) [106], seamless ligation cloning extract (SLICE) [107], or AQUA [108] where an exonuclease, DNA polymerase and DNA ligase function are included either as part of cell extract or within a cell.

While OE-PCR and Gibson assembly enable the seamless assembly of DNA sequences independent of restriction sites, both methods rely on homologous DNA sequences of 20–50 bp. This generally restricts the reuse of DNA coding for common receptor, actuator, and linker elements from existing, sequence verified DNA constructs and libraries. In addition, the longer the overhangs, the more expensive the synthesis of tailored oligonucleotides becomes. Alternatively, cloning strategies have been devised based on type
IIS restriction enzymes. These cut outside their recognition motif in a sequence-independent fashion to create tailored single-stranded DNA extensions. Crucially, unlike conventional restriction enzymes, the resulting single-stranded extensions are non-palindromic and thus facilitate the assembly of multiple DNA fragments in a directional manner.

Golden Gate cloning constitutes one of the most widely used DNA assembly methods based on type IIS restriction enzymes allowing for the directional and seamless assembly of multiple DNA fragments [109]. In the context of engineering synthetic protein switches, distinct structural motifs, linker elements, and functional domains are first amplified by PCR using synthetic oligonucleotides that introduce tailored DNA overhangs. These overhangs code for a type IIS restriction site and a short recombination motif that guide the ligation of multiple DNA fragments with complementary extension motifs. Individual DNA fragments are then fused following the combined action of a type IIS restriction enzyme and a DNA ligase. One key disadvantage of type IIS restriction enzyme-dependent cloning strategies is the need to remove any potential restriction sites in the coding sequence. While this does not pose a concern for synthetic DNA fragments, where restriction sites can be specifically omitted, this is not the case with genomic sequences and DNA constructs that are already available in the plasmid database of a lab.

Alternatively, USER Enzyme can be employed to create short single-stranded 3′ DNA extensions [110–114]. Here, single-stranded 3′ DNA extensions are created through the excision of uracil residues that are introduced via synthetic oligonucleotides at the PCR amplification step. The resulting 3′ DNA extensions subsequently guide the DNA ligase-dependent fusion of two or more DNA fragments. Scar sites are minimal as the only sequence requirement is a pair of A and T residues spaced apart by approximately two to six nucleotides. Similar to type IIS restriction sites, the single-stranded extensions of USER enzyme can be non-palindromic to enable the directional assembly of multiple DNA fragments.

Ultimately, the preferred DNA assembly procedure will be determined by a number of factors: This includes the architecture of a specific protein switch (e.g., whether it is modularly organized or integrated), the source of DNA (e.g., whether it is of genomic or synthetic origin), as well as any idiosyncrasies associated with the construction of a particular protein switch (e.g., whether linker regions feature repeat regions, secondary structures, or high GC content). In addition, the potential for automation and the use of commercial DNA synthesis and cloning services plays an increasingly important consideration in devising cost-effective and efficient DNA assembly processes and needs to be assessed individually for different types of synthetic protein switches.
Testing Synthetic Protein Switches

Historically, biotechnological innovation has extensively relied on experimental trial-and-error to adopt and reengineer existing biological functions toward specific applications. This particularly applies to the rational engineering of protein-associated functions which has been hampered by an insufficient understanding how the sequence of a protein relates to its function. Consequently, an increasing number of studies are breaking down the construction of synthetic protein switches into manageable substeps. This includes limited empirical optimization to engineer or optimize key functional properties such as the binding specificity of receptors and AI-domains, as well as their subsequent assembly into functional protein switches with tailored response functions. The latter is generally supported by medium- and high-throughput screening assays based on multi- and single-cell assay technologies.

The construction of modularly organized receptors and actuators, where allosteric transitions are primarily mediated by flexible linker regions, has raised the possibility of constructing synthetic protein switches from individual subcomponents based on structurally well-defined binding domains that either recognize the target ligand or modulate the output of the actuator. For instance, GFP and its engineered derivatives have a propensity of dimerizing with μM affinity which has been shown to enhance the sensitivity and dynamic range of FRET-based fluorescent sensors [115, 116]. Similarly, a number of enzymes feature naturally occurring, genetically encoded inhibitors that can be exploited for the construction of synthetic protein switches based on the autoinhibited β-lactamase module [56, 117, 118]. In the absence of structural information or the presence of naturally occurring receptor and AI-domains, highly specific protein-based binders that either recognize the target molecule or associate with the actuator to modulate its function can either be constructed de novo or sourced from natural sources and optimized using a variety of display technologies such as phage [119], yeast [120, 121], and various in vitro display technologies that either feature RNA [122–124] or DNA [125–129] as the coding nucleic acid.

Collectively, these systems display a protein either on the surface of either phage or yeast or in vitro directly on its coding nucleic acid maintaining a physical association between genotype (i.e., its coding nucleic acid) and phenotype (i.e., the protein binder that mediates its binding function). Depending on the type of display system, the target ligand can be immobilized on a solid surface retaining and enriching those phage or nucleic acids that code for a functional binder. Alternatively, the target ligand can be labeled with a fluorescent reporter molecule labeling those cells or μ-beads

4.1 Engineering Subcomponents Using Display Technologies

Historically, biotechnological innovation has extensively relied on experimental trial-and-error to adopt and reengineer existing biological functions toward specific applications. This particularly applies to the rational engineering of protein-associated functions which has been hampered by an insufficient understanding how the sequence of a protein relates to its function. Consequently, an increasing number of studies are breaking down the construction of synthetic protein switches into manageable substeps. This includes limited empirical optimization to engineer or optimize key functional properties such as the binding specificity of receptors and AI-domains, as well as their subsequent assembly into functional protein switches with tailored response functions. The latter is generally supported by medium- and high-throughput screening assays based on multi- and single-cell assay technologies.

The construction of modularly organized receptors and actuators, where allosteric transitions are primarily mediated by flexible linker regions, has raised the possibility of constructing synthetic protein switches from individual subcomponents based on structurally well-defined binding domains that either recognize the target ligand or modulate the output of the actuator. For instance, GFP and its engineered derivatives have a propensity of dimerizing with μM affinity which has been shown to enhance the sensitivity and dynamic range of FRET-based fluorescent sensors [115, 116]. Similarly, a number of enzymes feature naturally occurring, genetically encoded inhibitors that can be exploited for the construction of synthetic protein switches based on the autoinhibited β-lactamase module [56, 117, 118]. In the absence of structural information or the presence of naturally occurring receptor and AI-domains, highly specific protein-based binders that either recognize the target molecule or associate with the actuator to modulate its function can either be constructed de novo or sourced from natural sources and optimized using a variety of display technologies such as phage [119], yeast [120, 121], and various in vitro display technologies that either feature RNA [122–124] or DNA [125–129] as the coding nucleic acid.

Collectively, these systems display a protein either on the surface of either phage or yeast or in vitro directly on its coding nucleic acid maintaining a physical association between genotype (i.e., its coding nucleic acid) and phenotype (i.e., the protein binder that mediates its binding function). Depending on the type of display system, the target ligand can be immobilized on a solid surface retaining and enriching those phage or nucleic acids that code for a functional binder. Alternatively, the target ligand can be labeled with a fluorescent reporter molecule labeling those cells or μ-beads
that display a functional binder and enriching them by means of fluorescence activated cell sorting (FACS).

Beyond choosing a suitable display system, the second major consideration concerns the scaffold protein to construct tailored protein binders. Historically, the development of next-generation biologics has yielded a diverse repertoire of recombinant binding scaffolds as alternatives to monoclonal antibodies [130–132]. These typically comprise independently folding, single-chain protein domains and short peptide motifs with more or less defined structural propensities. Crucially, these newly developed binding scaffolds can be readily produced in *Escherichia coli*, fused to additional protein domains and generally display superior structural, folding, and thermodynamic properties that facilitate their purification, biophysical characterization, and integration into modularly organized synthetic protein switches.

In one recent example, an allosteric binding receptor was constructed by means of phage display fusing a circularly permuted PDZ domain with an engineered fibronectin (FN) scaffold that serves as an enhancer domain [57, 59]. The two domains are connected through a Gly-Ser rich linker, which is unstructured in the ligand unbound state, but forms a structurally well-defined sandwich complex in the ligand-bound state. Biophysical studies have also shown that formation of the sandwich complex is associated with a distinct movement of the receptor domain. This was subsequently exploited to create fluorescence and protease-based switches following recombination of the affinity clamp receptor with fluorescent proteins [58] and autoinhibited protease modules [60].

Arguably, the most technically challenging aspect in the construction of synthetic protein switches is to recombine individual subcomponents (e.g., the binding receptor, the actuator, and AI-domains) into fully functional protein switches with tailored response functions. Depending on the type of switch, this requires testing a varying number of designs over successive screening and selection cycles while looking to optimize their input-dependent switching behavior. Experimentally, this is the most labor-intensive step and, apart from designing a particular synthetic protein switch (*see Subheading 2*), the most creative one considering for every different actuator a tailored screening assay needs to be devised.

As a rule of thumb, the higher the throughput, the more technically challenging it becomes to establish a suitable screening assay. This particularly applies to synthetic protein switches that are ideally screened in positive and negative selection modes looking to identify those switches that display the largest differential function in the presence and absence of a desired target analyte. Considering the majority of synthetic protein switches actuate
their signal either through enzymes or FPs, the preferred readouts are based on spectroscopic assays monitoring changes in fluorescence, luminescence, or absorbance.

In addition, synthetic protein switches are usually composed of multiple protein domains. This constitutes a frequently underestimated factor that imposes constraints on the recombinant expression of a particular class of protein switches as well as their operating environment that both have to be accounted for in the design process. For instance, if a particular protein switch is designed to function intracellularly, its performance can be limited by cell intrinsic factors: e.g., incomplete translation or proteolytic cleavage of flexible linker regions can limit the expression of a full-length synthetic protein switch and ultimately the maximum induction ratio. This constitutes less of a concern if a synthetic protein switch is developed for in vitro applications where full-length proteins can be purified through N- and/or C-terminal purification tags.

Spectroscopic assays in combination with multiwell plate readers constitute one of the most ubiquitous assay formats used to monitor and measure binding or catalytic functions of several thousands of mutants by means of comparatively inexpensive and widespread laboratory equipment. Notably, spectroscopic assays that monitor changes in fluorescence or absorbance in multiwell plate assays formats allow for the time resolved measurement of protein function and the possibility to duplicate samples within a single plate. The latter greatly facilitates quantitative comparisons between synthetic protein switches in the presence and absence of a desired target analyte (e.g., binding ligand, cofactor, or any other target analyte that modulates the activity of the protein switch). Colony screens are conceptually similar to multiwell plate assays considering microbial colonies on an agar plate comprise thousands of mutants that can be screened on average in a cost-efficient manner. The only added complication is that assay readouts need to be spatially confined to individual colonies, for instance, through a FP or precipitating products of an enzyme-catalyzed reaction.

To assess the function of synthetic protein switches in high throughput in either multiwell or colony-based screening formats, experimental screening procedures need to be as simple as possible, ideally requiring only the sequential addition of reagents with no successive washing steps that can introduce comparatively large variabilities. Frequently, the target analyte or substrate cannot be coexpressed nor readily diffuses across the cell membrane, but needs to be added exogenously while a protein needs to be secreted or released into the lysate. The former imposes limitations on the functional folding of a protein, for instance, if a particular scaffold or enzyme naturally folds in the reducing environment of the cytoplasm, it may not efficiently export and fold in the periplasm of *Escherichia coli*. 

4.3 Multi-cell Screening in Colony- and Multiwell-Format
In practice, these considerations already prove challenging in the construction of allosterically regulated FP sensors by means of colony-based screening format: e.g., considering Ca\(^{2+}\) ions do not readily diffuse across the *Escherichia coli* inner plasma membrane, this means Ca\(^{2+}\)-specific single-FP sensors need to be exported to the periplasm to bind Ca\(^{2+}\) [133]. In comparison, genetically FRET-based FP sensors that are composed of two FP domains cannot be exported to the *E. coli* periplasm so that the plasma membrane needs to be selectively permeabilized to allow diffusion of Ca\(^{2+}\) [134]. Beyond throughput, it has also been realized that, for optimal performance, synthetic protein switches are preferably screened in their operating environment: e.g., Ca\(^{2+}\)-responsive single-FP sensors for imaging applications in neurobiology are increasingly screened in human cell lines such as HEK293 [135]. Similarly, HIF1-responsive cytosine deaminases originally screened and optimized in *Escherichia coli* have subsequently been shown to have off-target effects in human cells [136].

For higher throughput assays, fluorescent-activated cell sorting (FACS) can boost the screening capacity by several orders of magnitude, while the outcome of a screening and selection experiment can be holistically analyzed by means of next-generation sequencing. The key difference is that FACS-based selection procedures assay the function associated with a single cell as opposed to an average output of tens of millions of cells in multiwell plates or a colony. This imposes a number of technical challenges on FACS-based screening procedures: Firstly, the activity of a synthetic protein switch needs to be assayed either inside or directly on the surface of a cell. For actuators with catalytic functions, the substrate and/or the product thus need to be retained inside or attached on the surface of the cell. Secondly, asynchronies in cell growth and division as well as bursts in transcription and translation render the expression of recombinant proteins stochastic. As a result, expression levels and therefore the experimental signal usually vary by an order of magnitude across individual cells. To some extent, varying expression levels can be normalized over the size of a cell, e.g., by normalizing over forward scatter, which is an indicator of cell size, but does not allow for the same precision as multiwell plate screening assays. This poses challenges if the function of a synthetic protein switch only fractionally improves during every design-build-test cycle. Furthermore, positive and negative selections need to be performed in a sequential fashion that provides less precise readouts compared to side-by-side comparison in multiwell plates.

Unsurprisingly, the number of FACS-based screening procedures that have been successfully devised to construct synthetic protein switches is limited. In one recent example, trehalose-specific single-FP sensors were engineered in *Escherichia coli* by inserting allosteric, trehalose-specific PBPs into GFP [137].

### 4.4 Single-Cell Screening Based on FACS

For higher throughput assays, fluorescent-activated cell sorting (FACS) can boost the screening capacity by several orders of magnitude, while the outcome of a screening and selection experiment can be holistically analyzed by means of next-generation sequencing. The key difference is that FACS-based selection procedures assay the function associated with a single cell as opposed to an average output of tens of millions of cells in multiwell plates or a colony. This imposes a number of technical challenges on FACS-based screening procedures: Firstly, the activity of a synthetic protein switch needs to be assayed either inside or directly on the surface of a cell. For actuators with catalytic functions, the substrate and/or the product thus need to be retained inside or attached on the surface of the cell. Secondly, asynchronies in cell growth and division as well as bursts in transcription and translation render the expression of recombinant proteins stochastic. As a result, expression levels and therefore the experimental signal usually vary by an order of magnitude across individual cells. To some extent, varying expression levels can be normalized over the size of a cell, e.g., by normalizing over forward scatter, which is an indicator of cell size, but does not allow for the same precision as multiwell plate screening assays. This poses challenges if the function of a synthetic protein switch only fractionally improves during every design-build-test cycle. Furthermore, positive and negative selections need to be performed in a sequential fashion that provides less precise readouts compared to side-by-side comparison in multiwell plates.

Unsurprisingly, the number of FACS-based screening procedures that have been successfully devised to construct synthetic protein switches is limited. In one recent example, trehalose-specific single-FP sensors were engineered in *Escherichia coli* by inserting allosteric, trehalose-specific PBPs into GFP [137].
In another example, small molecule-dependent sensors were engineered based on ligand receptors that are proteolytically degraded, but stabilized upon binding their cognate ligand [138]. The resulting ligand sensors can, in turn, be fused either to a fluorescent protein or a transcription factor to regulate the activity of a reporter gene. Yet, this strategy heavily relies on screening millions of mutants in *Saccharomyces cerevisiae* using a suitable high-throughput screening procedure, while destabilizing mutations cannot be identified computationally. In another example, a protease-based screening strategy has been devised in *Saccharomyces cerevisiae* based on controlling the export of a reporter protein to the cell surface by cleaving off a protein export tag [139]. This screen was specifically applied to engineer the substrate specificity of TEV protease using sequential positive and negative selection cycles and could be readily adapted to screen the function of synthetic protein switches in the presence and absence of a target analyte.

Beyond FACS-based screening procedures, μ-droplet-based screening strategies have been developed for greater control of protein expression and reaction conditions in microfluidic devices [140–142]. While technically challenging and not widely available in the majority of molecular and cell biology oriented labs, μ-droplet screening strategies carry a number of advantages that are directly applicable to the construction of synthetic protein switches: Firstly, the function of a synthetic protein switch including catalytic functions can be assayed extracellularly, which facilitates the control of the reaction conditions. To this end, individual proteins can be expressed in *Escherichia coli* [143, 144], *Saccharomyces cerevisiae* [145, 146], or mammalian cells [147] and displayed on the cell surface or released following cell lysis. Droplet fusion technology can, in turn, be employed to deliver a substrate or a target analyte that is used to analyze a distinct protein function. μ-droplet screening strategies are, however, technically challenging and limited to laboratories with the specialist expertise, especially considering only few studies have successfully screened protein function using integrated devices that can fuse droplets, deliver reagents, incubate for defined time periods and assay protein function such as enzymatic activity [143–147].

4.5 Growth-Based Genetic Selection Assays

Beyond spectroscopic readouts based on synthetic or genetically encoded fluorescent, luminescent, and absorbant reporter molecules, selecting for the function of enzyme-based actuators can also be directly linked to the survival of a microorganism by means of growth-based selection assays. Both *Escherichia coli* and *Saccharomyces cerevisiae* are suitable for this purpose providing a range of antibiotic and auxotrophic markers. The most widely used selectable marker in *Escherichia coli* is based on β-lactamase that can also be assayed spectroscopically and has pioneered the design of synthetic protein switches by means of domain insertion
but also mutually exclusive binding interactions [56, 117, 118]. *Saccharomyces cerevisiae* also provides a powerful microorganism for devising high-throughput selection procedures based on genetic complementation. In one recent example, a light-regulated K+ channel was engineered by recombining a photo-responsive LOV2-Jα domain with the small viral K+ channel Kc [149]. To this end, synthetic K+ channels were screened for light-responsiveness in positive and negative selection modes following illumination with blue light or in the dark. The selection strategy was based on a mutant strain of *Saccharomyces cerevisiae* deficient in endogenous K+ channels. Similar genetic complementation strategies are conceivable to screen for metabolic functions in high throughput through auxotrophic complementation of metabolic enzymes in both *Escherichia coli* and *Saccharomyces cerevisiae*.

Ultimately however, the key technical challenge with growth-based selection assays is to control the reaction conditions, in particular, the reaction environment, the concentration of individual components and the selective pressure. In addition, the growth of a particular synthetic protein switch may not exclusively depend on its function, but a cell can both adapt genetically and biochemically to enhance growth irrespective of a given synthetic protein switch mutant.

5 Outlook

Synthetic protein switches are increasingly developed and applied both in basic research and biotechnology to monitor biological processes in an integrated and autonomous fashion. For now, due to our limited understanding to predictively engineer protein-associated functions, the construction of tailor-engineered protein switches has relied, to a significant extent, on empirical optimization based on high-throughput screening procedures. Suitable high-throughput screening and selection procedures are however technically challenging to establish, need to be tailored toward specific enzyme readouts, and are ideally amenable to positive and negative selection modes. This is further hampered by the vast size of protein sequence space which generally outstrips our capacity to screen and engineer protein-associated functions in high throughput. This particularly applies to engineering allostericity that constitutes one of the most complex and least understood protein functions. Computational strategies have therefore been limited to optimizing individual properties such as the thermodynamic stability or the binding specificity of an allosteric receptor, but will undoubtedly continue gaining importance as our molecular mechanistic understanding of artificially engineered protein switches is anticipated to improve.
Acknowledgments

This work was funded by the Hessen’s LOEWE Federal State iNAPO research network.

References


Chapter 2

Construction of Allosteric Protein Switches by Alternate Frame Folding and Intermolecular Fragment Exchange

Jeung-Hoi Ha and Stewart N. Loh

Abstract

Alternate frame folding (AFF) and protein/fragment exchange (FREX) are related technologies for engineering allosteric conformational changes into proteins that have no pre-existing allosteric properties. One of their chief purposes is to turn an ordinary protein into a biomolecular switch capable of transforming an input event into an optical or functional readout. Here, we present a guide for converting an arbitrary binding protein into a fluorescent biosensor with Förster resonance energy transfer output. Because the AFF and FREX mechanisms are founded on general principles of protein structure and stability rather than a property that is idiosyncratic to the target protein, the basic design steps—choice of permutation/cleavage sites, molecular biology, and construct optimization—remain the same for any target protein. We highlight effective strategies as well as common pitfalls based on our experience with multiple AFF and FREX constructs.

Key words AFF, Biosensor, Fluorescence, FRET, FREX, Protein design, Protein engineering

Abbreviations

AFF Alternate frame folding
CP Circular permutant
Fn3 Fibronectin 3
FP Fluorescent protein
FRET Förster resonance energy transfer
FREX Protein/fragment exchange
N Normal fold of protein
N’ Alternate fold of protein
POI Protein of interest
RBP Ribose-binding protein
WT Wild-type
1 Introduction

A biosensor is minimally composed of an input module, which interacts with the analyte, and an output module that reports on that interaction. Proteins excel at both roles. As input domains they are masters of molecular recognition, having the ability to bind targets tightly and specifically amidst a sea of similar-looking decoys. As output domains they possess a wide array of biological functions, among the most useful of which for biosensing are fluorescence and enzymatic activity. A major additional advantage of a protein-based biosensor is that it is genetically encodable for in vivo applications.

The main challenge in designing a protein-based biosensor is solving the problem of how to couple input and output domains, both physically and functionally, so that binding the analyte produces a detectable signal. Nature has given us some treasured, but rare clues in the form of proteins that undergo large-scale conformational changes in response to ligand binding (e.g., calmodulin, which has launched a family of fluorescent calcium sensors). The great majority of proteins, however, do not change their structure appreciably upon binding.

Calbindin D9k, fibronectin 3 (Fn3), and ribose binding protein (RBP) are three such examples in which the structures of the proteins in their ligand-free and ligand-bound states are similar. To address this challenge, we developed two methodologies for engineering a large, binding-dependent conformational change into each protein, which was then detected by placement of either chemical fluorophores or fluorescent proteins (FPs) [1–4]. The methods are known as alternate frame folding (AFF) and protein/fragment exchange (FREX).

AFF and FREX both use partial sequence duplication to give a protein of interest (POI) a mutually exclusive choice between folding to its normal native state (N) or to an alternate form (N′) that possesses native-like structure and function (Fig. 1). Step 1 (see Subheading 3) of the protocol for creating an AFF-modified POI (POI-AFF) involves choosing an appropriate N-terminal or C-terminal segment of the POI to duplicate. One or more amino acids are identified in the segment that, when mutated, abrogate binding of the POI to its target ligand. Step 2 (see Subheading 4) entails attaching the duplicate copy of the N-terminal or C-terminal segment to the C- or N-terminus of the POI, respectively, by means of a peptide linker. Fig. 1a illustrates that POI-AFF can either fold by using the normal order of amino acids to yield N, or by using a rearranged order of amino acids to generate a circularly permuted structure (N′). The relative thermodynamic stabilities of N and N′ are tuned in Step 3 (see Subheading 5) of the protocol such that POI-AFF is predominantly in state N (or N′) in the
absence of the target ligand, and chiefly in state $N'$ (or $N$) in its presence. Finally, fluorophores are incorporated at locations that consistently report on the $N \rightleftharpoons N'$ conformational change, independent of the choice of POI.

The FREX mechanism can be considered an intermolecular version of AFF, in which the duplicated segment is not covalently attached to the POI from which it was derived, but rather added in trans to the POI (Fig. 1b). The $N'$-state of a FREX-modified POI (POI-FREX) is the intermolecular complex of the POI and the fragment, which forms only in the presence of the target ligand and is detected by FRET between donor and acceptor fluorophores placed on the POI and fragment, respectively. The chief advantage

Fig. 1 Schematic of AFF (a) and FREX (b) switching mechanisms. Primary amino acid sequences are indicated by horizontal bars with folded protein structures represented below the sequences. For the two sequences in parentheses an N-terminal segment (containing a critical binding residue) is duplicated. The other two sequences, and the structures that result from their folding, represent the analogous case in which a C-terminal segment is duplicated. Wavy lines indicate the copy of the duplicate segment that is orphaned in $N$ and $N'$ conformations and is hence unfolded. The $N$-fold of POI-FREX is shown with a packing mutation in the green arrow that is swapped out by the wild-type residue from the orange arrow in the $N'$-fold.
of FREX and other two-component designs is that the ratiometric FRET change observed upon binding tends to be greater than that of single-component sensors, because FRET efficiency is typically reduced to near-zero values in the unbound state of the two-component sensors. The main limitation of FREX is that the POI and fragment should be present at close to equimolar concentrations to achieve maximum FRET response. The protocols for creating POI-AFF and POI-FREX sensors are very similar. We outline the protocol for AFF below, and enumerate the modifications for FREX after each step.

2 Gathering Ingredients: The POI

1. An available X-ray or other high-resolution structures of the POI or homolog thereof greatly facilitate the design process.
2. For AFF, the POI should ideally not contain any reduced Cys residues, as fluorophores are typically introduced by thiol-reactive chemistry. Cys in the POI may be tolerated vis-à-vis fluorescence labeling if they are buried and inaccessible to solvent. FREX can make use of fluorescent proteins for detection so reduced Cys residues are not inherently problematic. For both AFF and FREX, if oxidized Cys are present, the duplicate segment must be chosen (see Subheading 3) such that it is not crosslinked to the nonduplicated region of the POI by a disulfide bond.
3. Consider using the most stable variant of the POI available, e.g., one derived from a thermophilic organism. An axiom of protein folding is that it is far easier to destabilize a protein than to stabilize it. Accordingly, most of the modifications and tuning mutations employed herein either intentionally or unintentionally destabilize the POI. Starting with a stable template allows for greater design freedom.

3 Step 1 of AFF Protocol: Choosing the Segment of the POI to Duplicate

3.1 Identify a Binding Mutation

The only absolute requirement for the duplicate segment is that it contain at least one residue that, when mutated, greatly reduces affinity of the POI for its target ligand. Binding knockout mutations are often known from prior functional or genetic studies, and they may also be deduced from an existing crystal structure of the POI or homolog thereof. We have found that choosing a binding mutation close to the beginning or end of the amino acid sequence is advantageous, because this allows the duplicate segment to be short in length. Since the duplicated amino acids extend from POI-AFF as N- or C-terminal tails (Fig. 1a), shorter segments may
present less of a risk for aggregation or degradation, although the opposite may be true. In any case, a binding mutation near one of the termini gives one the freedom to experiment with both short and long duplicate segments.

For a binding mutation near the C-terminus, the duplicated segment ends at the C-terminus (Fig. 1a). It begins at a position N-terminal to the binding mutation and this position is chosen to be a surface loop or turn. The reason is that this loop becomes the permutation site of the N'-fold, i.e., the location at which the polypeptide chain is broken and new N- and C-termini are generated. Interrupting an alpha helix, beta strand, or buried hydrophobic region is expected to be more destabilizing than disrupting a surface loop, although there are examples of successful permutation sites at the former locations [5–10]. The same considerations apply to the case where the binding mutation is near the N-terminus, except the duplicated segment begins with the N-terminus and ends at a surface loop C-terminal to the binding mutation (Fig. 1a).

Inability to find a stable circular permutant (CP) is the most common failure point in the AFF protocol. Permutation almost always destabilizes a protein, and there is no reliable method for predicting the extent of destabilization for a given permutation site. The CP needs to be at least marginally stable ($\Delta G_{\text{unfold}} \geq 2–3$ kcal/mol), with more stable CPs requiring less optimization (see Subheading 5). Our approach for selecting permutation sites is to choose the first three to four surface loops either N-terminal or C-terminal to the binding mutation, depending on whether the binding mutation is closer to the C-terminus or N-terminus, respectively. Loops that are close to the binding/active site should be avoided for functional reasons, although xylanase [9] and beta lactamase [11] were permuted at several loops proximal to their active sites without major loss of activity. Fortunately, all but the smallest POIs will have many loops from which to choose and at least one will usually be stable and functional enough for the AFF design. For example, RBP (277 amino acids) has 13 surface loops (Fig. 2a). We created CPs at eight of these loops and all were stable, soluble, and functional. All were destabilized compared to wild-type (WT) RBP, however, and this finding demonstrates the advantage of starting with the most stable variant of the POI available.

The linker functions to physically bridge the original N- and C-termini of the POI. It effectively becomes a new surface loop of the CP. As such, the amino acid sequence should be hydrophilic and flexible enough to not impose any new constraints on the protein structure. We base our linkers on Gly/Ala/Ser repeats, although more advanced design criteria have been discussed [9, 11–15]. With regard to linker length, a rule of thumb is to measure the N-to-C distance ($C_\alpha-C_\alpha$) from the structure “as the
crow flies,” and calculate length using ~2.5 Å per amino acid. For POIs in which the line-of-sight between termini is blocked by structure, we use a figure of <2.0 Å per amino acid to account for the arc that the linker must take over the curved surface of the protein. For example, RBP has an N-to-C distance of ~40 Å (Fig. 2a), which we spanned with a linker consisting of 30 amino acids [4]. If the N-to-C distance is not known precisely, it is best to err on the side of length, as we have found that using linkers longer than necessary does not dramatically destabilize the CP, in contrast to using linkers that are too short [16].

3.4 Characterize CPs

At this stage it is important to express and purify candidate CPs. CPs that exhibit degradation, aggregation, or loss of function should be rejected. The CPs are purified and their relative thermodynamic stabilities (as well as that of the POI containing binding mutation) are determined using chemical or thermal denaturation techniques [17] (see Subheadings 3.4.1 and 3.4.2). The most stable CP is then selected for AFF gene construction (Subheading 4). As a final note, fluorophores are typically introduced into the AFF protein by means of two Cys residues introduced at the N-terminus and in the surface loop of the N-fold selected as the permutation site. Although these Cys do not generally affect structure or stability of WT or CP forms of the POI, it is prudent to incorporate them into the constructs at this point to most accurately represent the N and N’ folds of POI-AFF.

Fig. 2 Examples of viable circular permutants for AFF and fragments for FREX. (a) Stable and functional CPs were generated by cleaving the RBP sequences at the eight surface loops centered around the positions indicated by black spheres, and joining the original termini by a Gly/Ala/Ser-based linker of 30 amino acids. The ribose ligand is shown as black sticks. (b) The Fn3-FREX sensor was created by duplicating residues 48 to the C-terminus (black segment). The binding mutation site (Tyr87) contacts the target ligand (SH2 domain, sticks) and the tuning mutation site (Ile75) packs against hydrophobic residues in the two gray beta strands shown.
1. Prepare solution A and solution B of the desired buffer, pH, salt, etc. The two solutions are made identically except solid, ultrapure urea (final concentration of 8 M) or GdnHCl (final concentration of 6 M) is added to solution B prior to the aliquots of stock buffer, salt, etc.

2. Add identical aliquots of concentrated protein to solution A and solution B. The final protein concentration depends on the instrumentation used to monitor unfolding, but typical concentrations for fluorescence and circular dichroism (CD) are 1–20 μM.

3. Prepare at least 25 samples consisting of evenly spaced mixtures (in denaturant concentration) of solution A and solution B. For example, sample #1 is 100% solution A, sample #25 is 100% solution B, and samples #2–24 contain linearly increasing concentrations of denaturant. Use of a two-syringe Hamilton dilutor or a manual repeating pipet is recommended to minimize denaturant concentration error. Incubate samples at the desired temperature until equilibrium is reached (typically ≥2 h).

4. Scan samples using the instrument of choice (UV/Vis spectrophotometer, fluorimeter, CD spectrophotometer) at the desired wavelength. The observed signal ($\theta_{\text{obs}}$) follows a sigmoidal curve, as shown in Fig. 3 (left). Fit $\theta_{\text{obs}}$ to the linear extrapolation equation (Eq. 1):

$$\theta_{\text{obs}} = \left(\theta_U + s_U[D]\right) + \left(\theta_N + s_N[D]\right) \left(\exp(\Delta G_{\text{unfold}} - m[D])\right) / \left(1 + \exp(\Delta G_{\text{unfold}} - m[D])\right)$$

where [D] is denaturant concentration and parameters $\theta_N$ and $\theta_U$ are the signals of the native and unfolded forms of the protein at zero denaturant concentration, $s_N$ and $s_U$ are the slopes of the native and unfolded baselines, $\Delta G_{\text{fold}}$ is the stability of the protein at zero denaturant concentration, and $m$ is the cooperativity parameter. The midpoint of chemical denaturation ($C_m$; equal to $\Delta G_{\text{unfold}}/m$) is a particularly useful parameter for ranking the relative stabilities of related protein variants with similar $m$-values (e.g., the POI and mutants thereof), because it is more accurate and reproducible than $\Delta G_{\text{unfold}}$.

3.4.2 Obtaining Stability Parameters from Thermal Denaturation Curves

1. Prepare a solution of protein in the desired buffer. For CD-monitored denaturation, one should generally use dilute protein solutions (1–5 μM) and a long path length cuvette (1 cm) to minimize aggregation at higher temperatures.

2. To establish native and unfolded baselines, start the melt at a temperature at least 10 °C below the beginning of the unfolding transition, and continue the melt at least 10 °C after the transition is 90% complete. After the melt is finished, let the
sample equilibrate at the starting temperature and record the CD signal. Thermal denaturation is considered reversible if the signal returns to $\geq 90\%$ of its original value.

3. One can obtain $\Delta G_{\text{fold}}$ and melting temperature ($T_m$) parameters by fitting $\theta_{\text{obs}}$ to the v’ant Hoff equation (17) if denaturation is reversible. However, since thermal transitions are often irreversible, and because we are mostly interested in the relative stabilities of closely related protein variants, it is sufficient to report apparent $T_m$ values by interpolating the midpoints of melting curves obtained under identical solution and heating conditions. To do so, fit the linear portions of the native and unfolded baselines as shown in Fig. 3 (right) and calculate $\theta_N$ and $\theta_U$ at each experimental temperature. Interpolate the temperature at which $\theta_U - \theta_{\text{obs}} = \theta_{\text{obs}} - \theta_N$.

3.5 Modifications to Step 1 for FREX

The steps described in Subheadings 3.1 and 3.2 are identical in the FREX protocol. Just as there is no method for predicting CP stability as a function of permutation site, there is no means for predicting the affinity of two complementary protein fragments based on the location of the cleavage site. We therefore apply the same criteria in choosing the binding mutation and duplicate segment for FREX. The steps described in Subheadings 3.3 and 3.4 do not apply to FREX. At this design stage, however, one should identify a site in the POI for making tuning mutations. A tuning mutation is often needed for AFF and will always be necessary for FREX, since the duplicate fragment will not form a complex with the POI (at reasonable fragment concentrations) unless a...
A structural defect is introduced into the POI which is then “swapped out” by fragment binding. A reliable tuning mutation site consists of a large, buried hydrophobic residue (Leu, Ile, Phe, or Tyr) that is in the duplicated portion of the POI and packs against hydrophobic residues in the unduplicated portion (Fig. 2b). A single tuning site should (and usually can) be selected that will service all fragment lengths to be tested. Tuning consists of substituting progressively smaller hydrophobic residues (Val, Ala, Gly, and for extreme destabilization, a charged residue) into that site to progressively destabilize the POI.

4 Step 2 of AFF Protocol: Gene Construction

AFF genes are generated by constructing two half-genes corresponding to the N- and N'-frames, mutating them as needed, and then fusing them together. Keeping the frames separate is necessary because mutagenic PCR primers will bind to two locations in the full-length AFF gene due to sequence duplication. This redundancy also makes it difficult to obtain fully synthetic AFF genes from commercial sources.

A His6–8 sequence or other purification/expression tag can be introduced at this point. The tag should be expressed at the terminus of the protein that comprises the N'-fold, i.e., at the N-terminus of POI-AFF if a C-terminal fragment was duplicated, or at the C-terminus if an N-terminal fragment was duplicated. This arrangement will help select against purifying degradation products, because if a cellular protease cleaves POI-AFF it will likely do so by attacking the less stable N'-fold. The following steps describe construction of an AFF-POI in which a C-terminal fragment is duplicated and appended to the N-terminus; analogous steps are taken for an AFF-POI with a duplicated N-terminal fragment.

4.1 N-frame Half-Gene

Introduce binding, Cys, and tuning mutations into the WT POI gene as dictated in Subheading 3. Amplify the N-frame half-gene by annealing primer 1 (containing a stop codon) to the 3'-end of the above-modified POI gene, and primer 2 to the 5'-end (Fig. 4a). Primer 2 encodes for the C-terminal half of the desired linker and bears a restriction site of choice at its 5'-end. The restriction site necessarily becomes incorporated into the linker and for this reason we favor the Not I restriction site because it encodes for Ala-Ala-Ala.

4.2 N'-frame Half-Gene

Introduce an N'-frame tuning mutation into the WT POI gene if necessary. Design primer 3 such that it binds the POI gene at a position that defines the start of the duplicate polypeptide segment (and the N-terminus of POI-AFF) (Fig. 4a). Primer 3 begins with a codon for Cys. Primer 4 anneals to the 3'-end of the POI gene,
encodes for the N-terminal half of the linker, and ends with the same restriction site as in Subheading 4.1. PCR amplify to generate the $N'$-frame half-gene. Digest the two half-genes with the restriction enzyme. The products are now ready to be ligated and subcloned into an expression vector of choice (Fig. 4b).

4.3 Alternate Method: Fusion PCR

The amino acids imposed by the restriction site usually have little effect on the properties of the linker, especially when the linker is long. For short linkers or those that must be composed of a particular sequence (e.g., protease site), these leftover residues may be undesirable. The alternate method allows one to eliminate the restriction site signature altogether. Construct the $N$-frame PCR template as in Subheading 4.1 except use primer 5 instead of primer 2 (Fig. 4a). Primer 5 contains enough of the linker sequence to overlap with primer 6 by at least 20 base pairs; this can correspond to the full or partial linker sequence depending on linker length. Primer 5 does not contain a restriction site. Create the $N'$-frame template as in Subheading 4.2 except use primer 6 instead of primer 4. Primer 6 binds to the 3'-end of the POI gene and ends with a full or partial linker sequence, complementary to that in primer 5 with at least a 20 base pair overlap, again without a restriction site. PCR amplify the $N'$-frame template by annealing primer

**Fig. 4** Cloning strategy for constructing AFF genes. (a) Annealing sites and compositions of primers 1–6 are indicated below the POI gene. The restriction enzyme digestion/ligation and fusion PCR methods for connecting the $N$ and $N'$ half-genes are shown in (b, c), respectively.
3 and primer 6 to the POI gene. Mix the N-frame template, N'-frame template, primer 1, and primer 3, and generate the full-length AFF gene using the overlap extension PCR method (Fig. 4c) [18]. Primer 1 and primer 3 will each anneal to a second, undesirable site within the N'-frame and N-frame template, respectively, and this results in shorter PCR products. Purify the longest PCR product by agarose gel and subclone into the expression vector.

4.4 Modifications to Step 2 for FREX

The cloning steps for creating FREX constructs are simpler, since the duplicated DNA sequences are never physically joined and no linker is present. At this stage, we find it useful to fuse the gene of a donor and acceptor FP to the either end of the POI and fragment genes. In addition to providing a direct binding assay for sensor tuning (Subheading 5), an FP serves as a carrier protein to help the fragment express well, resist degradation, and stay soluble in cells.

5 Step 3 of AFF Protocol: Optimization

The objective of this step is to tune the thermodynamics of the sensor so that it is mainly in the N-form in the absence of ligand, and switches to the N'-form upon ligand binding. For the sake of discussion, we assume the most common outcome of Subheading 3, i.e., that the CP form of the POI was found to be less stable than the WT POI, and that the binding mutation was introduced into the N-frame. The ideal distribution of N:N' populations in the absence of ligand is ~10:1. This scenario, which corresponds to the N-fold being ~1.4 kcal/mol more stable than the N'-fold, achieves an optimal balance of near-maximum fluorescence signal change and minimal reduction of observed ligand-binding affinity. Which tuning mutation (if any) needs to be made to achieve this balance, and into which frame it should be placed, can be predicted from the results of Subheading 3.4. Introducing this mutation into POI-AFF during Subheading 4 can save time by requiring less subsequent optimization in Subheading 5.

5.1 Binding Positive Control

A common problem with POI-AFFs prior to their thermodynamic balancing is that the N-fold is so much more stable than the N'-fold that the ligand binds weakly or not at all. This snag is likely to be encountered if the POI was found in Subheading 3.4 to be much more stable than the CP (e.g., $\Delta T_m > 10$ °C or $\Delta \Delta G_{\text{unfold}} > 3$ kcal/mol; see Subheadings 3.4.1 and 3.4.2). We

\[ \text{Because a portion of the binding energy is used to drive the } N \rightarrow N' \text{ conformational change, the observed } K_j \text{ will be greater than the intrinsic } K_d \text{ of the POI by a factor of } (1 + K_N + K_{NC})/K_N, \text{ where } K = \exp(-\Delta G_{\text{unfold}}/RT) \text{ for the respective } N- \text{ and } N'-\text{folds} \ [3]. \]
therefore recommend that one initially perform a positive binding test using saturating levels of ligand by means of a rapid and not necessarily quantitative assay including size exclusion chromatography, affinity pull-down, or shift in $T_m$ or $C_m$. If no binding is observed, then one or more destabilizing mutations are introduced into the $N$-fold, e.g., substituting Gly, Ala, or Val at one or more hydrophobic packing sites identified in Subheading 3. A positive binding result indicates that the protein has adopted the $N'$-fold in the presence of ligand but does not reveal whether it had switched to the $N'$-fold from the $N$-fold, or was already in the $N'$-fold in the absence of ligand.

### 5.2 Stability Tuning

Once a binding interaction has been established, the final step is to optimize the sensor for maximum fluorescence response and binding affinity. Fluorescent donor and acceptor groups are attached to POI-AFF at its N-terminus and in the permutation loop of the $N$-frame. We favor attaching maleimide dyes to engineered Cys residues, although there are a variety of alternate chemistries and labeling strategies from which to choose. The important considerations are that one obtains close to a 1:1 donor:acceptor ratio and that the chosen dyes are sensitive to short-range distance changes (short Förster radii for FRET), as they will be close enough to be in contact in the $N'$-fold and of variable distance apart in the $N$-fold, depending on the length and residual structure of the duplicated segment. To achieve 1:1 donor:acceptor labeling one may have to experiment with different ratios of dyes in the labeling step. We have also obtained satisfactory results by labeling both positions with a single fluorophore (pyrene excimer formation and BODIPY-FL self-quenching).

To evaluate sensor performance, increasing amounts of ligand are added to a fixed concentration of fluorescently labeled sensor. Fitting the observed fluorescence ($\theta_{\text{obs}}$) to the one-site binding equation (Eq. 2) yields the dissociation constant ($K_d$) and fluorescence change ($\Delta \theta = \theta_{\text{bound}} - \theta_{\text{free}}$) as the parameters. $L_T$ and $P_T$ are total concentrations of ligand and protein.

$$\theta_{\text{obs}} = \theta_{\text{bound}} + (\theta_{\text{bound}} - \theta_{\text{free}})\left( (L_T + P_T + K_d)^{-1} - \left( (L_T + P_T + K_d)^{-2} - 4P_TK_d \right) \right)^{1/2} / (2P_T) \quad (2)$$

The three possible outcomes are that a binding curve: (i) is obtained with a fitted $K_d$ value close to that of WT POI; (ii) is obtained with a fitted $K_d$ significantly larger than that of WT POI; and (iii) cannot be generated because no fluorescence change is observed. Outcome (i) indicates that no further optimization is

---

5 Existing AFF sensors have thus far made exclusive use of chemical fluorophores rather than genetically encoded FPs. The reason is that inserting an FP into the permutation loop of the $N$-frame would likely destabilize the $N$-fold and necessitate extensive thermodynamic rebalancing. It may be possible, however, to insert a CP form of the FP (in which the close proximity of its N- and C-termini would be less perturbing to the POI), or to move the FP from the permutant loop to the C-terminus of POI-AFF.
necessary and the sensor is ready for use. Result (ii) suggests that the N-fold is too stable and that a moderately destabilizing mutation should be introduced into the N-frame. If $\Delta \theta$ is close to zero (outcome (iii)) then one can conclude that POI-AFF either never switched conformation (i.e., it was in the $N'$-fold even in the absence of ligand), or that it changed conformation and the fluorophores did not report on the change. In the former scenario the solution is to introduce tuning mutations into the $N'$-frame instead of the N-frame. The expectation is that $\Delta \theta$ will initially increase and then approach a maximum value as the $N'$-fold is progressively destabilized by packing mutations of increasing severity [3]. Any one of these variants can be used for sensing, although for the typical case in which a balance between high $\Delta \theta$ and low $K_d$ is desired, the mutant that yields outcome (i) is selected as the final optimized product.

### 5.3 Optional: Kinetic Tuning

The AFF sensing mechanism is reversible and the speed at which the forward and reverse conformational changes occur determines how rapidly the sensor can detect fluctuations in analyte concentration. It may be desirable in some applications to increase this response rate. For calbindin D$_{9k}$ and RBP, the rate limiting step of the $N \rightarrow N'$ and $N' \rightarrow N$ reactions appears to involve a local but not global unfolding event, most likely unfolding/dissociation of the copy of the duplicate segment that is docked to the shared segment of the protein. If this result is general, any mutation that destabilizes the interaction between the shared and duplicate regions has the potential to accelerate the conformational switch. If one wishes to attempt to engineer a faster AFF sensor, we recommend first following all of the thermodynamic balancing procedures described above, then introducing kinetic tuning mutation(s) in a manner that does not perturb the existing $N/N'$ balance or ligand-binding affinity. For example, consider a hypothetical 200-amino acid POI from which residues 150–200 were duplicated to generate POI-AFF. Further suppose that residues Phe130 and Leu170 pack tightly against each other. Kinetic tuning could consist of a single Phe130→Ala substitution in the shared region, or Leu170→Ala + Leu170'→Ala mutations in the respective duplicate segments. In either case the N-fold and N'-fold are in principle destabilized equally but the kinetic barrier to their interconversion may be lowered, depending on the extent to which the Phe130-Leu170/Leu170' interaction is formed in the transition state ensemble.

### 5.4 Modifications to Step 3 for FREX

As noted, it is simplest to fuse donor and acceptor FPs to the full-length POI and duplicate fragment at the cloning stage; however, if chemical fluorophores are desired then the labeling procedure is identical to that of AFF. One difference of FREX is that the two components of the sensor are on average very far apart in the absence of ligand (at reasonable protein concentrations), so donor/acceptor pairs with a wide range of Förster distances can be considered.
Thermodynamic tuning for FREX consists of binding experiments in which ligand is added to equimolar concentrations of fluorescently tagged full-length POI and fragment. For example, we optimized the Fn3-FREX sensor by screening the I75V, I75A, and I75G tuning variants of the POI in the background of the Y87A binding mutation (Fig. 2b). Fluorescence data are fit to Eq. 2. Our simulations have shown that it is possible, although highly unlikely, that a ternary complex will form in the absence of a packing mutation, as the entropic barrier for intermolecular folding is usually too large to be overcome by ligand-binding energy alone [3]. A more realistic concern is that the binary complex of full-length POI and fragment will form without ligand present. This occurs when the full-length POI is “overbalanced,” i.e., so destabilized by the tuning mutation that restoration of the native packing interaction by the fragment drives folding in the absence of ligand binding. It is not uncommon to span these extremes of stability by placing a set of packing mutations at a single well-chosen site. A nicely balanced FREX construct is characterized by unusually high Δθ (achieved by reducing FRET in the free state to near zero) and Kₐ similar to that of the WT POI. This condition was met in the Fn3-FREX sensor by the I75A tuning mutation.

6 Concluding Remarks

Both AFF and FREX involve a folding competition akin to a two-person game of molecular musical chairs: one copy of the duplicate segment is left standing and is at least partially unfolded at any given time. One must be alert to degradation and/or aggregation originating from the orphaned copy. Computational methods for predicting the stability and solubility of CPs, protein fragments, and complexes thereof will greatly facilitate the design of FREX and AFF sensors, as well as that of conformational switches based on other mechanisms.

Acknowledgments

This work was supported by NIH grant R01 GM115762 to S.N.L.

References


Chapter 3

Construction of Protein Switches by Domain Insertion and Directed Evolution

Lucas F. Ribeiro, Tiana D. Warren, and Marc Ostermeier

Abstract

A protein switch is a protein that changes between inactive (“off”) and active (“on”) states in response to a biomolecule or physical signal. These switches can be created by fusing two domains in such a way that the activity of the output domain is regulated by the input domain’s recognition of an input signal (such as the binding of a molecule, recognition of light). Here, we describe several methods for randomly fusing two domains to create domain insertion libraries from which protein switches can be identified by selections and/or screens.

Key words Protein switch, Domain insertion, Circular permutation, Directed evolution

1 Introduction

Biological systems are often described as complex circuits consisting of an interacting network of molecules. A key component of these networks are protein switches that couple cellular functions. These switches change their active state (output) in response to a physical signal or binding event (input). The ability to create switches by linking any desired input and output domains would enable the rewiring of cellular circuitry to suit a researcher’s objective. This rewiring has numerous potential applications such as cancer therapeutics [1], fluorescent biosensors [2–4], biomass degradation [5], and regulators of cell signaling elements [6].

Our directed evolution strategy for constructing protein switches involves fusion by domain insertion between two different proteins with the prerequisite input and output functions such that the behavior of the output domain is responsive to signal detection (e.g., a binding event) in the input domain (see Fig. 1). The exact fusion to construct to create a switch can be difficult to predict. Thus, a directed evolution approach is preferred in which an insert gene is randomly inserted into an acceptor gene, to create a library of gene fusions encoding fusion proteins. The insert gene
can encode for the input or output domain. These fusion proteins are then subjected to selections and/or screens to find a chimeric protein that possesses signal-dependent behavior. Using this strategy our lab has produced several examples of ligand-activated protein switches [1, 5, 7–18]. In one example, we created a ligand-activated enzyme with a 600-fold change in enzyme activity in response ligand binding [7].

One of the major steps in constructing protein switches by directed evolution is creating diversity in the ways you fuse the two domains. Random insertion libraries can be created using DNase I or S1 nuclease to generate a single double-stranded break throughout the plasmid that encodes the acceptor gene [7, 10, 12] (see Fig. 2a). However, some drawbacks of using these nuclease-digestion methods include the generation of insertions outside the gene-coding sequence and out-of-frame or inverted insertions, resulting in a significant fraction of the library encoding undesired and nonfunctional library members. Therefore, this strategy is best combined with a robust screening/selection system. In comparison to these nuclease methods, multiplex inverse PCR affords much greater control over the library composition. In inverse PCR, abutting primers are designed to “open up” a plasmid to facilitate domain insertion (see Fig. 2b). A separate inverse PCR reaction is performed for each desired insertion site, each of which requires a unique set of primers (see Fig. 2d). One feature of this method is the ability to use it for both random domain insertion (i.e., inverse PCR at each codon) or for insertions at sites chosen by rational design (i.e., achieved by inverse PCR only at designed positions).

Another layer of diversity can be created by circular permutation of the insert gene, thus increasing the overall diversity of the protein switch library. Conceptually, circular permutation is changing the intramolecular order of amino acids. Circular permutation

Fig. 1 Schematic depiction of the creation of protein switches by domain insertion. A protein switch is a fusion of two domains (by domain insertion) in such a way that the activity of the output domain is regulated by the input domain’s recognition of an input signal. (a) DNA sequences are depicted as lines and (b) their corresponding proteins as geometric shapes. A light gray color of the output domain indicates that the domain is inactive or less active. The signal that modulates the switch is depicted as a black triangle.
is performed such that the original N- and C-termini of the protein is linked by peptide linker of appropriate length and “opened up” at a different point in the gene to yield new N- and C-termini at which the insert gene can be inserted into the acceptor gene. Previous studies suggest that the sequence space accessed by combining circular permutation with random domain insertion is rich in active switches with very large differences in activity between their “on” and “off” states [7, 14]. Circular permutation libraries can be created by PCR on a plasmid containing a tandem repeat of the gene joined by DNA encoding the linker between the N- and C-termini (see Fig. 2c). The tandem repeat lacks a stop codon in the first repeat.

The PCR reactions for preparing the acceptor insertion sites (see Fig. 2b) and preparing the circularly permuted insert (see Fig. 2c) offer an opportunity for two types of additional diversity. First, forward and reverse primers can be combined in different combinations to delete or duplicate a codon (see Fig. 2d). Second, DNA encoding linkers of different lengths and compositions (both variable...
linkers and linkers of specific sequences) can be appended to the 5′ end of the primers for the inverse PCR reactions. These linkers between the two domains can be important for the switching property [5, 15, 16]; therefore, variable linkers and fixed composition linkers are often tested.

## 2 Materials

1. QIAPrep Spin Miniprep Kit (Qiagen).
2. Invitrogen PureLink Gel Extraction Kit (Invitrogen).
3. DNA Clean and Concentrator Kit (Zymo Research Corp., Irvine, CA).
4. DNase I (Promega).
5. 0.5 M Tris–HCl pH 7.5.
6. 50% glycerol.
7. 10 mM MnCl₂
8. 20 mg/mL BSA solution (100× stock, e.g., New England Biolabs).
9. 1 M EDTA.
10. T4 DNA Polymerase (e.g., New England Biolabs, Ipswitch, MA).
11. T4 DNA Ligase (2,000,000 U/mL) (e.g., New England Biolabs, Ipswitch, MA).
12. 10 mM dNTPs each.
13. Antarctic Phosphatase (e.g., New England Biolabs, Ipswitch, MA).
14. S1 Nuclease (e.g., Promega, Madison, WI).
15. 10× S1 Nuclease buffer (500 mM sodium acetate, 2800 mM NaCl, 45 mM ZnSO₄, pH 4.5).
16. Phusion High-Fidelity PCR Master Mix.
17. 1× TAE buffer: 40 mM Tris-base, 20 mM Acetic Acid, 1 mM EDTA, pH 7.6.
18. Nanodrop.
19. Gene Pulser Electroporation Cuvettes (Biorad, Hercules, California).
20. SOC medium: 2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM Glucose.
21. 245×245 mm Square Bioassay Dish (Corning, New York, NY).
22. LB Agar: 10% Tryptone, 5% Yeast Extract, 10% NaCl, and 15 g Bacto Agar per 1 L of medium.
3 Methods

3.1 Preparing Acceptor Vector Using DNase I

1. Prepare 100 μg of plasmid DNA containing the gene encoding the acceptor domain (see Note 1).

2. Prepare a 50 μL working solution of DNase I (1 U/mL) in a 0.5 mL microcentrifuge tube by combining 50 U DNase I, 25 mM Tris–HCl pH 7.5 and 50% glycerol. Store at −20 °C (see Note 2).

3. Prepare 1.2 mL of diluent solution in a 1.5 mL microcentrifuge tube with a final concentration of 50 mM Tris–HCl pH 7.5, 1 mM MnCl₂ and 0.5× of BSA. Vortex to mix and store at room temperature (see Note 3).

4. Before creating a library, determine the concentration of DNase I that will produce the most DNA vectors with a single double-stranded break. To determine this optimum concentration, combine 20 μg of plasmid DNA and 380 μL of the diluent solution in a 1.5 mL microcentrifuge tube. Mix and centrifuge briefly to collect all the liquid at the bottom of the tube. Aliquot 95 μL each into four 0.5 mL microcentrifuge tubes and incubate these tubes at 22 °C for 10 min.

5. After the incubation, add 5 μL of diluted DNase I to each tube at different concentrations (i.e., 1, 2, 4, and 8 mU) and incubate at 22 °C for 8 min.

6. To inactivate the DNase I, add 2.4 μL of 0.5 M EDTA to each tube and incubate at 75 °C for 10 min.

7. Purify each reaction using the Zymo DNA Clean and Concentrator kit or a PCR purification kit column. Elute with water.

8. Run 5–10 μL of each sample on a 0.8% agarose gel made with 1× TAE buffer to determine the concentration of DNase I that yields the best digestion results (see Notes 4 and 5).

9. Once the optimal DNase I dilution is determined, prepare 80 μg plasmid DNA in 1520 mL of diluent solution and aliquot 95 μL into 16 microtubes (i.e., 5 μg DNA per tube). Incubate for 10 min at 22 °C.

10. Add 5 μL of the correct DNase dilution to each tube in 30 s intervals (it will take 8 min to add DNase I to all tubes). After 8 min, stop the reaction as described in step 6 (see Note 6).

11. Combine all the reactions into a 15 mL Nalgene test tube and purify the DNA using four 25-μg columns from a Zymo DNA Clean and Concentrator Kit. Elute the DNA from each column in two 20 μL volumes of water pre-warmed to 65 °C. Combine the eluate from each spin column (160 μL total) (see Note 7).
12. Determine the DNA concentration using a NanoDrop and run 100 ng of DNA in an agarose gel to estimate the percent of linear DNA using a gel-imaging software (see Note 4).

13. Once the concentration is known, proceed to the repair step (see Note 8). Prepare multiple reactions consisting of DNA, T4 DNA ligase (160 cohesive end units/μg linear DNA), T4 DNA Polymerase (1 unit/μg linear DNA), 200 μM dNTPs, 1× T4 DNA Ligase buffer, and 1× BSA in water. Each reaction should be ≤50 μL. Incubate at 12 °C for 20 min.

14. Stop the reaction with 10 mM EDTA (final concentration) followed by incubation at 75 °C for 10 min.

15. Purify the DNA using a 25-μg Zymo DNA Clean and Concentrator Kit.

16. Separate the linear DNA from the supercoiled and nicked DNA via DNA gel electrophoresis. Run the gel at 90 V on a 0.8% agarose gel made with 1× TAE buffer for approximately 1.5 h or until the DNA is separated into three distinct bands (see Note 9).

17. Excise the band containing the linear DNA and purify using Invitrogen’s Pure Link Gel Extraction Kit according the manufacturer’s instructions.

18. Dephosphorylate the repaired linear plasmid using Antarctic Phosphatase. For every microgram of DNA, add 20 units of enzyme. Incubate for 1 h at 37 °C.

19. Heat inactivate the enzyme by incubating the reaction at 65 °C for 10 min.

3.2 Preparing Acceptor Vector Using S1 Nuclease

1. Prepare 50 μg of plasmid DNA containing the gene encoding the acceptor domain (see Note 1).

2. In a 1.5 mL microcentrifuge tube, add 26 μg of plasmid DNA, 1× S1 Nuclease buffer, and 125 units of S1 Nuclease to a final volume of 325 μL. Aliquot 25 μL each into 13 microtubes (i.e., 2 μg DNA per tube). Incubate at 37 °C for 20 min.

3. Combine all the reactions and purify using two 25-μg spin columns of a Zymo DNA Clean and Concentrator Kit. Elute the DNA from each column into two 20 μL aliquots of water prewarmed to 65 °C. Combine the eluate from each spin column (80 μL total) (see Note 7).

4. Determine the DNA concentration using a NanoDrop and electrophorese 100 ng of sample to estimate the percentage of linear DNA using gel-imaging software (see Note 4).

5. Once the concentration is known, proceed to the repair, gel purification, and dephosphorylation steps (see Subheading 3.1, steps 13–19).
1. Design one set of primers for each insertion site (see Fig. 2c and Note 10).

2. Purify the plasmid using the Qiagen Plasmid Miniprep Kit following the manufacturer’s instructions.

3. Measure the concentration of the DNA using a NanoDrop (see Note 11).

4. Dilute the template DNA to 10 ng/μL in DNase-free water. You will need 1 μL per PCR reaction.

5. Prepare 10 μM primer mixes in a 96-well plate (one primer mix for each primer pair). To each well, add 10 μL of 100 μM forward primer, 10 μL of 100 μM reverse primer, and 80 μL of water to bring the primer mix to 100 μL total (see Note 12).

6. Prepare a dilute DNA solution by mixing the template DNA and DNase-free water in a 1.5 mL microcentrifuge tube. For every PCR reaction, add 8 μL of water and 1 μL of the diluted DNA stock solution. A separate PCR reaction is required for each primer mix.

7. To each well of a 96-well PCR plate, add 9 μL of the diluted DNA solution, 1 μL of primer mix, and 10 μL of Phusion 2× Master Mix. Decrease the volume setting on the pipette and mix by pipetting up and down several times (see Note 13).

8. Centrifuge for 30 s at 720×g to collect all reaction components at the bottom of the wells.

9. Use the following guidelines to design your amplification program (see Note 14):
   - 1 cycle: 98 °C for 2.5 min.
   - 25 cycles: (1) 98 °C for 30 s, (2) T_opt for 30 s, and (3) 72 °C for 30 s/kilobases of your vector.
   - 1 cycle: 72 °C for 10 min.

10. Analyze 5 μL from 20 or more random samples using DNA gel electrophoresis with a 0.8% agarose gel at 90 V for 40 min to confirm successful amplification of most reactions.

11. After confirmation, pool together 5 μL from each PCR reaction. Using DNA electrophoresis and a 0.8% agarose gel, electrophorese your sample at 90 V for 1 h to allow separation of the template DNA and PCR product.

12. Excise the PCR product from the gel and purify using Invitrogen’s PureLink Gel Extraction Kit.

13. Check the concentration and quality of the DNA using a Nanodrop. If the 260/280 and 260/230 values are too low, purify the DNA sample using the Zymo Clean and Concentrate Kit and elute in 20 μL of DNase-free water (see Note 11).
3.4 Preparing Circular Permuted Insert

1. Construct a vector as shown in Fig. 2c by duplicating the insert gene by end-to-end fusion, connecting them with a linker (see Notes 15–17).

2. Design a forward primer starting at the first base of the selected codon and a reverse primer starting at the last base of the previous codon (the melting temperatures should be between 60 and 65 °C (see Notes 10 and 18–20).

3. The PCR protocol is the same as used for multiplex inverse PCR (see Subheading 3.3, steps 2–13).

3.5 Ligation of the Insert and Acceptor DNA

1. Starting with 200 ng of prepared vector, calculate the pmol of vector using the following formula:

\[ \text{pmol} = \left( \frac{\text{ng of DNA}}{\text{kilobase pairs of vector} \times 650 \text{ Da}} \right) \]

2. Using the same formula, calculate the pmol of insert required to achieve a 3:1 or 5:1 insert-to-vector molar ratio.

3. Add the calculated amount of prepared vector and insert to a thin-walled PCR tube. Add 2 μL of 10× T4 DNA Ligase Buffer and fill to 19 μL. Add 1 μL of T4 DNA Ligase, flick the tube to mix and centrifuge briefly to collect the reaction components at the bottom of the tube (see Notes 21 and 22).

4. Place the ligation reaction in a thermocycler. Incubate for 12–18 h cycling between: 25 °C for 30 s and 10 °C for 30 s.

5. After the incubation, purify the DNA using the Zymo Clean & Concentrator Kit. Elute the DNA into a 1.5 mL microcentrifuge tube in 10 μL of water. Concentrate your sample to approximately 4 μL using a vacufuge at room temperature (see Note 23).

3.6 Transformation

1. Turn on the Gene Pulser. Adjust the voltage to 2.5 kV.

2. Place the following on ice: high-efficiency (>10^8 cfu/μg DNA) electrocompetent cells, a 0.2 cm Gene Pulser Cuvette, a 0.5 mL microcentrifuge tube and your purified and concentrated ligation (see Note 24).

3. Once the cells are thawed, aliquot 40 μL of cells into the microcentrifuge tube with the ligation, ensuring the cells engulf the DNA sample. Mix well.

4. Pipette the cells into the cuvette (see Note 25).

5. Place the cuvette in the Gene Pulser and slide into the chamber, ensuring that the cuvette clicks into place.

6. Simultaneously hold down the pulse buttons until you hear the beep. Then, release the buttons.

7. Quickly, add 1 mL of SOC media to the cuvette.

8. Transfer the cell/SOC mix to a 1.5 mL microcentrifuge tube.
9. Incubate the sample at 37 °C for 1 h.
10. Plate dilutions of the cells on 10 cm dishes and plate the remainder on 245 × 245 mm plate bioassay dishes with LB agar (see Note 26).
11. Incubate the plates at 37 °C for 12–18 h.
12. The next day, count the number of colonies on the 10 cm dishes to determine the number of transformants (see Note 27).
13. You can further confirm the diversity of this “naive” domain insertion library by performing colony PCR on several transformants and/or growing 10 mL LB cultures of samples, purifying their DNA and analyzing the DNA sequence (see Note 28).

4 Notes

1. We have found that the Qiagen Miniprep kits have better yield and less genomic DNA contamination compared to the Midiprep or Maxiprep kits.
2. This working solution is good for at least 3 months.
3. If chilled on ice, solids may fall out of solution. If this happens, a fresh diluent must be prepared.
4. The slowest migrating band is nicked open plasmid. The fastest migrating band is supercoiled DNA and the band in the middle is linear DNA.
5. The optimal DNase I concentration is the one that produces a sharp band of linearized plasmid. To achieve this, 30–50% of the plasmid usually remains supercoiled. Overdigestion of the plasmid will result in smearing indicating deletions at the ends of the digestion products.
6. The 30 s staggered intervals are to insure that each tube will be stopped at exactly 8 min. We recommend that experiments be performed on the same day as the optimization. If the reactions will be completed on a different day, prepare fresh DNase dilutions.
7. The prepared DNA can be stored at −20 °C.
8. This step creates blunt ends by filling in the end or by chewing away any overhanging DNA. Note the amount of T4 DNA Ligase and T4 DNA Polymerase should be calculated based on the amount of linear DNA.
9. Do not load more than 2.5 μg per lane (of the medium-sized lanes). If possible, do not use UV light or ethidium bromide as these can cause DNA damage.
10. Your primers should adhere to the following guidelines as much as possible: (1) 40–60% GC, (2) at least 18 nucleotides
in length, (3) the last two nucleotides on the 3′ each of primer should be Cs or Gs, and (4) there should be no more than 3 Cs or Gs in the final five nucleotides of the 3′ end of the primer. All primers should be designed such that they have approximately the same melting temperature (ideally 60–65 °C). The 5′ end of the forward primer should begin at the codon downstream of the insertion site. The 5′ end of the reverse primer should begin at last nucleotide of the codon that is upstream of the insertion site. For example, if you want to insert a gene between codons 2 and 3 of the acceptor DNA, the 5′ end of your forward primer should begin at the first nucleotide in codon 3 and the 5′ end of your reverse primer should begin at the last nucleotide in codon 2. This will create an in-frame opening between codons 2 and 3 (see Fig. 2d). You can also duplicate and delete codons by shifting the primers upstream or downstream (see Fig. 2d). We have written a MATLAB script to design such primer pairs, which is freely available. After designing the primers, use a PCR simulation software such as AmpliTaq to determine the optimal annealing temperature ($T_{opt}$) for a few of the primer pairs. Use the lowest $T_{opt}$ value of the primer pairs. Also use the software to check for the likelihood of nonspecific primer annealing to the template or the formation of primer dimers. Alternatively, you can subtract 3–4 °C from the average melting temperature of all the primers as a good estimation for $T_{opt}$.

11. Check the quality of your DNA. A reading with 260/280 value of 1.8–2.0 and a 260/230 value of 1.8–2.3 will yield the best PCR results.

12. You can scale down the volume of the primer mix if necessary.

13. This process can be facilitated by using 12-channel basins and 12-channel pipettes.

14. You can minimize the formation of undesired PCR product by performing touch-down PCR using the following amplification program: (i) Preheat the thermocycler to 98 °C, (ii) 1 cycle: 98°C for 3 min, (iii) 15 cycles: (1) 98 °C for 30 s, (2) $T_{opt}$ + 15 °C for 30 s, and (3) 72 °C for 30 s/kg of GOI. The annealing temperature should decrease by 1 °C for each cycle, (iv) 10 cycles: (1) 98 °C for 30 s, (2) $T_{opt}$ for 30 s, and (3) 72 °C for 30 s/kg of GOI, (v) 72 °C for 10 min.

15. Circular permutation aims to change the position of N- and C-termini of the protein without disturbing its three-dimensional structure. In general, the N- and C-terminals of the original protein should be close enough to allow their union (<10 Å) [19].

16. When the gene (without a stop codon) is already available in cloning vector and a convenient restriction site is available at
the end of that gene, you can PCR amplify the second copy of
gene (with a stop codon) to add a suitable linker and directly
ligate this amplified gene at the 3′ end of the first gene copy
using the restriction site. When no restriction site is available,
the genes can be fused together by overlap PCR.

17. Linker connecting the original N- and C-termini: Linker
length and composition can interfere with the folding, topology,
and stability of the protein, so it must be rationally designed.
Here are some considerations. (a) Length—the distance
between the alpha carbons of the N- and C-termini residues
can be determined using Pymol or similar molecular visualiza-
tion software. Depending on the secondary structure of the
linker residues, one amino acid typically spans a distance of
1.5–3.5 Å. Linkers that bridge distances greater than 10 Å
require more careful design. Depending on the situation, the
protein’s surface properties must be taken into account. (b)
Composition—Flexible backbones can help accommodate
emerging conformational strains; thus, in most cases, glycine-
rich (Gly) bridges are preferred. Typically, a GlySerGlyGly
linker is sufficient to span a distance of 7–8 Å. However, linkers
of more diverse composition can be tested.

18. Inter-domain linker: In order to alleviate possible disturbances
cau sed by domain insertion, we usually add two residues to the
new termini of the permuted protein. This can be done using
degenerate primers with two codons of 5′-NNK-3′ at 5′ end of
the forward and reverse primers. Note that the reverse primer
must encode the reverse complement of 5′-(NNK)2-3′.

19. Reducing library size: For large proteins, purchasing compre-
hensive sets of primers can be expensive. To decrease primer
costs and the library size, you can create select circular permuta-
tions that focus on residues that are solvent accessible, flexible,
loosely packed, and between secondary structure elements.

20. Phosphorylated primers can be ordered from most manufac-
turers or unphosphorylated primers can be phosphorylated
using T4 Polynucleotide Kinase (New England Biolabs,
Ipswich, MA) following the manufacturer’s instructions.
Alternatively, you can phosphorylate the PCR product.

21. For the best results, do not exceed 10 ng of vector per µL of the
ligation reaction. Note that the ligase used is high concentra-
tion ligase (2,000,000 units/mL) for this blunt end ligation
reaction.

22. The ligation reaction can be scaled up. For the best results,
prepare a master mix of the ligation and split it into 20 µL ali-
quot s for the incubation.

23. Alternatively, you can elute in 6 µL of DNA but this may
decrease your yield.
24. You can use any high-efficiency competent cells that suit your system. We recommend NEB 5-alpha cells.

25. Check to make sure the cell/DNA sample reaches the bottom of the cuvette, that it does not have any bubbles and that it is making contact with both electrodes within the cuvette.

26. Make one large 245 × 245 mm plate (approximately 250 mL of LB agar) and at least three 10 cm petri dishes of LB agar (approximately 25 mL of LB agar/plate).

27. If you used multiplex inverse PCR to prepare your vector, you want the number of transformants to be >5 times the number of possible variants to have a high probability that the library contains all possible variants (assuming that each library member is expected to appear at the same frequency). For an in-depth discussion of this topic, see Bosley and Ostermeier [20].

28. Both DNase I and S1 Nuclease generate libraries with deletions (and occasionally duplications) distributed along the acceptor sequence. These deletions contribute to the sequence variability of the library, potentially generating relevant diversity for the creation of new properties in the chimeric proteins [11, 19]. Library construction using S1 is usually easier and results in smaller deletions since S1 nuclease digestion of plasmid DNA is halted after the first double-stranded break [10]. However, S1 digestion can be heavily biased to occur in inverted repeats regions in the plasmid, if they are present [8].

References


Part II

Peptide Switches
Catalytic Amyloid Fibrils That Bind Copper to Activate Oxygen

Alex Sternisha and Olga Makhlynets

Abstract

Amyloid-like fibrils assembled from de novo designed peptides lock ligands in a conformation optimal for metal binding and catalysis in a manner similar to how metalloenzymes provide proper coordination environment through fold. These supramolecular assemblies efficiently catalyze $p$-nitrophenyl ester hydrolysis in the presence of zinc and phenol oxidation by dioxygen in the presence of copper. The resulting heterogeneous catalysts are inherently switchable, as addition and removal of the metal ions turns the catalytic activity on and off, respectively. The ease of peptide preparation and self-assembly makes amyloid-like fibrils an attractive platform for developing catalysts for a broad range of chemical reactions. Here, we present a detailed protocol for the preparation of copper-containing fibrils and for kinetic characterization of their abilities to oxidize phenols.

Key words Peptides, Fibrils, Catalysis, Dioxygen activation, Phenol oxidation

1 Introduction

Metalloenzymes utilize metal ions to catalyze some of the most challenging chemical reactions, such as methane oxidation, lignin degradation, and nitrogen fixation [1–5]. Even more impressive is the fact that metalloenzymes are capable of performing these reactions in a very selective and stereospecific manner. The high efficiency and selectivity of natural metalloenzymes has inspired much effort to (re)design catalytic metalloproteins to improve selectivity for a substrate of choice and to catalyze novel chemical transformations [6]. Many of the current state-of-the-art catalyst design strategies take advantage of both existing and de novo designed proteins to optimize the metal coordination sphere and to provide a binding site for the substrate. Recently, this approach was taken to the next level by employing self-assembly of short peptides into fibrils. The fibrils provide stability and the appropriate coordination sphere for the metal ions necessary to facilitate chemical transformations [7–9]. Amyloid-like fibrils, assembled from short
(7-residues or less) de novo designed peptides, catalyze ester hydrolysis in the presence of Zn$^{2+}$. The most catalytically active peptide, Ac-IHIHIQI-CONH$_2$, was identified by varying hydrophobic core residues and the sequence of hydrophilic portion of the peptide. This peptide self-assembles into fibrils and binds zinc, the resulting material catalyzes hydrolysis of $p$-nitrophenyl esters with efficiency comparable to that of the natural enzymes by weight [7]. Peptides that do not form fibrils, but have similar sequences otherwise, showed significantly lower catalytic activity. The correlation between the propensity to form fibrils and the ability to catalyze chemical reactions suggests that the rigid fibril arrangement of functional groups in peptide assemblies locks the metal ion in a coordination environment that promotes catalysis. The strategy of using peptide self-assembly for design of catalysts can be further expanded to other chemical transformations (Fig. 1). We recently applied this design strategy to create a catalyst for oxygen activation [10]. We focused on Cu$^{2+}$ as a metal cofactor, because it is known to facilitate various redox reactions (e.g., C-H oxidation [11], epoxidation, etc. [12]), and is stable in an aqueous environment at

**Fig. 1** Formation of fibrils creates a coordination sphere optimal for metal ion binding (Cu$^{2+}$ or Zn$^{2+}$). These amyloid assemblies can catalyze ester hydrolysis and DMP oxidation.
neutral pH. We investigated the ability of self-assembling supramolecular catalysts to catalyze oxidative coupling of 2,6-dimethoxyphenol (DMP), a well-characterized reaction that produces a bright orange product [13, 14]. Several highly reactive peptides that are capable to efficiently oxidize DMP using dioxygen as a cofactor were identified. Just like in the case of zinc-binding supramolecular catalysts of ester hydrolysis, the oxygen-activating ability strongly correlated with the propensity of peptides to form a β-sheet structure.

Catalytic efficiency of the self-assembling peptides is impressive given the simplicity and adaptability of the design. A large number of short peptides can be screened in a high-throughput fashion. Additionally, mixing of different peptides prior to fibril formation provides access to a large number of different arrangements of functional groups in a single amyloid-like supramolecular assembly. Feasibility of this idea has been previously demonstrated: mixing peptides with different residues at position 6 produced fibrils more active than fibrils assembled from uniform peptides [7]. Simple variation of side chains in the hydrophilic positions of the peptide sequence allows for exploration of many different coordination arrangements that efficiently support the metal ion in all oxidation states required for catalysis [15]. Enzymes selectively recognize their substrates through multiple contacts with the protein’s side chains. This property could be potentially engineered into catalytic fibrils.

Here, we provide a detailed protocol for robust preparation and kinetic characterization of supramolecular peptide assemblies that bind Cu²⁺ and catalyze phenol oxidation. The assay we developed is based on the benchmarked reaction of DMP oxidation, often used to assess activity of laccase enzymes [16–19]. This assay provides a simple readout signal, is easy to perform, and is suitable for high-throughput screening in 96-well plates. We also present an optimized protocol for the preparation of heteropeptidic assemblies, where peptides with different primary sequences are incorporated into a single fibril.

2 Materials

All solutions need to be prepared using MilliQ water, HPLC grade organic solvents, and analytical grade reagents.

1. 1 M Hepes-KOH buffer, pH 8: Weigh out 9.53 g of Hepes free acid in a 50 mL Falcon tube. Add 30 mL of water to dissolve Hepes-free acid and adjust pH using 5 M KOH, and then add water to 40 mL (see Note 1).

2. 50 mM CuSO₄ solution in water: Weigh out 0.5 g of CuSO₄·5H₂O and add water to 40 mL. The final concentration
of this solution should be confirmed using ICP or by measuring absorbance of Cu-Zincon complex (see Note 2).

3. 10 mM HCl (see Note 3).

4. 80 mM 2,6-dimethoxyphenol (DMP): weigh out 20–30 mg of DMP solid is a microcentrifuge tube, add 1 mL of isopropanol, and invert the tube to dissolve the solid. Further dilute the resulting solution using isopropanol to a final concentration of 80 mM (see Note 4).

5. 8 M urea solution in water: weigh out 2.4 g of urea, then add water to 5 mL.

6. Isopropanol.

7. 96-well plates.

8. Multichannel pipette.

9. UV-Vis spectrophotometer.

10. Plate reader.

3 Methods

Peptides were synthesized by manual Fmoc solid-phase synthesis at elevated temperature using Rink Amide or Wang resin and Fmoc-protected amino acids using a previously reported protocol [20]. The crude peptides were precipitated and washed with cold methyl-tert-butyl ether, and purified using reverse phase High Performance Liquid Chromatography. The identities of the peptides were confirmed using MALDI-TOF mass spectrometer and the purity was evaluated on analytical HPLC instrument. Pure (>90%) lyophilized peptides were stored at −20 °C (see Note 5).

3.1 Preparation of Peptide Stock Solution at pH 2

1. Calculate the extinction coefficient of peptide at 214 nm (see Note 6).

2. Prepare a 1 mM stock solution of peptide in 10 mM HCl. Using the Beer-Lambert law, calculate the concentration of the peptide that will give an absorbance at 214 nm between 0.2 and 1. Dissolve a small amount of solid peptide in 1 mL of 10 mM HCl, dilute the stock of peptide using 10 mM HCl (measure at least 10 μL of peptide, perform serial dilutions for peptides with high $\varepsilon_{214}$), and measure the absorbance. If the concentration is above 1 mM, dilute the stock using 10 mM HCl. If the concentration is less than 1 mM, add more solid peptide. Some peptides do not dissolve completely. If this is the case, centrifuge the solution at 10000 × g for 5 min, carefully remove the supernatant, and then measure concentration. Once the stock is at 1 mM, perform dilutions and retake absorbance measurement at least one more time to ensure reproducibility.
1. Determine the path length in the 96-well plate for a volume of 200 μL (see Note 7).

2. Prepare two working buffers: (1) 25 mM Hepes-KOH, pH 8—mix 0.5 mL of 1 M Hepes-KOH and 19.5 mL of water in a 50 mL Falcon tube; (2) 28.4 mM Hepes-KOH, pH 8—mix 0.568 mL of 1 M Hepes-KOH and 19.4 mL of water.

3. Dilute 50 mM CuSO₄ into water to 10 mM and then further dilute the solution tenfold to make 1 mM stock of CuSO₄ in water (see Note 8).

4. Prepare 2 mL of 2 mM DMP solution in water: mix 1.825 mL of water, 125 μL of isopropanol, 50 μL of 80 mM DMP.

5. Prepare a mixture of peptide and Cu²⁺ in buffer: mix 40 μL of 1 mM peptide stock in 10 mM HCl and 20 μL of 1 mM Cu²⁺ in water, then add 440 μL of 28.4 mM Hepes-KOH buffer, pH 8 (see Note 9). This will give 500 μL of solution with 40 μM Cu²⁺ and 80 μM peptide. For a blank solution, mix 40 μL of 10 mM HCl, 20 μL of 1 mM Cu²⁺ in water, and 440 μL of 28.4 mM Hepes-KOH buffer, pH 8.

6. Fill reaction reservoir (trough) with 25 mM Hepes-KOH buffer, pH 8. Using a 12-channel pipette dispense 100 μL of this buffer into each well to be used of the 96-well plate.

7. Add 50 μL of Cu-peptide solution (40 μM Cu²⁺ and 80 μM peptide) or blank. Each sample should be analyzed at least three times (use three wells).

8. Set up an experiment on the plate reader to follow the absorbance at 476 nm, taking measurements every 10 s for 5 min.

9. Fill another trough with 2 mM DMP. Using a 12-channel pipette deliver 50 μL of 2 mM DMP into each well. Gently pipette up and down four times to mix (see Note 10).

10. Place the 96-well plate into the plate reader and start the measurements (see Note 11).

11. Plot the absorbance at 476 nm as a function of time (s) and determine the slope of the linear portion of each kinetic trace; this will give the change in absorbance per second. To calculate initial rate, divide the slope by the extinction coefficient of the DMP oxidation product (3,3′,5,5′-tetramethoxy biphenyl-4,4′-diol)—14800 M⁻¹ cm⁻¹ [16, 21]. The following formula illustrates how to calculate initial rate in units of μmol/min (1.35 is a coefficient for path length correction):

\[
\text{Rate (μM / min)} = \left( \text{slope} \times 1.35 \times 10^6 \times 60 \right) / 14,800
\]

12. Compare the initial rates of blank sample (Cu²⁺ in buffer) and Cu-peptide samples (see Note 12).
3.3 Mixing Peptides with Different Sequences

1. Prepare 4 mM peptide stocks in 10 mM HCl, as described in Subheading 3.1 (see Note 13).
2. Mix the two different peptide stock solutions (4 mM) in different ratios to a final volume of 40 μL.
3. Add 150 μL of 8 M urea to each sample, mix and incubate at room temperature for 15 min (see Note 14).
4. Add 40 μL of 2 mM Cu²⁺ in water.
5. Initiate fibril formation by adding 1.77 mL of 28.4 mM Hepes-KOH buffer, pH 8. Final concentrations of peptide and Cu²⁺ are 80 and 40 μM, respectively (see Note 15).
6. For the blank sample, mix 40 μL of 10 mM HCl, 150 μL of 8 M urea, 40 μL of 2 mM Cu²⁺, and 1.77 mL of 28.4 mM Hepes-KOH buffer, pH 8.
7. Set up kinetic assay as described in Subheading 3.2, steps 6–10.

4 Notes

1. Keep the solution of 1 M Hepes-KOH at 4 °C, warm up to room temperature before use. Tris buffer coordinates to Cu²⁺ and therefore should not be used in this assay. For screening at different pH, we also used MES (pH 6) and TAPS (pH > 8).
2. Atomic absorption standard for Cu²⁺ was initially used for screening. However, this solution (1 g/L = 15.73 mM) contains 3% nitric acid and slightly reduces the final pH of samples (pH 7.9 instead of pH 8). In addition, we found that the standard (Ricca) had high absorbance at λ below 220 nm (even when Cu²⁺ was diluted to 10 μM) and we avoided using it as a source of Cu²⁺ when measuring concentration of Cu-peptide or when preparing samples for circular dichroism analysis. To measure Cu²⁺ concentration in CuSO₄ stock, solution was diluted to 40 μM into 25 mM Hepes-KOH pH 8 and then Zincon in water (2 mM stock) was added to a final concentration of 50 μM. We measured the extinction coefficient for Cu-Zincon complex to be ε₆₂⁰ = 18145 M⁻¹ cm⁻¹ and dissociation constant K_d = 1.7 μM. The complex of Cu-Zincon has a maximum absorbance at 600 nm; however, ε₆₂⁰ was used to minimize contribution from unbound Zincon (λ_max = 470 nm).
3. First, dilute concentrated HCl (12 M, ACS grade) to make a 1 M HCl solution.
4. DMP solution (80 mM) in isopropanol is stable for at least 1 week at room temperature.
5. Peptides that fibrillate quickly need to be lyophilized as soon as HPLC fractions are collected and samples are prepared for subsequent analytical HPLC and MALDI-TOF analysis. Lyophilized peptides can be stored for up to a year at −20 °C with containers tightly sealed. Strongly aggregating peptides tend to become less soluble over time and their catalytic activity decreases.

6. Calculate extinction coefficients (M$^{-1}$ cm$^{-1}$) of peptides at 214 nm using the following equation [22], where $\varepsilon_{\text{peptide\_bond}}$ is 923 M$^{-1}$ cm$^{-1}$ and $\varepsilon_{\text{amino\_acid}(i)}$ are contributions from side chains (Table 1). For example, for a 7-residue peptide with an acyl cap (Ac) on the N-terminus and an amide on the C-terminus, the number of peptide bonds is 7 to account for the contribution of an amide cap.

$$\varepsilon_{\text{peptide}} = \varepsilon_{\text{peptide\_bond}} * \eta_{\text{peptide\_bond}} + \sum_{i=1}^{20} \varepsilon_{\text{amino\_acid}(i)} * \eta_{\text{amino\_acid}(i)}$$

7. Prepare colored solution and measure UV-vis spectrum in a cuvette with 1 cm path length. Pipette 200 μL of the same solution and measure absorbance at the wavelength of choice using plate reader. Calculate correction coefficient using the following equation:

$$\text{Correction coefficient} = \frac{\text{Absorbance measured by UV-vis (1 cm)}}{\text{Absorbance measured by plate reader}}$$

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>$\varepsilon_{214}$, M$^{-1}$ cm$^{-1}$</th>
<th>Amino acid</th>
<th>$\varepsilon_{214}$, M$^{-1}$ cm$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine (A)</td>
<td>32</td>
<td>Leucine (L)</td>
<td>45</td>
</tr>
<tr>
<td>Arginine (R)</td>
<td>102</td>
<td>Lysine (K)</td>
<td>41</td>
</tr>
<tr>
<td>Asparagine (N)</td>
<td>136</td>
<td>Methionine (M)</td>
<td>980</td>
</tr>
<tr>
<td>Aspartic acid (D)</td>
<td>58</td>
<td>Phenylalanine (F)</td>
<td>5200</td>
</tr>
<tr>
<td>Cysteine (C)</td>
<td>225</td>
<td>Proline (P)</td>
<td>2675</td>
</tr>
<tr>
<td>Glutamine (Q)</td>
<td>142</td>
<td>Serine (S)</td>
<td>34</td>
</tr>
<tr>
<td>Glutamic acid (E)</td>
<td>78</td>
<td>Threonine (T)</td>
<td>41</td>
</tr>
<tr>
<td>Glycine (G)</td>
<td>21</td>
<td>Tryptophan (W)</td>
<td>29050</td>
</tr>
<tr>
<td>Histidine (H)</td>
<td>5125</td>
<td>Tyrosine (Y)</td>
<td>5375</td>
</tr>
<tr>
<td>Isoleucine (I)</td>
<td>45</td>
<td>Valine (V)</td>
<td>43</td>
</tr>
</tbody>
</table>
Different plate readers could have different correction coefficients even when the same plate and the same volume are used.

8. 50 mM solution of CuSO₄ in water is stable for at least a year; the 1 mM stock was prepared immediately before kinetic experiment.

9. It is important to mix the peptide and Cu²⁺ at a low pH, before fibrils are formed. First, mix peptide (pH 2 stock) and Cu²⁺ solution in water, then add Hepes-KOH buffer, pH 8. Using 28.4 mM Hepes-KOH will give a 25 mM final concentration of buffer in the sample. The change in pH from 2 to 8, and the presence of Cu²⁺, triggers fibril formation. Always make Cu-peptide samples in pH 8 buffer immediately before the experiment; we observed reduced catalytic activity over time due to fibrils binding to plasticware. Some peptides need prolonged incubation time to form fibrils. Optimal incubation time should be established for each peptide family. The easiest approach would be to screen for peptide activity using freshly prepared Cu-peptide samples, then incubate leftover sample for 24 h and repeat the screen.

10. Avoid formation of bubbles as they will interfere with absorbance measurements.

11. Final concentrations of the reagents in the well at the onset of the reaction are: 10 µM Cu²⁺, 20 µM peptide, 500 µM DMP, 2.2% isopropanol, 18.8 mM Hepes-KOH, 0.26 mM dioxygen (the solubility of dioxygen in air-saturated water at 25 °C and 760 Torr [23]). With DMP substrate we observed that after a minute, the kinetic trace at 476 nm starts to curve; therefore losing a few data points in the beginning of the reaction can have a significant effect on the calculated rate. If peptides are very active, it is recommended to reduce the concentration of DMP to 200 µM (final concentration) or lower.

12. Conditions described above use stoichiometric amount of Cu²⁺ needed to saturate all binding sites in fibrils. Dividing the rate calculated for Cu-peptide sample by the rate of the blank sample will show rate enhancement of the reaction in the presence of peptide.

13. Some peptides might not be soluble enough in 10 mM HCl to make 4 mM solution. In this case, peptides can be weighed out and then dissolved in 8 M urea to make 4 mM solution without measuring absorbance at 214 nm (urea absorbs at 214 nm).

14. Urea breaks fibrils and this step ensures that the two peptides mix properly. Without urea treatment, peptides in 10 mM HCl might exist as protofibrils and would not mix well enough to result in uniform fibril composition.

15. To initiate fibril formation, urea needs to be diluted at least 10-fold.
Acknowledgment

This work was supported in part by the NSF ADVANCE (HRD-1008643) fellowship to O.V.M.

References

oxidase CueO from *Escherichia coli*—deletion of the methionine-rich helical region covering the substrate-binding binding site. J Mol Biol 373:141–152


Part III

Fluorescent and Bioluminescent Sensors
Chapter 5

Ancestral Protein Reconstruction and Circular Permutation for Improving the Stability and Dynamic Range of FRET Sensors


Abstract

Small molecule biosensors based on Förster resonance energy transfer (FRET) enable small molecule signaling to be monitored with high spatial and temporal resolution in complex cellular environments. FRET sensors can be constructed by fusing a pair of fluorescent proteins to a suitable recognition domain, such as a member of the solute-binding protein (SBP) superfamily. However, naturally occurring SBPs may be unsuitable for incorporation into FRET sensors due to their low thermostability, which may preclude imaging under physiological conditions, or because the positions of their N- and C-termini may be suboptimal for fusion of fluorescent proteins, which may limit the dynamic range of the resulting sensors. Here, we show how these problems can be overcome using ancestral protein reconstruction and circular permutation. Ancestral protein reconstruction, used as a protein engineering strategy, leverages phylogenetic information to improve the thermostability of proteins, while circular permutation enables the termini of an SBP to be repositioned to maximize the dynamic range of the resulting FRET sensor. We also provide a protocol for cloning the engineered SBPs into FRET sensor constructs using Golden Gate assembly and discuss considerations for in situ characterization of the FRET sensors.

Key words Ancestral protein reconstruction, Phylogenetic analysis, Protein engineering, Thermostability, Circular permutation, Förster resonance energy transfer, Fluorescence, Biosensor

1 Introduction

Optical biosensors, including those that rely on Förster resonance energy transfer (FRET sensors), allow robust, noninvasive quantification of small molecule dynamics in biological systems [1]. FRET is a physical phenomenon whereby excitation of a donor fluorophore results in nonradiative energy transfer to an acceptor

*These authors contributed equally.
fluorophore in a distance-dependent manner, which results in fluorescence of the acceptor fluorophore. FRET sensors can be constructed by fusing a specific pair of fluorescent proteins to the termini of a suitable recognition domain, which undergoes a conformational change when it binds the small molecule of interest. This ligand-dependent conformational change alters the distance between the fluorescent proteins, causing an observable change in FRET efficiency, which can be observed as a change in the ratio of the fluorescence intensities of the donor and acceptor fluorophores. The dynamic range of the sensor is the maximum change in FRET efficiency between the unbound and ligand-bound states.

In addition to having a large dynamic range, an ideal sensor would be specific for the target ligand and responsive over the physiological concentration range of the ligand; importantly, the sensor must also be stable for extended periods under the experimental conditions (e.g., temperature and pH) required for the biological system of interest. These properties of the sensor are largely determined by the choice of recognition domain.

The solute-binding protein (SBP) superfamily is one set of recognition domains that is commonly used in FRET sensors for small molecules [2–5]. SBPs exhibit high affinity and specificity toward a diverse array of ligands, and they undergo a large conformational change upon ligand binding, which can be transduced into an optical signal in a FRET sensor construct. However, the low thermostability of existing SBPs can limit the utility of the resulting sensors in biological environments, especially when destabilizing modifications to the recognition domain, such as circular permutation or specificity-switching mutations, are necessary to improve the dynamic range or specificity of the sensor [6].

Ancestral protein reconstruction (APR) is one method that has been shown to produce consistently thermostable proteins, with denaturation temperatures up to 40 °C greater than comparable mesophilic proteins [7–11]. We have previously used this method to reconstruct a thermostable SBP as a recognition domain for a robust FRET sensor for L-arginine [6]. The main steps involved in APR are: (1) collection and alignment of a sequence dataset representative of a protein family; (2) inference of a phylogenetic tree describing the evolutionary relationships between the protein sequences; (3) probabilistic reconstruction of the ancestral protein sequences; (4) synthesis and cloning of genes encoding the ancestral proteins; and (5) expression, purification, and biophysical analysis of the ancestral proteins.

It has been argued that the high thermostability of reconstructed ancestral proteins reflects the environment of the ancient, thermophilic organisms from which they originate [7, 8, 12]; thus, reconstruction of the ancestor of a sufficiently ancient protein family (for instance, one that predates the divergence of the major bacterial kingdoms) is a viable method for engineering a highly thermostable
protein [13]. Notably, computational evolutionary simulations [14] and the observation that ancestral proteins share similarities with consensus proteins in terms of both sequence and thermostability [8, 15, 16] suggest that the high thermostability of ancient proteins might also originate from bias in the maximum-likelihood method commonly used for APR, specifically, that more probable (often stabilizing) residues are chosen at every site, while less probable (often destabilizing) residues are neglected [14]. The implication of this potential bias for protein engineering is that APR may produce thermostable proteins regardless of their hypothetical age.

One problem that may be encountered in the construction of SBP-based FRET sensors is that the conformational change of the SBP may not translate into a change in distance between the N- and C-termini, which is often the case when the termini are located on the same domain of the bilobal SBP. A small change in the distance between the termini of the SBP results in a small change in FRET efficiency between the bound and unbound states of the sensor, which limits its dynamic range. In this circumstance, the dynamic range of the sensor can be improved by circular permutation, a method that allows the termini of a protein to be relocated [2, 6]. Circular permutation is achieved by rearranging the sequence of the SBP such that the original termini are connected by a flexible linker and new termini are created by disconnecting the sequence at a different location, which is chosen to maximize the distance change between the termini upon ligand binding. Since circular permutation is often destabilizing [17], the use of APR to obtain a thermostable SBP and the use of circular permutation to improve the dynamic range of the resulting FRET sensor are complementary.

Here, we show how APR can be used to create thermostable SBPs for the construction of robust FRET sensors. In this example, the maximum-likelihood statistical framework is used for phylogenetic analysis and reconstruction of ancestral sequences, although a variety of alternative methods are also available [18–20]. We focus on the reconstruction of ancestral SBPs, but similar approaches could be used to improve the thermostability of any protein, provided that a suitable sequence dataset can be obtained. We also show how the dynamic range of the sensor derived from an ancestral SBP can be improved by circular permutation prior to the insertion of the SBP into the FRET sensor construct.

2 Materials

2.1 Software

3. FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree).
Prepare all solutions using ultrapure water.

1. Synthetic genes encoding the circularly permuted ancestral proteins (see Subheading 3.7 for design).
2. Primers to amplify the circularly permuted genes with flanking *SapI* sites for Golden Gate assembly (see Subheading 3.7 for design).
3. High-fidelity DNA polymerase (e.g., Phusion Hot Start II Polymerase, Thermo Scientific).
4. dNTPs.
5. Thin-walled PCR tubes.
6. PCR thermocycler.
7. PCR purification kit (e.g., Wizard® SV Gel and Clean-Up System from Promega).
8. pDOTS4 or pDOTS10 plasmid [6] (see Note 1).
9. *SapI* (10 U/μL) (e.g., from New England Biolabs).
10. T4 DNA ligase (400 U/μL) with 10× T4 DNA ligase buffer (e.g., from New England Biolabs).
11. Electrocompetent *E. coli* TOP10 cells (Invitrogen).
12. Electroporation cuvettes.
14. YenB media: 7.5 g/L yeast extract, 8 g/L nutrient broth.
15. 5 mL culture tubes.
16. Luria-Bertani (LB) agar plates containing ampicillin (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar, 100 mg/L ampicillin).
17. Materials for colony PCR and analysis by gel electrophoresis.
18. PCR master mix with loading dye for gel electrophoresis (e.g., 2× PCR Super Master Mix from Biotool).
19. Primers for sequencing (see Subheading 3.8).

20. 1% agarose gel made with SB buffer (46 g/L boric acid, 8 g/L sodium hydroxide), stained with, e.g., GelRed™ (Biotium).

21. DNA ladder (e.g., GeneRuler 1 kb DNA ladder, Thermo Scientific).

22. Agarose gel electrophoresis apparatus.

---

### 3 Methods

#### 3.1 Sequence Collection

Given a reference SBP with the desired binding specificity, reconstruction of an ancestral SBP with improved thermostability and the same binding specificity requires a dataset of protein sequences that are homologous to the reference SBP (see Note 2). These sequences can be identified using a BLAST (Basic Local Alignment Search Tool) search of protein sequence databases, such as the NCBI database of nonredundant protein sequences or the UniProt-KB database. The protein sequences should be selected to maximize sequence diversity and phylogenetic diversity, but should have the same binding specificity as the reference SBP (see Notes 3 and 4). The sequences should be evenly distributed throughout sequence space; in other words, clusters of sequences with very high identity (>90%) and outlier sequences with low identity to any other sequence in the dataset should be avoided. A small set of outgroup sequences should also be selected for rooting the phylogenetic tree (see Note 5). In the case of SBPs, a suitable outgroup would be an SBP homologous to the reference SBP, but with a different binding specificity; for example, in our recent work, anionic amino acid-binding proteins served as an outgroup for cationic and neutral amino acid-binding proteins [6].

As a guideline, between 50 and 250 sequences should be selected for the phylogenetic analysis. If too many sequences are included, the phylogenetic analysis will be computationally intensive, while if too few sequences are included, there may be insufficient data to reconstruct the ancestral sequences accurately, or the sequence diversity within the dataset may be insufficient to make the ancestral proteins more thermostable than the reference SBP.

#### 3.2 Multiple Sequence Alignment

Before the sequences can be used for phylogenetic analysis, they must be collated in a multiple sequence alignment. In the phylogenetic analysis and subsequent reconstruction of ancestral sequences, it is assumed that each column in the multiple sequence alignment corresponds to a set of residues that originated from a common ancestor. If the alignment is incorrect, this assumption is violated; hence, the quality of the multiple sequence alignment is a critical factor in the success of the project. The number of gaps should be minimized by careful editing of the alignment to remove poorly
conserved N- and C-terminal regions and insertions that are unique to individual protein sequences, since PAML (the program used for reconstruction of ancestral protein sequences) will attempt to reconstruct one ancestral residue for each column in the alignment.

1. Align the protein sequences from Subheading 3.1 using the MUSCLE server [21] (see Note 6).
2. Open the alignment in SeaView [22].
3. Inspect the alignment carefully. If necessary, make manual adjustments to improve the alignment, particularly in regions with many gaps.
4. Trim the alignment, removing insertions that would not be present in the ancestral proteins of interest (see Note 7). Remove poorly conserved N- and C-terminal extensions, including the N-terminal signal peptides in the case of SBPs.
5. Save the alignment in PHYLIP interleaved format.

### 3.3 Model Selection

Phylogenetic analysis and reconstruction of ancestral protein sequences using the maximum-likelihood method requires a probabilistic model of protein evolution, which is needed to evaluate the likelihood of different evolutionary scenarios. The main component of this evolutionary model is an empirical substitution matrix that specifies the frequency of each amino acid and the substitution rate between each pair of amino acids. Some modifications are also available to make the basic evolutionary model more realistic; for example, different sites in the protein are often allowed to evolve at different rates by modeling rate heterogeneity using the discrete-gamma model (denoted +G) [23].

The first step in maximum-likelihood phylogenetic analysis is to choose the evolutionary model most appropriate for the given sequence dataset, which can be achieved using programs such as ProtTest [24] (see Note 8). ProtTest evaluates different models using the Akaike information criterion (AIC), a metric that accounts for the likelihood of the tree inferred using the model (the goodness of fit) and the number of model parameters (to penalize overfitting).

1. Open the ProtTest graphical interface and load the alignment file.
2. Select “Compute likelihood scores” under the Analysis menu.
3. Select the substitution matrices to evaluate. At a minimum, select the general-purpose JTT, WAG, and LG matrices; most of the remaining matrices are based on specific types of proteins (for example, MtArt is based on sequence data from invertebrate mitochondrial proteins).
4. Ensure that the +I, +G, +I+G, and empirical amino acid frequency options are selected to evaluate these modifications.
5. Select “BIONJ tree” as the base tree for likelihood calculations (see Note 9).

6. After the computation is complete, select “Results” under the Selection menu. Identify the model with the lowest AIC, which will be used for the subsequent analysis.

3.4 Phylogenetic Tree Inference

Given an alignment of protein sequences and a probabilistic model of protein evolution, a phylogenetic tree can be inferred using the maximum-likelihood method. The goal of phylogenetic inference using ML is to identify the tree topology (branching pattern) and parameters (for example, branch lengths) that optimize the likelihood function. Likelihood is defined in this context as the probability of observing the data—that is, the probability that the present-day sequences would have evolved—given the tree topology, tree parameters, and an evolutionary model. In practice, the maximum-likelihood tree is inferred by first estimating the correct tree using a computationally simple method such as the BIONJ method; heuristic tree-searching algorithms are then applied to modify the topology of the tree and increase its likelihood until convergence on the maximum-likelihood tree is achieved.

Inference of a phylogenetic tree for the purpose of APR should be considered an iterative process; problematic sequences should be identified and removed, and alternative sequences should be added before repeating the phylogenetic analysis. The final tree should be biologically plausible, supported by high bootstrap values, and robust to variations in the evolutionary model used.

1. Open the PhyML [25] web interface and upload the sequence alignment.

2. Select “Amino-Acids” for data type.

3. Under “Substitution Model,” specify the substitution model chosen in Subheading 3.3. +I denotes an estimated proportion of invariant sites; +G denotes an estimated gamma-shape parameter, and +F denotes empirical amino acid frequencies.

4. Under “Tree Searching,” select “BIONJ” for the starting tree and “NNI + SPR” for the type of tree improvement (see Note 10).

5. Under “Branch Support,” select the “aLRT SH-like” branch support test. Alternatively, for the final tree, select a bootstrap calculation with 100 replicates.

6. Once the analysis has completed, open the maximum-likelihood tree in FigTree. The tree can be rooted by selecting the branch connecting the ingroup and outgroup sequences and clicking the “Reroot” button.

7. Inspect the tree to identify problems such as long branches and low branch support values, and rectify these problems by repeating the multiple sequence alignment and phylogenetic analysis using a modified sequence dataset if necessary (see Notes 11 and 12).
8. Assess the robustness of the ML tree to variations in the substitution model by repeating the tree inference in PhyML using alternative substitution matrices that scored highly in Subheading 3.3. The resulting trees should be very similar to the maximum-likelihood tree.

3.5 Reconstruction of Ancestral Sequences

1. Ensure that the alignment file and tree file are properly formatted for input into PAML [26] (see Note 13).

2. Set up the program by modifying the control file (“codeml.ctl”), as shown in Table 1.

3. Run the program from the command line using the command codeml.

Table 1
Variables in the control file for ancestral protein reconstruction using PAML

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>seqfile</td>
<td>Filename of alignment</td>
</tr>
<tr>
<td>treefile</td>
<td>Filename of tree</td>
</tr>
<tr>
<td>outfile</td>
<td>Filename of output</td>
</tr>
<tr>
<td>noisy</td>
<td>9</td>
</tr>
<tr>
<td>verbose</td>
<td>2</td>
</tr>
<tr>
<td>runmode</td>
<td>0</td>
</tr>
<tr>
<td>seqtype</td>
<td>2</td>
</tr>
<tr>
<td>clock</td>
<td>0</td>
</tr>
<tr>
<td>aaRatefile</td>
<td>Filename of rate matrix (e.g., lg.dat for the LG matrix)</td>
</tr>
<tr>
<td>model</td>
<td>2</td>
</tr>
<tr>
<td>Mgene</td>
<td>0</td>
</tr>
<tr>
<td>fix_alpha</td>
<td>0 to use +G model, else 1</td>
</tr>
<tr>
<td>alpha</td>
<td>0.5 to use +G model, else 0</td>
</tr>
<tr>
<td>Malpha</td>
<td>0</td>
</tr>
<tr>
<td>ncatG</td>
<td>Number of rate categories to use for the +G model</td>
</tr>
<tr>
<td>getSE</td>
<td>0</td>
</tr>
<tr>
<td>RateAncestor</td>
<td>1</td>
</tr>
<tr>
<td>Small_Diff</td>
<td>.5e-6</td>
</tr>
<tr>
<td>cleandata</td>
<td>0</td>
</tr>
<tr>
<td>method</td>
<td>0</td>
</tr>
</tbody>
</table>

Deviations from the settings in the default control file are indicated in bold. The remaining variables in the control file are not applicable to amino acid-based analysis and can be removed.
4. Open the result file (“rst”) and scroll down to the text “tree with node labels.” Copy the tree into a new file and open the file in FigTree. Identify the ancestral nodes of interest (see Note 14) and record the node labels that identify them.

5. Search the result file for the text “node #x” (where x is the number identifying the ancestral node) to find the maximum-likelihood ancestral sequence associated with that ancestral node.

6. Open the alignment file in SeaView and add the ancestral protein sequences to the alignment. Edit the ancestral sequences to remove any remaining insertions that are artifacts of the reconstruction process (see Note 7).

3.6 Characterization of Ancestral Proteins

Once the ancestral protein sequences have been obtained, they are back-translated into nucleotide sequences and codon-optimized for expression in *Escherichia coli*, and the genes are synthesized. The synthetic genes are cloned into expression vectors, and the ancestral proteins are expressed in *E. coli* and purified. The thermostability of the ancestral proteins can be assessed using methods such as circular dichroism spectroscopy, differential scanning fluorimetry (DSF), or differential scanning calorimetry. The binding specificity of the ancestral proteins can be assessed using methods such as isothermal titration calorimetry (ITC), DSF, or (in some cases) fluorescence spectroscopy. If the ancestral SBPs have high thermostability and the desired binding specificity, they are strong candidates for circular permutation and incorporation into FRET sensor constructs.

3.7 Design of Circularly Permuted SBPs

The circular permutation of SBPs has been described by Okada et al. [2]. It is critical to identify sites for the new N- and C-termini that will produce a sensor with high dynamic range without negatively affecting the binding affinity or stability of the protein.

1. Using a crystal structure or homology model of the ancestral SBP (created using the Phyre2 server [27], for example), select the positions of the new N- and C-termini. The new termini must be located on different lobes of the SBP to maximize the change in their relative positions due to the ligand-induced conformational change. This can be achieved by deleting a section of the protein that links the two lobes (i.e., a hinge strand) (see Note 15). The result is a theoretical protein with the original N- and C-termini, and additional N*- and C*-termini, i.e., two protein fragments.

2. Design a linker sequence to fuse the original N- and C-termini of the SBP. Each repeat of a flexible (GGS)_n linker is approximately 11.4 Å in length. Measure the distance between the N- and C-termini of the SBP; the linker should have enough (GGS)_n repeats to bridge this distance.
3. Theoretically construct the circularly permuted SBP sequence by concatenating the three protein fragments: the N*-terminus to C-terminus fragment, the flexible linker, and the N-terminus to C*-terminus fragment.

4. Back-translate the circularly permuted SBP sequence to obtain a nucleotide sequence codon-optimized for expression in E. coli, and synthesize the gene (see Note 16).

5. Design primers to amplify the gene with flanking SapI sites for insertion into pDOTS4 or pDOTS10. The forward primer contains the sequence 5′-GCTCTTCAATC-3′ followed by the first 10–15 nucleotides of the circularly permuted gene. The reverse primer contains the sequence 5′-GCTCTTCCGAG-3′ followed by the first 10–15 nucleotides of the reverse complement of the circularly permuted gene (see Note 17). The length of the primers should be chosen such that the melting temperatures of the primers are approximately 65 °C.

Carry out all procedures at room temperature unless otherwise specified.

1. Using standard PCR with a high-fidelity DNA polymerase (e.g., Phusion Hot Start II), amplify the SBP gene using the cloning primers from Subheading 3.7. Purify the PCR product using a PCR purification kit, according to the manufacturer’s instructions.

2. In a thin-walled PCR tube, chilled on ice, mix 2 μL of 10× T4 DNA Ligase buffer, 1 μL 400 U/μL T4 DNA ligase, 1 μL 10 U/μL SapI, 100 ng pDOTS4 or pDOTS10, a fivefold molar excess of the purified PCR product from the previous step, and water to give a final volume of 20 μL.

3. Perform the Golden Gate assembly reaction using the following thermocycling protocol: 33 cycles of (37 °C for 2 min, then 16 °C for 3 min); 55 °C for 10 min; 80 °C for 5 min; then hold at 4 °C.

4. Transform one aliquot of E. coli TOP10 cells with 2.5 μL of the Golden Gate assembly reaction mixture by electroporation (see Note 18). After electroporation, resuspend the cells in 1 mL YenB and incubate with shaking at 37 °C for 1 h. Plate 100 μL of the culture on an LB/ampicillin agar plate and incubate at 37 °C overnight (see Note 19).

5. Use colony PCR and Sanger sequencing of the resulting PCR products to confirm correct insertion of the gene into the vector (see Note 20).

3.8 Cloning into a FRET Sensor Construct

3.9 Characterization of the FRET Sensors In Vitro and In Situ

After expression of the FRET sensor in E. coli and subsequent purification (see Notes 21 and 22), the dynamic range of the sensor and the affinity of the sensor for its ligand should be measured.
using fluorescence titrations. ITC can also be used to accurately measure the dissociation constant ($K_d$) of the sensor and the proportion of the sensor that is active.

The approach for in situ characterization of the FRET sensor will depend on where the sensor will be used to measure the resting level of its ligand, its dynamic changes or distribution. For instance, this could be intra- or extracellularly, in cell culture systems, or more complex tissue such as acute slices or in vivo. While intracellular ligand measurements typically require the sensor to be expressed by the cell of interest using an appropriate vector, extracellular sensor localization can be achieved by expression using plasma membrane targeting sequences and anchoring motifs [4]. However, sensor expressed endogenously and targeted to the membrane can be exposed to intracellular ligands during transit to the membrane. Both sensor populations, surface-presented and intracellular, will generate fluorescence that cannot be easily separated using diffraction-limited microscopy techniques, thus potentially limiting the interpretation of data. An alternative approach to target FRET exclusively to extracellular space relies on a biotin-streptavidin anchoring strategy [6, 28]. In our previous work, a biotin tag was incorporated into the FRET sensor using the pDOTS10 plasmid. Extracellular immobilization of the sensor was then achieved by linking the sensor to surface proteins, biotinylated using commercially available streptavidins and NHS-ester activated biotinylation reagents, via streptavidin [6].

Classically, FRET sensor imaging is performed using single photon excitation and quantification of donor/acceptor fluorescence intensity, donor lifetime, or emission spectra [4]. In thicker preparations like acute tissue slices and in vivo, two-photon excitation using near-infrared pulsed lasers has superior performance and provides good optical access to deeper structures [29] at the potential cost of reduced spectral separation of FRET donor/acceptor excitation (e.g., ECFP/EYFP) [30]. Using pulsed lasers, typically in the MHz frequency range (e.g., Ti:sapphire lasers around 80 MHz), also enables fluorescence lifetime imaging using time-correlated single photon counting (TCSPC-FLIM) of FRET donor fluorescence, a powerful tool to study binding of a ligand to a FRET sensor. Both modes of excitation allow monitoring of sensor fluorescence at high spatial and temporal, micrometer and millisecond, resolution.

### 4 Notes

1. pDOTS4 and pDOTS10 are mother plasmids for the cloning of SBPs to generate FRET sensors [6]. The pDOTS4 backbone is based on a pRSET plasmid [4] (pRSET FLIPE-600n, Addgene #13537, courtesy of Wolf Frommer) containing a 6× His-Tag at the N-terminus. For the construction of
pDOTS4, the native SapI site was first removed by site-directed mutagenesis. To adapt this plasmid for Golden Gate assembly [31], the ybeJ coding sequence in pRSET FLIPE-600n was removed and replaced with SapI-NotI-SapI restriction sites. The SBP gene can be introduced into the pDOTS4 plasmid using Golden Gate assembly (see Subheadings 3.7 and 3.8); the SBP will be fused to a 6× His-Tag and ECFP at the N-terminus and to Venus fluorescent protein at the C-terminus. To generate pDOTS10, a biotin tag followed by a GSGG linker, synthesized de novo (Epoch Life Science) based on the sequence from the PinPoint™ Xa-1 plasmid, was amplified by PCR and cloned into pDOTS4 between the 6× His-Tag and ECFP using the BamHI restriction site. The SBP gene can be introduced into the pDOTS10 plasmid using Golden Gate assembly (see Subheadings 3.7 and 3.8); the SBP will be fused to a 6× His-Tag, biotin tag, GSGG linker, and ECFP at the N-terminus, and to Venus fluorescent protein at the C-terminus. pDOTS4 and pDOTS10 will be made available via Addgene.

2. Usually, protein sequences are used for ancestral reconstruction rather than nucleotide sequences because they are more highly conserved and therefore retain more phylogenetic signal over longer evolutionary distances; protein sequences are under stronger selection than nucleotide sequences because many nucleotide substitutions are synonymous and have low fitness costs. However, nucleotide- or codon-based phylogenetic analyses are also possible.

3. Information about sequence-function relationships from crystal structures of the reference SBP or other experimental data is useful for identifying protein sequences that are likely to have the same or similar binding specificity. If the structure of the reference SBP is unknown, it may be possible to predict binding site residues using the structures of homologous SBPs; these residues should be conserved when the binding specificity is conserved.

4. Several strategies can be used to increase the diversity of a sequence dataset: PSI-BLAST, which performs iterative BLAST searches, can be used to find sequences more distantly related to the reference SBP; multiple SBP sequences can be used as input for the BLAST search (provided that they are actually homologous); programs such as CD-HIT [32] can be used to filter redundant sequences out of the dataset; BLAST searches can be restricted to certain taxonomic groups to obtain sequences from phylogenetically diverse organisms. Phylogenetic diversity in the sequence dataset is desirable to enable reconstruction of more ancient (and more thermostable) ancestral proteins.
5. The root of a phylogenetic tree is the node corresponding to the ancestor of all sequences in the tree. Rooted trees show the position of the root and the direction of evolution, whereas the unrooted trees produced by the maximum-likelihood method contain no information about the direction of evolution. This problem can be circumvented by the inclusion of outgroup sequences, which are known to be more distantly related to the sequences of interest (ingroup sequences), in the phylogenetic analysis. The root of the tree will be located somewhere on the branch joining the ingroup and outgroup sequences.

6. Various alignment programs with different specialities are available. In addition to MUSCLE, these include MAFFT [33], T-COFFEE [34], and PRANK [35]. In particular, PRANK may be useful for APR, since its phylogeny-aware algorithm is able to distinguish independent insertion and deletion events that would result in erroneous inferences of homology in conventional multiple sequence alignment programs [36].

7. Programs for automatic curation of multiple sequence alignments are available (such as Gblocks [37]), but these programs may be too zealous in removing poorly aligned regions of the alignment for this particular application. Although insertions unique to individual sequences or a small cluster of sequences can be safely removed, care should be taken not to delete columns that might contain residues present in the ancestral proteins of interest. Since phylogenetic information is needed to make this judgement, the alignment should be edited conservatively before the reconstruction of ancestral sequences. Artifactual insertions in the ancestral sequences can be removed afterward, once the phylogeny is known.

8. More recent implementations of maximum-likelihood methods may give superior performance in terms of computational speed and/or accuracy. Alternative software options include IQ-TREE [38] for model selection, IQ-TREE or RAxML [39] for tree inference, and FastML [40] for reconstruction of ancestral sequences.

9. The assumption implicit in this choice is that the BIONJ tree is a good enough approximation to the maximum-likelihood tree that the substitution model of best fit will be identical for both trees. It is more accurate to use the maximum-likelihood tree as the base tree for the likelihood calculations, but the computation will be much slower.

10. If these settings are used, the tree search will be deterministic, that is, the same alignment will always produce the same tree. This can be problematic, since the tree search may stall at a locally optimal tree and fail to converge on globally optimal tree (the true maximum-likelihood tree). It is therefore good
practice to repeat the analysis in PhyML with randomized starting trees; if the same tree topology is recovered each time, it is more likely that the globally optimal tree has been recovered.

11. Long branches on a phylogenetic tree arise from the inclusion of highly divergent sequences and cause a systematic error known as long branch attraction. If possible, long branches should be broken up by including additional sequences that have higher identity to the divergent sequences; otherwise, the divergent sequences may need to be removed.

12. Phylogenetic trees inferred using maximum-likelihood can be validated in several ways. First, the topology of the tree should be robust to variations in the evolutionary models used in the calculation; the phylogenetic analysis should be repeated using alternative substitution matrices that scored highly in Subheading 3.3, and the resulting tree topologies should be very similar to the maximum-likelihood tree. Second, statistical support for individual branches in the tree should be assessed using the bootstrap method, which involves repeating the phylogenetic analysis many times (usually at least 100) using multiple sequence alignments generated by random resampling of the complete alignment, to determine whether the same tree topology can be reproduced even when some of the data are removed or duplicated. The bootstrap value of a branch is the proportion of the bootstrap trees that contain the same bifurcation of sequences as denoted by that branch. Branches with bootstrap values <70% should be viewed with suspicion. Finally, if the tree describes the evolution of orthologous sequences, it should be consistent with established species-based trees (although this criterion is less applicable to SBPs due to the prevalence of horizontal gene transfer).

13. In our experience, the file formatting requirements of PAML are the most common cause of any errors encountered while running the program. The alignment must be in PHYLIP interleaved format. This format must be specified by adding the letter “I” (for “interleaved”) to the end of the first line (after the number of sequences and number of columns). There must also be two spaces between each sequence name and the corresponding sequence. The tree must be in Newick format. PAML can only parse trees obtained from PhyML if they do not contain branch support values. Trees with branch support values can be reformatted using the program Retree in the PHYLIP package; open the tree in Retree and write the (un rooted) tree to a new file without further modification.

14. Ideally, multiple ancestral nodes between the reference SBP and the last common ancestor of the sequences of interest would be selected and characterized experimentally.
15. Previous examples of circularly permuted SBPs [2, 6] can be used as a guide to identify potential sites for circular permutation, provided that the proteins share sufficient homology. Otherwise, select sites in disordered loops of the protein (as indicated by high B-factors in the crystal structures of related proteins), and avoid the removal of any residues that contribute to secondary structure. Multiple sites may need to be tested. As a further check, the CPred server [41] can be used to evaluate the viability of potential sites for circular permutation.

16. Ensure that no SapI sites (GCTCTTC) are present in the synthetic gene, as these will prevent successful cloning into pDOTS4 or pDOTS10.

17. Cloning into the pDOTS4 or pDOTS10 plasmids by Golden Gate assembly results in addition of an N-terminal isoleucine residue and a C-terminal leucine residue to the cloned gene, which should be taken into account if the linkers between the fluorescent proteins and recognition domain are optimized.

18. Competent cells with high transformation efficiency (10⁷–10⁸ cfu per μg DNA) are required for this step. If the transformation efficiency of the competent cells is low, plate the entire volume of electroporated cells by pelleting the cells (18000 × g on a benchtop centrifuge for 1 min), discarding the supernatant, resuspending the pellet in 100 μL YenB media, and plating the resulting suspension.

19. If no colonies are observed, the ligation by T4 DNA ligase may have been unsuccessful; if the colonies only contain empty vectors, digestion of the template gene by SapI may have been unsuccessful.

20. We recommend sequencing in two reactions, first using a T7-specific primer (forward) in combination with a SBP-specific primer (reverse), then using an SBP-specific primer (forward) with a T7-specific primer (reverse), because the length of the full construct exceeds the length of most Sanger sequencing reads.

21. High levels of expression can be achieved using pDOTS4 or pDOTS10 by incubation of E. coli BL21(DE3) transformants in auto-induction media at 20 °C for 72 h. Expression can be monitored during growth by removing a 1 mL aliquot from the culture, centrifuging the sample to pellet the cells, resuspending the pellet in 1 mL water, and measuring the fluorescence spectrum of the resulting cell suspension. With excitation at 433 nm, an emission peak at 476 nm corresponding to ECFP and an emission peak at 525 nm corresponding to Venus should be apparent. Since maturation of the Venus fluorophore is slower than maturation of the ECFP fluorophore, the Venus peak may not be observed until after 48 h.
22. The FRET sensors are expressed with N-terminal hexahistidine tags, allowing purification by nickel affinity chromatography. In our experience, further purification by size-exclusion chromatography can also improve the dynamic range of the sensor. Finally, extensive dialysis may be required to remove any endogenously bound ligands, as sensor molecules that are already ligand-bound do not exhibit a ligand-dependent change in FRET, reducing the observed FRET efficiency.

Acknowledgments

Research was funded by Human Frontiers Science Program Young Investigator Award (HFSP to H.J., C.H., and C.J.J., grant number: RGY0084/2012), German Academic Exchange Service (DAAD-Go8) Travel Fellowship (to C.H. and C.J.J.), NRW-Rückkehrerprogramm (to C.H.), and German Research Foundation (DFG, SFB1089 B03, SPP1757 HE6949/1-1 and HE6949/3-1, all to C.H.).

References

Chapter 6

Method for Developing Optical Sensors Using a Synthetic Dye-Fluorescent Protein FRET Pair and Computational Modeling and Assessment


Abstract

Biosensors that exploit Förster resonance energy transfer (FRET) can be used to visualize biological and physiological processes and are capable of providing detailed information in both spatial and temporal dimensions. In a FRET-based biosensor, substrate binding is associated with a change in the relative positions of two fluorophores, leading to a change in FRET efficiency that may be observed in the fluorescence spectrum. As a result, their design requires a ligand-binding protein that exhibits a conformational change upon binding. However, not all ligand-binding proteins produce responsive sensors upon conjugation to fluorescent proteins or dyes, and identifying the optimum locations for the fluorophores often involves labor-intensive iterative design or high-throughput screening. Combining the genetic fusion of a fluorescent protein to the ligand-binding protein with site-specific covalent attachment of a fluorescent dye can allow fine control over the positions of the two fluorophores, allowing the construction of very sensitive sensors. This relies upon the accurate prediction of the locations of the two fluorophores in bound and unbound states. In this chapter, we describe a method for computational identification of dye-attachment sites that allows the use of cysteine modification to attach synthetic dyes that can be paired with a fluorescent protein for the purposes of creating FRET sensors.

Key words Synthetic dye, Optical sensor, Computational modeling, Förster resonance energy transfer

1 Introduction

Optical sensors have allowed for the investigation of physiological processes such as neurotransmission with both spatial and temporal resolution. FRET-based optical sensors are particularly useful, as they are capable of giving quantitative recordings independent of sensor concentration due to the ratiometric signal output of the sensor and the concentration independence of FRET donor lifetimes [1, 2]. Contemporary sensors typically use fluorescent proteins (FPs). However, some have used synthetic fluorescent
dyes rather than FPs as the signaling component of the sensor. As there is much greater control over the precise location of synthetic fluorophores, synthetic dyes can theoretically allow for even small conformational changes to produce a measurable FRET signal. While a large conformational change for a given protein is always desirable for sensor construction, it is often a necessity when using FPs, meaning that synthetic dyes are potentially applicable across a much larger range of proteins, rather than being restricted to those with large distance-based conformational changes.

Sensors that use synthetic dyes are either intensity-based sensors (non-FRET), which require multiple synthetic components, or must be developed through extensive high-throughput screening [3, 4]. For example, in addition to the use of two synthetic dyes, the Snifit-type sensor design requires the development and synthesis of a tethered competitive ligand that can occupy the binding active site, which necessarily introduces an extra design phase of engineering and screening [3]. On the other hand, the EOS-type sensor developed by Namiki et al. only requires a single synthetic component (a dye) [4], but still requires exhaustive screening of different residues to find a location that gives a strong signal. In addition, EOS sensors do not produce ratiometric output and are therefore not quantitative.

It is possible to improve on one or more of these design components when creating a synthetic dye-based sensor. Specifically, it has been shown that it is possible to create FRET sensors that are a combination of one FP and one synthetic dye [5], which is an improvement over both single fluorophore and two dye sensors as it is ratiometric and requires one fewer site-specific modification. Additionally, rather than using brute force high-throughput screening of all residue locations to identify a dye labeling site, we have expedited sensor development through the use of computational screening, which can reduce the number of possible dye labeling sites to a subset with a higher likelihood of yielding a functional sensor. We have used these computational techniques in tandem with synthetic dyes (via thiol-maleimide labeling of residues) to develop optical sensors with large dynamic ranges.

2 Materials

Whenever possible, prepare all stock solutions and buffers in ultrapure water (MilliQ). For reagents that have poor solubility in water, dissolve them with the smallest possible proportion of organic solvent (i.e., 5% DMSO would be preferable to 10% DMSO) as some proteins may have poor stability in organic solvent. All solutions and reagents should be prepared as fresh as possible as some reagents will have short lifetimes when in solution, even at −20 °C. All buffer solutions should be filtered through a membrane filter (pore size 0.45 μm or smaller) after preparation.
2.1 Software

1. Python 2.7.11 (www.python.org).
2. Bash 4.2.46 (www.gnu.org/software/bash/).
3. GROMACS 5.1.2 (www.gromacs.org) [6].
4. MARTINI 2.1 (http://md.chem.rug.nl/) [7].
5. PyMol 1.7.6.0 (https://sourceforge.net/projects/pymol).
6. GAWK 4.0.2 (www.gnu.org/software/gawk/).
7. DSSP 2.2.1 (http://swift.cmbi.ru.nl/gv/dssp/).
9. Curl 7.29.0 (https://curl.haxx.se/).

2.2 Cloning and Protein Purification Components

1. The gene of interest in an expression vector (see Subheading 3.3).
2. Primers containing the mutation of interest (see Subheading 3.3).
3. High-fidelity DNA polymerase kit.
4. Gibson assembly kit.
5. Thin-walled PCR tubes.
6. PCR thermocycler.
7. PCR purification kit.
8. Competent cells for cloning (Top10).
10. Materials for colony PCR and analysis by gel electrophoresis:
    (a) PCR master mix.
    (b) T7 primers.
    (c) 1% agarose gel made with SB buffer (46 g/L boric acid, 8 g/L sodium hydroxide), with a visualizing stain.
    (d) DNA ladder mix.
    (e) Agarose gel electrophoresis apparatus.

2.3 Dye Labeling Components

1. Buffer solution, Phosphate buffer (see Note 1): 0.05 M Sodium phosphate, 0.2 M NaCl. Adjust the pH to what is appropriate for both the protein of interest and the chemistry that is needed to label the protein with the synthetic dye (using either hydrochloric acid or sodium hydroxide).
2. TCEP stock solution: dissolve 0.1437 g of Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) in 1 mL of buffer solution (501 mM stock solution) (see Note 2).
3. Dye stock solution: Add a suitable solvent to the synthetic dye to achieve a stock solution with a final concentration of 10 mM. In the case of the Alexa Fluor 532 C5 Maleimide, 1 mg was dissolved in 123 μL of phosphate buffer (10 mM stock solution) (see Note 3).
4. Protein solutions: proteins should be concentrated as much as practical (ideally at least 500 μM) and exchanged or dialyzed into the same buffer solution as used to prepare the reagent stock solutions (see Note 4).

3 Methods

All experimental (noncomputational) work should be performed at 4 °C unless otherwise specified.

3.1 Computational Screening and Residue Selection

First, prepare models of the target fusion protein in both its bound and unbound conformations. This involves modeling both the solute-binding and fluorescent domains with the appropriate linker. Start with crystal structures or high-quality homology models of the desired binding core in both conformations and the fluorescent protein. Ensure no nonstandard residues exist in the models, as these are not parameterized in the MARTINI forcefield (see Note 5). Reconstruct any residues missing from the crystal structures at the termini by which they are fused. Then, construct the linker sequence as expressed experimentally (see Note 6). Complete the model by fusing the two domains (see Note 7). Save both SBP-linker-FP models as .PDB files.

Create a directory for each conformation, copy the appropriate .PDB file into each, and also copy simulations.sh into each (see Note 8). The script simulations.sh automatically prepares and runs a number of MARTINI simulations. It depends on all of the software in Subheading 2.1 except MARTINI, which is downloaded automatically by the script, provided all software dependencies are available in your $PATH (see Note 9). The script can be configured by making a copy of it in each conformation directory and editing the leading “CONFIG” portion. Most defaults should be appropriate; however, the input_file, system_name and linker_residues variables should be set for your system (see Note 10). Configure and run the script, read and follow the directions given, then repeat for the other conformation.

Run the second script “process-data.py,” giving as the first two arguments the locations of each set of output files. The residue number of the central residue of the FP fluorophore and the range of residues that should be checked for sensor generation should also be given as arguments, respectively, with the –f and –r switches (see Note 11). process-data.py reads the given .PDB files, and predicts dynamic ranges for sensors formed by labeling each residue in the given range with a dye with configurable Forster distance. This output is stored by default as comma-separated values with appropriate headers in sensor_predict.csv and can be visualized graphically using a program such as Excel.
Select residues that are appropriate for cysteine mutagenesis (or mutagenesis to an appropriate residue). Note that this prediction does not account for any disruption of binding core function associated with chemical labeling. Therefore, a residue that yields the largest predicted dynamic range may not necessarily yield the best sensor, as the residue may have some structural or functional importance, which may be disrupted with mutagenesis. Residues with side chains oriented toward the solvent, or that are not a part of a structural motif should be selected preferentially. In the hypothetical data set example (Fig. 1), residue 200 is predicted to yield a large dynamic range upon labeling with a dye. Suppose, however, that for this hypothetical protein residue 200 is both not exposed to solvent and has its sidechain oriented toward the binding site of the protein (Fig. 2). Mutating and labeling this residue may abolish protein function, as the sterically bulky dye excludes the substrate from the active site. In contrast, although residue 100 has a smaller predicted dynamic range than residue 200, it is located on a flexible loop that does not have significant contact with other structural features of the binding protein. This makes this location preferable to residue 200, as labeling the residue is less likely to impact the correct function and dynamics of the binding domain, while still providing an excellent dynamic range.

3.2 Cloning and Purification of the Mutant Proteins

1. The sensor construct (SBP fused with the fluorescent protein) should be first cloned into an expression vector (with a T7 promter). The sequence to be cloned should match the sequence used to model the sensor exactly, with the exception
that there can be a histidine tag at the C-terminus of the fluorescent protein to facilitate protein purification.

2. Next, cysteine mutants of this sensor should be created at the residue locations identified by the computational screening (see Note 12). Any unwanted surface cysteines should be mutated to alternative residues to avoid nonspecific labeling (see Note 13). Although many cloning methods are suitable to introduce mutations, our preferred method is Gibson assembly.

3. In order to create the cysteine mutants through Gibson assembly, first synthesize or order a set of complementary primers (forward and reverse primers encoding the same sequence) that encompasses the residue of interest, with the total length of the primer between 30 and 50 nucleotides. The nucleotides coding the residue of interest should be changed to encode a cysteine residue, all other residues should match the template DNA exactly.

4. For the PCR, these primers will then be paired with the T7 promoter/terminator primers for PCR amplification. The T7 promoter forward primer is paired with the reverse mutagenic primer, while the T7 terminator reverse primer is paired with the mutagenic forward primer.
5. The PCRs using these primer sets should result in two fragments, with one fragment overlapping with the T7 promoter region and the mutagenic region, and the other fragment overlapping with the mutagenic region and the T7 terminator region. The mutagenic regions of these fragments will overlap and anneal during Gibson assembly. Gel purification is highly recommended to improve cloning efficiency, even if the agarose gel electrophoresis of the PCR product shows a single clear band.

6. Linear vector for use in the Gibson assembly reaction can be made through PCR using complementary primers of the T7 regions. Specifically, a T7 terminator forward primer and a T7 promoter reverse primer should be used. The linearized vector produced from this PCR reaction will have T7 regions that can overlap with the T7 regions of the gene fragments for the Gibson assembly reaction.

7. The PCR fragments (for both the sensor and vector) can then be combined into the Gibson master mix (in equimolar ratios). The typical incubation time for the master mix is 50 °C for 1 h.

8. The Gibson assembly mixture should then be transformed into competent cells that are designed for DNA cloning (e.g., TOP10 cells). If electrocompetent cells are used, the Gibson mix can be diluted with water to prevent arcing.

9. After confirming that the cloning is successful through sequencing, the gene should be transformed into an expression cell type (e.g., BL21DE3).

10. If the sensor construct contains a histidine tag, it can then be purified through nickel affinity chromatography.

### 3.3 Dye Labeling and Purification

1. To a volume of 849 μL buffer add 100 μL of the concentrated protein solution (assuming the concentration of the protein solution is 500 μM) and 1 μL of the TCEP solution. Wait approximately for 5 min to allow this solution to equilibrate to room temperature and for any disulfides to be reduced. This reaction can be performed in an Eppendorf tube or an equivalent (see Note 14).

2. To the reaction mixture add 50 μL of the dye stock. The final composition of the reaction mixture should contain approximately 50 μM of protein, 500 μM TCEP, and 500 μM dye. The reaction should be allowed to proceed overnight (16 h) at 4 °C in the dark with constant agitation (see Note 15).

3. After the reaction period, centrifuge the reaction mixture at high speed (18,000 × g, 5 min) to separate out any precipitated dye or protein (see Note 16).

4. The mixture should then be purified through gel filtration, with a desalting column usually being sufficient (see Note 17).
5. After one round of purification with the desalting columns, the protein mixture should be re-concentrated and buffer exchanged using centrifugal protein concentrators, which will remove trace TCEP and further remove unreacted and free dye.

6. The protein should then undergo a final desalting step to ensure that no free dye or TCEP remains and labeling efficiency can be evaluated using the following equation.

7. Moles of dye per mole protein = \( \frac{A}{\varepsilon \times C} \)

where for a given sample “A” is the absorbance at the peak excitation wavelength of the dye in use, “\( \varepsilon \)” is the extinction coefficient of the dye, and “\( C \)” is the concentration of protein.

## 4 Notes

1. Phosphate buffer might not be compatible with all proteins; substitute with an appropriate buffer if needed, but ensure that the pH range is compatible with the dye system in use. In the case of thiol-maleimide conjugations, the desired pH is between 7.0 and 7.5.

2. TCEP, or reducing agents in general, are typically only necessary for thiol based conjugation. TCEP is far preferable to other reducing agents such as DTT, as under normal conditions TCEP will not interfere with the maleimide-thiol reaction and also does not have issues with odor.

3. Sometimes there may be trace precipitate or undissolved dye in the stock solution. This is not a cause for concern so long as the precipitate is suspended homogenously prior to use. This trace precipitate will dissolve slowly over time as the reagent is consumed.

4. The protein should be purified to the highest degree possible, as other proteins could be labeled by the dye, which can affect the observed dynamic range of the protein sample. If additional purification is needed, size exclusion chromatography can be performed either before or after the labeling reaction.

5. For fluorescent proteins, this generally means mutating the fluorophore back to the three autocatalytic residues that form it; other proteins may incorporate nonstandard residues as a result of chemical modification for crystallization, etc. Unrecognized residues can be mutated to glycine by opening the .PDB file in a text editor, deleting the side chain atoms, and changing the residue names to GLY. They can then be mutated to the appropriate residue with PyMOL’s mutagenesis wizard.
6. We found this easiest to do by working out from both domains and allowing the constructed extended linker model to meet in the middle. Residues can be added to an existing model in PyMOL in Editing mode, accessed by clicking the mouse key shortcuts box in the lower right corner of the viewer window while in the default Viewing mode. Once in Editing mode, select the N-terminal nitrogen or C-terminal carbonyl carbon by clicking on it, then add residues by holding Alt and typing the single-letter code associated with the desired amino acid.

7. In PyMOL’s Editing mode, select both atoms that should be bonded in the product, then run PyMOL’s fuse command. For example if a model with an ECFP on the N terminus of the protein is desired, start by selecting both the amine of the N terminus of the protein and the carbonyl carbon of the C terminus of the ECFP. Then, while both atoms remain selected, enter “fuse” as a command input, which will generate an approximate model of the fusion protein. The fuse command may sometimes orient the proteins poorly; make sure to rotate the proteins as such that they do not overlap in physical space and the linker is fully extended. This can be done in editing mode by holding shift and right-clicking on a bond to rotate the associated torsion angle.

8. For example, you may be working in a subfolder of your home directory called ~/dyes. You should create new directories for each conformation, perhaps ~/dyes/open and ~/dyes/closed. You add the script and starting structure to each directory, and are left with the following files:

```
~/dyes/open/open-start.pdb
~/dyes/open/simulations.sh
~/dyes/closed/closed-start.pdb
~/dyes/closed/simulations.sh
```

9. The script will then generate folders for setup, each run, and the trimmed, fitted trajectories as pdb files:

```
~/dyes/open/setup/
~/dyes/open/run1/ , ~/dyes/open/run2/ ,
~/dyes/open/run3/ etc.
~/dyes/open/results/
```

10. The version numbers given in the materials are known to work; other versions will probably work as well but have not been tested. Note that MARTINIZE.py, which is downloaded and run by simulations.sh, is not compatible with Python 3, and thus Python 2 must be available on your system for the setup steps. The typical name for the DSSP executable varies from system to system; simulations.sh can be told the correct name for your system either by editing the dssp_name variable or by passing the correct name as an argument.
with the –dssp_name switch. Finally, GROMACS versions prior to 5.0 are not supported.

11. By default, simulations.sh prepares, equilibrates, and runs all simulations it is asked to without taking a break. However, if it is terminated, it can restart from the beginning of the last step it finished successfully, so it may be used (with some modification) as a resubmit script for systems like PBS. If preferred, the -o switch can be supplied to the script, which will perform the setup steps and generate individual run folders, but will not perform the full-length production runs. The full-length simulations can then be run on whatever hardware is appropriate by simply running gmx grompp on the provided .MDP files and starting structures.

The script first prepares the MARTINI coarse-grained force field topology, and converts the provided .PDB file to a coarse-grained model with the script Martinize. This includes construction of an elastic network around both protein domains, which keep their conformations constant and allow sampling of linker collapse. It then energy minimizes and solvates the model, including addition of neutralizing ions, antifreeze MARTINI water, and experimental salt concentration if desired. It performs a second energy minimization, and then equilibrates the system with thermostat and barostat in several rounds of progressively weaker backbone-restrained MD. Force constants of 1000, 500, 100, 50, and 10 are used by the script. The production run involves 30 replicate 200 ns simulations. New velocities are generated for each of these runs in a final unrestrained equilibration step. These simulations take approximately 90 min each for a 600 residue fusion model running on dual Tesla K40s with 32 cores. Finally, it outputs the last 500 ns of each simulation, sampled in 1 ns intervals, as a PDB file to the results directory.

For instance, if step 2 was performed in directories called ~/dyes/open and ~/dyes/closed, and the script is being run in ~/dyes, per the example in Note 8, a typical call for a fusion model with the binding protein occupying residues 242–586 and the fluorophore at residue 63 might be:

```bash
python process-data.py open/results/*.*.pdb closed/results/*.*.pdb -f 63 -r 242-586
```

12. While cysteine mutagenesis is a functional method of chemically labeling a protein with a synthetic dye, alternative chemistries are possible through the use of unnatural amino acid incorporation. This can allow for biorthogonal labeling of proteins in vivo, and in principle, it can allow for the sensor to be genetically encoded in an organism capable of utilizing the required unnatural amino acid.
13. There are two conserved cysteine residues in the GFP family; we have had success with the mutations C49S and C71V with minimal fluorescence loss, as described by Suzuki et al. [8].

14. The reaction volume and reagents should be scaled appropriately, relative to the concentration of the protein stock solution.

15. When agitating the solution, care should be made that the solution does not begin to form froth or foam as this can lead to precipitation of protein.

16. Filtration can be used instead of centrifugation; however, there is typically some volume/yield loss when filtering small volumes.

17. PD-10 columns (GE healthcare) are usually sufficient to separate free dye from the protein. If the purity of the protein is a concern, the protein should be first buffer exchanged to remove excess TCEP and then purified with SEC (GE healthcare, Hiloade 26/600 superdex 200pg, adequate for most proteins). This should separate labeled protein from free dye and any contaminant proteins.

References

Chapter 7

Rational Design and Applications of Semisynthetic Modular Biosensors: SNIFITs and LUCIDs

Helen Farrants, Julien Hiblot, Rudolf Griss, and Kai Johnsson

Abstract

Biosensors are used in many fields to measure the concentration of analytes, both in a cellular context and in human samples for medical care. Here, we outline the design of two types of modular biosensors: SNAP-tag-based indicators with a Fluorescent Intramolecular Tether (SNIFITs) and LUCiferase-based Indicators of Drugs (LUCIDs). These semisynthetic biosensors quantitatively measure analyte concentrations in vitro and on cell surfaces by an intramolecular competitive mechanism. We provide an overview of how to design and apply SNIFITs and LUCIDs.

Key words Biosensors, Protein engineering, Protein switches, Intramolecular ligands, Self-labeling proteins, SNAP-tag, CLIP-tag, Polyproline linkers, Therapeutic drug monitoring, Cell-surface biosensors

1 Introduction

1.1 SNIFITS and LUCIDs: Semisynthetic Modular Biosensors

Small molecule analytes are central to many biological systems. The ability to quantify such analytes is important in basic research to better understand anabolic, metabolic, and signal relay processes. It is important to quantify small molecules not only in basic research, but also in healthcare, where it is crucial to be able to monitor the levels of drugs and metabolites for therapeutic purposes. The use of biosensors is one method for this quantification, and has many different applications [1, 2].

Several natural proteins bind small molecule analytes. If such a binding protein undergoes a conformational change upon analyte binding, it can be used as part of a biosensor system [3, 4]. For example, it can be sandwiched in between two fluorescent proteins that are Förster resonance energy transfer (FRET)-partners [5, 6] or between two bioluminescent resonance energy transfer (BRET)-partners [7]. In both cases, the change in distance and orientation between the two resonance energy transfer (RET) partners is translated into changes in the optical properties of the system, which
can be used to give a quantitative output. However, many binding proteins do not undergo a sufficiently large conformational change when they bind to the analyte of interest. For these binding proteins, modular biosensors can be used to give larger artificial conformational changes (see [8], and references therein). Our laboratory has introduced two such examples: SNAP-tag-based indicators with a Fluorescent Intramolecular Tether (SNIFITs) and LUCiferase-based Indicators of Drugs (LUCIDs).

SNIFITs and LUCIDs are semisynthetic sensors composed of (1) a recombinant fusion protein and (2) a synthetic intramolecular tether (Fig. 1). The fusion protein is composed of a binding protein for an analyte of interest, a RET donor, and a self-labeling protein (SLP). The intramolecular tether is covalently attached to the SLP by a specific bio-orthogonal reaction, and contains a competitive ligand for the binding protein. In the absence of the analyte, the intramolecular ligand binds to the binding protein forming a “closed” state of the sensor. In the presence of the free analyte, the intramolecular ligand is displaced from the binding protein leading to an “open” state of the sensor. Since a RET donor is inserted into the recombinant fusion protein, and a RET acceptor is incorporated into the intramolecular tether, the conformation change can be translated into an optical readout. The equilibrium between the closed and the open states of the sensor can then be measured quantitatively.

**Fig. 1** The Architecture of SNIFITs and LUCIDs. An intramolecular tether (dashed line) containing a RET acceptor and ligand for the binding protein is covalently attached to SNAP-tag, a self-labeling protein (SLP). SNAP-tag is fused to a RET donor and a binding protein. Displacing the intramolecular ligand by an analyte of interest leads to an overall change in the sensor geometry, and the RET efficiency between a RET donor and RET acceptor. The RET donor can be a synthetic dye covalently attached to a second orthogonal SLP, a fluorescent protein, or a luciferase.
The intramolecular tether is covalently attached to the recombinant fusion protein via an SLP. SLPs, such as SNAP-tag [9, 10], CLIP-tag [11], and Halo-tag [12], react selectively with various bio-orthogonal chemical groups for applications in imaging, as well as for purifying proteins and modifying protein function [13–15]. SNIFITS and LUCIDs use SNAP-tag, a derivative of the $O^6$-alkylguanine-DNA alkyltransferase, which reacts specifically with $O^6$-benzylguanine (BG) and chloropyrimidine (CP) derivatives (Fig. 2a). These moieties are incorporated at one end of the intramolecular ligand. A second orthogonal SLP, such as CLIP-tag, can be used to attach a RET donor fluorophore to the recombinant fusion protein. CLIP-tag is a further development of the $O^6$-alkylguanine-DNA alkyltransferase that reacts specifically with $O^2$-benzylcytosine (BC) (Fig. 2b).

The optical readouts of SNIFITS and LUCIDs are FRET and BRET, respectively. FRET is generally preferred for cell imaging. The FRET donor can be either a fluorescent protein or a small-molecule synthetic dye, where synthetic dyes have better quantum yields, less photo-bleaching, and narrower excitation/emission spectra, than fluorescent proteins [16, 17]. For applications on cell surfaces, cell-impermeable synthetic dyes should be used, as cell-permeable dyes will also label the sensor in the secretory pathway. BRET is preferred when bulk measurements are made in complex biological mixtures such as serum. Luciferases generate light upon reacting with a chemical substrate [18, 19] bypassing

![Fig. 2 Mechanisms of SLPs used in the work of SNIFITS and LUCIDs. SNAP-tag reacts selectively with $O^6$-benzylguanine moieties while CLIP-tag reacts with $O^2$-benzylcytosine (BC). The SLPs are used to covalently attach the intramolecular tether and sometimes a fluorescent molecule (here annotated as R), to the fusion protein.](image)
the need to excite the donor fluorophore. This eliminates problems with the autofluorescence of other molecules in the samples, and eliminates the photo-toxicity and bleaching effects that may take place when using FRET. Nevertheless, the photon count emitted from the luciferase reaction is lower than that observed when a fluorophore is excited, which makes FRET the readout method of choice for applications in cell biology.

To quantify the response of the sensor to an analyte of interest, the light intensity is measured at the emission wavelengths of both RET partners (after excitation of the RET donor for SNIFITs or after the addition of the luciferase substrate for LUCIDs). The ratio of these two light intensities ($I_{\text{donor}}/I_{\text{acceptor}}$) is plotted against the free analyte concentration (Fig. 3). At low analyte concentrations, the sensor is mostly in the closed state and the ratio is low (lower plateau). At high analyte concentrations, the equilibrium is shifted to the open state of the sensor, and the ratio is high (upper plateau). The sensor’s dynamic range refers to the change in signal magnitude in the absence of analyte compared to under saturating

![Fig. 3 A typical titration curve from a RET-based biosensor. The sensor is described by the lowest and highest possible ratios of the RET donor and RET acceptor emission ratios (lower and upper plateaus), the dynamic range and the $c_{50}$](image-url)
concentrations of analyte, and is calculated as the ratio between the upper and the lower plateaus. The concentration of analyte that causes half the maximum signal change is defined as the $c_{50}$. The $c_{50}$ for a certain analyte can be tuned by altering the affinity of the intramolecular ligand for the binding protein. The analyte response range can thus be altered to fall within physiological concentrations of the analyte.

Some common design principles can be followed to construct a SNIFIT or a LUCID for a specific analyte. Over the years, we have gathered experience from our design of biosensors from analytes ranging from sulfonamides and neurotransmitters to cancer therapeutics such as methotrexate. Here, we give an overview of how to design these biosensors, followed by an overview of their practical applications.

A binding protein is required to display sufficient affinity and specificity for the analyte of interest. Since SNIFITs and LUCIDs are based on competition with another ligand, the dissociation constant of the binding protein for the analyte of interest must be lower compared to the desired response range.

Another important factor when choosing the binding protein is the availability of structural information, ideally in complex with a potential intramolecular ligand. This is not only crucial for the geometrical optimization of the sensor, but also for the choice of derivatization points of the ligand. In this regard, it can be advantageous if the binding protein is small, monomeric, and stable. Yet, if no suitable natural binding protein is available for the analyte of interest, computational methods can be used to design a protein with tailor-engineered binding properties [20].

The synthetic component of SNIFITs and LUCIDs consists of three parts: (1) a BG group that serves as the attachment site for SNAP-tag, (2) a fluorophore as RET acceptor, and (3) an intramolecular ligand (Fig. 4). These three parts are connected by oligoethylene glycol (EG) linkers that must be sufficiently long not to affect the binding of the intramolecular ligand and the closing of the sensor. We routinely use an (EG)$_{11}$ linker between the BG group and the fluorophore, and an (EG)$_{2}$ linker between the fluorophore and the intramolecular ligand.

Three main points must be considered when designing the intramolecular ligand of the sensor: (1) its affinity must be strong enough for the sensor to be predominantly closed in the absence of analyte, (2) its affinity determines the sensor’s analyte response range, and (3) its dissociation kinetics determine the response kinetics of the sensor.
The response of the sensor can be described in two steps: unbinding of the intramolecular ligand ($K_1$, Eq. (1)) followed by binding of the analyte ($K_2$, Eq. (2)) (Fig. 5). Equation (1) shows that the ratio between the closed ($S_{(closed)}$) and open ($S_{(open)}$) state of the sensor in the absence of analyte is directly proportional to the affinity of the intramolecular ligand. Equation (3) describes the relationship between the sensor’s apparent affinity for the analyte ($K_{tot}$) and the analyte’s affinity for the receptor protein ($K_{d,analyte}$) (see [21] for a more detailed discussion).

The effective molarity ($M_{eff}$) of the intramolecular ligand in SNIFIT and LUCID sensors is in the order of 100 μM [22]. The
intramolecular ligand should therefore have a $K_d$ of 10 $\mu$M or lower, so that the closed state of the sensor is favored over the open state by at least a factor of 10 when the analyte is not present. Since the analyte competes with the intramolecular ligand for binding to the same site, the apparent affinity of the sensor for the analyte is inversely proportional to its affinity for the intramolecular ligand. In this way, the response range of the sensor can be tuned to the desired concentration. However, if the intramolecular ligand’s $K_d$ value is very low, its unbinding kinetics tend to be slow, which limits the sensor’s temporal resolution.

1.2.3 Sensor Geometry

To achieve a large dynamic range, the RET efficiency should be high in the closed state and low in the open state of the sensor. Since the RET efficiency depends strongly on distance, this can be achieved by modifying the sensor geometry. The modular architecture of SNIFITs and LUCIDs allows to control the distance between the RET partners in the closed (see D1 in Fig. 6) and open states (see D2 in Fig. 6) independently of each other.

It may be difficult to obtain close proximity between the RET donor and RET acceptor in the closed state of the sensor (i.e., aiming to minimize D1 in Fig. 6) which strongly depends on the structural characteristics of a binding protein that is available for a given analyte. If the terminus of the binding protein through which it is fused to the RET donor is close to the analyte binding site, the distance D1 will be small and the construct does not have to be optimized further by protein engineering. Yet, if neither of the two termini are suitable, it is sometimes possible to circular permute the binding protein to create new termini closer to the binding site of the analyte (Fig. 7). Circular permutation of the binding protein may, however, be cumbersome, and may affect the properties of the binding protein.

Fig. 6 Sensor geometry optimization. D1 refers to the distance in space between the RET donor and acceptor in the closed state, and can be optimized by changing the linker length L1. L1 refers to the distance along the polypeptide chain covered by parts of the synthetic tether, the ligand-binding domain and the connecting linker (highlighted in blue). Similarly, D2 refers to the distance in space in between the RET donor and acceptor in the open state and can be optimized by increasing L2. L2 refers to the distance along the polypeptide chain covered by parts of the synthetic tether, the SNAP tag, and the connecting linker (highlighted in orange).
The distance $D_2$ between the RET donor and RET acceptor in the open state of the sensor can be controlled by modifying the linker $L_2$ which connects the SNAP-tag and the RET donor. Since simple GGS linkers are conformationally flexible, they are not optimal as spacers [23]. Instead, polyproline linkers can be used, as they form rigid helical structures [24]. In our experience, a 30-proline linker is sufficient to reduce RET efficiency in the open state [25].

A list of all SNIFIT and LUCID sensors developed to date is given (Table 1) while the following protocol provides a step-by-step summary how to construct and apply individual sensors.

### 2 Materials

#### 2.1 Bacterial Expression and Purification of SNIFITS and LUCIDs

1. RosettaGami DE3 pLysS electrocompetent *E. coli* cell.
2. LB medium: 5 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract pH 7.5.
3. 100 mg/mL ampicillin stock solution (1000×), filter-sterilized.
4. LB-agar plates supplemented with 100 μg/mL ampicillin.
5. 1 M isopropyl-β-thiogalactoside (IPTG) stock (1000×), filter-sterilized.
6. Sonicator or French press.

---

*Fig. 7* An example of the binding protein dihydrofolate reductase from *E. coli* with bound ligand. The crystal structure of eDHFR (PDB ID: 1RA3) is represented in *gray cartoon*, the methotrexate analyte is represented in *gray sticks*. The site of derivatization for the intramolecular ligand is highlighted by an *arrow*. The sites for circular permutation are shown as *dashed lines*.
Table 1
Summary of SNIFIT and LUCID sensors published to date

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Use</th>
<th>Design</th>
<th>Intramolecular ligand</th>
<th>Mechanism</th>
<th>Dynamic range</th>
<th>C50</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCA inhibitors (μM) and Zn2+ (nM)</td>
<td>SNIFIT in vitro and cell surface</td>
<td>SNAP-mCherry-HCA</td>
<td>Sulfonamides</td>
<td>FRET</td>
<td>390% (iv)</td>
<td>220–600 μM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SNAP-CLIP-HCA</td>
<td></td>
<td>dye-fluorescent protein Cy5-mCherry</td>
<td>230% (cells)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SNAP-PP-CLIP-HCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamate (high μM)</td>
<td>SNIFIT in vitro and cell surface</td>
<td>SNAP-PP15-CLIP-iGluR5-S1S2</td>
<td>Glutamate</td>
<td>FRET</td>
<td>90% (iv)</td>
<td>12 μM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dye-dye</td>
<td>56% (cells)</td>
<td></td>
</tr>
<tr>
<td>γ-Aminobutyric acid (GABA) (μM–mM)</td>
<td>SNIFIT cell surface</td>
<td>SNAP-CLIP-GB1α and GB2 coexpressed</td>
<td>GABA&lt;sub&gt;B&lt;/sub&gt; receptor antagonist CGP 51783</td>
<td>FRET</td>
<td>80% (cells)</td>
<td>100 μM</td>
</tr>
<tr>
<td>ACh and AChE inhibitors (nM–mM)</td>
<td>SNIFIT in vitro and cell surface</td>
<td>AChE-CLIP-(GPGGA)&lt;sub&gt;G&lt;/sub&gt;-SNAP</td>
<td>Decamethonium or edrophonium-C13</td>
<td>FRET</td>
<td>65%</td>
<td>7 nM–10 mM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dye-dye</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cy5-Cy3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methotrexate (nM–mM)</td>
<td>LUCID</td>
<td>SNAP-PP30-NLuc-cpDHFRL24G5</td>
<td>Trimethoprim/methotrexate</td>
<td>BRET</td>
<td>1340%</td>
<td>0.75–85 μM</td>
</tr>
<tr>
<td>Tacrolimus and sirolimus (nM)</td>
<td>LUCID</td>
<td>SNAP-PP30-NLuc-FKBPI2</td>
<td>Two-headed ligand</td>
<td>BRET</td>
<td>460%</td>
<td>17 nM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cy3-Nluc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclosporine A (nM)</td>
<td>LUCID</td>
<td>SNAP-PP30-NLuc-cphuman cyclophilin R149G5</td>
<td>Mixed urea</td>
<td>BRET</td>
<td>192%</td>
<td>500 nM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cy3-Nluc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topiramate (μM)</td>
<td>LUCID</td>
<td>SNAP-PP30-NLuc-HCA</td>
<td>Parasulfonamide</td>
<td>BRET</td>
<td>591%</td>
<td>7.3 μM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TMR-Nluc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digoxin</td>
<td>LUCID</td>
<td>DIGI03-NLuc-PP30-SNAP-</td>
<td>Progesterone</td>
<td>BRET</td>
<td>485%</td>
<td>22 nM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dye-dye</td>
<td></td>
<td>TMR-Nluc</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7. Ni²⁺-NTA agarose slurry.
8. Propylene column.
9. Ni²⁺-NTA lysis buffer: 50 mM KH₂PO₄, pH 8, 150 mM NaCl, 5 mM imidazole.
10. Phenylmethylsulfonyl fluoride (PMSF): 100 mM in ethanol.
11. Lysozyme: 10 mg/mL in water.
12. Ni²⁺-NTA wash buffer: 50 mM KH₂PO₄, pH 7.5, 150 mM NaCl, 10 mM imidazole.
13. Ni²⁺-NTA elution buffer: 50 mM KH₂PO₄, pH 7.5, 150 mM NaCl, 500 mM imidazole.
15. Streptavidin-resin in propylene column, such as the Strep-Tactin column.
16. Streptavidin wash buffer: 100 mM Tris–HCl, pH 8, 150 mM NaCl, 1 mM EDTA.
17. Streptavidin elution buffer: 100 mM Tris–HCl, pH 8, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin.
18. Centrifugal filter device with appropriate molecular weight cutoff.
19. Storage buffer: 50 mM HEPES, 50 mM NaCl, pH 7.2.

2.2 In Vitro Labeling and Characterization

1. Labeling buffer: 50 mM HEPES, 50 mM NaCl, pH 7.2, 1 mg/mL bovine serum albumin (BSA).
2. Fluorescein or tetramethylrhodamine (TMR) labeled ligand stock solution in DMSO, for characterizing the binding strength by means of fluorescence polarization.
3. BG-fluorophore-ligand conjugate stock solution in DMSO, for labeling SNAP-tag based sensors with an intramolecular tether.
4. BC-fluorophore-ligand conjugate stock solution in DMSO, for labeling CLIP-tag-based sensors with an intramolecular tether.
5. Nonbinding multiwell plates (black or white).
6. Plate reader with filters for fluorescent polarization and fluorescent measurements.

2.3 Cell-Surface Expression

1. Glass coverslips (Ø 15 mm).
2. Ethanol.
3. Poly-L-lysine solution: 0.7 mg/mL in sterile water.
4. HEK293 cells.
6. Opti-MEM® I reduced serum medium.
7. Dulbecco’s Modified Eagle Medium (DMEM) Glutamax medium supplemented with 10% fetal bovine serum.
8. Hank’s Balanced Salt Solution (HBSS) supplemented with 10 mg/mL of BSA.

3 Methods

3.1 Generation of Fusion Proteins for Expression in E. coli

Standard methods of cloning can be used to construct the expression plasmids for the recombinant fusion proteins. We initially utilized restriction-enzyme cloning and ligation methods, but have recently started to use newer ligation methods such as Gibson assembly [26].

We advise using a pET51b(+) (Novagen) expression vector that contains a double-purification system of a His-tag and a Strep-tag. Other purification systems can be used, but it is important to use affinity purification tags on both the N- and C-termini of the protein to prevent purifying truncated versions of the sensor. We provide here a general protocol for the expression of sensors, but modification of the method may be necessary if the binding protein is unstable or insoluble.

1. Clone the sensor into a bacterial expression vector using standard cloning methods (see Note 1).
2. Transform RosettaGami DE3 pLysS E. coli cells for protein production, plate cells onto selective LB agar medium, and let colonies grow at 37 °C overnight.
3. Pick a single colony and use it to inoculate 2 mL of selective LB medium and allow the bacteria to grow overnight, with shaking at 220 rpm at 37 °C.
4. Inoculate 1 L selective LB medium with 1 mL of the preculture and incubate (220 rpm at 37 °C) until an OD600nm of 0.6-0.8 is reached.
5. Cool the incubator to the expression temperature and induce the expression of the protein depending on the promoter used (see Note 2).
6. After expression, harvest the cells by centrifugation and resuspend the bacteria in a total of 30 mL of Ni²⁺-NTA lysis buffer supplemented with 1 mM PMSF and 0.25 mg/mL of lysozyme.
7. Lyse cells by sonication on ice.
8. Centrifuge at 40,000 × g for 10 min at 4 °C to remove cell debris and insoluble proteins.
9. Incubate the supernatant with 1 mL Ni²⁺-NTA agarose slurry under agitation for 1 h at 4 °C. Then centrifuge gently and transfer the resin onto a propylene column.
10. Wash the Ni\textsuperscript{2+}-NTA agarose with at least six column volumes of cold Ni\textsuperscript{2+}-NTA wash buffer.

11. Elute the protein with 500 μL aliquots of Ni\textsuperscript{2+}-NTA elution buffer. Estimate the protein concentration of each fraction using Bradford Reagent and/or absorbance at 280 nm.

12. Take aliquots of the eluate of the Ni\textsuperscript{2+}-NTA column to follow and analyze the purification process by SDS-PAGE (see Note 3).

13. Load all of the eluted fractions that contain protein onto a StrepTactin-column.

14. Wash the column using the Strep-tag washing buffer. Wash with six column volumes.

15. Elute the protein using 500 μL fractions of Strep-tag elution buffer (for a 1 mL column). Estimate the protein concentration of each fraction using a Bradford quantification and/or absorbance at 280 nm.

16. Take aliquots of each step and run an SDS-PAGE to verify the purity of the protein.

17. Concentrate the protein using molecular-weight dependent filtration devices, with the recommendation of the provider, and change the buffer to storage buffer. Dilute the sample in a 1:1 ratio with glycerol 87% and store it at –20 °C at a concentration between 20 and 200 μM.

3.2 Chemical Synthesis of the Intramolecular Tether

The synthesis of the intramolecular tether heavily depends on the chemical properties of the intramolecular ligand. In one approach, each component of the intramolecular tether (i.e., the BG-reactive group, fluorophore, and intramolecular ligand) can be synthesized separately in a modular approach and the tether subsequently assembled using standard chemical methods such as amide bond formation, click chemistry, and alkylation. For examples, see [22, 25, 27–30]. When making the intramolecular tether, it is worth making also a ligand that only carries a fluorophore and can subsequently be used to measure the $K_d$ by means of fluorescence polarization measurements.

3.3 Determining the Affinity of the Intramolecular Ligand

Many methods can be used to determine the affinity of the receptor protein for both the intramolecular ligand and the free analyte. We routinely use fluorescence polarization to determine both of these parameters.

1. Prepare serial dilutions of the receptor protein in labeling buffer that contains 20 nM fluorescein or tetramethylrhodamine (TMR)-labeled ligand.

2. Pipette 100 μL of each dilution into the wells of a black 96-well plate.

3. Incubate the plate for 10 min at room temperature.
4. Record the fluorescence polarization signal using a suitable plate reader.

5. Plot the fluorescence polarization response against the protein concentration, and fit the curve to a single-site binding isotherm to determine the $K_d$ of the receptor protein for the derivatized ligand:

$$F = F_{\text{min}} + \frac{F_{\text{max}} - F_{\text{min}}}{K_{d,\text{fluorescent ligand}}} \frac{1}{1 + \frac{[\text{protein}]}{K_{\text{d,fluorescent ligand}}}}$$

6. Dilute the protein to a concentration close to the $K_d$ for the labeled ligand in the same labeling buffer with 20 nM fluorescein- or TMR-labeled ligand; then add serial dilutions of the free ligand analyte.

7. Record the fluorescence polarization signal using a suitable plate reader.

8. Determine the concentration of unbound protein at each analyte concentration from the equation of the single-site binding isotherm (using the values for $K_{d,\text{fluorescent ligand}}$, $F_{\text{min}}$, and $F_{\text{max}}$ determined above):

$$[\text{protein}_{\text{free}}] = K_{d,\text{fluorescent ligand}} \frac{F_{\text{min}} - F}{F - F_{\text{max}}}$$

9. Plot the concentration of free protein against the analyte concentration and fit the curve to a single-site binding isotherm to determine the $K_d$ of the receptor protein for the free analyte.

### 3.4 In Vitro Labeling

1. Label the recombinant fusion protein with synthetic molecules. Common concentrations are 1 μM sensor protein with 2 μM BG-based substrate for SNAP-tag, and 10 μM BC-based substrate for CLIP-tag in labeling buffer (see Note 4).

2. For SNIFITs, remove the excess labeling reagents using three washes with labeling buffer in centrifugal filtering devices, or using a small gel filtration column. This is not needed for LUCIDs.

3. Quantify the amount of labeled protein by estimating the concentration of dye from its specific absorbance and comparing it to the protein absorbance at 280 using a spectrophotometer, such as the nanodrop.

4. The quantitative labeling of the sensor, assuming fully functional SLPs, can also be verified by SDS-PAGE and in-gel fluorescence scanning. To this end, incubate the labeled sensor with an additional fivefold excess of a BG- or BC-based derivative labeled with a second fluorophore for 30 min at room temperature, and then separate the samples by SDS-PAGE.
3.5 In Vitro Titrations

1. Prepare a suitable dilution of the sensor in labeling buffer and add 98 μL into each well of a 96-well plate. Use nonbinding black 96-well plates for fluorescence and white plates for luminescence measurements. We find that 50 nM sensor is suitable for SNIFITs and 5 nM for LUCIDs.

2. Prepare serial dilutions of the analyte of interest in DMSO, and add 2 μL of the dilutions to each well with the sensor.

3. Incubate for 1 h at room temperature, or until the sensor is fully open (see Note 5).

4. Record spectra on a spectrophotometer exciting the FRET donor, or add luciferase substrate to give an adequate concentration. We usually use a 400-fold dilution in storage buffer of Nano-glo® substrate from Promega stock with LUCIDs.

5. Plot the intensity ratio \( R \) between the RET donor and RET acceptor against the sensor concentration, and fit it to a competitive single binding site isotherm.

\[
R = \frac{I_{\text{donor}}}{I_{\text{acceptor}}}
\]

\[
R = R_{\text{zero}} + \frac{R_{\text{sat}} - R_{\text{zero}}}{1 + \left[ \frac{[\text{Analyte}]}{e_{50}} \right]}
\]

3.6 Application of SNIFITs and LUCIDs on the Cell Surface

1. Clone the sensor of interest into a pDisplay mammalian expression vector using standard cloning protocols. We usually use transient transfection for sensor evaluation on cell surfaces, but semistable cell lines can also be prepared.

2. Sterilize glass coverslips (Ø 15 mm) by dipping them into an ethanol bath and briefly holding them in a Bunsen burner flame.

3. Place the sterile coverslips into 12-well cell culture plates.

4. Coat the coverslips by adding 1 mL poly-L-lysine solution (0.7 mg/mL in sterile water) per well. Incubate the plate overnight at 37 °C.

5. Remove the poly-l-lysine and wash the cover slips once with 1 mL sterile water.

6. Seed HEK293 cells to around 30% confluency in DMEM Glutamax medium supplemented with 10% fetal bovine serum, and grow at 37 °C and 5% CO₂ in a humidified incubator.

7. The next day, transfect the cells with the plasmid using a transfection agent of choice. We routinely use Lipofectamine 2000. Dilute the target DNA to a final concentration of 0.8 μg/mL in Opti-MEM® I reduced serum medium and add Lipofectamine 2000 to a final concentration of 1 μL/mL. Allow to stand at room temperature for 20 min to form a pre-complex. Dilute a
115

further tenfold and add 100 μL of the Lipofectamine-DNA mixture to each well and allow to stand for 3–6 h at 37 °C and 5% CO₂.

8. Optional step: Exchange the medium after 4–6 h to remove the transfection reagent.

9. 24–48 h after transfection, label the sensor by replacing the medium with HBSS supplemented with 10 mg/mL BSA and 2 μM BG-based and 10 μM BC-based fluorophore-ligand label. In order to avoid using large amounts of the synthetic compounds, the glass coverslips can be transferred from the 12-well plate onto a sheet of parafilm and 100 μL of the labeling solution carefully pipetted onto the edge of the coverslip.

10. After 10 min of labeling at room temperature, aspirate the labeling solution and wash the glass coverslip four times with HBSS.

11. For perfusion experiments, transfer the coverslip into a flow cell. We use a Warner imaging chamber RC-20 and gravity-fed perfusion at a flow rate of 0.5 mL/min. If no perfusion system is available, the analyte can be added to the cells by pipetting (see Note 6).

12. Image the sensor using a standard wide-field or confocal fluorescence microscope.

4 Notes

1. The polyproline region can be difficult to clone, as it is highly repetitive. We have cloned the region by annealing 5′-phosphorlated oligonucleotides and ligating them into a restriction-digested plasmid.

2. We use pET vector expression systems that are induced by IPTG (0.5–1 mM final concentration) and expressed at 16 °C overnight.

3. You can follow the sensor in the different fractions by labeling it with a fluorescent substrate and doing an in-gel fluorescent scan.

4. We recommend keeping the DMSO concentration below 5%, as higher concentrations may denature the protein. The duration of labeling depends on the intramolecular ligand, but we find that the reaction is finished after 1 h.

5. The kinetics of opening may differ between sensors, depending on the affinity of the intermolecular tether for the binding protein.

6. We find that measuring the half-life of $t_{\text{opening}}$ and $t_{\text{closing}}$ on cell surfaces is often limited by the perfusion speed of the perfusion chamber.
References


Chapter 8

Ultrasensitive Firefly Luminescent Intermediate-Based Protein-Protein Interaction Assay (FlimPIA) Based on the Functional Complementation of Mutant Firefly Luciferases

Yuki Ohmuro-Matsuyama and Hiroshi Ueda

Abstract

We recently developed a protein-protein interaction assay, FlimPIA (Firefly luminescent intermediate-based Protein-protein Interaction Assay) based on the catalytic mechanism of firefly luciferase (Fluc) that can be divided into two half-reactions: the adenylation step and the oxidative luminescent steps. We engineered two mutant Fluc enzymes named “Donor” and “Acceptor” where the oxidative luminescent activity of the Donor is almost eliminated and the adenylation activity of the Acceptor is suppressed. When the Donor and the Acceptor are each fused to one of two interacting partners, and put together to interact, the Donor and the Acceptor come sufficiently close such that the Acceptor can react with the luciferyl-adenylate intermediate (LH₂-AMP) produced by the Donor, and thus emit luminescence.

FlimPIA can be used in vitro and in cultured cells. Owing to recent improvements, it has several advantages in terms of signal/background ratio, detectable size of interacting partner, and sensitivity over conventional protein-protein interaction assays based on Förster/fluorescence resonance energy transfer and protein-fragment complementation performed in vitro. Here, we describe a protocol to make use of the latest version of FlimPIA which shows even lower background and higher signal than previously described ones.

Key words Protein-protein interaction, Firefly luciferase, Substrate channeling, Reaction intermediate, Sensitivity, Detectable distance, Signal-background ratio

1 Introduction

Development of protein-protein interaction (PPI) assays is considered highly important in various fields, such as molecular and cellular biology, diagnostics, and drug screening. Recently, we developed a novel protein-protein interaction assay that we termed FlimPIA (Firefly luminescent intermediate-based Protein-protein Interaction Assay) based on the following unique detection principle [1].
The method is based on the multiple (two half) reactions catalyzed by firefly luciferase (Fluc). Fluc is frequently applied as a reporter enzyme in biological assays because of its sensitivity, dynamic range, rapidity, and easy-to-handle measurement. In Fluc-catalyzed reactions, firefly \(\delta\)-luciferin (LH\(_2\)) is converted to luminescent oxyluciferin (OxL). This reaction can be subdivided into two half-reactions, namely the adenylation step and the oxidative luminescent steps (Fig. 1a). In the former step, a luciferyl-adenylate intermediate (LH\(_2\)-AMP) is produced from LH\(_2\) and ATP, and in the latter OxL is produced via proton abstraction at the C4 carbon of LH\(_2\)-AMP. We made two mutant Flucs named the “Donor” and the “Acceptor.” The Donor is a triple mutant H245D/K443A/L530R based on the most commonly used \textit{Photinus pyralis} residue.

Fig. 1 Principle of FlimPIA (a) Chemical reactions catalyzed by Fluc. Fluc produces excited state oxyluciferin (OxL) from \(\delta\)-luciferin (LH\(_2\)) by a two-step catalysis, the adenylation step, and the following oxidative steps. (b) Working mechanism of FlimPIA. The adenylation activity of the Donor is suppressed, on the other hand, the oxidative activity of the Acceptor is suppressed. Donor and Acceptor are enough close that the Acceptor use luciferyl-AMP (LH\(_2\)-AMP) produced by Donor when binding domains are interacting. Adapted with permission from Fig. 1c in Ref. 2. Copyright John Wiley and Sons.
number. For this mutant, the oxidative luminescent reaction is synergistically suppressed by the H245D/K443A mutations while the adenylation reaction is enhanced by the L530R mutation. As a result, the Donor can efficiently produce LH₂-AMP, but can only poorly produce OxL. On the other hand, the Acceptor that originally featured a K529Q mutation had a lower adenylation activity while maintaining full oxidative luminescent activity. Thus, the Acceptor cannot produce LH₂-AMP efficiently on its own.

To perform the interaction assay, the Donor and the Acceptor are each fused to one of two interacting proteins of interest, respectively. When they do not interact with each other, the Donor and the Acceptor are far apart, and the luminescence intensity remains low. However, when they are brought into close proximity by the interacting protein pair, the Acceptor can react with LH₂-AMP that was produced by the Donor. As a result, the luminescence that is emitted is increased (Fig. 1b). The splitting of two half reactions is made possible partly due to the conformational change of the Fluc enzyme, which accompanies rotation of the C-terminal domain.

So far, several attempts to improve the PPI assay FlimPIA have been achieved, and several advantages over conventional PPI assays have been discovered [1–3]. First, FlimPIA generally works well in vitro as well as in cultured cells. Second, compared to conventional Fluc-based PPI assays based on protein-fragment complementation assays (PCA), the FlimPIA probes are more thermostable and reproducibly give higher luminescence. Third, the distance that can be detected between the probes is longer than the distance between the probes for Förster/fluorescence resonance energy transfer (FRET)-based PPI assays and PCA. In practice, this means that larger interacting proteins can be detected by FlimPIA. Fourth, the sensitivity is approximately tenfold higher than Fluc-based PCA both in terms of the detection limit and signal intensity in vitro. Finally, using a recently discovered double mutant R437K/K529Q as a new Acceptor significantly improved signal/background (S/B) ratio as well as the signal intensity under optimized assay conditions. The mutation of residue R437 which is located in the hinge region connecting the N-terminal and C-terminal domains was found to selectively lower adenylation activity. The obtained maximum S/B ratio (> 40) is higher than that obtained with Fluc PCA under the same experimental reaction conditions. In this chapter, a protocol to generate and assay this latest version of FlimPIA is described [2].

2 Materials

Prepare all solutions using ultrapure water (made by purifying deionized water to attain resistivity of 18 MΩ cm at 25 °C). For all the reagents, use them of the highest grade available.
1. 100 mM ATP stock solution made by adjusting pH with sodium hydroxide (NaOH) to 7.2, and stored at −80 °C.
2. 20 mM Firefly d-luciferin (LH2) (available from Biosynth, Staad, Switzerland) stock solution made by adjusting pH by NaOH to 7.3, and stored at −20 °C.
3. QuikChange Site-Directed Mutagenesis kit and/or QuikChange Multi Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA).
4. In-Fusion HD cloning kit (Takara-Bio, Otsu, Shiga, Japan).
5. E. coli JM109 competent cells.
6. LB agar: 5 g/L yeast extract, 10 g/L tryptone, 10 g/L NaCl and 15 g/L agar.
7. LB liquid medium: 5 g/L yeast extract, 10 g/L tryptone, and 10 g/L NaCl.
8. LBA agar plate: LB agar supplemented with 100 μg/mL ampicillin.
9. LBA liquid medium: LB liquid medium supplemented with 100 μg/mL ampicillin.
10. E. coli SHuffle T7 express lysY competent cells (New England Biolabs Japan, Tokyo, Japan).
11. 10 mM Rapamycin (e.g., LKT Laboratories, St. Paul, MN) stock solution dissolved in methanol or dimethyl sulfoxide, stored at −20 °C.
12. 1 M Isopropyl-β-thiogalactopyranoside (IPTG) stock solution for inducing protein expression.
13. White half area 96-well plate (Costar 3694, Corning-Costar, Osaka, Japan).
14. A plasmid encoding Photinus pyralis Fluc cDNA such as pGEX-Ppy or pGEM-luc (Promega, Madison, WI).
15. pET32b (Novagen, Merck, Darmstadt, Germany).
16. TALON metal affinity resin and TALON disposable gravity column (Clontech, Takara-Bio).
17. Toyopearl HW-40S gel filtration resin (Tosoh, Tokyo, Japan).
18. Synthetic genes coding for binding domains such as E. coli codon-optimized human FK506-binding protein 12 (FKBP12) and FKBP-rapamycin-associated protein (FRB), appended with NcoI/SfiI and NotI sites at the 5′ and 3′ ends, respectively.
19. TALON extraction buffer: 300 mM sodium chloride (NaCl), 50 mM sodium phosphate, adjusted to pH 7.0.
20. TALON elution buffer: 150 mM of imidazole in TALON extraction buffer.
21. MOPS buffer: 100 mM MOPS, 10 mM MgSO₄, pH 7.3 (see Notes 1 and 2).
22. 20 mM coenzyme A trilithium salt stock solution dissolved in water, and stored at −20 °C.
23. For the determination of enzyme kinetics, use a multiwell luminometer such as AB-2350 equipped with a sample injector (ATTO, Tokyo, Japan).
24. For the calculation of enzyme kinetics, use a software that can perform nonlinear curve fitting, such as Kaleida Graph 4.1 (Synergy Software, Reading, PA).
25. For FlimPIA with faster kinetics, use a tube-based luminometer such as NU-2600 equipped with a sample injector and a stirrer (Microtech Nichion, Chiba, Japan).
26. Optional: For the measurement of LH2-AMP concentration, use a fluorescence spectrometer such as FP-8500 (Jasco, Tokyo, Japan).

3 Methods

3.1 Construction of the Binding Domain-Fused Expression Vectors (Fig. 2)

1. Amplify the DNA fragment encoding FLuc derived of Photinus pyralis (Ppy) by PCR using pGEX-Ppy [4] or pGEM-luc vector as a template, and primers NotG4SBack (with G4S linker and a NotI site) and XhoFor (with an XhoI site).
2. Digest the amplified fragment with restriction enzymes NotI and XhoI, and subclone into the multicloning site of pET32b between NotI and XhoI sites.
3. Prepare a pair of DNA fragments each encoding a binding domain (BD) of your interest, which are appended with restriction sites NcoI and NotI by PCR. Digest the product with the same enzymes, and ligate each with the plasmid prepared in 2 that was digested with NcoI and NotI (see Notes 3 and 4). Here, the procedure is exemplified with FKBP12 and FRB that are denoted as BD1 and BD2, respectively (see Note 5).
4. Transform JM109 competent E. coli cells with the ligation mixture.
5. Culture the cells on LBA agar plates and incubate overnight at 37 °C.
6. Pick a colony and culture it in 4 mL LBA medium and incubate overnight at 37 °C.
7. Extract the plasmid, and confirm the nucleotide sequence of the entire open reading frame.

3.2 Construction of BD1/Donor Expression Vector (Fig. 2a)

1. To make H245D/K443A/L530R/E354K mutants as the Donor, conduct a series of site-directed mutagenesis with QuikChange Site-directed mutagenesis kit as per manufacturer’s instructions. Use pET32/BD1/Fluc as the template and primers H245D, K443A, L530R, and E354K (Table 1) in combination with their complementary strand primers (see Notes 6 and 7).
2. After thermal cycling, add 1 μL of DpnI (~10 U) to the reaction mixture to digest the methylated template by incubating for 1 h at 37 °C. Transform JM109 competent cells with a part of the digested template.

3. Culture the cells on LBA agar plates.

4. Pick several colonies and culture each colony in 4 mL LBA medium.

5. Extract the plasmid and confirm the correct DNA sequence.

### 3.3 Construction of the BD2/Acceptor Expression Vector (Fig. 2b)

1. To make a K529Q/E354K mutant, perform QuikChange Site-directed mutagenesis using pET32/BD2/Fluc as a template and primer pairs K529Q and E354K (as in Subheading 3.2, step 1) in combination with their complementary strands (see Notes 7 and 8).
2. Follow the steps 2–5 in Subheading 3.2.

3. Amplify the DNA fragment using primers FlucEcoBack and FlucPacR437KF or using the plasmid constructs generated in steps 1 and 2 as the template.

4. Digest the plasmid constructs generated in steps 1 and 2 with EcoRI and PacI.

5. Recombine the fragments of steps 3 and 4 using an In-Fusion HD cloning kit as per manufacturer’s instructions.

3.4 Expression and Purification of the BD1/Donor and the BD2/Acceptor Sensitive Protein-Protein Interaction Assay

1. Transform *E. coli* SHuffle T7 Express lysY competent cells with pET32/BD1/Donor and pET32/BD2/Acceptor (*see Note 9*).

2. Incubate the transformed cells on a LBA agar plate at 30 °C for 24 h.

3. Pick a colony, and culture it in 4 mL LBA medium at 30 °C.

4. Centrifuge 0.4 mL of the resultant cell culture and collect the cell pellet (*see Note 10*).

5. Resuspend the cell pellet with 100 mL of LBA medium in a 500 mL flask, and culture at 30 °C.

6. When the OD600 reaches 0.4–0.6, add IPTG to a final concentration of 40 μM and culture the cells at 16 °C.

7. Centrifuge and collect the cells, and resuspend the cell pellet with 10 mL of the TALON extraction buffer.

8. Sonicate the cells on ice using an ultrasonifier for 2 min (50% interval) five times (*see Note 11*).

9. Centrifuge the sonicated solution at 11000 × *g* for 20 min at 4 °C.

10. Add 200 μL of TALON Metal Affinity Resin to the supernatant, and rotate or shake gently for 20 min at 25 °C.

11. Centrifuge at 700 × *g* for 2 min at 4 °C, and resuspend the resin with 10 mL of TALON extraction buffer.

12. Rotate or shake gently for 10 min at 4 °C.

13. Repeat steps 11 and 12 two times.

14. Centrifuge at 700 × *g* for 2 min and collect the resin.

15. Suspend the resin with 3 mL of TALON extraction buffer, and transfer the suspension into a TALON 2 mL Disposable Gravity Column.

16. Allow to drain until it reaches the top of the resin bed.

17. Wash the column with 2 mL of TALON extraction buffer.

18. Elute the enzyme with 500 μL of TALON elution buffer (*see Note 12*).
3.5 Chemical Synthesis and Purification of LH₂-AMP (See Note 13) [5, 6]

1. Dissolve 15 mg of adenylic acid (AMP) and 5 mg of LH₂ in 1 mL of dimethyl sulfoxide.
2. Dissolve 100 mg of dicyclohexylcarbodiimide in 1 mL of dimethyl sulfoxide.
3. Mix the solutions of 1 and 2.
4. Incubate for 10 min at room temperature (see Note 14).
5. Add 5 mL of acetone to stop the reaction.
6. Centrifuge and carefully remove the supernatant.
7. Add 3 mL of acetone.
8. Centrifuge and carefully remove the supernatant.
9. Repeat steps 7 and 8.
10. Dissolve the precipitate in 1.5 mL of the solution of acetic acid (10 mM) and NaCl (40 mM).
11. Apply to the column containing the resin of Toyopearl HW-40S (bed volume ~2 mL).
12. Collect the fraction by adding 1 mL of diluted hydrochloric acid (pH 4.5) (see Notes 15 and 16).
13. Repeat step 12 10–15 times.
14. Measure the fluorescence spectrum of each fraction in steps 12 and 13 with excitation wavelength of 327 nm, and collect the fractions that emit fluorescence peaking at around 535 nm (see Note 17) [7]. When the peak values of LH₂-AMP and x nM of LH₂ are y and z, respectively, the concentration of LH₂-AMP can be determined as follows (see Note 18).

\[
\text{The concentration of LH₂-AMP (nM) } = \frac{x \times y}{z \times 0.45}.
\]

3.6 LH₂ Kinetic Determination of BD1/Donor and BD2/Acceptor (See Note 19)

1. Prepare substrate solutions with varied concentrations 1–2000 μM of LH₂ in the MOPS buffer containing 20 mM of ATP (see Note 20).
2. Dispense 50 μL of the solution to a well of 96-well plate.
3. Measure the bioluminescence intensity after injection of 50 μL of 50 nM of the probe (BD1/Donor or BD2/Acceptor) diluted in the MOPS buffer with 0.1-s interval for 10 s.
4. Use the maximal light intensity when each concentration of LH₂ was applied and perform nonlinear curve fitting to derive \( V_{\text{max}} \) and \( K_m \) using Kaleida Graph 4.1.

3.7 ATP Kinetic Determination of BD1/Donor and BD2/Acceptor (See Note 19)

1. Prepare solutions with varying concentrations (3–2000 μM) of ATP in MOPS buffer containing 10 mM of LH₂ (see Note 20).
2. Follow the steps 2–4 in Subheading 3.6.
1. Prepare solutions with varying concentrations (2–10,000 nM) of LH$_2$-AMP in MOPS buffer.

2. Follow the steps 2–4 in Subheading 3.6.

3.8 LH$_2$-AMP Kinetic Determination of BD1/Donor and BD2/Acceptor (See Note 21)

3.9 FlimPIA

1. Suspend equimolar BD1/Donor and BD2/Acceptor (20–500 nM) with varying concentrations of rapamycin in the MOPS buffer (see Note 3).

2. Dispense the mixture (50 μL) to an assay tube for a luminometer.

3. Measure the light intensity after injecting 450 μL of 1.1 mM ATP, 83 μM of LH$_2$, and 1.1 mM coenzyme A with 10-msec interval for few seconds (Fig. 3, see Notes 22–25).

4 Notes

1. MgSO$_4$ is essential for the adenylation step (Fig. 1a).

2. Emission intensity and color of light is strongly influenced by pH.

3. A BD should be fused to the N-terminus of the Donor or the Acceptor. When the BD pair is fused to the C-terminus of the Donor and the Acceptor, a smaller increase in light intensity in response to an interaction is observed. Fluc consists of structurally distinct N- and C-terminal domains. Furthermore, the
C-terminal domain rotates as the reaction proceeds from the
adenylation to the oxidative luminescent steps. As a result,
when the BD is fused to the C-terminus, its rotation might be
hampered.

4. Thioredoxin (Trx), which is originally encoded in the pET32
vector, is fused to the N-terminus of BD. Trx is used to increase
the protein solubility in the bacterial expression process.

5. FRB associates with FKBP12 in a rapamycin-dependent man-
ner [8–10]. Rapamycin is known to act as immunosuppres-
sant drug, anticancer agent, and accelerating agent for
autophagy.

6. H245 and K443 are essential for the oxidative reaction of Fluc,
and the mutations H245D and K443A selectively suppress the
oxidative activity [5, 11, 12]. The mutation L530R stabilizes
bound ATP in Fluc to accelerate adenylation activity [13]. E354K
is a mutation that confers additional thermostability [14].

7. QuikChange Multi Site-Directed Mutagenesis kit is also
useful.

8. K529 is a residue that stabilizes phosphates of bound ATP, and
the K529Q mutation selectively suppresses the adenylation
activity [15]. The hinge region residue R437 is located close
to the active site of Fluc in the adenylation conformation.
The mutation R437K further decreases the distance between
the residue and LH2, and thus suppresses the adenylation
activity [2].

9. SHuffle T7 Express lysY is an engineered E. coli B strain in
which DsbC is constitutively expressed. DsbC works as a chap-
erone to promote the folding of protein.

10. To remove secreted β-lactamase from the precultured E. coli,
the medium of the preculture should not be added to the main
culture.

11. From this step the samples must be kept on ice or at 4 °C.

12. The purified enzyme can be stored at −80 °C in the buffer
containing 15% glycerol for several months.

13. The synthesis and use of LH2-AMP is optional. It is used to mea-
sure LH2-AMP kinetics of BD1/Donor and BD2/Acceptor.

14. The color of the solution turns yellow when the reaction is
completed.

15. LH2-AMP is stable in an acidic condition.

16. The solution for elution should be cooled because LH2-AMP
is unstable.

17. The peak of the fluorescent spectrum of LH2-AMP is at
approximately 535 nm, and the peak of the spectrum of LH2 is
at 529 nm; therefore, the fractions containing LH₂-AMP can be separated from the fractions containing LH₂.

18. LH₂-AMP can be stored at −80 °C for several months.

19. The adenylation activity of BD1/Donor should be fully maintained while that of BD2/Acceptor should be suppressed as much as possible. Therefore, these kinetics are advised to be determined before performing FlimPIA.

20. When the kinetics for one substrate are determined, the concentration of the other substrate should be maintained under saturating conditions.

21. The oxidative luminescent activity of BD1/Donor should be suppressed, while that of BD2/Acceptor should be fully maintained. Therefore, these LH₂-AMP kinetics are preferably determined before performing FlimPIA.

22. The maximum S/B ratio is usually observed 1 s after injecting substrates. The longer reaction time gives rise to higher background luminescence due to the accumulation of LH₂-AMP in the reaction mixture.

23. When the substrate solution is injected into the reaction mixture, rapid mixing gives reliable result of light emission. Therefore, we recommend using a tube-based luminometer such as NU-2600 with stirring the mixture at top speed. However, a 96-well luminometer also gives a reasonably good result.

24. Dehydroruniferyl-AMP (L-AMP), which is spontaneously produced from LH₂-AMP, acts as a competitor of LH₂-AMP and inhibits light emission. Coenzyme A converts L-AMP to dehydroruniferyl-coenzyme A, which does not act as a competitor of LH₂-AMP.

25. We previously used higher concentration (20 mM) of ATP. The ATP-based $K_m$ of the Acceptor is increased due to K529Q mutation, while that of the Donor should remain unchanged. In the optimized FlimPIA, the final concentration of ATP is reduced to 1 mM, so that the residual adenylation activity of the Acceptor is selectively suppressed to lower the background signal.

Acknowledgment

We thank B. Branchini for providing inspiration for the project. This project was supported by SENTAN, JST, Japan, and partly by Kikkoman Co., Japan, and the “Leave a Nest” Microtech Nichion award.
References


Chapter 9

Quantitative and Dynamic Imaging of ATM Kinase Activity

Shyam Nyati, Grant Young, Brian Dale Ross, and Alnawaz Rehemtulla

Abstract

Ataxia telangiectasia mutated (ATM) is a serine/threonine kinase critical to the cellular DNA-damage response, including DNA double-strand breaks (DSBs). ATM activation results in the initiation of a complex cascade of events facilitating DNA damage repair, cell cycle checkpoint control, and survival. Traditionally, protein kinases have been analyzed in vitro using biochemical methods (kinase assays using purified proteins or immunological assays) requiring a large number of cells and cell lysis. Genetically encoded biosensors based on optical molecular imaging such as fluorescence or bioluminescence have been developed to enable interrogation of kinase activities in live cells with a high signal to background. We have genetically engineered a hybrid protein whose bioluminescent activity is dependent on the ATM-mediated phosphorylation of a substrate. The engineered protein consists of the split luciferase-based protein complementation pair with a CHK2 (a substrate for ATM kinase activity) target sequence and a phospho-serine/threonine-binding domain, FHA2, derived from yeast Rad53. Phosphorylation of the serine residue within the target sequence by ATM would lead to its interaction with the phospho-serine-binding domain, thereby preventing complementation of the split luciferase pair and loss of reporter activity. Bioluminescence imaging of reporter expressing cells in cultured plates or as mouse xenografts provides a quantitative surrogate for ATM kinase activity and therefore the cellular DNA damage response in a noninvasive, dynamic fashion.

Key words ATM, Bioluminescence, Complementation, In vivo, Kinase activity, Live cell, Molecular imaging, Reporter, Split-luciferase

1 Introduction

Protein kinases constitute one of the largest gene families, comprising ~2% of the human genome. It is estimated that approximately 30% of all cellular proteins are phosphorylated on at least one residue. Thus, protein kinases have key roles in many fundamental processes of cellular signaling in cancer as well as normal cells. Biochemical methods have been widely used to investigate whether or not a protein kinase of interest is active. Although biochemical methods are robust in vitro, they generally do not provide information about protein kinase activity in specific subcellular compartments; nor do they provide information about activity changes at the single-cell level. We and others have developed
optical imaging reporters to measure the kinase activity of various oncologically important kinases ([1–11], Table 1) and have utilized these reporters in subsequent studies that lead to the identification of new inhibitors and discovery of novel signaling mechanisms [12, 13].

Bioluminescence is a chemical reaction where light is emitted by a living organism. Luciferases are a large family of light-generating enzymes that catalyze the oxidation of a substrate, generically called luciferin, to yield oxyluciferin with the concomitant production of light. For in vivo bioluminescence imaging of malignancy, tumor cells or cancer-related genes are tagged with a reporter gene that encodes a light-generating enzyme, luciferase [14–16]. When this reporter is in the presence of the substrate it emits a blue to yellow-green light with an emission spectra peaking at a wavelength between 490 and 620 nm [14]. An extremely sensitive cooled charged-coupled device (CCD) camera or a photomultiplier detects any low light that is emitted during the bioluminescence reaction. Due to its extreme sensitivity, broad dynamic range and exceptionally large signal-to-noise ratio, this type of noninvasive imaging permits a real-time analysis of an ample amount of various biological events [15]. Although there are more than 30 luciferase-luciferin systems that were derived independently of each other, the most frequently used luciferase for in vivo molecular imaging is the ATP-dependent firefly (Photinus pyralis) luciferase [17]. The reason for this is that 30% of the light produced by firefly luciferase has an emission spectra above 600 nm, a region in which the signal attenuation by the absorbing and scattering properties of live mammalian tissue is at a minimum [15, 17]. Recently, a very bright and smaller luciferase (NanoLuc; NLuc) from deep sea shrimp (Oplophorus gracilirostris) has been successfully used for dual luciferase imaging in a mouse model [18].

<table>
<thead>
<tr>
<th>Kinase</th>
<th>Substrate</th>
<th>Peptide sequence</th>
<th>Phospho peptide-binding domain</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT</td>
<td>FOXO4/AFX1</td>
<td>QSRPRSCTWPLRPPEKKK</td>
<td>FHA2</td>
<td>[9]</td>
</tr>
<tr>
<td>ATM</td>
<td>CHK2</td>
<td>LETVSTQELYSI</td>
<td>FHA2</td>
<td>[12]</td>
</tr>
<tr>
<td>EGFR</td>
<td>EPS15</td>
<td>KPANFSAYPSEEDMIE</td>
<td>SH2</td>
<td>[2]</td>
</tr>
<tr>
<td>FADD KINASES</td>
<td>FADD</td>
<td>QNRSGAMSPMSWNSDASTSEAS</td>
<td>FHA2</td>
<td>[3]</td>
</tr>
<tr>
<td>GSK3β/CKIα</td>
<td>β-CATENIN</td>
<td>SYLDSGIHSGATTTAPSLSG</td>
<td>FHA2</td>
<td>[4]</td>
</tr>
<tr>
<td>c-MET</td>
<td>PYK2</td>
<td>LSESCSIESDIYAEIPDETNR</td>
<td>SH2</td>
<td>[10]</td>
</tr>
<tr>
<td>TGFβR</td>
<td>SMAD2</td>
<td>LTQMGPSVRCSSMS</td>
<td>FHA2</td>
<td>[6]</td>
</tr>
</tbody>
</table>
A significant advantage of cell-based bioluminescent kinase reporter is its adaptability for high-throughput screening. Bioluminescence generated in luciferase assays offers higher sensitivity than FRET-based systems due to amplification of the signal. In addition, luciferase is less susceptible to inference from nonspecific fluorescence of compounds. Thus, bioluminescence-based assays are highly suited for high-throughput screening. Furthermore, luciferase activity can be monitored dynamically and noninvasively, allowing bioluminescence-based cell assays to provide a unique method for identifying specific compounds that interact with the target in the correct cellular compartment and under normal cellular physiological conditions of that compartment (pH, concentrations of specific ions, etc). Reporters wherein the firefly luciferase enzyme has been divided into two halves (N-Luc and C-Luc) were originally developed to study protein-protein interaction [19]. These split-luciferase reporters were based on either the inter-molecular or intra-molecular complementation of the luciferase fragments to generate signal in response to cellular cues.

Ataxia Telangiectasia Mutated (ATM) is a member of the PI3-like family of serine/threonine kinases. It is a very large 370 KDa protein encoded by human chromosome 11q22-23. It plays a critical role in repair of DNA double-stranded breaks (DSBs) thereby maintaining genomic stability. These processes include, but are not limited to, DNA replication, DNA repair, cell cycle progression, apoptosis, and senescence. ATM exists in its inactive form as a non-covalently linked dimer where the kinase domain of one monomer is bound to the internal domain of another monomer covering the S1981 residue. In response to DSBs, the kinase domain of one monomer phosphorylates S1981 of the other interacting ATM resulting in subunit dissociation, ATM activation, and recruitment to DNA break sites [20]. Ionizing radiation-induced ATM activation results in the activation of a large number of ATM substrates [21–26] including P53, MDM2, SMC1, KAP1, BRCA1, γH2AX, and CHK2. The activated ATM triggers a sequence of events including cell cycle arrest, allowing time for the repair of the damaged DNA in sync with circadian rhythm [27]. If damaged DNA is left unrepaired it can lead to cell death, genomic instability, cancer, and/or other pathologies [28]. The 2015 award of the Nobel Prize in Chemistry for the discovery of DNA repair mechanisms highlights the importance of this pathway. Because of the important role ATM plays in cancer, therapeutics have been devised to target it [29].

In vitro kinase assays using purified substrate and kinase are routinely used to evaluate kinase activity. For traditional cell-based studies, immunohistological and biochemical techniques have been utilized for evaluating the kinase activity of ATM, such as counting pATM foci, γH2AX foci, immunofluorescence, or immunoprecipitation-western blotting [26, 30, 31]. Johnson,
You, and Hunter [11] described a fluorescence resonance energy transfer (FRET)-based biosensor for monitoring ATM kinase activity in live cells. Although this reporter provides direct measurement of the ATM kinase activity, it has a limitation of usability in mouse model due to tissue penetration and autofluorescence in the CFP-YFP range. In this book chapter, we provide detailed methods of use for the recently developed split firefly-based bioluminescence reporter to noninvasively, dynamically, and sensitively measure ATM kinase activity in live cells and mouse models [7].

2 Materials

2.1 Molecular Biology

1. DNA encoding open reading frame for phospho-protein-binding domain (e.g., Rad53p FHA2 domain).
2. Full-length coding sequence for the Firefly luciferase (FLuc) or plasmids coding for the N-terminal luciferase (amino acids 1–416) and C-terminal luciferase (amino acids 398–550) fragments.
3. Expression vectors with constitutive promoters for expression in mammalian cells (e.g., pEF vector from Clontech).
4. Optional: Expression vectors and packaging plasmids for generating lentiviral particles.
5. High fidelity DNA polymerases (e.g., Pfu DNA polymerase), dNTPs, oligonucleotide primers for cloning and sequencing, PCR reaction buffers, PCR Thermocycler, restriction endonuclease, DNA ligase, site-directed mutagenesis kits, high efficiency competent cells, antibiotic, bacterial growth media (LB, SOC), agar plates, plasmid DNA extraction kits, DNA gel purification, and sequencing kits.

2.2 Cell Culture

1. HEK293T cells or other readily transfectable cell lines.
2. Desired cell line(s) for biologic question of interest (i.e., D54, U87).
3. Fetal bovine serum (FBS).
4. Complete growth medium with FBS: Growth media with 10% FBS, and 1% penicillin/streptomycin.
5. Serum-free medium: Growth media without FBS and penicillin/streptomycin.
6. Trypsinization medium: 0.05% Trypsin-EDTA.
7. 1000× penicillin stock solutions: 10,000 Units/mL penicillin.
8. 1000× streptomycin stock solution: 10 mg/mL streptomycin.
9. 1000× geneticin/G418 stock solution: 50 mg/mL geneticin/G418.
10. Sterile 1× phosphate-buffered saline (PBS) solution.
11. Transfection reagents.

2.3 **Cell Imaging**

1. Black-walled or white-walled 96-well clear-bottom plates for tissue culture.
2. Sterile low adherence pipette tips with barrier filter.
3. Firefly luciferase substrate: 4 mg/mL d-luciferin 40× stock solution in PBS stored in dark colored vials at −80 °C. Alternatively, GloSensor c-AMP reagent (Promega) can be used to monitor firefly luciferase activity.
4. ATM inhibitor: 1000× stock solution 3 M caffeine.
5. ATM inhibitor: 1000× stock solution 20 mM KU-60019.
6. ATM inhibitor: 1000× stock solution 20 mM KU-55933.
7. ATM inhibitor: 1000× stock solution 20 mM CGK733.
8. Live cell bioluminescence imaging system with very high sensitivity and required software package for data generation and analysis (IVIS, Envision Xcite multi-label Plate Readers from Perkin Elmer; or similar system).
9. Liquid handling instrument (Biomek Nx from Beckman or similar).
10. Plate handling robot (Plate Handler II robot from Perkin Elmer or similar).
11. Optional: Cell culture incubator compatible with high-throughput instruments (LiCONiC StoreX STX44 IC precision incubator from Liconic Instruments or similar) for high-throughput assays.

2.4 **Animal Imaging**

1. Appropriate mouse strain for desired experimental system such as immunocompromised mice (nude, SCID, or NSG) for human tumor xenografts.
2. Optional: Small animal shaver such as Wahl trimmer.
3. 40 mg/mL luciferin stock in PBS, store in tightly sealed dark tubes at −20 or −80 °C.
4. 28–30 gauge insulin syringe for intra-peritoneal (IP) luciferin injection in mice.
5. Bioluminescence imaging instrument (IVIS or similar instrument) with a heated platform and isoflurane anesthesia injection and controller systems.
6. Additional accessories such as nose cones, animal partitions, black paper sheets, ear tags, markers, 70% alcohol, and 10% bleach or similar solution for disinfecting bench surfaces.
3 Methods

3.1 Construct Firefly Luciferase Complementation-Based ATM kinase Activity Reporter

1. Select a substrate such as CHK2 and determine the length of the substrate sequence that can be used for the construction of the reporter (see Note 1). We typically select 12–20 amino acid long substrate sequences with the target residue/s at the center of the sequence where possible (Table 1). For the construction of the ATM kinase reporter (ATMR), we selected a 12-residue sequence derived from CHK2 (Figs. 1 and 2).

2. Add a 5–7 amino acid long linker sequence at both the ends of the substrate sequence. We typically use GGSGG as the linker in our kinase reporters. For Ser/Thr kinases, attach a phospho-peptide-binding domain such as FHA2 (residues 420–582) [32]. For Tyr kinases, attach a SH2 domain (residue 374–465 of mouse shc2 [33, 34]). Use appropriate N-terminal (N-Luc) and C-terminal (C-Luc) firefly luciferase fragment pairs [19] at the flanks.

Fig. 1 The DNA-coding sequence and translated amino acid sequence for all the domains of the ATM kinase activity reporter. In frame short linker sequences (linker) inserted between each functional domain provide flexibility for the intramolecular domain interaction in the chimeric reporter molecule. N-Luc denotes amino acids 1–416 of firefly luciferase and C-Luc denotes amino acids 398–550. The target peptide sequence was derived from the CHK2 coding sequence (amino acids 63–74). The Ser/Thr phospho-peptide-binding domain (FHA2) comprises amino acids 420–522 of Rad53P protein.
3. Generate chimeric reporter constructs using appropriate molecular biology procedures. We generally incorporate linkers, substrates, and restriction enzyme sites in primers. We typically confirm the clones by sequencing, construct expression by Western blotting and functional bioluminescence assays before proceeding to generate stable cell lines.

4. Generate relevant control constructs wherein the phospho-target residue/s within the substrate is mutated to neutral amino acids. We typically mutate Ser/Thr or Tyr to Ala in mutant reporters by site-directed mutagenesis (Fig. 2a).

5. Express complementation reporters in appropriate vectors for mammalian cells. Vectors should be selected with markers, such as antibiotic resistance genes or co-expressed fluorescent proteins that are suitable for generating stable cell lines. We typically use pEF vector in combination with the geneticin/G418 antibiotic for stable clone selection.
3.2 Cell-Based Bioluminescence Imaging of ATM Kinase Activity

1. We typically do all of our cell-based and in vivo bioluminescence assays using the reporter expressing stable cell lines. We carefully select cell lines that represent an appropriate cellular and biological context for our studies (see Note 2). Cells are transfected with the reporter plasmids and allowed to grow under the antibiotic selection media. We typically pick 12–24 single-cell clones using sterilized filter paper discs and choose the best clones by measuring bioluminescence in response to specific kinase inhibitors (see Notes 3 and 4).

2. The three best reporter expressing stable cell lines (clones) are expanded and frozen at low passages for future use. Cells are maintained in 10 cm dishes with complete growth media containing serum and the appropriate amount of G418 (see Note 5).

3. Stable cell lines are plated overnight in black-walled or white-walled, clear-bottom 96-well plates for live cell assays. Cell density should be 2500–10,000 cells per well in 100 μL complete growth medium with serum (see Notes 6 and 7).

4. Cell culture media is removed and ATMR expressing cells are treated with different concentrations of ATM inhibitors such as caffeine, KU-60019, KU-55933, or CGK733 in serum-free media (100 μL per well).

5. After 5 min, 2.5 μL d-luciferin (black-walled plates) or cAMP-Glo reagent (white-walled plates) is added by multichannel pipette into each well for a working concentration of 100 μg/mL of firefly luciferase substrate.

6. Black-walled 96-well plates are imaged on the IVIS imaging system as soon as possible after adding luciferin. Typically, bioluminescence is acquired for 30–60 s at medium binning. For a time-course, the images are acquired with 3–10 min delay between the reads (Fig. 3a, b).

7. The white-walled plates are read on the Envision system right after the addition of the substrate. Generally, each well of the plate is read for 0.01–1.0 s. For a longer time-course activity measurement, a delay of 15–60 min between each read is set (Fig. 3c). For each read, the robot takes the plate out from the incubator, loads it on the reader where the plate is read, and is transferred back to the incubator until the next time point (see Note 8).

8. Quantify bioluminescence acquired on IVIS system by region-of-interest (ROI) analysis using Living Image software. The bioluminescence data from Envision system is automatically saved in quantitative form in tab-delimited file format.

9. Since radiation activates ATM within minutes, the bioluminescence activity of ATMR can be evaluated within 15 min after irradiation.
10. All the bioluminescence measurements should be validated by Western blotting in ATMR expressing cell lines in parallel experiments.

3.3 In Vivo Imaging of ATM Kinase Activity

1. D54-ATMR cells are expanded, trypsinized, and suspended in serum-free media at $40 \times 10^6$ cells/mL. 50 μL of this suspension is injected into each flank ($2 \times 10^6$ cells) in nude mice using a 22-gauge needle. We usually wait until the tumor reaches 60–100 mm$^3$ size (3–4 weeks) before starting the experiments.

2. We acquire baseline bioluminescence measurements 3–6 h before starting the treatment (Fig. 4a). Each mouse is injected with 100 μL d-luciferin (4 mg/mL stock prepared in sterile
PBS; 400 μg per mouse) anesthetized with 1–2% isoflurane for 5 min (see Note 9).

3. Transfer mice to the bioluminescence instrument, where they are maintained under anesthesia, and acquire bioluminescence. We typically acquire data on five mice at once isolated by a plastic separator. Generally, a 15–30 s acquisition at medium sensitivity is sufficient. We typically acquire data for 10–20 reads with a 1–5 min delay between the reads to cover the bioluminescence peak from all the tumors in each of the mice.

4. Treat the mice with appropriate inhibitors such as KU-55933 (both 25 mg/kg), or activators such as radiation (5 Gy) and monitor bioluminescence over time. Vehicle control (DMSO) or sham-irradiated mice should be used as control (Figs. 3c and 4b; see Note 10).

Fig. 4 In vivo measurement of ATM kinase activity in mouse tumor xenograft model. (a) CD-1 nude mice harboring D54-ATMR WT tumor xenografts were injected with luciferin and bioluminescence was acquired as described 3 h before treatments. (b) The animals were injected with KU-55933 (25 mg/kg) or vehicle control (DMSO) and bioluminescence was acquired 1, 4, 8, and 24 h posttreatment. The ATM reporter fold activation upon ATM inhibition is plotted over mock treatment. (c) Similarly, mouse harboring D54-ATMR WT tumor xenografts were whole body irradiated with 5 Gy of radiation or sham irradiated and bioluminescence was measured for up to 24 h. About 70% decrease in the reporter activity was observed 8 h post irradiation.
5. Remove mouse from imaging instrument and monitor for complete recovery from anesthesia.

6. Quantify imaging data by region-of-interest (ROI) analysis of bioluminescence produced by the tumor, using units of photon flux (Fig. 4a; see Notes 11–17).

3.4 Conclusions

The method described herein is an adaptation of the traditional protein complementation assay for the detection of protein-protein interaction in live cells. Instead of monitoring the interaction of two proteins through the use of split reporter molecules, we have adapted the assay such that the interaction between the “bait” and the “prey” occurs in response to the activity of a specific kinase. The kinase can be a serine/threonine- or a tyrosine-kinase. The reporter has also been engineered such that increased complementation (and therefore reporter activity) occurs in response to decreased kinase activity. This approach is therefore very well suited for high-throughput screens for kinase inhibitor libraries since a positive hit would be detected as an increase in bioluminescence activity, thereby less likely to result in false positives. We have also used analogous reporters for whole genome siRNA screens. As an example, a reporter for TGF-β receptor serine/threonione kinase activity was used in a human kinome siRNA screen to yield a number of novel genes as regulators of the TGF-β receptor function [12]. Regulation of the molecular events that lead to the activation and/or inactivation of the ATM kinase activity is yet to be defined; therefore, it is anticipated that analogous whole genome siRNA screens against the ATMR will most likely yield new insights into the role of novel genes in the regulation of the cellular response to DNA damage.

4 Notes

1. Substrate sequence for the construction of the reporter should be decided based on literature searches. It is imperative to have good antibodies available to detect changes in the substrate phosphorylation as this will help in validating the kinase reporter bioluminescence data by biochemical techniques.

2. The selection of a specific cell line for creating a reporter-expressing stable line should be based on the specific questions being interrogated and the intrinsic activity and detectability of the kinase and the substrate in the cell line. This should be determined by experimentation as well as literature searches. We selected the D54 glioblastoma cell line because ATM inhibition in glioblastoma may sensitize them to IR and chemotherapy [35]. D54 cells express wild-type ATM protein and responds to ATM inhibitors and irradiation as seen by Western
immunoblotting using antibodies against ATM and its substrate CHEK2.

3. We generally select for clones that exhibit low to moderate bioluminescence after adding luciferin. Since it is an activatable reporter that shows an increase in the light with inhibition of the ATM kinase, clones expressing the reporter in very high abundance may not yield high signal to background and thus may show a limited fold activation in response to inhibitors.

4. For picking up the clones, complete growth media with antibiotic is aspirated from the tissue culture dishes containing single-cell clones. Sterilized filter paper discs pre-wetted in trypsin are transferred to clones using sterilized forceps in a laminar flow hood. After 4–5 min, the discs are lifted from the tissue culture plate and swirled around in a 24-well plate containing complete growth media with half the concentration of antibiotic used for clone selection. Generally, 12–24 single-cell clones are picked. Forceps are sterilized either by heat or by dipping them in ethanol for 5–10 min. Make sure to let the ethanol evaporate before using the forceps to pick up the clones.

5. For expanding and maintaining of stable cell lines, we use half the concentration of G418 than that was used for selection. The G418 concentration for stable clone selection should be empirically decided.

6. The N-Luc and C-Luc fragments used in the construction of the complementation-based ATM reporter are derived from a firefly luciferase that has been optimized to work in mammalian cells at the physiological temperature of 37 °C. Therefore, all the bioluminescence acquisitions should be performed at 37 °C.

7. Since this reporter is based on the complementation of the light-generating enzyme luciferase, it works only in live cells under physiological conditions. Methods wherein cell lysate is used to measure the bioluminescence signals are incompatible with this reporter system and will not be able to yield any detectable change in signal to background (under different treatment conditions).

8. The Envision system is built with ultra-sensitive luminescence detection technology; thus 0.01–0.1 s measurement for each well is usually sufficient. The detection time can be increased if the signal is very weak and desired signal-to-noise is not reached with lower detection times. Furthermore, the bioluminescence signal from cAMP-Glo reagent is stable for a prolonged period of time, allowing us to measure the kinase activity of ATM for longer periods.

9. In our hands, we find that 400 μg luciferin/mouse gives us the best signal-to-background bioluminescence readings for split firefly luciferase-based kinase reporters. For smaller tumors or
tumors generated from cell lines expressing very low levels of reporter, the amount of luciferin can be increased to 150 mg/kg body weight (i.e., 3.5–4 mg/mouse with 20–25 g average mouse weight).

10. One may need to empirically determine the optimal inhibitor concentration for the best signal-to-noise bioluminescence detection in vivo. We usually test two to three different concentrations of the drugs in mouse tumor xenograft model to find the optimal concentration that gives highest fold change over vehicle control without being toxic to animals.

11. We create separate region-of-interest (ROI) for each tumor based on its size and shape. We also make sure that this ROI does not overlap with the ROI of any other tumors. We copy-paste the same ROI for each tumor for counting total photon flux for all the time points. We may move the position of the ROI so that it covers the tumor (because the position of the same animal with the same tumor might be slightly different between different time points) but do not change the overall shape or size of any ROI. This removes the chances of including background photon counts emanating from mouse skin.

12. We generally use four to five mice in an experimental group for bioluminescence data acquisition on the IVIS Spectrum system. Based on the number of animals in a group, we choose a stage level (distance of the CCD camera and the subject) and use the same stage level for the whole experiment. Changing the distance between the CCD camera and the subject height between different time points or reads would lead to different photon counts which will be difficult to analyze.

13. We usually use medium sensitivity settings with 15–60 s data collection time for each read. One should acquire photon flux without any saturated pixels. All the saturated pixels (above $10^8$ counts) show the same value, and therefore, cannot be accurately analyzed.

14. We typically use the maximum photon flux emitted from each tumor separately for all the calculations. Each tumor may show the maximum emittance at different periods of time; therefore, we usually perform sequential reads with a delay between each read which allows us to collect the maximum photon flux for all the tumors.

15. We typically use the maximum photon count for each tumor for each time for the analysis. If a tumor does not respond at any time-point, we omit that reading from the analysis. We analyze and plot combined values of all the tumors from one treatment group. Photon counts from vehicle (DMSO) or sham treated animals are set as onefold and the ratios of other treatments are counted as fold change from vehicle/sham treatment.
16. Due to the sensitivity of the imaging system, one typically gets some background bioluminescence in nude and SCID mice. It is important to shave and NAIR® the mice if using a mouse strain with hair to reduce the background and determine the correct size, shape, and position of a ROI.

17. Bioluminescence data acquired in a mouse xenograft model should be validated by biochemical methods such as Western blotting or immunohistochemistry (IHC). For validation of the bioluminescence data for ATMR, tumor tissue should be analyzed with pATM and pCHK2 antibodies after control, KU-55933, KU-60019, or radiation treatments.

Acknowledgments

We would like to acknowledge members of the Center for Molecular Imaging at the University of Michigan for their valuable input and support for the studies. This work was supported by the National Institutes of Health grant R01CA193690 (AR), P01CA087634 (BDR, AR and SN) as well as a P30CA046592 award to the University of Michigan Cancer Center.

References


Molecular Imaging of ATM Kinase Activity
Part IV

β-lactamase Sensors
Chapter 10

Creation of Antigen-Dependent β-Lactamase Fusion Protein Tethered by Circularly Permutated Antibody Variable Domains

Hiroto Iwai, Miki Kojima-Misaizu, Jinhua Dong, and Hiroshi Ueda

Abstract

Antibody-based molecular switches that are able to recognize a range of exogenous antigens can be highly useful as a versatile biosensor. However, regulating the catalytic activity of enzymes by antibodies is still hard to achieve. Here, we describe a design method of unique antibody variable region Fv introduced with two circular permutations, called Clampbody. By tethering the Clampbody to a circularly permuted TEM-1 β-lactamase (BLA), we successfully constructed a genetically encoded molecular switch Cbody-cpBLA that shows antigen-dependent catalytic activity.

Key words TEM-1 β-lactamase, Antibody variable region, Circular permutation, Allosteric regulation, Immunosensor

1 Introduction

Regulating the catalytic activity of enzymes by exogenous substances is potentially useful for the detection of biomarkers, environmental pollutants, and for enzyme-based prodrug therapies [1, 2]. To create the molecular switches that can recognize a wide variety of molecules, methods to control the catalytic activity of an enzyme by means of antibody domains are considered highly important [3].

To date, there have been a number of fusion proteins of antibody domains and enzymes whose catalytic activity was regulated by the binding antigen-dependent binding to their cognate antibody [4, 5]. Recently, we reported a fusion protein consisting of a circularly permuted TEM-1 β-lactamase (cpBLA) and Fv, called Fv-cpBLA [6]. This protein switch was made on a cpBLA that had been successfully used to make a maltose-dependent switch when fused with a circularly permuted maltose-binding protein (cpMBP) [7–9]. Fv-cpBLA showed antigen-dependent catalytic activation...
and antigen-dependent growth-enhancement of *E. coli* expressing the fusion protein in the presence of ampicillin. However, the increase of the specific activity was only up to 25% and there was a room for improvement.

More recently, aiming to transduce the antigen-dependent binding signal more effectively to the catalytic domain, we tried to optimize Fv-cpBLA by recombining the antibody domains to cpBLA more directly instead of using commonly employed flexible linkers [10]. The distance between the N- and C-termini of antibody variable domain V\(_H\) and V\(_L\) is between 30 and 40 Å. This is considered too long and may disrupt the optimal orientation of protein fragments or the conformation of circularly permuted enzymes. Therefore, based on the previous report [11], we focused on the 3 and 3b loops located far from antigen-binding loops but near the domain interface between V\(_H\) and V\(_L\) (Fig. 1a). Thus, we

![Fig. 1 Creation of Clampbody and its fusion to cpBLA.](image)

(a) Schematic structure of single chain Clampbody (sc-Cbody). cpV\(_H\) and cpV\(_L\) chains are shown in magenta and cyan, respectively. The (G\(_4\)S\(_3\)) linkers connecting their native N- and C-termini are drawn as dashed lines. Two cysteine residues were inserted at the N- and C-termini of cpV\(_H\) to promote correct folding via disulfide linkage. The SG\(_4\) linker connecting cpV\(_H\) and cpV\(_L\) is drawn as dotted line. (b) Schematic structure of Cbody-cpBLA, in which the structure of cpBLA (PDB code 4DXB) is shown in green. (c, d) Expression and purification of Cbody-cpBLA. (c) CBB-stained SDS - PAGE. (1) MW marker, (2) concentrated culture supernatant, (3) soluble fraction from the cell lysate, (4) insoluble fraction from the same, (5) solubilized insoluble protein, (6) purified and refolded protein. (d) Western blot to detect the His tag. The lanes are the same as in (c). The bands for Cbody-cpBLA are indicated by arrows. Reproduced from Iwai et al. [10] with permission from American Chemical Society.
designed a novel antibody Fv domain format whose termini were located at these sites, and named this doubly circularly permuted antibody Fv as Clampbody. In addition, we also designed a fusion protein comprising Clampbody and cpBLA by connecting cpVH and cpVL to the N- and C-termini of cpBLA [7] with a minimum number of linker residues between them (Clampbody-cpBLA, Cbody-cpBLA) (Fig. 1b).

We demonstrated the specific antigen binding of Clampbody, and more importantly, the antigen-dependent catalytic activity of novel molecular switch Cbody-cpBLA. Interestingly, we also found that the antigen-dependency of Cbody-cpBLA became more significant in reaction buffers containing some denaturant or detergent such as urea and Triton X-100. From these results, we reasoned that Cbody-cpBLA was stabilized by the binding of the antigen peptide, and therefore showed antigen-dependent catalytic activity. This would be a good example of the allosteric regulation driven by ligand-induced stabilization or thermodynamic shift, which is collectively called “ensemble” model [12–15]. By optimizing the reaction condition, Cbody-cpBLA could detect their antigen peptide specifically at single nM level. Considering the variety of available substrates, Cbody-cpBLA would be applied to biosensors, prodrug therapies, and possibly to bacterial growth regulation. We also propose Clampbody as a useful tool in antibody-based sensing proteins such as protein complementation assay [16], fluorescence resonance energy transfer-based assay [17, 18], and transduction assay of reaction intermediate [19, 20]. Taking advantage of the proximity of their termini, Clampbody might be fused with many protein domains and fragments, and therefore be able to control the distance and orientation of fused protein(s) in an antigen-dependent manner.

2 Materials

Prepare all solutions using ultrapure water (made by purifying deionized water to attain resistivity of 18 MΩ cm at 25 °C). For all the reagents, use them of the highest grade available. The chemicals and other reagents, unless otherwise indicated, were obtained from Sigma (St. Louis, MO), Nacalai Tesque (Kyoto, Japan), or Wako Pure Chemical Industries (Osaka, Japan).

2.1 Construction of Clampbody Genes

1. Oligonucleotides (see Table 1) (e.g., Greiner Japan, Tokyo, Japan; Eurofins genomics, Tokyo, Japan).

2. High-fidelity thermostable DNA polymerase (e.g., KOD FX, Toyobo, Osaka, Japan; PrimeSTAR® Max DNA Polymerase, Takara-Bio, Shiga, Japan).
<table>
<thead>
<tr>
<th>Name</th>
<th>DNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NcoVHrev</td>
<td>5′-GATGGCCATGGCCGAGG ACAGCTGGAG-3′</td>
</tr>
<tr>
<td>VHSpefor</td>
<td>5′-TGGTACTAGTTGAGGAGACGGTGACCGT-3′</td>
</tr>
<tr>
<td>AgeVLrev</td>
<td>5′-CAAAACCGGTGATATTGTGCTGACCCAA-3′</td>
</tr>
<tr>
<td>VLNotfor</td>
<td>5′-CGTGGCGGCCGCGCGTTTTATTTCCAGC-3′</td>
</tr>
<tr>
<td>NcoCcpVHrev</td>
<td>5′-GATGGCCATGGCCGACCAAGGCTGATGGGAC-3′</td>
</tr>
<tr>
<td>VHGSfor</td>
<td>5′-CGGACCCGCCACCCGG TGCACTCCGCTTGACCGCTCACAAGGCTCCACCCGCTGACCCGAGGCTGACCCG-3′</td>
</tr>
<tr>
<td>G4SVHrev</td>
<td>5′-GGTGAGGCCGAGGTGCGGCTTGGCAGGTCGAGGTC-3′</td>
</tr>
<tr>
<td>cpVHCSpefor</td>
<td>5′-GGGACCTAGTACATGGGCTGATTGACCCAGTG-3′</td>
</tr>
<tr>
<td>AgecpVLrev</td>
<td>5′-GATACCCGCCATGGGGTCCAGTCTCCGCTGACCCGAGGTGCGGCTGACCCGAGGCTGACCCGAGG-3′</td>
</tr>
<tr>
<td>VLG4Sfor</td>
<td>5′-AGATCCGACGAGCGCGACGAGCGAATCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG-3′</td>
</tr>
<tr>
<td>G4SVLrev</td>
<td>5′-GGTGAGGCCGAGGTGCGGCTTGGCAGGTCGAGGTC-3′</td>
</tr>
<tr>
<td>cpVLNotfor</td>
<td>5′-GGTGCGGCGCGCGGTTCAGGTATGGACCCCAAATG-3′</td>
</tr>
<tr>
<td>MCScpBLArev</td>
<td>5′-CCCTCCATGGGTAAGGAGCTCAGAGCAGGTAACACTAGTAAATGACGAGACATCAAACCAGCTGACCCAGGA-3′</td>
</tr>
<tr>
<td>cpBLAMCSfor</td>
<td>5′-GGTGCGGCGGCGGTTCAGGTATGGACCCCAAATGACGAGACATCAAACCAGCTGACCCGCTGACCCAGG-3′</td>
</tr>
<tr>
<td>InfVHSpeRG13rev</td>
<td>5′-GTCTCCTCAACTGGTACAGGATTTATGTTATGATGAGTAAATC-3′</td>
</tr>
<tr>
<td>RG13OL1for</td>
<td>5′-GACACCTCTTAAGACCCCAATGTTAATCAGGTA-3′</td>
</tr>
<tr>
<td>RG13OL2rev</td>
<td>5′-GGTTCAGGAGGTGGTCACAGAGACCGTTGCTATCCCACGTCAGCCGCTG-3′</td>
</tr>
<tr>
<td>RG13AgeVLInffor</td>
<td>5′-GCACAATATCACCCGCTGACGCGGAGGGGCGCAG-3′</td>
</tr>
<tr>
<td>VHBGPEAAAK2Spefor</td>
<td>5′-GGGACCTAGTCTTGGGAGCGCCCGCTTCTAGGCCGCGCCCTGCTGAGACGCTGACCCGCTGACCCGAGG-3′</td>
</tr>
<tr>
<td>AgeEAAAK2EcoVLrev</td>
<td>5′-GATACCCGGTGAGCCGCGGCTAAAGAGGCGGGCTGCCAGAAGTATCGTGCTGACCACATAC-3′</td>
</tr>
<tr>
<td>TSG4TGrev</td>
<td>5′-CTAGTGGAGGTGGGCGGGA-3′</td>
</tr>
<tr>
<td>TSG4TGfor</td>
<td>5′-CCGGTACCCGACCCTCAA-3′</td>
</tr>
</tbody>
</table>
3. DNA purification kits (e.g., Wizard® SV Gel and PCR Clean-Up System, Promega, WI).
4. Restriction enzymes (e.g., New England Biolabs, Beverly, MA).
5. DNA ligation kit (e.g., Ligation high ver. 2, Toyobo).
6. BGP C-terminal peptides (e.g., Genscript, Piscataway, NJ).
7. pET26/Fv-cpBLA (previously described [6]).
8. pIT2(13CG2) (previously described [21]).
9. pET30b(+) (Novagen, Darmstadt, Germany).
11. In-Fusion® HD cloning kit (Clontech, CA).
12. Tris-EDTA buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 7.4.

2.2 Protein Expression

1. SHuffle T7 express lysY competent cells (New England Biolabs).
2. LB medium: 1% tryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.0.
3. LBK medium: LB containing antibiotics (50 μg/mL kanamycin).
4. Agar.
5. 1 M Isopropyl-β-d-thiogalactopyranoside (IPTG).
6. Supersonic homogenizer (e.g., VP-15S Taitec, Saitama, Japan).
7. 2-mercaptoethanol.
8. 0.5 M EDTA, pH 8.0.
9. Guanidium hydrochloride (Gdn-HCl).
10. TALON extraction buffer: 300 mM NaCl, 50 mM sodium phosphate, adjusted to pH 7.0.
11. TALON metal affinity resin and TALON disposable gravity column (Clontech, CA).
12. TALON elution buffer: TALON extraction buffer supplement with 200 mM imidazole.

2.3 Refolding of Clampbodies

2. Dialysis buffer: 50 mM Tris–HCl, 200 mM NaCl, 1 mM EDTA, pH 8.0.

2.4 Western Blot of Clampbodies

1. SDS-PAGE gradient gels (5 ~ 20%) (e.g., e-PAGEL, ATTO, Tokyo, Japan).
2. SDS running buffer: 25 mM Tris–HCl, 192 mM l-glycine, 1 g/L sodium dodecyl sulfate.
3. PVDF membrane (e.g., Clear Blot Membrane –P plus, ATTO).
4. Transfer buffer: 25 mM Tris–HCl, 192 mM l-glycine.
5. Tris buffered saline (TBS): 50 mM Tris–HCl, 150 mM NaCl, pH 7.4.
6. TBS with Tween (TBST): TBS supplemented with 0.05% of Tween 20.
7. TBST with skim milk (TBST-S): TBS-T supplemented with 5% skim milk.
8. HRP-conjugated anti-His$_6$ antibody (e.g., Roche Diagnostics Japan, Tokyo, Japan).
10. Chemiluminescence imager (e.g., LAS-4000 mini Luminoimager, Fujifilm, Tokyo, Japan).

2.5 ELISA for Clampbodies

1. Transparent 96-well microplates (e.g., Greiner 655,061).
2. 10 μg/mL streptavidin (e.g., type II, Wako, Osaka, Japan).
3. Phosphate buffered saline (PBS): 1.47 mM KH$_2$PO$_4$, 8.1 mM Na$_2$HPO$_4$, 2.7 mM KCl, 137 mM NaCl, pH 7.4.
4. PBST: PBS containing 0.1% Tween 20.
5. Blocking solution (e.g., Immunoblock, DS Pharma, Osaka, Japan).
6. 1 μg/mL N-terminally biotinylated BGP-C11 peptide in PBST.
7. TMBZ reaction buffer: 100 μg/mL TMBZ (3,3′,5,5′-tetramethylbenzidine), 0.006% H$_2$O$_2$, 100 mM NaOAc, pH 6.0.
8. 10% H$_2$SO$_4$ to stop the reaction.

2.6 Enzyme Assays for Clampbodies

1. Black 96-well microplates (e.g., Corning Costar 3693).
2. Transparent 96-well microplates (e.g., Greiner 655061).
3. Dimethyl sulfoxide (DMSO).
4. Fluorocillin™ Green 495/525 β-Lactamase Substrate, soluble product (Thermo Fisher Scientific, Waltham, MA). 1 mM stock solution in DMSO can be stored for several months at −30 °C.
5. Nitrocefin (e.g., Calbiochem, Darmstadt, Germany). 20 mM stock solution in DMSO can be stored for several months at −30 °C.
6. PBS containing 1% Triton X-100.
7. PBS containing 3 M urea.
8. PBS containing 20 mg/mL bovine serum albumin (BSA) (e.g., Sigma-Aldrich, St. Louis, MO).
9. A fluorescence microplate reader (e.g., GENios Pro Tecan, San Jose, CA).
10. A microplate reader (e.g., SH-1000, Corona Electric, Hitachi, Ibaraki, Japan).
3 Methods

3.1 Construction of the \( V_H \) and \( V_L \) genes for Anti-BGP C-Terminal Peptide Antibody

To create a Clampbody as an antigen-dependent switch, any antibody for small molecule antigen, whose variable region \((Fv = V_H + V_L)\) is stabilized by the bound antigen, will suffice. Here, we employ the \( V_H \) and \( V_L \) domains of antibody KTM219 [22], which recognizes the C-terminal peptide of the human bone Gla protein (BGP) as a model. First, clone the \( V_H \) and \( V_L \) genes by the following procedure.

1. Amplify the DNA fragment encoding \( V_H \) fragment of the antibody by PCR using the primers NcoVHrev and VHSpefor, pET26/Fv-cpBLA as a template, and KOD FX DNA polymerase and the DNA fragment encoding \( V_L \) fragment of the antibody by PCR with AgeVLrev and VLNotfor as primers.

2. Purify the resulting DNA fragments by Wizard® SV Gel and PCR Clean-Up System as per manufacturer’s instructions.

3. Ligate these fragments into pCR4Blunt-TOPO as per manufacturer’s instructions.

3.2 Construction of the cp\( V_H \) Gene

Make a construct for a circularly permuted \( V_H \) (cp\( V_H \)), whose new termini are generated between Pro41 and Gly42 (as described according to Kabat numbering) while the native N- and C-termini are connected via a flexible linker \((G_4S)_3\). Considering the possible lesser stability of \( V_H \) fragment, two Cys residues are added at the N- and C-termini of cp\( V_H \) to promote the correct folding and increased stability through the disulfide bridge.

1. Amplify the DNA fragment encoding Gly42 to Ser113 of \( V_H \) with a linker peptide \((G_4S)_3\) at the C-terminus by PCR using the primers NcoCcpVHrev and VHG4Sfor, pET26/Fv-cpBLA as a template, and KOD FX DNA polymerase.

2. Amplify the DNA fragment encoding Glu1 to Pro41, described according to Kabat numbering, with the linker at the N-terminus by PCR using G4S3VHrev and cpVHCSpefor as primers.

3. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

4. Perform overlap-extension PCR with the gel-purified PCR products (20 ng each) and KOD FX for 15 cycles without primers and 35 cycles with NcoCcpVHrev and cpVHCSpefor primers.

5. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

6. Ligate into pCR4Blunt-TOPO.
Similarly, make a construct for the circularly permuted V₇ (cpV₇) with new termini between Pro40 and Gly41. In this case, two Cys residues at the N- and C-termini of cpV₇ are not added to avoid unfavorable interdomain crosslinking at the later stage.

1. Amplify the DNA fragment encoding Gly41 to Arg108 of V₇ with a linker peptide (G₄S)₃ at the C-terminus by PCR using the primers AgecpVLrev and VLG4Sfor, pET26/Fv-cpBLA as a template, and KOD FX DNA polymerase.

2. Amplify the DNA fragment encoding Asp1 to Pro40 with the linker at the N-terminus by PCR with G₄S₃VLrev and cpVL-Notfor as primers.

3. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

4. Perform overlap-extension PCR with the gel-purified PCR products (20 ng each) and KOD FX DNA polymerase for 15 cycles without primers and 35 cycles with AgecpVLrev and cpVLNotfor primers.

5. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

6. Ligate into pCR4Blunt-TOPO.

To make a fusion protein comprising Clampbody and cpBLA, first make a construct for the cpBLA with known 3D structure that Guntas et al. reported [23, 24].

1. Amplify the DNA fragment encoding Trp227 to Trp286 of TEM-1 β-lactamase with a short peptide linker (GSGGS) at the C-terminus by PCR using the primers InfVHSpeRG13rev and RG13OL1for, pIT2/31IJ3 as a template, and PrimeSTAR® Max DNA Polymerase.

2. Amplify the fragment encoding His24 to Gly226 with the linker at the N-terminus by PCR with RG13OL2rev and RG13AgeVLInffor as primers.

3. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

4. With the gel-purified PCR products (20 ng each), perform overlap-extension PCR with PrimeSTAR® Max DNA Polymerase for 15 cycles without primers and 35 cycles with InfVHSpeRG13rev and RG13AgeVLInffor primers.

5. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

Make a construct for a fusion protein comprising Clampbody and cpBLA by connecting cpV₇ and cpV₇ to the N- and C-termini of cpBLA [7] with a minimum number of amino acid residues (Clampbody-cpBLA, Cbody-cpBLA) (Fig. 1b, c).
1. Amplify the DNA fragment encoding \textit{Neol}, \textit{SpeI}, \textit{AgeI}, and \textit{NotI} sites by PCR using the primers MCScpBLArev and cpBLAMCSfor, pET26/Fv-cpBLA as a template, and KOD FX DNA polymerase.

2. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System.

3. Digest the gel-purified PCR product (2 μg) with \textit{Neol} HF and \textit{NotI} HF.

4. Purify the resulting DNA fragments using the Wizard® SV Gel and PCR Clean-Up System and insert into pIT2(13CG2) digested with the same, resulting in pIT2/MCS.

5. Digest the cpV\textsubscript{H} gene with \textit{Neol} I and \textit{Spe} I.

6. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System and inserted into pIT2/MCS digested with the same.

7. After the cpV\textsubscript{H} insertion, digest the cpV\textsubscript{L} gene with \textit{Age} I and \textit{Not} I.

8. Purify these fragments using Wizard® SV Gel and PCR Clean-Up System and inserted into cpV\textsubscript{H}-inserted pIT2/MCS.

9. Digest these plasmids with \textit{Neol} and \textit{NotI}.

10. Purify these fragments by Wizard® SV Gel and PCR Clean-Up System and insert into pET30b(+) digested with the same, resulting in pET30b/cpVH-cpVL.

11. Insert the cpBLA(RG13) gene into pET30b/cpVH-cpVL with In-Fusion® HD Cloning Kit according to the manufacturer, resulting in pClamp-cpBLA.

### 3.6 Construction of \textit{scCb} Expression Vectors

Prepare a single-chain protein construct where cpV\textsubscript{H} and cpV\textsubscript{L} are connected by a short linker G\textsubscript{4}S (single-chain Clampbody, scCb) to evaluate the antigen-binding activity of the Clampbody.

1. Anneal TSG4TGrev and TSG4TGfor in Tris-EDTA buffer to prepare the DNA fragment encoding the SG\textsubscript{4} linker.

2. Insert the fragment into pClamp-cpBLA digested with \textit{AgeI} and \textit{SpeI}, resulting in pClamp-SG4.

### 3.7 Expression and Purification of Clampbodies [6, 25]

When Clampbody and Cbody-cpBLA are expressed in \textit{E. coli}, the majority of the expressed proteins will accumulate in inclusion body even if cultured at lower temperature. In such cases, perform the following procedures to solubilize the insoluble Clampbodies.

1. Transform SHuffle T7 Express lysY competent cells with expression vectors (see \textbf{Note 1}).
2. Incubate the transformed *E. coli* on a LB medium containing antibiotics (50 μg/mL kanamycin) (LBK) and 1.5% agar at 37 °C for 24 h.

3. Pick one colony up and culture it in 4 mL LBK medium at 37 °C.

4. Inoculate the 1 mL of cultured cells to 250 mL of LBK medium in a 1000 mL flask, and culture in 37 °C.

5. When the OD<sub>600</sub> reaches 0.4–0.6, add to a final concentration of 0.4 mM of IPTG, and culture the cells at 16 °C overnight.

6. Centrifuge at 10,000 × *g* for 15 min at 4 °C to collect the cells (see Note 2).

7. Resuspend the pellet with 30 mL of TALON extraction buffer and sonicate the *E. coli* on ice using a supersonic homogenizer VP-15S for 2 min (50% interval) three times.

8. Centrifuge the sonicated solution at 10,000 × *g* for 60 min at 4 °C (see Note 3).

9. Resuspend the pellet with 25 mL of distilled water containing 1% Triton-X100 and 1 mM EDTA.

10. Centrifuge at 75,000 × *g* for 15 min at 4 °C. Resuspend pellet with 25 mL of distilled water containing 1% Triton-X100 and 1 mM EDTA.

11. Repeat step 11 to wash the pellet.

12. Resuspend the pellet with 10 mL of TALON extraction buffer containing 6 M guanidium hydrochloride (Gdn-HCl) to solubilize it.

13. Centrifuge at 75,000 × *g* for 15 min at 4 °C and collect the supernatant carefully.

14. Add 200 μL of TALON Metal Affinity Resin to the supernatant, and rotate or shake gently for 20 min (see Note 4).

15. Centrifuge at 700 × *g* for 2 min, and resuspend the resin with 10 mL of TALON extraction buffer containing 6 M Gdn-HCl.

16. Rotate or shake gently for 10 min.

17. Repeat steps 11 and 12 twice.

18. Centrifuge at 700 × *g* for 2 min, and collect the resin.

19. Suspend the resin with 3 mL of TALON extraction buffer containing 6 M Gdn-HCl, and put the suspension into a TALON 2-mL Disposable Gravity Column.

20. Allow to drain until it reaches the top of the resin bed.

21. Wash the column with 2 mL of TALON extraction buffer containing 6 M Gdn-HCl.

22. Elute the protein with 1.5 mL of TALON elution buffer containing 5.4 M Gdn-HCl.
Refold the purified denatured Clampbodies by stepwise dialysis procedure as below.

1. Add 1.05 μL of 2-mercaptoethanol and 3 μL of 0.5 M EDTA to 1.5 mL of purified solubilized protein.
2. Rotate gently at room temperature for 30 min to reduce disulfide bonds.
3. Inject protein solution into Slide-A-Lyzer Dialysis Cassettes, 3.5 K MWCO, 3 mL.
4. Dialyze protein solution by putting the Slide-A-Lyzer Dialysis Cassettes into 1 L of Dialysis buffer containing 3 M Gdn-HCl. Incubate at 4 °C for overnight (see Note 5).
5. Dialyze protein against Dialysis buffer containing 2 M Gdn-HCl at 4 °C overnight.
6. Dialyze protein against Dialysis buffer containing 1 M Gdn-HCl at 4 °C overnight.
7. Dialyze protein against Dialysis buffer containing 0.5 M Gdn-HCl at 4 °C overnight.
8. Dialyze protein against Dialysis buffer without Gdn-HCl at 4 °C overnight.
9. Collect the contents in the cassette and centrifuge at 20,000 × g for 5 min at 4 °C.
10. Collect the supernatant (see Note 6). Measure the concentration of purified protein with a Pierce™ BCA protein assay kit as per manufacturer’s instructions.

After the preparation of Clampbodies, perform SDS-PAGE and Western blotting to confirm the purity and the amount of the proteins (Fig. 1c, d).

1. Analyze the proteins by SDS-PAGE using a 5–20% gradient gel and SDS running buffer.
2. Transfer the protein from the gel to a PVDF membrane with Transfer buffer.
3. Incubate the membrane in TBST-S at 4 °C overnight.
4. Wash membrane three times with TBST.
5. Incubate the membrane with HRP-conjugated anti-His6 antibody 4000-times diluted in TBST at 4 °C for 1 h.
6. Wash membrane three times with TBST.
7. Soak membrane in Amersham ECL Prime as per manufacturer’s instructions.
8. Detect the luminescence using a LAS-4000 mini Luminoimager.
Perform following ELISAs to confirm that the refolded Clampbodies retain the antigen-binding activity derived of the parental $V_{H}$ and $V_{L}$ domains (Fig. 2a, b).

1. Apply 50 μL per well of 10 μg/mL streptavidin in microplate wells at 4 °C overnight.
2. Discard the supernatant carefully, and then apply 200 μL/well of PBST supplemented with 20% Immunoblock at room temperature for 1 h.
3. Discard the supernatant carefully, and wash wells with PBST three times.
4. Apply 50 μL per well of 1 μg/mL biotinylated BGP-C11 peptide in PBST containing 5% Immunoblock at room temperature for 1 h.
5. Discard the supernatant carefully, and wash wells with PBST three times.
6. Apply 50 μL per well of 100 nM protein samples in PBST containing 5% Immunoblock at room temperature for 1 h. Use approximately 1 μM protein stock and dilute it in PBST containing 5% Immunoblock to make 100 nM working solution immediately before the assay.
7. Discard the supernatant carefully, and wash wells with PBST three times.
8. Apply 50 μL per well of HRP-conjugated anti-His$_{6}$ antibody 4000-times diluted in PBST containing 5% Immunoblock at room temperature for 1 h.
9. Discard the supernatant carefully, and wash wells with PBST three times.

10. Apply 50 μL per well of TMBZ reaction buffer at room temperature. After several minutes, add 50 μL per well of 10% H₂SO₄ to stop the reaction.

11. Detect the absorbance at 450 nm with a reference absorbance at 650 nm.

3.11 Clampbody Activity Assay with Fluorocillin

Measure catalytic activity of Cbody-cpBLA using a fluorogenic substrate fluorocillin (Fig. 3). The antigen-dependency of Cbody-cpBLA becomes more significant in reaction buffers containing appropriate amount of denaturant or detergent such as urea and Triton X-100.

**Fig. 3** Antigen-dependent catalytic activity of Cbody-cpBLA. The solution of 70 nM Cbody-cpBLA, 1 μg/mL BGP-C10, and 1 μM fluorocillin was mixed in (a) PBS, (b) PBS containing 1 M urea, and (c) PBS containing 0.1% Triton X-100 and 250 μg/mL BSA, incubated at 30 °C, and read for fluorescence. Background fluorescence of the samples without Cbody-cpBLA was subtracted for each condition. Averages of three samples with an error bar of 1 SD are shown. Blue circle indicates the catalytic activity in the presence of 1 μg/mL BGP-C10, and red triangle indicates in the absence of peptide. (d) Dose — response curves for BGP-C10 and BGP-C10dV peptides in the reaction buffer as in (c). The reaction rates at each peptide concentration are shown as the relative value to the rate in the absence of peptides. Statistical analysis was conducted using the two-tailed unpaired Student’s t test: (α) p < 0.05, (αα) p < 0.01. Reproduced from Iwai et al. [10] with permission from American Chemical Society.
1. Prepare BGP-C10 (NH$_2$-EAYRRFYGPV-COOH) and BGP-C10dV (NH$_2$-EAYRRFYGP-COOH) as antigen and non-antigen peptide, respectively.

2. Mix 25 µL per well of 140 nM protein samples in PBS and 25 µL per well of antigen peptides in PBS. Incubate for 30 min at room temperature in a black microplate. Use approximately 1 µM protein stock and dilute it in PBS immediately before the assay.

3. Apply 50 µL per well of 2 µM fluorocillin in PBS containing urea, TritonX-100, Tween-20 or BSA, at varying concentrations. Use 1 mM fluorocillin stock solution in DMSO and dilute it in PBS immediately before the assay.

4. Immediately read for their fluorescence intensity at 535 nm with 485 nm at 30 °C for 2 h using a fluorescence plate reader GENios Pro.

5. Fit the initial rates of reactions to the Michaelis-Menten equation as follows:

\[
\nu = \frac{V_{\text{max}}[S]}{K_m + [S]}
\]

by the nonlinear least squares algorithm of Kaleidagraph 4.1 (Synergy Software, Reading, PA, USA).

### 3.12 Clampbody Activity Assay with Nitrocefin

Evaluate the catalytic activity of Cbody-epBLA using a colorimetric substrate nitrocefin, as shown in Fig. 4.

1. Mix 25 µL per well of 140 nM proteins in PBS and 25 µL per well of antigen samples in PBS. Incubate for 30 min at room temperature in transparent microplate. Use approximately 1 µM stock protein and dilute this in PBS immediately before the assay.

2. Apply 50 µL per well of 0–500 µM nitrocefin in PBS containing 2.4 M urea. Use 20 mM nitrocefin stock solution in DMSO and dilute it in PBS immediately before the assay.

3. Immediately read for their absorbance at 486 nm at 30 °C for 1 h using a microplate reader SH-1000.

4. Calculate the catalytic parameters following the same procedures in Subheading 3.11.

### 4 Notes

1. SHuffle T7 Express lysY is an engineered *E. coli* B strain to express proteins containing disulfide bonds in the cytoplasm. Constitutively expressed disulfide bond isomerase DsbC promotes the correction of mis-oxidized proteins into their correct form, and also works as a chaperone that can assist in the folding of proteins that do not require disulfide bonds.
2. Culture supernatant (2 mL) can be precipitated with 10% trichloroacetic acid, and washed twice with cold acetone, and resuspended in 20 μL of PBS and used for SDS-PAGE.

3. The supernatant including soluble protein should be kept for SDS-PAGE.

4. Before adding to the solubilized protein, TALON metal affinity resin should be washed twice with TALON extraction buffer containing 6 M Gdn-HCl.

5. Keep stirring the buffer using a magnetic stirrer.

6. The purified enzyme in the buffer containing 15% glycerol can be stored for several months at −80 °C. The concentration of the proteins in stock solution was typically 1 μM in our hands.

Acknowledgment

We thank Tatsuya Shinoda of Kyowa Medex Co. for generously allowing the use of the genes derived from KTM-219 IgG. This study was supported by Grant-in-Aid for Scientific Research (Grants 24360336 and 15H04191 to H.U. and Grant 46420793 to J.D.) from JSPS, Japan, and partly by Strategic International Collaborative Research Program, Japan Science and Technology Agency (JST).
References


Chapter 11

Protein and Protease Sensing by Allosteric Derepression

Hui Chin Goh, Farid J. Ghadessy, and Saurabh Nirantar

Abstract

Peptide motifs are crucial mediators of protein-protein interactions as well as sites of specific protease activity. The detection and characterization of these events is therefore indispensable for a detailed understanding of cellular regulation. Here, we present versatile and modular sensors that allow the user to detect protease activity and protein-peptide interactions, as well as to screen for inhibitors using chromogenic, fluorescent, or luminescent output.

Key words Protease sensors, Protein-protein interaction, Drug screening

1 Introduction

Sensing the binding of macromolecules to and cleavage of linear peptide motifs is of great importance in elucidating the many regulatory processes in the cell. Several interaction-mediating domains like the PDZ or SH2 families bind their cognate peptide to enable colocalization of interacting proteins [1–3]. Similarly, processes as diverse as viral maturation and blood clotting depend on the recognition of a specific peptide sequence by their target proteases such as thrombin or tobacco etch virus (TEV) protease [4, 5].

In vitro studies with such peptides are very useful in detecting and characterizing these functions. Typically, in vitro characterization of such events relies on techniques such as fluorescence polarization (FP) [6], enzyme-linked immunosorbent assays (ELISA) [7], surface plasmon resonance (SPR) [8], and Forster resonance energy transfer (FRET) [9]. While powerful, these methods require substantial instrumentation and preparation and suffer from a number of limitations. These include interference from autofluorescent molecules in case of FP measurements, multiple washing and incubation steps in case of ELISAs, or low throughput for SPR.

Protein engineering-based methods have been developed to address this need. These typically involve the insertion of a relevant peptide or protein domain into a read-out capable enzyme such
that binding of the peptide or cleavage thereof leads to a change in parameters like enzyme activity, thus detecting the event [10–12]. However, these systems are typically custom-made and require substantial empirical optimization. There is therefore a need for simpler, modularly organized systems that allow the user to study these processes.

We have developed such a system based on auto-inhibited enzymes. The basic structure of a sensor is depicted in Fig. 1. Briefly, it comprises a read-out enzyme (e.g., Tem1 β-lactamase) fused via a flexible peptide linker that incorporates the peptide of interest, to a domain that binds and inhibits the read-out enzyme. The flexible peptide retains the inhibitor in close proximity to the read-out enzyme resulting in a high local concentration, and allows the inhibitor domain to reach out and bind the read-out module, thereby inhibiting it.

For the detection of proteases, cleavage of the peptide of interest in the flexible linker allows the dissociation of the inhibitor domain and its diffusion into the bulk solution leading to derepression and signal generation by the read-out enzyme [13].

In the case of peptide-protein binding events, the peptide of interest is included in the flexible linker such that binding of a large

![Fig. 1](image)

**Fig. 1** The sensor is a recombinant fusion protein comprising of a read-out capable enzyme (gray), and an inhibitor of the enzyme (black) connected by a linker. The peptide of interest is placed in the linker region. The close proximity of the enzyme and inhibitor enabled by being fused together via the linker leads to a high effective local concentration and strong inhibition. **Left:** Binding of the peptide of interest to a macromolecule (white hexagon) will interfere sterically with enzyme-inhibitor binding if the linker design is optimized. This leads to stabilization of the open state and turnover of substrate (white stars) leading to signal generation (gray stars). **Right:** Upon cleavage of the peptide by a suitable protease, the inhibitor is free to diffuse throughout the bulk solution, leading to derepression of the enzyme and signal generation.
macromolecule like a protein domain to the peptide in the linker sterically hinders binding of the inhibitor to the read-out enzyme (Fig. 1). Thus, the presence of a peptide-interacting macromolecule stabilizes the open state of the read-out enzyme and increases its activity. It is important that the flexible linker is sufficiently long to allow the inhibitor to bind to the read-out enzyme, but not long enough to allow a macromolecule to bind to the peptide without sterically hindering the interaction between the read-out enzyme and the inhibitor. In this case, the exact linker length requires empirical optimization. Sensitivity can also be enhanced in the case of bivalently binding macromolecules such as antibodies [13]. For instance, this can be achieved by placing a second protein-interacting peptide sequence in the sensor that can bind one arm of the antibody molecule and thus tether the other arm in close proximity of the peptide in the linker. In this way, its local concentration is increased and thus leads to increased sensitivity.

We have also developed another, more modular peptide-protein interaction detection methodology [14]. This involves a similar architecture, but instead of relying on the macromolecule sterically clashing with the read-out enzyme and inhibitor binding, the macromolecule binding to its cognate peptide now sterically interferes with proteolytic cleavage at a proteolytic cleavage site in close proximity to the peptide of interest. Thus, macromolecule binding prevents sensor activation by the protease. This system is particularly useful for drug screening as peptide-protein antagonist drugs will restore protease access and lead to signal generation [14].

Below is a detailed description of the construction and use of these sensors.

2 Materials

2.1 Cloning of Expression Constructs

1. Expression constructs can either be synthesized commercially (Genscript) or custom cloned.

2. A plasmid coding for Tem1-BLIP D49A (see Note 1) inserted into the NdeI and XhoI sites of a pET28a vector with a C-terminal His6-tag (see Note 2). Here, BLIP stands for β-lactamase inhibitor protein while D49A refers to a mutation in BLIP that reduces affinity for Tem1 [15]. The coding DNA sequence for a fusion of thermostable Tem1 β-lactamase enzyme [16] and BLIP which is used as a template for inserting the desired peptide sequence between Tem1 and BLIP, and if necessary at the N- or C-termini as well, is given (see Note 3). Also, it is advisable to avoid AmpR cassette containing vectors (see Note 4).

3. Standard reagents for high-fidelity PCR amplification: 10× reaction buffer, 10 mM each dNTP solution, and a thermo-
stable DNA polymerase such as Pfu Turbo polymerase (e.g., Agilent).

4. Standard desalted oligonucleotide primers for PCR. The primer stock solution is 10 μM. For illustration, primers used for inserting a proteolytic cleavage site for SplB (available from ThermoFisher; also known as WELQut protease) between Tem1 and BLIP are shown: SplB-ins forward: 5′-TGGGAACTGCAGGGTTCTAGTGGTGCGAAAATT-3′ and SplB-ins reverse: 5′-ACCCTGCAGTTCCCAACCAGGCACCCTACTACCGCCCCAATG-3′.

5. PCR purification kit (e.g., QIAquick PCR purification kit, QIAGEN).

6. In-Fusion cloning kit (Clontech). The user may substitute another cloning method of choice instead.

7. Competent *E. coli* cells for transforming DNA: One Shot TOP10 Chemically Competent Cells *E. coli* (e.g., Invitrogen).

8. LB-Kan agar plates: 5 g/L yeast extract, 10 g/L peptone, 10 g/L NaCl, and 15 g/L agar supplemented with 50 μg/mL kanamycin.

9. LB-Kan liquid medium: 5 g/L yeast extract, 10 g/L peptone, and 10 g/L NaCl supplemented with 50 μg/mL kanamycin.

10. Commercial plasmid miniprep kit (e.g., Qiagen) or self-made reagents.

11. Oligonucleotide primers suitable for sequencing newly constructed plasmids. If using the pET28a vector, petF2 (5′-CATCGGTGATGTCGGCGAT-3′) and pETR (5′-CGGATATAGTTCCTCCTTTCAGCA-3′) are suitable for sequencing from the N- and C-terminal ends, respectively.

12. DNA Sequencing facility and suitable sequence analysis software (e.g., DNAStar Lasergene Suite or Snapgene).

### 2.2 Expression of Recombinant Proteins

1. SHuffle T7 Competent *E. coli* (NEB). (*See also* Note 5).

2. LB-Kan agar plates: 10 g/L yeast extract, 20 g/L peptone, 20 g/L sodium chloride, and 15 g/L agar supplemented with 50 μg/mL kanamycin. Other non-ampicillin plates are also suitable for use with alternative vector systems.

3. LB-Kan liquid medium: 10 g/L yeast extract, 20 g/L peptone, and 20 g/L sodium chloride supplemented with 50 μg/mL kanamycin.

4. Stock solution for inducing protein expression: 1 M IPTG.

5. Erlenmeyer flasks and incubator-shaker for bacterial cell culture.

6. Centrifuge for pelleting bacterial cells (e.g. Beckman ultracentrifuge).
2.3 Purification of Expressed Protein

1. Lysis and Binding buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 30 mM imidazole, pH 7.4.
2. Elution buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 500 mM imidazole, pH 7.4.
3. Sonicator or other means of cell lysis.
4. Syringes and 0.45 μm filter to clarify cell lysate.
5. 1 mL Nickel Sepharose His Trap column (GE Healthcare).
6. SDS-PAGE electrophoresis equipment to verify protein expression and quality.

2.4 Biosensor Assay

1. 96- or 384-well transparent bottom plate (Greiner Bio) for absorbance reading. For alternative readouts based on fluorescence or luminescence, other types of plates are required.
2. Phosphate buffered saline (PBS) as diluent: 10 mM Na2HPO4, 2 mM KH2PO4, 137 mM NaCl, and 2.7 mM KCl, pH 7.4.
3. 1 mM nitrocefin substrate stock solution: 1 mg nitrocefin (Merck) dissolved in 100 μL DMSO and 1.9 mL PBS as per manufacturer’s instructions.
4. Multi-well plate reader capable of reading OD at 492 nm. Target protease or target protein binder whose cleavage or binding site is present in the biosensor linker. By way of example, enterokinase (NEB), TEV protease, and anti-HA tag antibody have been successfully detected [13].

3 Methods

3.1 Plasmid Construction

1. After the gene coding for Tem1-BLIP (see Note 3) has been synthesized, it may be used as a standard template for inserting suitable peptides as required. We generally use custom gene synthesis to obtain the backbone, but the plasmid may also be obtained from our lab upon request. For cloning, we typically use inverse PCR with a reverse primer binding at the 3’ end of the Tem1 ORF and a forward primer binding at the 5’ end of the BLIP ORF while the 5’ tails encode the desired peptide and linker amino acids. Furthermore, the 5’ fifteen bases of the primers are complementary to each other to enable Infusion cloning. Alternatively, if the sequence of the insert is too long, the desired plasmid can be constructed by means of serial PCR using overlapping primers. We frequently employ primers that have been gel purified, especially if the oligonucleotide is greater than 40 bases to avoid single-base dropouts and frameshifts.
2. A typical PCR reaction based on Pfu Turbo polymerase typically contains in a total volume of 50 μL (Pfu Turbo polymerase is
typically only added to the PCR when the denaturation temperature has been reached to reduce mispriming): 5 µL 10× reaction buffer.
1 µL 10mM each dNTP solution.
1 µL primer SplB-ins forward (10uM stock).
1 µL primer SplB-ins reverse (10uM stock).
3 µL DMSO.
1–10 ng DNA template.
1 µL Pfu Turbo polymerase.
H₂O added to a final volume of 50 µL.
3. The backbone is typically amplified using the following PCR cycling protocol: 95 °C for 3 min, [95 °C for 10 s, 52 °C for 10 s, 68 °C for 7 min] × 30 cycles, 72 °C for 5 min.
4. Following PCR, purify the reaction using a commercial PCR purification kit as per manufacturer’s instruction, and verify the resulting DNA product by analytical agarose gel electrophoresis. The DNA product should be ~7 kb when using Tem1-BLIP inserted into the pET28a vector backbone.
5. Add a suitable amount of purified PCR product into an infusion cloning reaction according to manufacturer’s instruction. The complementary 5′ fifteen bases on both forward and reverse primers will enable an inter-molecular infusion reaction that will lead to the formation of the desired plasmid. Dilute 5× with water and transform into One Shot TOP10 Chemically Competent Cells as per manufacturer’s instructions before plating them on LB-Kan agar plates and incubating overnight.
6. Inoculate a few of the resulting colonies in LB-Kan liquid medium overnight and prepare plasmid DNA using a commercial miniprep kit according to manufacturer’s instructions.
7. Sequence the resulting plasmids using primers pETF2 and pETR and ensure the sequence is correct.

3.2 Protein Expression and Purification

1. Transform the verified plasmid into SHuffle T7 Competent E. coli as per manufacturer’s instructions. Inoculate the resulting colonies overnight at 30 °C. The next day, dilute this culture at 1% into 800 mL LB-Kan liquid medium at 30 °C. Induce with IPTG (0.4mM final) when OD reaches 0.5–0.8 followed by overnight incubation at 16 °C. Pellet the cells by centrifugation. The cells can be frozen at −80 °C for later processing.
2. Resuspend the cells in 10 mL Lysis and Binding buffer in a 50 mL Falcon tube and place on ice. Lyse cells by sonication for 10 min using the following pulsing sequence: 5 s sonication at amplitude 35 followed by 10 s rest. Centrifuge at 4 °C at 10,000 × g for 10 min and collect the supernatant. Aspirate the
supernatant into a syringe and attach a 0.45 μm filter to the end. Filter the supernatant to remove large particles. Replace the filter if it becomes clogged.

3. Equilibrate a 1 mL Nickel Sepharose His Trap column with Lysis and Binding buffer by washing the column with 10 column volumes. Load the clarified protein lysate through the column. Keep an aliquot of the lysate and flow-through for analysis. Wash the column with 10 column volumes of Lysis and Binding buffer and collect an aliquot for analysis. Elute with 10 mL of Elution Buffer. Load equal volume fractions of all collected aliquots on a SDS-PAGE gel to estimate the efficiency of purification. Add glycerol to a final concentration of 50% and store at −20 °C.

4. Alternatively, if using an FPLC instrument, equilibrate column with 10 column volumes of Lysis and Binding buffer. Clarified lysate is loaded into the system at a flow rate of 1 mL/min. Wash column with 10 column volumes of Lysis and Binding buffer. Samples are eluted using gradient length of 25. The entire run can be done at a flow rate of 1 mL/min with the pressure kept below 0.30 MPa.

3.3 Biosensor Assays

3.3.1 Protease and Macromolecule-Binding Assay

1. The assays are generally carried out at room temperature in a 384-well transparent bottom plate.

2. The exact concentration of the sensor and protease or macromolecular binder must be determined empirically. Before starting the assay, test a range of sensor concentrations with varying concentrations of protease or macromolecular binder to determine the point of optimal sensitivity.

3. When setting up multiple reactions, it is desirable to make a master mix containing the common components. The typical reaction in a total volume of 25 μL is set up as follows.

4. Add a suitable volume of PBS that will eventually make up a total reaction volume of 25 μL (i.e., we add PBS first to prevent small volumes of other reagents from sticking to the side of the well or remaining stuck to the pipette tip).

5. Add 2 μL of diluted sensor to a final concentration of approximately 5 nM and mix with 2–5 μL of various dilutions of protease or macromolecular binder. The final optimal concentration of components needs to be determined empirically. For enterokinase or TEV protease, the threshold of detection is approximately in the mid-picomolar to low-nanomolar range, respectively.

6. Initiate the reaction by adding 5–7 μL 1 mM nitrocefin substrate stock solution to a final concentration of 250 μM (see Notes 6 and 7).

7. After the addition of nitrocefin, immediately insert the plate into a plate reader and measure the OD at 492 nm every 2 min.
We typically measure enzyme activity by subtracting the OD492 value of read 3 from read 5 to determine substrate turnover/time (see Note 8).

3.3.2 Protease Exclusion Assay

1. The protease exclusion assay is carried out as above with the important exception that the macromolecular binder is added to the sensor and allowed to bind before the protease is added. This prevents premature cleavage of the sensor by protease before the macromolecule has an opportunity to bind and protect the sensor.

2. The typical reaction is set up as follows (see Note 9):

3. Add a suitable volume of PBS that will eventually make up a total reaction volume of 25 μL (see step 4 in Subheading 3.3.1).

4. Add 2 μL of diluted sensor to a final concentration of approximately 5 nM and mix with 2–5 μL of a macromolecular binder. This should be in excess of the sensor concentration to fully block all sensor molecules. The exact ratio will depend on the Kd and should be confirmed empirically. Ideally, these components should be part of a master mix.

5. Add 2 μL of the macromolecular antagonist. The concentration of the antagonist should be at least as much as the macromolecule to allow complete dissociation.

6. After adding the macromolecular binder, incubate for 5–10 min to allow an equilibrium to be reached (see Notes 10 and 11).

7. Add 2 μL of protease. The optimal concentration needs to be determined empirically, e.g., for enterokinase we found ~0.15 nM to be suitable.

8. Initiate the enzyme reaction by adding 5–7 μL 1 mM nitrocefin substrate stock solution to a final concentration of 250 μM (see Notes 6 and 7).

9. Analyze the reaction by plotting the rate of substrate turnover against the concentration of the antagonist or blocking macromolecule.

4 Notes

1. In some cases, the affinity of BLIP D49A for Tem1 may be modulated for improved performance. So reverting BLIP to WT or introducing additional mutations like W150A may be helpful. Wang et al. [19] and Reichmann et al. [20] provide a spectrum of mutations that could be tested for this purpose.

5. This concept can be extended to enzyme-inhibitor pairs other than Tem1-BLIP. When constructing alternative enzyme-linker-inhibitor ORFs, we would recommend placing purifica-
tion tags like His6-tag on the same side as the inhibitor. This ensures that if truncated fragments are present in the cell lysate, it is the inhibitor and not free enzyme which is pulled down. Free enzyme can lead to a high background.

3. The DNA sequence coding for Tem1-BLIP is given below. For expression, it may be inserted into pET28a NdeI XhoI sites:

\[
3'-\text{aaaaaaCATATGCACCCGGAAACCCTTGTTAAAAGTCA}
\]
\[
\text{AAGATGCGGAAACCAGCTGGGTGACCGCGTGGGC}
\]
\[
TATATTGAACTGGATCTGAACTCAGGCAAAATCCTG
\]
\[
GAATCGTTTCGTTCTGAAGAACGCTTCCCGATG
\]
\[
ATGTCAACCTTTAAGTTCTGTCTGCTGGT
\]
\[
CTATCCTGTGCCTGATTGTGCGGGTGCAAGGA
\]
\[
CAACCTGGACATCCAGCTGGGTGGACCCGAAAA
\]
\[
CACCTGACGCGACCGGATGCAGGTGCACGGAACTGT
\]
\[
GTAGCGCAGCAATTACCATGTCCTGATAAACACGGCT
\]
\[
CGAATCTGTCTGTACCCAGATCGGTGTTGT
\]
\[
GCAGAATCCTGACGTCCTGCTGAGCCGGTTGCT
\]
\[
GTTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
\[
CCGCTGCTGCTGCGATCGGCTGCTCAGGCTG
\]
\[
GGTTTATCGCGAGATGAAAAAGTGTGCGAGGCAAA
\]
\[
CGTGTTTCCTCCGCAGCGGATCTGCTGCAGGAAC
\]
\[
TGATGTCCGATGATGCAATCGGGCGGGGTTG
\]
4. pET28a or other non-ampicillin-based resistance marker plasmids are recommended because Tem1 is produced by the ampicillin resistance gene. This can interfere with or contaminate the recombinant sensor. Also ampicillin is a competitive inhibitor of the Tem1 enzyme, so it should be ensured that the reaction is free of it.

5. These cells are recommended as Tem1 and BLIP are disulfide bond containing proteins and need an oxidizing environment for their formation. SHuffle T7 Competent *E. coli* are modified to have an oxidizing environment in their cytoplasm, so disulfide bond containing protein can be expressed therein [17]. Alternatively, expression in the periplasm or re-folding from denatured proteins can be considered.

6. Nitrocefin is light sensitive and light exposure should be minimized.

7. Nitrocefin is known to be turned over in highly reducing or oxidizing environments [18]. We have found that the presence of strong reducing agents like DTT or β-mercaptoethanol in the storage buffer of some macromolecules can lead to rapid enzyme-independent turnover of nitrocefin. This should be tested during the initial optimization. If this is found to be the case, the relevant buffer should be exchanged before adding to the reaction.

8. The plate reader we use (Perkin Elmer Envision) takes 1–2 reads to stabilize the signal. We have been using read 3 as our initial measurement point. Also, we typically allow the reaction to proceed for about 30–60 min to ensure that it is complete.

9. It is highly recommended that a master mix comprising PBS, the sensor, and the macromolecular binder is set up as this saves time and allows time for the sensor and macromolecule to bind each other. We typically set up the reaction on ice to minimize potential thermal unfolding.

10. It is recommended that in each assay, suitable controls are included. These can include the protease and the sensor, but without the sterically interfering macromolecular binder to confirm that both sensor and protease are functional.

11. Some of the potential inhibitors we tested were colored and had a substantial OD492 value. This can generally be dealt with by normalizing the sensor reaction against a control reaction with the potential inhibitor and other components except the sensor itself, or by subtracting the OD at read 1 from all subsequent reads.
References

Chapter 12

DNA-Specific Biosensors Based on Intramolecular β-Lactamase-Inhibitor Complex Formation

Wouter Engelen and Maarten Merkx

Abstract

Synthetic protein switches that sequence-specifically respond to oligonucleotide-based input triggers provide valuable tools for the readout of oligonucleotide-based biomolecular systems and networks. Here, we discuss a highly modular approach to reversibly control the DNA-directed assembly and disassembly of a complex between TEM1-β-lactamase and its inhibitor protein BLIP. By conjugating each protein to a unique handle oligonucleotide, the enzyme-inhibitor pair is noncovalently assembled upon the addition of a complementary ssDNA template strand, resulting in inhibition of enzyme activity. Hybridization of an input-oligonucleotide that is complementary to a target recognition sequence in the ssDNA template strand results in the formation of a rigid dsDNA helix that mechanically disrupts the enzyme-inhibitor complex, hereby restoring enzyme activity. Following this noncovalent approach allowed straightforward tuning of the ssDNA template recognition sequence and target oligonucleotide lengths with only a single set of oligonucleotide-functionalized enzyme and inhibitor domains. Using a fluorescent substrate, as little as 10 pM target oligonucleotide resulted in a distinguishable increase in enzyme activity.

Key words β-lactamase, Reporter enzyme, DNA detection, Synthetic biology, Protein-DNA conjugation, Biosensor

1 Introduction

Synthetic protein switches are extensively used in synthetic biology, molecular imaging, and molecular diagnostics to study biological processes both in vitro and in vivo by responding to the presence of a specific input molecule. Ideally, molecular recognition (input) and signal generating functions (output) are part of separate domains in these switches, as such a modular organization allows straightforward exchange of input or output function [1]. A successful design approach that ensures efficient translation of input activation to output modulation, is to design switches that are able to adopt two conformational states; one in which two output domains can interact intramolecularly and one where their interaction is prohibited. The dynamic response of such a protein switch is
determined by the relative affinities of the output domain interaction and the magnitude of the input-induced thermodynamic change, e.g., the strength of ligand binding [2]. Examples of output domains that have been applied for these approaches include fluorescent or luminescent proteins domains forming FRET or BRET pairs. To ensure efficient energy transfer between them in either the on or off state, fluorescent domains can be used that have an increased tendency to form intramolecular complexes, or their close proximity can be promoted by so-called helper domains [3]. Our group used these self-associating fluorescent protein domains to develop robust FRET sensor protein for a range of analytes, including metal ions, bile acids, and antibodies [4–7]. In addition, by replacing the self-associating fluorescent domains by an enzyme-inhibitor pair with similar affinity, the optical readout of the antibody FRET sensor could be replaced by a more sensitive enzymatic readout [8].

The modular organization of protein switches based on mutually exclusive domain interactions does not only allow exchange of different output functions, but is also well suited to combine with oligonucleotide-based biomolecular systems and networks. Most approaches reported thus far to control enzyme activity by DNA are based on the templated assembly of split enzyme fragments [9, 10]. Whereas split reporter enzymes have the advantage of providing a low background activity (at least when used intermolecularly), split enzymes generally suffer from poor thermodynamic stability and protein complementation is often not reversible. Therefore, our group recently introduced an alternative approach that is based on the reversible assembly of a complex between the reporter enzyme TEM1-β-lactamase and its inhibitor domain BLIP [11–13]. This design exploits the inherent difference in mechanical properties between ssDNA and dsDNA to mechanically disrupt and spatially separate the enzyme-inhibitor pair upon binding of an input oligonucleotide to the input-binding module. To facilitate straightforward exchange of input sequences, a noncovalent approach is followed by conjugating both TEM1-β-lactamase and BLIP to individual, 21 nucleotide handle oligonucleotides. By appending anti-handle sequences to each end of the synthetic ssDNA target recognition sequence, a ternary complex is formed upon mixing of the three protein switch components (Fig. 1). Formation of this ternary complex results in a large increase in effective concentration and hereby induces an intramolecular interaction between the enzyme-inhibitor pair. Hybridization of the complementary target oligonucleotide to the input module results in the formation of a rigid double-helix which mechanically disrupts the enzyme-inhibitor interaction, hereby switching on enzyme activity.

Since this system is based on two well-folded and stable protein domains, it is more robust compared to the assembly of split-protein fragments and readily reversible. Moreover, because of the
catalytic conversion of substrate to product, the enzyme-based DNA detection is more sensitive than direct optical readout of conformational switching sensors such as molecular beacons [14]. Here, we provide detailed protocols for the expression, purification, and bioconjugation of the TEM1-β-lactamase- and BLIP-oligonucleotide conjugates, as well as their assembly into an easily tunable system for oligonucleotide detection in molecular diagnostics and the readout of DNA-based molecular circuits. While the DNA-directed protein switch described in this chapter is originally designed to respond to ssDNA, the range of input molecules could be extended to small molecules and proteins by intelligent redesign of the input module using conformation-switching aptamers. Since plasmids for expression of both protein components will be made accessible via AddGene, we hope that these protocols will enable the TEM1-β-lactamase/BLIP system to be widely used. In addition, both the design principles and the bioconjugation methods described here for the TEM1-β-lactamase/BLIP system should be generally applicable to other enzyme-inhibitor pairs, provided that they interact with a similar affinity.

2 Materials

2.1 Protein Expression and Purification

1. Plasmid pET29a_TEM1-β-lactamaseE104D_CtermC (encoding periplasmic leader sequence, His-tag, Thrombin cleavage site, TEM1-β-lactamase(E104D), cysteine, Strep-tag) (see Notes 1 and 2).

Fig. 1 Design of a modular DNA-directed biosensor based on intramolecular TEM1-β-lactamase–inhibitor complex formation. To implement full modularity, a multi-component strategy is used where the input and output modules are noncovalently assembled. Therefore, the enzyme and inhibitor proteins are conjugated to oligonucleotide handles that are complementary to anti-handle sequences appended to the target recognition sequence, allowing the assembly of the enzyme-inhibitor pair in a ternary complex ("OFF” state). Hybridization of a complementary input oligonucleotide to the target recognition sequence results in the formation of a rigid double-helix, hereby mechanically disrupting the enzyme-inhibitor pair, i.e., the enzyme is switched on (“ON” state).
2. Plasmid pET29a_BLIP_NtermC (encoding periplasmic leader sequence, His-tag, Thrombin cleavage site, cysteine, GSHG linker, \(\beta\)-lactamase inhibitor protein (BLIP)) (see Notes 1 and 2).

3. Competent *Escherichia coli* BL21 (DE3) cells.

4. LB medium: 10 g peptone, 5 g yeast extract and 10 g NaCl in 1 L dH2O supplemented with 30 mg/L kanamycin.

5. 100 mM isopropyl \(\beta\)-D-1-thiogalactopyranoside (IPTG) stock solution. Filtered to sterilize (0.2 \(\mu\)m).


7. High-speed centrifuge.

8. Osmotic shock solution: 30 mM Tris–HCl, 1 mM EDTA, 20% sucrose at pH 8.0.

9. Low salt solution: 5 mM MgSO4.

10. 10× Tris–HCl buffer: 200 mM Tris–HCl at pH 7.4.

11. His-tag binding resin.


13. Bind buffer: 20 mM Tris–HCl, 500 mM NaCl, 30 mM imidazole at pH 7.9.

14. Wash buffer: 20 mM Tris–HCl, 500 mM NaCl, 60 mM imidazole at pH 7.9.

15. Elution buffer: 20 mM Tris–HCl, 500 mM NaCl, 400 mM imidazole, 2 mM TCEP at pH 7.9.

16. Storage buffer: 20 mM Tris–HCl, 150 mM NaCl, 2 mM TCEP at pH 7.9.

17. Amicon centrifugal filter unit (NMWL = 10 kDa).

18. Nanodrop UV-VIS spectrophotometer.

### 2.2 Maleimide Functionalization of Amine-Modified Oligonucleotides

1. Oligonucleotides modified with a primary amine at their 3′- or 5′-terminus (e.g., Integrated DNA Technologies).

2. Dimethylsulfoxide (DMSO).

3. Ethanol (cooled to \(-30\) °C).

4. Phosphate buffered saline (PBS): 100 mM sodium phosphate, 150 mM NaCl at pH 7.2.

5. 5 M NaCl.


### 2.3 Conjugation of Maleimide-Functionalized Oligonucleotides to Proteins

1. Ligation buffer: 100 mM sodium phosphate at pH 7.0.

2. Desalting column: PD-10 (e.g., GE Healthcare, Cat No: 17-0851-01).

3. Strong anion-exchange spin column, mini (e.g., Thermo Scientific, Cat No: PI90008).
4. Buffer A: 20 mM Tris–HCl, 50 mM NaCl at pH 8.0.
5. Buffer B: 20 mM Tris–HCl, 1 M NaCl at pH 8.0.

1. Complexation buffer: 50 mM sodium phosphate, 100 mM NaCl at pH 7.0.
2. 2x sample buffer: 125 mM Tris–HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.01% bromophenol blue.
3. 30% (w/v) acrylamide/bisacrylamide solution (29:1, w/w).
5. 10% (w/v) ammonium persulfate solution in water.
6. Isopropanol.
7. Running buffer: 187 mM glycine, 19 mM Tris–HCl, 3.5 mM SDS.
8. Protein molecular weight marker, prestained.
9. Coomassie Brilliant Blue.

2.4 Semi-Native SDS-PAGE of Reconstituted DNA-Directed Enzyme-Inhibitor Complex

1. PBS+: 50 mM sodium phosphate, 100 mM NaCl, 1 mg/mL bovine serum albumin (BSA) at pH 7.0.
2. 50 μM ssDNA template stock solution dissolved in H₂O (Table 1).
3. 50 μM ssDNA target stock solution dissolved in H₂O (Table 1).
4. 500 μM nitrocefin, chromogenic β-lactamase substrate (e.g., EMD Millipore, VWR, Cat No: 80,017–707) dissolved in PBS.
5. Multiwell plate reader (with absorbance mode).
6. Transparent 96-well plates.

2.5 Enzyme Activity Assay of Reconstituted DNA-Directed Enzyme-Inhibitor Complex

1. PBS+: 50 mM sodium phosphate, 100 mM NaCl, 1 mg/mL BSA at pH 7.0.
2. 50 μM ssDNA template stock solution dissolved in H₂O (Table 1).
3. 50 μM ssDNA target stock solution dissolved in H₂O (Table 1).
4. 20 μM CCF2-FA, fluorescent β-lactamase substrate (e.g., Invitrogen, Cat No: K1034) dissolved in PBS.
5. Multiwell plate reader (with fluorescence mode).
6. Black 96-well plate.

2.6 Enzyme Activity Assay at Different Input Oligonucleotide Concentrations

DNA Sensors Based on β-Lactamase
### Table 1
Sequences of handle oligonucleotides, ssDNA templates, and ssDNA targets

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODN1</td>
<td>5′-TGTCACCGATGAAACTGTCTA-NH₂-3′</td>
</tr>
<tr>
<td>ODN2</td>
<td>5′-H₂-N-GTGATGTTGGAGGAGGA-3′</td>
</tr>
<tr>
<td>Template 0</td>
<td>5′-TTCCCTCTACCACCTACATCATAGAAGTTTTTCATCGGTGACA-3′</td>
</tr>
<tr>
<td>Template 10</td>
<td>5′-TTCCCTCTACCTACATCAACCGACAATAGAGTTTTTCATCGGTGACA-3′</td>
</tr>
<tr>
<td>Template 20</td>
<td>5′-TTCCCTCTACCTACATCACACCGCAATAGAGTTTTTCATCGGTGACA-3′</td>
</tr>
<tr>
<td>Template 30</td>
<td>5′-TTCCCTCTACCTACATCACCACCGCAATAGAGTTTTTCATCGGTGACA-3′</td>
</tr>
<tr>
<td>Template 40</td>
<td>5′-TTCCCTCTACCTACATCACCACCGCAATAGAGTTTTTCATCGGTGACA-3′</td>
</tr>
<tr>
<td>Template 50</td>
<td>5′-TTCCCTCTACCTACATCACCACCGCAATAGAGTTTTTCATCGGTGACA-3′</td>
</tr>
<tr>
<td>Target 10</td>
<td>5′-TGCGGTGTTGGT-3′</td>
</tr>
<tr>
<td>Target 20</td>
<td>5′-GTGGTGGTGGTGGTGTTGTGGTG-3′</td>
</tr>
<tr>
<td>Target 30</td>
<td>5′-GGTGTTGGTGGTGGTGTTGTGGTG-3′</td>
</tr>
<tr>
<td>Target 40</td>
<td>5′-TTGGTGGTGGTGGTGGTGTTGTGGTG-3′</td>
</tr>
<tr>
<td>Target 50</td>
<td>5′-TGGTGTTGGTGGTGGTGGTGTTGTGGTG-3′</td>
</tr>
</tbody>
</table>
3 Methods

To facilitate hybridization-induced assembly of the enzyme-inhibitor pair the oligonucleotide handles have to be conjugated to the individual protein domains. In particular, cysteine is an attractive target for protein conjugation due to its low prevalence in natural proteins and ease of insertion using site-directed mutagenesis. In contrast to non-natural amino acids, which allow fully orthogonal protein conjugation chemistries, the incorporation of a non-native cysteine typically does not result in a significant decrease in expression yield (~0.3 mg/L culture for both TEM1-β-lactamase and BLIP).

To allow straightforward protein expression while ensuring site-specific conjugation of the oligonucleotide handles to the proteins (yielding homogeneous ODN-protein conjugates), thiol-maleimide coupling was used targeting a genetically inserted cysteine on the proteins. To install the maleimide moiety on the oligonucleotide handles, the heterobifunctional crosslinker Sulfo-SMCC, consisting of an NHS-ester, cyclohexyl spacer, and maleimide, was coupled to a 5′- or 3′-amine-modified oligonucleotide for BLIP and TEM1-β-lactamase, respectively. Subsequent ethanol precipitation of the maleimide-functionalized oligonucleotides allowed the removal of non-reacted Sulfo-SMCC. Finally, the oligonucleotides were conjugated to the cysteine inserted in the proteins (Fig. 2).

Fig. 2 (a) Crystal structure of a complex between TEM1-β-lactamase and BLIP (PDB: 3C7V). (b) ODN-protein conjugation strategy. Amine-modified oligonucleotides are reacted with the NHS-ester of the heterobifunctional crosslinker Sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (Sulfo-SMCC) installing a thiol-reactive maleimide moiety on the oligonucleotides. After the removal of the excess crosslinker by ethanol precipitation, the maleimide-activated oligonucleotides are reacted with a cysteine in the proteins. Adapted with permission from ref. 13. Copyright 2015 American Chemical Society
Conversion yields (typically 50–90%) can be estimated by SDS-PAGE analysis. Subsequent Ni²⁺-affinity chromatography and anion-exchange chromatography to remove excess oligonucleotides and unconjugated proteins, respectively, resulted in pure ODN-protein conjugates.

Formation of the intermolecular ternary complex is performed by mixing ODN1-TEM1-β-lactamase, ssDNA template, and ODN2-BLIP in a 1:1.2:2 molar ratio. Using a slight excess of ssDNA template and ODN2-BLIP ensures that no free ODN1-TEM1-β-lactamase or ODN1-TEM1-β-lactamase-template remains, both of which would contribute to background activity. In order to confirm the formation of homogeneous ternary complexes upon hybridization of the individual components, the mixtures were analyzed by semi-native PAGE. Quantitative conversion to the ternary complex was observed upon the addition of the ssDNA template to a mixture of the ODN-enzyme and ODN-inhibitor conjugates. A gel-shift to higher molecular weight after the addition of a complementary input oligonucleotide indicates the formation of the desired quaternary complex. Whereas semi-native PAGE suggests successful formation of the ternary DNA-directed enzyme-inhibitor complex, enzyme-activity assays were performed to test whether the enzyme-inhibitor interaction is indeed induced upon hybridization of both ODN-protein conjugates to the ssDNA template strand. To this end, the complexes (containing templates with target recognition sequences of 0 to 50 nucleotides) were diluted to a final concentration of 1 nM. After the addition of the colorimetric substrate nitrocefin, substrate conversion was monitored by measuring the optical absorbance of the formed product as a function of time. Whereas a strong inhibition of enzyme activity was observed with all target recognition sequence lengths, inhibition efficiencies increased as the target recognition sequence length increases. This observation suggests that a certain degree of flexibility between the protein partners is preferable to allow more efficient intramolecular complex formation. In contrast, full recovery of enzyme activity was only achieved upon the addition of input sequences of >40 nucleotides, suggesting that a double-helix of at least 40 nucleotides is required to overcome flexibility in the protein termini and SMCC-linker and spatially separate the interacting protein partners.

Making use of an enzyme as output domain provides the advantage over conventional fluorescent domains that substrate molecules are catalytically converted, hence the output signal accumulates. This allows the use of much lower concentrations of the sensor complex, hereby increasing the sensitivity of the input oligonucleotide detection. When using 100 pM of the DNA-specific biosensor together with the more sensitive fluorescent substrate CCF2-FA, as little as 10 pM of input oligonucleotide results in a distinguishable increase in enzyme activity.
1. Transform competent *E. coli* BL21 (DE3) cells separately with the pET29a_TEM1-β-lactamaseE104D and pET29a_BLIP plasmids *(see Note 3).*

2. For each construct, inoculate 20 mL of LB medium supplemented with 30 mg/L kanamycin with a single colony of transformed bacteria and incubate overnight at 37 °C in a shaking incubator at 250 rpm.

3. Transfer the seed culture to 2 L of LB medium supplemented with 30 mg/L kanamycin.

4. Grow the bacteria in a shaking incubator at 200 rpm at 37 °C. At an optical density at $\lambda = 600$ nm of 0.6, add 2 mL of IPTG stock yielding a final concentration of 100 μM IPTG and incubate at 16 °C for 20 h.

5. Harvest the cells by centrifugation at $10,000 \times g$ for 10 min at 4 °C.

6. Resuspend the pelleted cells in 300 mL osmotic shock solution and incubate for 10 min at room temperature.

7. Pellet the cells by centrifugation at $12,000 \times g$ for 20 min.

8. To extract the periplasmic fraction, resuspend the pelleted cells in 300 mL ice-cold low salt solution and incubate for 10 min on ice prior to centrifugation at $40,000 \times g$ for 40 min at 4 °C.

9. Add 10× Tris–HCl buffer to the supernatant (i.e., the periplasmic fraction) to yield a final concentration of 20 mM Tris–HCl.

10. Charge a 2 mL His-bind column (gravity flow) with Ni$^{2+}$ ions by washing the column with 10 column volumes of charge buffer and subsequently equilibrate with 10 column volumes bind buffer.

11. Load the periplasmic fraction on the His-bind column and allow it to enter the column bed completely. Wash the column with 10 column volumes of wash buffer and subsequently elute the protein of interest with 10 column volumes of elution buffer.

12. Concentrate the eluted proteins to a final volume of ~2.5 mL using the centrifugal filter unit.

13. Exchange the buffer to storage buffer by gel filtration using a PD-10 desalting column and determine the concentration by measuring the optical absorption at $\lambda = 280$ nm with the Nanodrop UV-VIS spectrophotometer and calculated extinction coefficients *(see Note 4).*

14. Flash-freeze the purified proteins in liquid nitrogen and store at −80 °C.
3.2 Maleimide Functionalization of Amine-Modified Oligonucleotides

1. Dissolve the amine-modified oligonucleotides in PBS to a concentration of ~1 mM.
2. Dissolve 2 mg of Sulfo-SMCC (1 vial) in 300 μL DMSO to yield a stock concentration of ~20 mM.
3. Mix an equal volume of Sulfo-SMCC with the amine-modified oligonucleotides yielding a 20-fold molar excess of Sulfo-SMCC in 1:1 PBS:DMSO.
4. Incubate the mixture at room temperature under continuous shaking at 850 rpm for 2 h.
5. To precipitate the oligonucleotides, add 10% (v/v) 5 M NaCl and 300% (v/v) ethanol precooled to −30 °C, and store the mixture at −30 °C for at least 1 h.
6. Pellet the precipitated oligonucleotides by centrifugation at 19,000 × g rpm for 15 min at 4 °C.
7. Dissolve the oligonucleotides in PBS to a concentration of ~1 mM and repeat steps 5 and 6 of the ethanol precipitation procedure twice (see Note 5).
8. Finally, dry the pelleted oligonucleotides to air overnight at room temperature.

3.3 Conjugation of Maleimide-Functionalized Oligonucleotides to Proteins

1. Immediately prior to oligonucleotide conjugation, desalt the proteins to ligation buffer to remove TCEP by gel filtration using a PD-10 desalting column and dilute to ~50 μM (see Note 6).
2. Add the appropriate volume of protein solution to the dried oligonucleotide to yield a threefold molar excess of maleimide-functionalized oligonucleotide.
3. Incubate the mixture at room temperature under continuous shaking at 850 rpm for 2 h.
4. To remove excess oligonucleotides load the reaction mixture on a 1 mL His-bind column charged with Ni²⁺ ions and equilibrated with bind buffer. Wash with 10 column volumes of bind buffer.
5. Elute the oligonucleotide-protein (ODN-protein) conjugates in 500 μL fractions using 10 column volumes of elution buffer.
6. Analyze the reaction efficiency and purity of the elution fractions by 12% SDS-PAGE.
7. Combine all elution fractions that contain ODN-protein conjugates.
8. Exchange the buffer to buffer A by gel filtration using a PD-10 desalting column.
9. Prewash a strong anion-exchange spin column by applying 400 μL buffer A and spinning at 2000 × g for 5 min and discarding the flow-through.

10. Repeatedly add 400 μL of desalted ODN-protein conjugates and centrifuge for 5 min at 2000 × g until the entire sample is loaded on the anion-exchange spin column. Discard the flow-through.

11. Remove non-conjugated protein by adding 400 μL buffer A and centrifuging for 5 min at 2000 × g. Discard the flow-through. Repeat this step ten times.

12. Elute the purified ODN-protein conjugate by adding 200 μL buffer B and centrifuging for 5 min at 2000 × g. Repeat this step ten times.

13. Analyze the elution fractions by 12% SDS-PAGE and combine all fractions containing the purified ODN-protein conjugates (Fig. 3).

14. Quantify the concentrations of the ODN-protein conjugates by measuring the absorbance at λ = 260 nm and using extinction coefficients of the oligonucleotides.

15. Aliquot the purified ODN-protein conjugates in 20 μL fractions and flash-freeze in liquid nitrogen. Store the ODN-protein conjugates at −80 °C.

![SDS-PAGE analysis of the conjugation and anion exchange purification of a maleimide-functionalized oligonucleotide to a cysteine in (a) TEM1-β-lactamase and (b) BLIP on a 12% SDS-PAGE gel stained with Coomassie blue. (rxn: reaction mixture, buffer A wash: first five wash fractions with buffer A (low ionic strength) from the anion-exchange spin column, buffer B elution: first eight elution fractions with buffer B (high ionic strength) from the anion-exchange spin column)](image)

Fig. 3 SDS-PAGE analysis of the conjugation and anion exchange purification of a maleimide-functionalized oligonucleotide to a cysteine in (a) TEM1-β-lactamase and (b) BLIP on a 12% SDS-PAGE gel stained with Coomassie blue. (rxn: reaction mixture, buffer A wash: first five wash fractions with buffer A (low ionic strength) from the anion-exchange spin column, buffer B elution: first eight elution fractions with buffer B (high ionic strength) from the anion-exchange spin column.
1. Prepare a 0.75-mm-thick 12% resolving gel by mixing 4 mL 30% (w/v) acrylamide, 2.5 mL 1.5 M Tris–HCl pH 8.9, 3.3 mL dH₂O, 100 μL 10% (w/v) SDS, 100 μL 10% (w/v) ammonium persulfate and 6 μL TEMED. Cast the gel up to 2 cm below the top and cover with a layer of isopropanol. Allow the gel to polymerize for 30 min.

2. Prepare a stacking gel by mixing 830 μL 30% (w/v) acrylamide, 630 μL 1 M Tris–HCl pH 6.8, 3.4 mL dH₂O, 50 μL 10% (w/v) SDS, 50 μL 10% (w/v) ammonium persulfate, and 3 μL TEMED. Pour the mixture on the top of the resolving gel and insert the comb. Allow the gel to polymerize for 30 min.

3. To ensure efficient complex formation and inhibition of the enzyme in the assembled complex, sequentially mix ODN1-TEM1-β-lactamase, ssDNA template (with a target recognition sequence length of 0, 10, 20, 30, 40, or 50 nucleotides) and ODN2-BLIP in a 1:1.2:2 ratio at low micromolar concentrations in complexation buffer and incubate for 30 minutes at room temperature.

4. Mix the hybridized samples prepared in step 2 with an equal volume of 2× sample buffer and load it on a semi-native 12% SDS-PAGE gel prepared in step 1. Run the gel for 15 min at 90 V followed by 1 h at 125 V.

5. Remove the gel from the glass plates and wash with dH₂O for 5 min. Subsequently stain the gel with Coomassie blue for 30 min. Then destain the gel with dH₂O until clear bands appear and image the gel (Fig. 4).

3.5 Enzyme Activity Assay of Reconstituted DNA-Directed Enzyme-Inhibitor Complex

1. Mix ODN1-TEM1-β-lactamase, ODN2-BLIP, and ssDNA template (with target recognition sequences of 0, 10, 20, 30, 40, and 50 nucleotides) in PBS⁺ at low micromolar concentrations in a ratio of 1:2:1.2 and incubate for 30 min at room temperature.

2. Dilute the mixtures in PBS⁺ to a final concentration of 50 nM ODN1-TEM1-β-lactamase.

3. In a 96-well plate, mix 4 μL of the diluted complexes with 152 μL PBS⁺ and 24 μL target oligonucleotides (from 100 nM stock solution in PBS⁺) and incubate for 1 h at room temperature.

4. Add 20 μL nitrocefin (from 500 μM stock solution in PBS) yielding a final concentration of 1 nM DNA-directed enzyme-inhibitor complex, 12 nM target oligonucleotide, and 50 μM nitrocefin.

5. Directly place the plate in the platerader and proceed with measuring the optical absorbance in time of the formed product at 486 nm.
6. Derive the hydrolysis rate by calculating the slope of the absorbance in time of the initial, linear regime. Correct for background hydrolysis of nitrocefin in the absence of enzyme (Fig. 5).

3.6 Enzyme Activity Assay at Sub-nanomolar Input Oligonucleotide Concentrations

1. Mix ODN1-TEM1-β-lactamase, ODN2-BLIP, and ssDNA template (with a target recognition sequence length of 40 nucleotides) in PBS at low micromolar concentrations in a ratio of 1:2:1.2 and incubate for 30 min at room temperature.

2. Dilute the mixture in PBS to a final concentration of 5 nM ODN1-TEM1-β-lactamase.

3. In a 96-well plate, mix 4 μL of the diluted complex with 152 μL PBS and 24 μL target oligonucleotide (concentrations ranging from 100 pM to 20 nM) and incubate for 1 h at room temperature.

4. Add 20 μL CCF2-FA (20 μM in PBS), yielding a final concentration of 100 pM DNA-directed enzyme-inhibitor complex, 10 pM to 2 nM target oligonucleotide, and 2 μM CCF2-FA. Incubate for 45 min at room temperature.
5. Place the plate in the platereader and monitor the formation of the fluorescent product at 409 nm excitation and 447 nm emission.

6. Derive the hydrolysis rate by calculating the slope of the fluorescence in time within the linear regime (Fig. 6).

4 Notes

1. Previous work has shown that conjugation efficiencies greatly depend on the steric availability of the inserted cysteine. In case of low conversion efficiencies insert a flexible GGSGG linker between the protein and cysteine [10].

2. Plasmids encoding for TEM1-β-lactamase and BLIP will be made available via AddGene.

3. Based on previous work using the intramolecular complex formation between TEM1-β-lactamase and BLIP as output module for genetically encoded protein switches, we here use the E104D mutant of TEM1-β-lactamase ($K_i = 1500$ nM) as the wild-type proteins bind too strongly ($K_i = 0.5$ nM), which might result in intermolecular complex formation and oligomerization of the DNA-directed protein switch.

4. TCEP is added to the storage buffer to reduce intermolecular disulfide bonds that render the genetically inserted cysteine
nonreactive toward maleimides. TCEP can also reduce native, intramolecular disulfide-bonds. However, the BLIP and TEM1-β-lactamase proteins did not show any (irreversible) reduction of intramolecular disulfides.

5. To ensure efficient conjugation of maleimide-functionalized oligonucleotides to the proteins of interest, it is crucial to perform at least three rounds of ethanol precipitation after reacting the amine-modified oligonucleotides with Sulfo-SMCC to ensure that all excess Sulfo-SMCC is removed since this will compete for conjugation to the protein. A single round of ethanol precipitation has shown to significantly reduce final conjugation efficiencies.

6. To avoid the formation of nonreactive disulfide-linked protein dimers directly proceed with mixing the desalted proteins (i.e., after removing TCEP) with the maleimide-functionalized oligonucleotides. Various reports have published mixed results with TCEP interfering with maleimide conjugation. In our experience, TCEP significantly lowers conversion yields and is therefore removed prior to the reaction with the maleimide-functionalized oligonucleotide.
Acknowledgments

The authors would like to thank Dr. Brian Janssen for co-developing the methods described in this chapter. This work was supported by an ERC Starting grant (ERC-2011-StG-280255) and by NanoNextNL, a micro and nanotechnology consortium of the government of The Netherlands and 130 partners.

References


Part V

Proteolytic Sensors
Chapter 13

Engineering and Characterizing Synthetic Protease Sensors and Switches

Viktor Stein and Kirill Alexandrov

Abstract

Proteases are finding an increasing number of applications as molecular tools and reporters in biotechnology and basic research. Proteases are also increasingly incorporated into synthetic genetic signaling circuits equipping cells with tailored new functions. In the majority of cases however, proteases are employed in constitutively active forms which limits their utility and application as molecular sensors. The following chapter provides a detailed experimental protocol for converting constitutively active proteases into regulated protease receptors. Such receptors can potentially sense, transduce, and amplify any molecular input, thereby opening up a range of new applications in basic research, biotechnology, and synthetic biology.

Key words Protein switches, Protein engineering, Proteases, Diagnostic reagents

1 Introduction

1.1 Proteases as Versatile Tools in Biotechnology and Basic Research

Proteases constitute one of the most abundant classes of enzymes that execute key physiological functions across all kingdoms of life [1]. Proteases have evolved to function inside as well as outside cellular environments taking on diverse roles such as regulating cell death, controlling blood coagulation, digesting nutrients, and remodeling the extracellular environment. Chemically, proteases catalyze the irreversible cleavage of a peptide bond. Their substrate specificity ranges from peptide motives as short as two amino acids, to complex tertiary interactions that are mediated by structurally distinct protein domains. Given the central importance of proteases to many physiological processes and the irreversible nature of the proteolytic cleavage, protease function is frequently tightly regulated. One common mechanism relies on the expression of inactive zymogens that require additional posttranslational processing by an activating protease that cleaves off an active site-directed inhibitor or triggers complex conformational rearrangements allowing a protease to transition into a catalytically active conformation.
Proteases are also increasingly applied in biotechnology as therapeutic and diagnostic reagents, or in basic research as molecular reporters and tools in the production of recombinant proteins. In addition, proteases are finding increasing application in synthetic biology as components of synthetic signaling circuits that equip cells with new tailor-engineered functions [2, 3]. The key advantage here is the fact that, by virtue of cleaving a peptide bond, they can be readily interfaced with any protein-associated biological function. Notably, in comparison to alternative posttranslational protein modifications such as protein phosphorylation, proteolysis constitutes a drastic conformational change as one polypeptide is split into two with the two resultant polypeptides transitioning from an intra- to an intermolecular state. This transition is associated with a significant change in free energy that can be readily exploited to engineer protease-inducible actuators: e.g. to activate protein function by cleaving off an autoinhibitory (AI) domain, or to inhibit protein function by cleaving functionally important residues. Consequently, a capacity to engineer synthetic protease receptors in a systematic fashion constitutes a powerful tool to engineer biological signaling transduction systems. The following protocol chapter gives a detailed guide how to engineer and characterize synthetic protease switches with custom response functions focusing on four elementary types of signaling sensing and transducing mechanisms: This includes autoinhibited protease transducers, allosteric protease receptors, integrated sensing and amplification circuits, as well as proximity-dependent protease sensors (Fig. 1).

Elementary protease transducers, allosteric protease receptors, and proximity-dependent protease sensors are engineered by recombinating naturally occurring proteases with binding receptors and AI-domains that bind and block the active site of the protease and thus shut down its activity in the basal state. Individual components can either be derived from natural sources or be artificially engineered. Given the modular organization of components, the construction of synthetic protease switches typically occurs in a step-wise fashion. This means, protease transducers, their cognate AI-domains, and binding receptors can be first engineered and characterized individually before they are assembled into functional switches, while empirically optimizing the structure and length of the linkers connecting individual domains.

In our initial proof-of-concept studies, we chose established proteases based on members of the Nuclear Inclusion a (NIA) family. The most prominent members are Tobacco Etch Virus (TEV) and Tobacco Vein Mottling Virus (TVMV) protease that have found widespread applications as tools in molecular and cell biological research [4]. In addition, we chose the NS3 protease of Hepatitis C Virus (HCV) which has been extensively studied as a

1.2 Choice of Proteases as Elementary Signal Transducers
drug target, but is also increasingly adopted to develop tools for basic research. Notably, a range of peptide-based, active site directed inhibitors have been devised for HCV that can serve as AI-domains and readily be identified from the existing literature [5].

In general terms, proteases are preferably well characterized both biochemically and structurally. The availability of large protein families is also advantageous as it allows to infer biochemical and structural information from phylogenetic data. Notably, for viral proteases whose native function is to process the viral polyprotein, the putative cleavage site can be directly identified from the genomic sequence requiring little experimental profiling of proteolytic cleavage sites. The availability of biochemical and X-ray structural data also suggests that it is possible to produce and engineer a particular protease through the recombinant route in *E. coli*.

---

**Fig. 1** Schematic summary of different types of synthetic protease sensors and amplification circuits based on modularly organized autoinhibited protease modules. (a) An elementary autoinhibited protease transducer P1 can be engineered by connecting a competitive autoinhibition domain that can be specifically cleaved off by an activating protease P2. (b) An allosterically regulated protease receptor P1 can be engineered by replacing the cleavage site for an activating protease with an allosteric binding receptor R1-R2 that undergoes a large conformational change upon binding its cognate ligand L. (c) To accelerate the response time and improve sensitivity, the signal generated by the primary allosterically regulated protease sensor P1 can be transduced to cleave and activate a secondary protease amplifier P2. (d) Alternatively, activation of the secondary amplifier can be based on the ligand L induced colocalization of a primary transducer P1 with a secondary amplifier P2 to engineer a proximity-dependent protein-protein interaction sensor.
This greatly facilitates their subsequent purification, in vitro characterization, and engineering by means of rational design or high-throughput screening approaches.

The second consideration concerns structural features of the protease: Ideally, the N- and C-termini of the protease are located in close proximity of the active site. In this way, an AI-domain is displayed at a high effective concentration such that it can effectively compete with the substrate and thus shut down the activity of the protease transducer in the basal state. Conversely, its activity can be strongly induced after the AI-domain is irreversibly cleaved off by an activating protease or reversibly dislodged from the active site through a binding event. The exact mechanism of activation ultimately depends on the molecular recognition elements in the linker connecting the AI-domain with the protease transducer.

1.3 Choice of Binding Receptors as Molecular Recognition Elements

Similar considerations concern the choice of ligand-binding receptors that serve as molecular recognition elements to detect the target analyte. Binding receptors can either recognize two distinct epitopes with separate binding sites on the same target analyte or feature an allosteric receptor that undergoes a significant conformational change upon binding the target analyte. In case of the latter, binding receptors that undergo cooperative binding interactions upon associating with their target analyte are preferred as this overcomes potential problems with multivalent interactions and the subsequent formation of oligomeric complexes. For instance, we exploited a binding receptor based on artificially engineered affinity clamps consisting of a circularly permutated PDZ domain attached to an enhancer domain that recognizes the PDZ domain exclusively in its ligand bound form and thus forms a sandwich complex upon binding its cognate ligand [6, 7].

From a protein engineer’s perspective, it is highly desirable to recombine autoinhibited protease modules with distinct families of binding receptors that display similar structural and biophysical characteristics, yet diverse ligand specificities. In this way, the protein engineering process is rendered as generally applicable as possible. In particular, the orientation of the N- and C-termini is critical if a distinct class of binding receptors is to be repeatedly recombined with an autoinhibited protease module with little empirical optimization to yield synthetic protease sensors with diverse ligand specificities. Ideally, a distinct class of binding receptors can be sourced from naturally occurring receptor families. Alternatively, binding receptors can be engineered using standardized workflows as for example recombinant antibody-like binders that can be derived using a variety of display technologies.

1.4 General Considerations on the Construction of AI-Domains

AI-domains are engineered using a combination of rational, structure-guided protein engineering in combination with high-throughput screening. Notably, for protease-based signal transducers, the proteolytic cleavage product provides a good molecular lead to
engineer genetically encoded, active site-directed competitive inhibitors that can be directly fused to the protease transducer. Depending on their size, individual cleavage products can either be appended to the N- or C-terminus. Considering the N-terminal cleavage product of TVMV comprises five of six amino acids of the core substrate motif, this means the product-based competitive inhibitor needs to be appended to the C-terminus. In many cases, appending a product-based inhibitor is however insufficient to ensure a high enough level of competitive autoinhibition. As a result, its binding affinity needs to be improved using a combination of structure-guided protein engineering and high-throughput combinatorial screening, for instance, by introducing additional charge and shape complementarity between the AI-domain and the protease transducer or by restricting the orientation of the connecting linker.

In addition, a key concern is that recombinant protein expression leads to premature termination of translation and therefore truncated protein products with no AI-domain. This may be further aggravated by limited nonspecific proteolysis of exposed and/or flexible linkers during protein expression and/or subsequent cell lysis. This is particularly problematic for modularly organized protease switches that rely on C-terminal AI-domains as this limits the extent to which a synthetic protein switch is autoinhibited in the basal state. For instance, if 10% of a purified protein is truncated with no AI-domain, its maximum activation is limited to tenfold. For C-terminally positioned AI-domains, it is therefore paramount to purify full-length protein through an affinity tag that is located C-terminal of the AI-domain. This also requires bridging the P1-P1’ junction through a dipeptide motif that binds, but at the same time cannot be cleaved. This is however not trivial for proteases. Bridging the P1-P1’ junction through an uncleavable dipeptide motif also allows to place the AI-domain in the N-terminal position and engineer additional affinity features beyond the core cleavage product to enhance binding. An N-terminal AI-domain also facilitates the detection of weakly enhancing mutations, especially if their contribution toward affinity is incremental and thus difficult to resolve if the induction is limited by prematurely truncated protein expression products in cell lysates. Consequently, it is recommended to first engineer dipeptide motifs that can bridge the P1-P1’ junction before improving the affinity of AI-domain for its cognate protease receptor.

2 Materials

2.1 General Reagents

1. 100 mg/mL carbenicillin (1000× stock).
2. 50 mg/mL kanamycin (1000× stock).
3. 34 mg/mL chloramphenicol (1000× stock).
4. LB medium: 10 g peptone, 5 g yeast extract, and 10 g NaCl in 1 L dH2O.
5. LB agar plates: 10 g peptone, 5 g yeast extract, 10 g NaCl, and 15 g agar in 1 L dH2O.

2.2 Construction of DNA Libraries with USER Enzyme

1. Chemically competent Top10 E. coli cells (or equivalent) for DNA cloning.
2. Pfu Cx DNA polymerase (Agilent).
3. 10× Pfu Cx reaction buffer.
4. 10 mM dNTPs each.
5. Primer-For: 5'-GCTGAAGTCTTACGAGGAAGAGTTGGC-3' (T_M = 70.5 °C).
6. Primer-C-Ter: 5'-ACGGTTUTCGGACCTACACCG-3' (T_M = 72.1 °C).
7. Primer-Library: 5'-AACCGUGCGCTTTNNNNNNNGGAACCCACCACCATCAT-3' (T_M = 70.3 °C).
8. Primer-Rev.: 5'-CGTTGTAAACGACGGCCAGTG-3' (T_M = 70.3 °C).
10. 10× T4 DNA ligase buffer (NEB).
11. 400 U/μL T4 DNA ligase (NEB).
12. 1 U/μL USER Enzyme Mix (NEB).
13. Restriction enzymes NcoI and BamHI to digest the pRK793 vector backbone.
14. Gibson Assembly Kit.
15. Wizard SV Gel and PCR Clean-Up System.
16. Miniprep plasmid DNA purification system.

2.3 High-Throughput Expression of Synthetic Protease Receptors

1. Chemically competent BL21(DE3)-RIL E. coli cells (or equivalent) for protein expression.
2. Plasmid 05665 featuring the bacteriophage λ SRRz autolysis cassette [8, 9] under the control of the tetracycline inducible promoter.
3. 96-deep-well-plates to grow E. coli in high throughput.
4. 50× 5052 medium: 25% glycerol (w/v), 2.5% glucose (w/v) and 10% α-lactose (w/v) [10].
5. 20× NPS: 0.5 M (NH4)_2SO4, 1 M KH2PO4 and 1 M Na2HPO4 pH 6.75.
6. 1 M MgSO4.
2.4 Expression and Purification of TVMV and HCV Protease Switches

1. Terrific broth (TB): 1.2% (w/v) tryptone, 2.4% (wt/vol) yeast extract, 0.04% glycerol, 0.17 M KH₂PO₄, and 0.72 M K₂HPO₄.
2. TB-based autoinduction medium: TB supplemented with 0.2% lactose, 0.05% glucose, and 2 mM MgCl₂.
3. Phosphate buffered saline (PBS): 10 mM Na₂HPO₄, 2 mM KH₂PO₄, 137 mM NaCl, and 2.7 mM KCl, pH 7.4.
4. Washing and binding buffer: 20 mM Na₂HPO₄ and 20 mM imidazole pH 8.0 supplemented with 300 mM NaCl for TVMVThr-AI protease transducers, 500 mM NaCl for HCVTVMV-AI protease transducers, and 1 M NaCl for TVMV-FN3-PDZ-AI allosteric receptors.
5. Elution buffer: 20 mM Na₂HPO₄ and 500 mM imidazole pH 8.0 supplemented with 300 mM NaCl for TVMVThr-AI protease transducers, 500 mM NaCl for HCVTVMV-AI, and 1 M for TVMV-FN3-PDZ-AI allosteric receptors.
6. Protein storage buffer: 50 mM Tris–HCl and 10% (v/v) glycerol, pH 8.0 supplemented with 1 mM EDTA and 2 mM DTT for TVMVThr-AI protease transducers, 1 M NaCl, 1 mM EDTA, and 2 mM DTT for TVMV-FN3-PDZ-AI allosteric receptors, and 500 mM NaCl and 2 mM β-mercaptoethanol for HCVTVMV-AI protease transducers.
7. 0.25-μm nitrocellulose filters.
8. Ni-NTA columns (e.g., 5 mL HisTrap FF Crude from GE Healthcare).
9. Centrifugal filters with 10-kDa cutoff.
10. PD-10 desalting columns.

2.5 Assay Synthetic Protease Receptors

1. Protease Assay Buffer: 50 mM Tris–HCl, 100 mM NaCl, 50 μg/mL BSA and 2 mM DTT, pH 8.0 (see Note 1).
2. Summary of protease reagents and their working concentrations is given in Table 1.
3. Summary of protease peptide substrate, ligand peptides, and AI domains used to characterize individual protease receptors is provided in Table 2.
4. 8 mM TVMV peptide substrate dissolved in DMSO and stored at −80 °C.

5. 20 μM TVMV peptide substrate diluted in protease assay buffer and stored on ice.
6. 8 mM HCV peptide substrate dissolved in DMSO and stored at −80 °C.
7. 20 μM HCV peptide substrate diluted in protease assay buffer and stored on ice.
8. 8 mM TVMV-DD peptide substrate dissolved in DMSO and stored at −80 °C.
9. 800 μM TVMV-DD peptide substrate diluted in protease assay buffer and stored on ice (see Note 2).
10. 1 U/μL thrombin dissolved in 50 mM Tris–HCl pH 8.0 and stored at −80 °C.
11. 1 mM Ligand B1 dissolved in 50 mM Tris–HCl pH 8.0 and stored at −80 °C.
12. 1 mM Ligand B2 dissolved in 50 mM Tris–HCl pH 8.0 and stored at −80 °C.
13. 100 mM AI domain NH₂-EYVRFAPGST-COOH as a synthetic peptide dissolved in DMSO and stored at −80 °C.
14. 200 μg/mL rapamycin stock solution dissolved in ethanol and stored at −80 °C.
15. Black 96-well plates for fluorescence measurements.
16. Optical adhesive film (to cover black 96-well plates for fluorescence spectroscopy).

### 2.6 Equipment

1. PCR thermocycler.
2. Agarose gel electrophoresis equipment.
3. SDS-PAGE equipment.
4. Fluorescence multiwell plate reader (e.g., Biotek Synergy 4).
5. Shaking incubator to grow E. coli in liquid medium.
6. Static incubator to grow E. coli on agar plates.
7. One shot cell disruptor for E. coli cell lysis.

### 3 Method

#### 3.1 Cloning

**Autoinhibited Proteases Modules**

Viral protease receptors are expressed as fusion proteins with maltose-binding protein (MBP) based on pRK793 [11]. This construct has previously been developed to facilitate the expression of the NIa protease from Tobacco Etch Virus (TEV) where MBP acts as a molecular chaperone to enable more efficient recombinant expression of NIa proteases in E. coli [12, 13]. To construct focused libraries, a variety of DNA library generation and DNA assembly techniques are available. For instance, combinatorial DNA libraries with degenerate codons coding for different amino acids are suitable to bridge the P1-P1′ junction or improve the affinity between the
AI-domain for its cognate protease transducer. In contrast, truncated DNA libraries are typically applied to optimize the length and structure of the linkers connecting functional domains. An exemplary protocol to engineer dipeptide motifs that can bind across the P1-P1’ junction without being cleaved is exemplified based on USER Enzyme-dependent DNA assembly [14].

1. Devise a cloning strategy for a focused DNA library to engineer a competitively autoinhibited protease module taking into account structural and biochemical data from the existing literature (Fig. 2). For instance, in case of TVMV, the design of uncleavable P1-P1’ junctions for the AI-domain is based on substrate mapping data for the related TEV protease which suggests that proline in the P1’ position prevents cleavage [15]. It is however unclear to what extent sequences with a proline in the P1’ position cannot bind or can bind, but cannot be cleaved by TEV protease. Similarly, it is not clear to what extent this holds true for other members of the NIa potyvirus protease family including the NIa protease of TVMV [2], but this needs to be screened and tested experimentally.

2. Commercially synthesize the gene coding for the autoinhibited protease module of interest (Fig. 2). For TVMV, include a cleavage site for autoproteolytic processing followed by TVMV protease, a linker including a thrombin cleavage site, its cognate substrate sequence (which serves as a lead structure for engineering competitive AI-domains), and an affinity purification tag and insert into pRK793 via NcoI and BamHI.

3. Generate a DNA fragment coding for the N-terminal portion by means of PCR using Pfu Cx DNA polymerase, 1× Pfu Cx reaction buffer, 10 mM dNTPs each, 1 ng template DNA coding for the elementary protease module (step 2), and primers Primer-For and Primer-C-Ter.

4. Generate a DNA fragment coding for the C-terminal portion by means of PCR using Pfu Cx DNA polymerase, 1× Pfu Cx reaction buffer, 10 mM dNTPs each, 1 ng template DNA coding...
(a) Synthetic DNA Fragment Coding for Elementary Autoinhibited Protease Module

gctgaagtctttgcggaaagatccagccaggctcatgaatgccggtgatcaccggtgacggctccggtgactgacttgttttgcggatcgcgggtctgaatccgtctctgttcccagcgacctctttgctgaagggcgtgcggcggatttttaatccgatctctgcttgccgtatgcctgactagctgctttctggtatgccgtgcgtactgcggtgatcaacggccgcagcggtcgtcagactgctgatgaaagcccctgaaagacgcgcagactaattcgatcacaagtttgtaaaagccaggctcg

5'-GCTGAAGTCTTACGAGGAAGAGTTGGCG-3'  Tm: 70.5 °C

5'-ACGGTUTCGCGACCTACACCG-3'  Tm: 72.1 °C

Sequence Annotation
- Restriction sites NcoI and BamHI, and USER Enzyme recombination sites are highlighted in black
- 50 bp homologous recombination sites with vector backbone are underlined
- TVMV protease is dotted-underlined
- TVMV cleavage sites are double-underlined
- Thrombin cleavage site is wave-underlined
- His affinity purification tag is thick-underlined
- P1-P1' junction is highlighted in grey

(b) Primer Pair N-terminal TVMV Fragment

5'-GCTGAAGTCTTACGAGGAAGAGTTGGCG-3'

5'-ACGGTUTCGCGACCTACACCG-3'

(c) Primer Pair C-terminal TVMV Fragment with Randomized P1-P1' Junction

5'-AACCGUGCGCTTTNNNNNNGGAAGCACCCACCACCACCATCAT-3'

5'-CGTTGTAAAACGACGGCCAGTG-3'

Tm: 70.3 °C

(d) USER Enzyme Dependent Recombination Site with Randomized P1-P1' Junction

GVGRERENTEVTTRFX_XGSTHHHHHH

5'-AACCGUGCGCTTTNNNNNNGGAAGCACCCACCACCACCATCAT-3'

3'-GCCACATCCAGCCGCTTGCGG-5'

Tm: 70.3 °C
for the elementary protease module (step 2), and primers Primer-Library and Primer-Rev.

5. Mix equimolar amounts of the N- and C-terminal coding DNA fragments generated in steps 3 and 4 together with 1 U USER Enzyme per μg of total DNA and 200 U T4 DNA ligase per μg DNA in 1× T4 DNA ligase buffer. Incubate for 60 min at 37 °C.

6. Purify the assembly reaction using a suitable PCR purification kit.

7. Check the quality of USER-mediated enzyme DNA assembly reaction by means of agarose gel electrophoresis. Typically, one major band corresponding to the desired DNA assembly product along with the two minor bands of the DNA assembly substrates is visible.

8. Optional: Purify the desired DNA fragment by means of agarose gel electrophoresis (see Note 3).

9. Perform a Gibson Assembly reaction to recombine the DNA insert library into a suitable expression vector as per manufacturer’s instructions. We typically use the vector pRK793 backbone based on the pMAL series cut with NcoI and BamHI (see Note 4).

10. Transform 5 μL of the Gibson Assembly mix into 50 μL chemically competent E. coli cells and plate on LB agar plates supplemented with 100 μg/mL carbenicillin.

11. To check the quality of the library, inoculate single colonies in 5 mL LB medium supplemented with 100 μg/mL carbenicillin, purify plasmid DNA and sequence.

3.2 Engineering Autoinhibited Protease Modules: High-Throughput Screening

Once a DNA library has been generated, it needs to be screened experimentally to identify mutants with the desired properties. The following generic protocol can be used to screen the function of autoinhibited protease modules, sensors, and switches aiming to optimize the binding strength of the AI-domain, or the length and structure of the connecting linkers.

1. Transform the library of autoinhibited TVMV mutants into chemically competent BL21(DE3)-RIL cells hosting the autolysis plasmid 05665 (see Note 5).

2. Plate transformed cells on LB agar plates supplemented with 100 μg/mL carbenicillin, 50 μg/mL kanamycin, and 34 μg/mL chloramphenicol.

3. Following overnight incubation at 37 °C, inoculate single colonies into 96-deep-well plates filled with 1 mL minimal PA-5052 autoinduction medium [10] supplemented with 100 μg/mL carbenicillin, 50 μg/mL kanamycin, and 34 μg/mL chloramphenicol.
4. Grow cells for 72 h at 30 °C at 320 rpm until an OD₆₀₀ of ~10 is typically reached. Store 96-deep-well plates at 4 °C for several days to regrow plasmids in step 11.

5. **Optional**: Spin down 200 μL cell suspension and resuspend cells in fresh PA-5052 autoinduction medium, but with no trace metals (see Note 6).

6. Transfer 100 μL aliquots of *E. coli* cell cultures from half a 96-well plate (48 samples) and dilute twofold in 100 μL autolysis medium and incubate for 90 min at 30 °C with agitation at 200 rpm.

7. Transfer 10 μL aliquots of cell lysates and duplicate into separate halves of a black 96-well plate filled with 140 μL protease assay buffer.

8. Supplement all of the duplicates with 1 μL of 1 U/μL thrombin to cleave the linker connecting TVMV to its AI-domain and incubate for 30 min.

9. Initiate the reaction by adding 50 μL of 20 μM TVMV peptide substrate solution giving rise to 5 μM TVMV peptide substrate in the final reaction.

10. Monitor the reaction progress in a 96-well-plate fluorescence reader by measuring the release of the quenched fluorophore 7-methoxycoumarinyl-4-acetyl at 405 nm following excitation at 330 nm.

11. Evaluate the induction of TVMV protease activities by comparing the time-course of the reaction in the presence and absence of thrombin.

12. Regrow *E. coli* from step 3 that code for protease transducer mutants that display a high induction ratio in step 9 in 5 mL LB medium supplemented with 100 μg/mL carbenicillin.

13. Purify and sequence plasmid DNA that codes for the protease transducer mutants with the desired properties.

### 3.3 Purifying Synthetic Protease Switches

To characterize engineered protease switches biochemically under defined reaction conditions, it is recommended to purify individual protease switches to homogeneity by means of affinity chromatography.

1. Transform plasmids coding for the desired synthetic protease switch into chemically competent BL21(DE3)-RIL cells, plate onto LB agar plates supplemented with 100 μg/mL carbenicillin, and 34 μg/mL chloramphenicol and incubate overnight at 37 °C.

2. Inoculate 250 mL TB-based autoinduction medium supplemented with 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol with a single colony.
3. Grow cells over the course of 48 h at 30 °C at 200 rpm expressing the protein by means of autoinduction (see Note 7). Cells usually reach OD_{600} values of ∼14.

4. Harvest cells by centrifugation at 4500 × g and wash once with 500 mL PBS. At this stage, cells can either be stored at −20 °C or directly processed.

5. Resuspend cells in washing and binding buffer and lyse using a one-shot cell disruptor.

6. Centrifuge samples at 25,000 × g and pass the supernatant through a 0.25-µm nitrocellulose filter to remove any large-scale debris and genomic DNA that could clog up the affinity purification column in the subsequent step.

7. Load the lysate onto 5 mL HisTrap FF Crude Ni-NTA columns equilibrated with washing and binding buffer.

8. After loading the cell lysate allow the His_{60}-tagged protease switch to bind to the column and then wash the column with 40 column volumes of washing and binding buffer.

9. Elute the His_{60}-tagged protease switch with an imidazole gradient from 20 to 500 mM over 40 column volumes. Throughout the procedure maintain a flow rate of 5 mL/min. The His_{60}-tagged protease switch typically elutes around 100 mM imidazole.

10. Pool fractions that contain the His_{60}-tagged protease switch, concentrate the pooled fractions using centrifugal filters with a 10 kDa cutoff and transfer into storage buffer with a suitable concentration of NaCl by means of gel filtration using disposable PD-10 desalting columns (see Note 8).

11. Synthetic protease switches can either be assayed directly or stored in aliquots at −80 °C. Aliquotting samples and flash-freezing in liquid nitrogen is highly recommended to prevent deterioration of the protein sample over successive freeze-thaw cycles.

### 3.4 Quantification of Synthetic Protease Switches

Once purified, individual protease receptors can be assayed and characterized under defined reaction conditions: e.g., to assess their maximum induction ratios, determine the dissociation constants of allostERIC protease receptors for their cognate ligands, measure the inhibition constants of AI-domain peptides, or assess the purity of protease transducers by means of kinetic analysis. The preferred assay reaction volume is 200 µL while individual assay components are preferably added in increments of 50 µL based on 4× working stock solutions of individual reagents.

#### 3.4.1 Determine the Induction of Autoinhibited Protease Transducers and Receptors

1. To assess the maximum induction ratio of elementary autoinhibited protease transducers, preincubate 10 µM of TVMV-Thr-AI or 10 µM of HCV_{TVMV}-AI for 10 min in 100 µL protease assay buffer in the presence and absence of an activating protease
target (i.e., for TVMVThr-AI include 0.005 U/μL thrombin and for HCVTVMV-AI include 500 nM TVMV). Typical reaction volumes are 50–100 μL supplemented with 1–2 μL of 1 U/mL thrombin.

2. Mix 100 μL protease assay buffer with 50 μL of 2 μM of TVMVThr-AI or 2 μM of HCVTVMV-AI as pretreated ± thrombin or TVMV from step 1. Generally perform dilutions in protease assay buffer.

3. Alternatively, to assess the maximum induction ratio of allosteric protease receptors upon binding their cognate ligands, preincubate 50 μL of 1 μM TVMV-FN3-PDZ-AI together with 50 μL of protease assay buffer and 50 μL of 4 μM ligand B1 or 40 μM ligand B2 (see Note 9).

4. Alternatively, to assess the $K_D$ value of allosteric protease receptors for their cognate ligands, preincubate 50 μL of 40 nM of TVMV-FN3-PDZ-AI with 50 μL protease assay buffer and 50 μL of varying concentrations of either ligand B1 or ligand B2. To reliably determine the $K_D$, the highest concentration of ligand B1 or ligand B2 in the final reaction should be at least 100 times the anticipated $K_D$ (see Note 9).

5. Depending on the type of protease switch, initiate the reaction by adding 50 μL of either TVMV or HCV protease peptide substrate giving rise to a final concentration of 5 μM peptide substrate in a final reaction volume of 200 μL. Generally include a no enzyme control in 200 μL protease assay buffer to account for excitation-dependent bleaching of the fluorophore in the protease peptide substrate.

6. Monitor the reaction progress using a fluorescence multiwell plate reader by monitoring the release of 7-methoxycoumarinyl-4-acetyl from the quenched protease peptide substrate with $\lambda_{ex/em}$ of 330 and 405 nm. Synthetic protease switches based on TVMV or HCV at a concentration of 100 nM or more generally yield a good fluorescence signal over the course of 90 min. In contrast, synthetic protease switches assayed at a concentration of 10 nM or less are preferably resolved over the course of 900 min. To prevent excessive bleaching of the fluorophore, no more than 45 time points should be measured under each condition. For prolonged periods of time, it is also recommended to cover the 96-well plate with optical adhesive film to prevent evaporation of the reaction assay.

7. When assessing maximum induction ratios, extract initial rates from the change in fluorescence (after subtracting the background fluorescence with no enzyme). Then calculate the induction ratio of individual TVMV- and HCV-based protease switches in the inactive and active states by dividing the initial rate in the presence and absence of the target analyte.
8. When performing ligand titrations, extract initial rates from the change in fluorescence (after subtracting the background fluorescence with no enzyme) and plot against the ligand concentration. Then determine the apparent $K_{D}$s of allosteric protease receptors by a nonlinear regression fit of the curve to Eq. 1.

$$
Y = V_0 + \left( V_{MAX} - V_0 \right) \times \frac{\sqrt{([\text{Sensor}] + [\text{Ligand}] + K_D)^2 - 4 \times [\text{Sensor}] \times [\text{Ligand}]}}{2 \times [\text{Sensor}]}
$$

1. Analysis by SDS-PAGE frequently cannot resolve prematurely truncated or partially cleaved synthetic protease transducers considering the small size of 10–15 amino acids that the AI-domain and the short His$_6$ affinity purification tag comprise at the C-terminus of both the TVMV- and HCV-based transducers. In addition, any cleavage and truncation products will in the majority of cases be heterogeneous and not concentrate in a single band, but spread out, rendering their detection even more difficult. Instead, the quality of a purified protein prep can be assessed indirectly by determining the kinetic parameters of purified protease switches in their autoinhibited and activated states. Notably, any residual activity associated with partially autoinhibited protease switches should display elevated $K_M$ compared to its inactive state.

2. Preincubate 10 μM of TVMV$^\text{Thr}$-AI in the presence and absence of 0.05 U/μL thrombin for 1 h in protease assay buffer. Typical reaction volumes are 50 to 100 μL supplemented with 1–2 μL of 1 U/μL thrombin.

3. Set up a dilution range of varying concentrations of the protease substrate TVMV-DD in 100 μL protease assay buffer in black 96-well plates. For the reliable, quantitative determination of $K_M$ values, set up duplicates for every substrate concentration. The highest substrate concentration in the final assay should be at least three times the anticipated $K_M$ for the protease substrate. For instance, in case of fully active TVMV with an anticipated $K_M$ of 80 μM, set up substrate concentrations from 5 to 400 μM based on 100 μL of a 2× TVMV-DD working stock solution.

4. In addition, prepare an equivalent dilution range of the protease substrate TVMV-DD in 200 μL protease assay for a non-enzyme control to account for excitation-dependent bleaching of the fluorophore.

5. Initiate the reaction by adding a suitable amount of TVMV-based transducer in a total volume of 100 μL. The optimal amount of TVMV will depend on the activity state of individ-
ual TVMV switches, as assessed in Subheading 3.4.1. For instance, for fully active TVMV switches, 100 nM is suitable while for largely predominantly autoinhibited TVMV switches, 10 μM is required to reliably measure the initial rate.

6. Monitor the reaction progress using a fluorescence 96-well-plate reader by monitoring the release of 7-methoxycoumarinyl-4-acetyl from the quenched TVMV peptide substrate with λex/em of 330 and 405 nm. Monitoring the reaction progress over the course of 90 min is usually sufficient to reliably extract data for quantitative analysis.

7. Extract initial rates from the change in fluorescence (after subtracting the background fluorescence in the no enzyme control at each substrate concentration) and plot against the substrate concentration.

8. Calculate the kinetic parameters by nonlinear regression fit to Michaelis Menten kinetics, as depicted in Eq. 2.

\[
Y = V_{\text{Max}} \times \frac{[\text{Substrate}]}{[\text{Substrate}] + K_M}
\]  

9. For autoinhibited TVMV protease transducers, the K_M should increase relative to the activated transducer. If not, this means the majority of residual activity associated with a particular protein prep is due to prematurely truncated or cleaved protein.

1. To estimate the strength of binding interaction of an Al-domain to its cognate protease transducer, preincubate 50 μL of 2 μM uninhibited TVMV^{T214I, V216W}-transducer mutant with 50 μL protease assay buffer and 50 μL varying concentrations of the Al-domain peptide NH_{2}-EYVRFAPGST-COOH. For the reliable, quantitative determination of K_i values, set up duplicates for every Al-domain peptide concentration. The optimal concentration range needs to be determined empirically with the highest concentration of the Al-domain peptide at least three to four times higher than the anticipated K_i.

2. Initiate the reaction by adding 50 μL of 20 μM TVMV-DD substrate solution (to a final assay concentration of 5 μM) in the presence of varying concentrations of the Al-domain peptide NH_{2}-EYVRFAPGST-COOH. Monitoring the reaction progress over the course of 90 min is usually sufficient to reliably extract data for quantitative analysis.

3. Monitor the reaction progress using a fluorescence 96-well-plate reader by monitoring the release of 7-methoxycoumarinyl-4-acetyl from the quenched TVMV-DD peptide substrate with λex/em of 330 and 405 nm.
4. Extract initial rates from the change in fluorescence (after subtracting the background fluorescence in the no enzyme control at each substrate concentration) and plot against the concentration of the AI-domain peptide.

5. Determine the $K_i$ by a nonlinear regression fit of the curve to Eq. 3 with the $K_M$ and the TVMV-DD substrate concentration set to 65 μM and 5 μM for TVMV, respectively (see Note 10).

$$
Y = V_{\text{Max}} \times \frac{[\text{Substrate}]}{[\text{Substrate}] + K_M \times \left(1 + \frac{[\text{Inhibitor}]}{K_i}\right)}
$$

3.5 Assembling Integrated Signal Sensing and Amplification Circuits

Signaling switches based on autoinhibited protease modules carry the advantage that they can be readily assembled into synthetic protease circuits where a primary sensor can be connected to cleave and activate a secondary amplifier based on an alternative autoinhibited protease module. This opens the possibility to assemble individual protease sensors and transducers into integrated signal sensing and amplification circuits with accelerated response times and improved sensitivity for their target analyte.

3.5.1 Integrated Signal Sensing and Amplification Circuits

1. Incubate 50 μL of the 40 nM SH3$^{\text{Dom}.}$-TVMV-FN3-PDZ-AI allosteric receptor with 50 μL of 20 μM HCV peptide substrate (final assay concentration of 5 μM) and 50 μL of varying concentrations of either ligand B1 or ligand B2. The optimal concentration of the SH3$^{\text{Dom}.}$-TVMV-FN3-PDZ-AI allosteric receptor needs to be determined empirically and is preferably chosen in the low nM range to limit excessive, nonspecific activation of the secondary amplifier HCV$^{\text{TVMV}.}$-AI-SH3$^{\text{Pep}}$. Similarly, the strength of the SH3-dependent binding interaction can be optimized by employing SH3-binding peptides of different interaction strength [16]. The concentrations of ligand B1 and ligand B2 can be chosen as in Subheading 3.4.1.

2. Initiate the reaction by adding 50 μL of 800 nM HCV$^{\text{TVMV}.}$-AI-SH3$^{\text{Pep}}$ giving rise to a final concentration of 200 nM HCV$^{\text{TVMV}.}$-AI-SH3$^{\text{Pep}}$.

3. Monitor the progress of the reaction using a fluorescence 96-well-plate reader by monitoring the release of 7-methoxycoumarinyl-4-acetyl from the quenched HCV peptide substrate with $\lambda_{\text{ex/em}}$ of 330 and 405 nm.

4. For the quantitative analysis of $K_D$ values, extract initial rates from the change in fluorescence (after subtracting the background fluorescence in the no enzyme control at each substrate concentration) and plot against the concentration of ligand B1 or ligand B2. Crucially, for cascaded systems, the initial rates need to be extracted from the exponential phase before all of
the secondary amplifier HCV\textsuperscript{TVMV-AI-SH3\textsuperscript{ Pep}} has been cleaved and activated to accurately determine the strength of binding interactions.

1. **Optional**: Preincubate 50 μL of 10 μM FKBP12-\textit{TVMV\textsuperscript{Thr-AI}} with 1 μL of 1 U/μL thrombin to cleave off the AI domain. For reduced signal-to-noise ratios, the AI domain can remain attached (see **Note 11**).

2. Incubate 50 μL of 400 pM FKBP12-\textit{TVMV\textsuperscript{Thr-AI}} together with 50 μL of 20 μM HCV substrate (giving rise to a final concentration of the HCV substrate 5 μM) and 50 μL of varying concentrations of rapamycin. With an anticipated $K_D$ in the sub-nM range for the FKBP12-rapamycin interaction, the preferred concentration of the primary FKBP12-\textit{TVMV\textsuperscript{Thr-AI}} sensor in the final assay is 100 pM to prevent titrating the amount of FKBP12-\textit{TVMV\textsuperscript{Thr-AI}} and not the strength of the FKBP12-rapamycin binding interaction. Similarly, to reliably determine its $K_D$ for FKBP12, rapamycin is applied between 0.01 and 10 nM.

3. Initiate the reaction with 50 μL of 800 nM HCV\textit{TVMV-AI-FRB}. The preferred concentration of the secondary amplifier HCV\textit{TVMV-AI-FRB} in the final assay is 200 nM, but can be optimized empirically depending on the signal-to-noise ratio and the desired response rate.

4. Monitor the reaction progress using a fluorescence 96-well-plate reader by detecting the release of 7-methoxycoumarinyl-4-acetyl from the quenched HCV peptide substrate with $\lambda_{ex/em}$ of 330 and 405 nm.

5. For the quantitative analysis, extract initial rates from the change in fluorescence (after subtracting the background fluorescence in the no enzyme control at each substrate concentration) and plot against the substrate concentration. Crucially, for cascaded systems, the initial rates need to be extracted from the exponential phase before all of the secondary amplifier HCV\textit{TVMV-AI-FRB} has been cleaved and activated to accurately determine the strength of FKBP12-rapamycin-binding interaction.

### Notes

1. Depending on specific protease switch being assayed, the NaCl concentration in the protease assay buffer needs to be optimized and tailored toward each application. For instance, SH\textsubscript{3\textsuperscript{ Dom}-\textit{TVMV-FN3-PDZ-AI}} allosteric receptors display higher induction ratios at 1 M NaCl compared to 100 mM NaCl. In contrast, the activity of HCV-based switches is
inhibited at higher NaCl concentrations and should be kept at 100 mM NaCl or less.

2. For the reliable determination of enzyme kinetics, the protease substrate needs to be fully dissolved at high substrate concentrations. If a protease substrate is not fully soluble at the desired substrate concentration and the substrate solution appears cloudy, its solubility can be improved by appending negatively charged residues.

3. In the construction of DNA libraries, it is recommended to purify all DNA fragments for the DNA assembly reaction by means of agarose gel electrophoresis. In our experience, the Wizard SV Gel and PCR Clean-Up System kit provides the highest quality of DNA fragments free of impurities that can potentially interfere with subsequent enzymatic or transformation steps. If necessary, individual DNA fragments can be additionally cleaned by means of ethanol precipitation.

4. Alternatively, suitable vector backbones can be prepared by means of PCR independent of any restriction site. Due to the large size of the pRK793 backbone and the difficulties associated with PCR amplification and susceptibility to point mutations, preparing the vector backbone by restriction enzyme digest is preferred in this case.

5. Plasmid 05665 features a bacteriophage λ derived SRRz autolysis cassette [9, 17] under the control of a tetracycline inducible promoter. This allows releasing the recombinantly expressed protease switches into the lysate under relatively mild conditions. The use of detergent-based cell lysis procedures is not recommended as it interferes with the protease assay.

6. The induction of autolysis by means of the SRRz autolysis cassette with tetracycline is most efficient in fresh PA-5052 autoinduction medium. In addition, the presence of glucose, which inhibits the expression of genes under the control of the LacO promoter, in fresh PA-5052 autoinduction medium is irrelevant at this stage as the protease transducer is already expressed. Depending on the type of protease transducer, including trace metals may interfere with protease function and should therefore be omitted.

7. Alternatively, proteins can also be expressed by means of Isopropyl β-D-1-thiogalactopyranoside (IPTG). In this case, cells are grown to an OD$_{600}$ of 0.6–0.8 before being induced with 0.1–1 mM IPTG and expressed for 3–12 h at 30 °C. Optimal induction conditions need to be determined empirically for each type of protease sensor.

8. Depending on the synthetic protease sensor, it is critical to include the correct amount of NaCl in the storage buffer. Notably, the allosteric protease receptors are not stable in the absence of NaCl, but form a precipitate.
9. To measure the maximum induction ratio, the binding interaction needs to be saturated. Given the AI-domain creates a decoy interaction, the affinity of an allosteric receptor for its cognate protease will be reduced.

10. For alternative protease sensors, the $K_m$ needs to be determined empirically.

11. Depending on whether the TVMV-based protease transducer is partially autoinhibited, the signal-to-noise ratios and the response time of the proximity-dependent protease switches can be optimized. For instance, at high sensitivity and low concentrations of FKBP12-TVMV-AI in the sub-nM range, non-specific cleavage and activation of the secondary amplifier is negligible and it is preferable to employ uninhibited TVMV. In contrast, at high concentrations of FKBP12-TVMV-AI, non-specific activation is elevated and it is therefore recommended to employ partially autoinhibited TVMV modules. Employing partially autoinhibited TVMV modules generally does not interfere, to a significant extent, with cleavage considering the TVMV cleavage site in the secondary amplifier is presented at a high local concentration in the ternary complex.

Acknowledgments

This work was funded by the Australian Research Council Discovery Project Grant DP1094080 to KA and in part by National Breast Cancer Foundation Innovator Grant. This research was also supported by Movember through Australia’s Prostate Cancer Foundation Research Program to KA and VS.

References


Chapter 14

Characterizing Dynamic Protein–Protein Interactions Using the Genetically Encoded Split Biosensor Assay Technique Split TEV

Jan P. Wintgens, Moritz J. Rossner, and Michael C. Wehr

Abstract

Dynamic protein–protein interactions (PPIs) are fundamental building blocks of cellular signaling and monitoring their regulation promotes the understanding of signaling in health and disease. Genetically encoded split protein biosensor assays, such as the split TEV method, have proved to be highly valuable when studying regulated PPIs in living cells. Split TEV is based on the functional complementation of two previously inactive TEV protease fragments fused to interacting proteins and provides a robust, sensitive and flexible readout to monitor PPIs both at the membrane and in the cytosol. Thus, split TEV can be used to analyze interactomes of receptors, membrane-associated proteins, and cytosolic proteins. In particular, split TEV is useful to assay activities of relevant drug targets, such as receptor tyrosine kinases and G protein-coupled receptors, in compound screens. As split TEV uses genetically encoded readouts, including standard reporters based on fluorescence and luminescence, the technique can also be combined with scalable molecular barcode reporter systems, allowing the integration into multiplexed high-throughput assay approaches. Split TEV can be used in standard heterologous cell lines and primary cell types, including neurons, either in a transient or stably integrated format. When using cell lines, the basic protocol takes 30–96 h to complete, depending on the complexity of the experimental question addressed.

Key words Protein–protein interaction, Split TEV, Biosensor, Split biosensor assay, RTK, GPCR, Phosphorylation-dependent interactions, Dose-response assay, Compound profiling

1 Introduction

1.1 Genetically Encoded Biosensor Assays for Studying Protein–Protein Interactions

Studying protein–protein interactions (PPIs) is paramount to understand cellular signaling and responses thereof. The majority of PPIs are precisely regulated both in time and space, and mediate distinct cellular activities, including differentiation, proliferation, apoptosis, and inflammation. In terms of signaling, cell surface receptors and highly interlinked intracellular proteins, or hubs, are of special interest, as these proteins integrate key signaling activities to regulate global cellular responses. For cytosolic proteins, pivotal associations with a strong implication on downstream signaling are
frequently found at sub-membranous localizations. Here, signal transduction activities are commonly initiated through the dynamic formation of protein complexes that are composed of multiple PPIs [1]. For cell surface receptors, receptor tyrosine kinases (RTKs) and G protein-coupled receptors (GPCRs) are intensively studied protein classes, as deregulated signaling caused by these receptors is implicated in various human diseases, such as cancer and neurodevelopmental diseases [2, 3]. RTKs and GPCRs represent major drug targets, supporting their significant role in clinical research and drug discovery [4]. Both the associations between two proteins as well as the activities of RTKs and GPCRs can be reliably studied in split biosensor assays (SBA, also called protein complementation assays [5]). SBA are particularly suitable to monitor interactions of cell surface receptors and cytosolic proteins with a sub-membranous localization as SBA facilitate an analysis in the natural habitat of these proteins [6].

RTKs are single transmembrane receptors, with a total number of 58 encoded in the human genome [3]. RTKs respond, with few exceptions, to extracellular cues, and have an extra-cellular ligand-binding domain, a cytosolic tyrosine kinase domain, and a cytosolic tail, which contains interaction motifs that establish, once phosphorylated through ligand-driven activation, docking sites for cytosolic adapter proteins. Upon ligand binding, RTKs can homodimerize or heterodimerize, leading to a conformational change that transmits the signal into cytosol, where the kinase domains phosphorylate target tyrosine residues in trans [7]. In turn, Src homology 2 (SH2) domain- and phosphotyrosine binding (PTB) domain-containing adapters are recruited to phosphorylated docking sites, a process that can be robustly and precisely monitored by SBA [8–10]. An established interaction between an RTK and an adapter protein initiates downstream signaling activities, as these adaptors link RTK activation to downstream signal transduction pathways, such as the MAP kinase or PI3K/AKT signaling cascades.

GPCRs are the largest class of receptors encoded in the human genome, comprising more than 800 receptors in total [11]. Like RTKs, deregulated GPCR signaling is strongly implicated in various human diseases. Strikingly, 30–40% of all marketed drugs target GPCRs, making them the largest class of druggable receptors supporting their prominent role in drug discovery [4, 12]. GPCRs are seven-transmembrane receptors with three intracellular and three extracellular loops of varying length, an extracellular N-terminus, and a cytoplasmic C-terminal tail, which can associate with effector proteins in an activity-dependent manner [12]. Heterotrimeric G proteins are the major binding partners for GPCRs and trigger defined pathway responses, such as cAMP-mediated or Calcium-dependent pathway activities. Prolonged activation of GPCR activation results in its desensitization, which
is induced by the regulated recruitment of G protein-coupled receptor kinases (GRKs) that phosphorylate residues on the C terminus of the GPCR and thus provide a platform for β-Arrestin binding [13]. The binding of β-arrestin sterically obstructs G protein coupling and triggers the internalization of the GPCR. The regulated association between a given GPCR and β-arrestin can be exploited to reliably monitor GPCR activities and is regularly used by SBA and related proximity assays [14–17].

To assess dynamic interactions among proteins, various approaches for genetically encoded SBA are available, including, but not limited to, GFP and variants thereof [18–20], luciferases (firefly [21, 22], Renilla [23, 24], Gaussia [25], click beetle luciferases [17, 26, 27]), β-lactamase [28], β-galactosidase [29, 30], ubiquitin [8, 31, 32], and the tobacco etch virus (TEV) protease [33]. In addition to SBA, proximity assays were developed, such as bioluminescence resonance energy transfer (BRET) assays comprising firefly luciferase and GFP [34, 35] and the Tango method that uses a full-length TEV protease fused to an interacting partner like β-Arrestin to monitor GPCR activities [16].

Here, we present a detailed protocol for the split TEV technique, a genetically encoded SBA that can be flexibly applied to sensitively monitor various types of regulated PPIs in living cells [33] (and reviewed in [6]). The technique is based on the interaction-induced fragment complementation of the TEV protease. Split TEV can be used to assess dynamic interactions at the membrane and in the cytosol, thus allowing various combinations among potentially interacting candidate proteins (Fig. 1). Notably, interactions in the nucleus cannot be monitored using the presented technique, as the method was specifically designed to analyze PPIs at the membrane and in the cytosol. The technique integrates various reporter systems, including fluorescent and luminescent reporters, making it implementable to many research laboratories. As split TEV flexibly allows using readouts of choice, scalable transcriptional barcode reporters that are amenable to multiplexed high-throughput formats and next-generation sequencing may also be used. Combining these technologies will enable assessing drug target activities and cellular response profiles in parallel, thereby opening up new avenues in drug discovery. Here, we centre on luciferase reporters, as these are widely used both in single interaction assays and in high-throughput applications.

For split TEV assays, an optimized form of the TEV protease is dissected into an N-terminal (NTEV, amino acids 1–118) and a C-terminal fragment (CTEV, amino acids 119–221, with a Ser→Pro substitution at residue 219 for enhanced assay stability). The S219P mutation renders the TEV protease refractory to autocatalysis [36]. In addition, the optimized form is truncated after amino acid 221 to remove the inhibitory C-terminal tail, which
blocks the active site of the TEV protease [37]. NTEV and CTEV fragments are fused to candidate proteins, preferably to the C-terminal end of the protein candidates. For NTEV, N-terminal fusions are also functional. Protein Interaction-induced reassembly of the NTEV and CTEV moieties leads to the reconstitution of TEV proteolytic activity, which activates TEV-specific reporters. These can be either of proteolytic or of transcriptional type, and they maybe fluorescent or luminescent-based. Notably, the luciferase-based transcriptional reporters proved to be most sensitive and robust as the readout is functionally uncoupled from the interaction event itself and background readings were commonly lower compared to other options. Transcriptional reporters based on luciferase exhibit another feature as they comprise three levels of signal amplification: In addition to the proteolytic cleavage and a transcriptional amplification step, enzyme-based (i.e., luciferase) reporters allow a third step of amplifying the initial signal resulting in a robust and sensitive readout.

1.2 Split TEV Assays for Membrane and Membrane-Associated Proteins

Fig. 1 Types of protein–protein interactions that can be monitored by split TEV. The candidates in a split TEV assay can be membrane proteins, membrane-associated proteins, and soluble proteins in the cytosol. Proteins localized to these subcellular areas can be tested for regulated interactions in all combinations, which are (a) membrane | membrane, (b) membrane | membrane-associated, (c) membrane | cytosolic, (d) membrane-associated | membrane-associated, (e) membrane-associated | cytosolic, and (f) cytosolic | cytosolic

For monitoring dynamic PPIs at the membrane, a membrane protein or membrane-associated protein candidate is fused to the NTEV fragment along with a TEV protease cleavage site (tevS, encoded by the amino acids ENLYFQ’G; the TEV protease cleaves between Q and G) and the artificial transcriptional co-activator...
GAL4-VP16 (GV), resulting in a NTEV-tevS-GV tag (Fig. 2a). As this hybrid tag contains a transcriptional co-activator, it is critical that the candidate protein is a membrane or membrane-associated protein, as a nuclear localization will readily lead to high background readings due to increased GV activity. The second candidate protein is fused to the CTEV fragment. This candidate may be a membrane, membrane-associated, a cytosolic protein, or...
even a protein that shuttles between the cytosol and the nucleus. PPI-induced reconstituted TEV protease activity cleaves off GV, which translocates into the nucleus and activates a firefly luciferase reporter gene, which is placed under the control of upstream activating sequences that contain DNA sequences recognized by GV.

Split TEV assays can be used to monitor ligand-dependent interactions of cell surface receptors, such as RTKs and GPCRs, thus allowing to assess their activity. For instance, the Neuregulin1-induced association between ERBB4, an RTK of the ERBB family, and the regulatory subunit alpha of the PI3K, PIK3R1, was determined in a dose-response assay (Fig. 3a), as well as its inhibition by lapatinib (Fig. 3b). Sample constitutive interactions assayed using split TEV are shown for the membrane-associated protein KIBRA (Fig. 3c). For assessing GPCR activities, we have recently reported a detailed protocol on the split TEV-based GPCR/β-Arrestin2 recruitment assay, which uses a truncated version of β-Arrestin2 and a modified NTEV-tevS-GV tag for improved sensitivity [38]. We therefore refer the reader to this specified GPCR split TEV protocol. Notably, the split TEV method proved to be more sensitive than the full-TEV Tango approach when assaying selected GPCR activities [15].

1.3 Split TEV Assays for Soluble Proteins

Interactions between candidate proteins that stay strictly cytosolic and do not shuttle into the nucleus may be monitored using the hybrid NTEV-tevS-GV tag (see Note 1). For candidate proteins that show a certain degree of nuclear localization, candidates are fused to the NTEV fragment only (Fig. 2b). The other candidate protein is fused to CTEV. When pursuing the sole NTEV/CTEV tagging strategy, PPI-induced TEV protease activity leads to the activation of a cytosolically localized transcriptional TEV reporter (denoted GV-2ER), which contains a central GV unit that is trapped by two ERT2 domains, each fused to GV via a tevS. ERT2 domains are modified oestrogen receptor domains, that do not respond to endogenous oestrogen, but to 4-hydroxytamoxifen. Liberated GV translocates into the nucleus and activates the firefly luciferase reporter. MST1 dimer formation based on the C-terminal SARAH domain is shown as an example for soluble protein interactions (Fig. 3d). Dynamic interactions like AKT-induced and phosphorylation-regulated BAD/14–3-3 and rapamycin-regulated FKBP/FRB have also been described using split TEV [9, 33].

1.4 Assay Controls

When studying PPIs using split TEV, tagged proteins of interest are commonly introduced into cells at rather high expression levels, either transiently by transfection or by stable integration. Therefore, assay components should be expressed at rather low levels, possibly close to those of endogenous counterparts. Further, each assay should be individually validated by specific
Characterising Dynamic PPIs Using Split TEV

**Fig. 3** Examples of constitutive and regulated split TEV assays. (a, b) Dose response split TEV assays for the ERBB4 receptor using (a) the agonist EGF-like domain (EGFid) and (b) the antagonist lapatinib. ERBB4–NTEV-tevS-GV and PIK3R1–CTEV fusions were transiently transfected into PC12 cells and treated for 20 h using the indicated compound concentrations. (a) ERBB4/PIK3R1-transfected assay cells were stimulated with increasing concentrations of EGFid. The EC₅₀ value was calculated at 3.06 ng/ml EGFid. (b) ERBB4/PIK3R1-transfected assay cells were stimulated with increasing concentrations of lapatinib, followed 1 h later by the addition of a constant stimulus of 10 ng/ml EGFid. The IC₅₀ value was calculated at 0.42 μM lapatinib (b). Six replicates per condition, error bars represent SEM. (c) Constitutive split TEV assays for the kinase MST1. MST1 dimerizes through its C-terminally located SARAH domain (lane 2). Note that MST1-ΔC, which contains residues 1–432 only and lacks the SARAH domain, does not interact with MST1 full-length, and thus only background signals are produced. Likewise, the missense mutation L444P disrupts the MST1 dimer formation [39]. (d) MST1 is not suitable for the hybrid tag NTEV-tevS-GV. MST1–NTEV-tevS-GV is cleaved and produces high background readings (lane 4). By contrast, MST1 interacting proteins MST2, SAV1 and RASSF1A (R1A) yield low background readings. Readings are compared to a strong PPI (KIBRA::KIBRA) (lane 1) and a background control (KIBRA::RASSF6 (R6)) (lane 2). RLU relative luciferase units; six replicates per condition, error bars represent SD

Interaction controls. Therefore, it is strongly advisable to include positive controls (i.e., proteins that do interact with the candidate of interest) and negative controls (i.e., proteins that do not bind to the candidate of interest) for each assay designed.
The split TEV technique can be applied to study constitutive and regulated interactions that are modulated by ligands, agonists, antagonists, or a combination thereof. Experimental setups for constitutive split TEV assays can be comfortably completed within 30 h (Fig. 4a). Dynamic interaction split TEV assays using agonists or agonist/antagonist combinations typically require up to 4 days, depending on the individual experimental design (cell seeding parameters, starving conditions, stimulation paradigms, etc.) (Fig. 4b, c). The effects of compound actions, either stimulatory or inhibitory, are optimally assessed in dose response assays, which allow the calculation of EC$_{50}$ and IC$_{50}$ values (concentrations of agonist/antagonist compounds at half-maximal stimulatory/inhibitory response).

In summary, the split TEV technique is a powerful and sensitive tool providing a robust readout to monitor dynamic PPIs within their natural context. The split TEV method is characterized by various features as it enables (1) the detection of dynamic

---

**Fig. 4** Suggested timelines for split TEV assays. Cells are plated on the day before (or alternatively in the morning) and transfected with assay plasmids. (a) Constitutive assay. After transfection, cells are incubated for 20 h before lysis. (b) Agonist assay: After transfection, plasmids are allowed to express for 20 h, followed by a medium change to starve the cells in low-serum media. The following day, an agonist is added for 6–20 h, depending on the setup of the assay. (c) Antagonist assay: Transfection and medium change are performed as in (b). Before an agonist is added, cells are treated with an antagonist for 1 h. The stimulation time is dependent on the assay setup.
interactions between full-length candidate proteins in living cells or in lysates thereof, (2) the detection of protein interactions localized at the membrane, sub-membrane compartment, and the cytosol, (3) the detection of protein interactions in heterologous mammalian cells, primary cell types including primary neurons and astrocytes, (4) the detection of regulated RTK and GPCR activities induced by extracellular stimuli, (5) the analysis of regulated, ligand or compound concentration-dependent interactions in dose-response assays assessing actions of agonists and antagonists, and (6) the setup as protein interaction readouts for high-throughput screening (HTS) purposes [9, 15, 33, 40–42]. Therefore, split TEV may also be adapted to industrial HTS setups to screen for RTK and GPCR activities, for example in drug discovery programs.

Below, we describe experimental setups for split TEV assays for membrane, membrane-associate and cytosolic proteins using firefly luciferase as a final reporter gene as this readout is widely used across laboratories, and allows adaptation to HTS approaches.

## 2 Materials

### 2.1 Plasmids

1. For membrane, membrane-associated and strictly cytosolically localized proteins:

   ORF(X)-NTEV-tevS-GV plasmid. The ORF of candidate X is fused to NTEV (amino acids 1–118), tevS (ENLYFQG), and GV (Fig. 2a). Use molecular cloning to generate candidate X fusion vectors (see Note 2). For a detailed description on GPCRs, see the protocol by Wehr et al. [40].

   For cytosolic protein candidates that shuffle between cytosol and nucleus:

   ORF(X)-NTEV plasmid. The ORF of candidate X is fused to NTEV (amino acids 1–118) only. In addition, the pGV-2ER plasmid is required as cytosolic TEV reporter (Fig. 2b).

2. ORF(Y)-CTEV. The ORF of candidate Y is fused to CTEV (amino acids 119–221, including a S219P mutation). Use molecular cloning to generate the fusion vector (see Note 2).

3. Firefly luciferase reporter plasmid. Use an upstream activating enhancer sequences (UAS)-driven firefly luciferase (e.g., pGL4.31[luc2P/GAL4UAS/Hygro], Promega, catalog no. C9351) (see Note 3).

4. Renilla luciferase control plasmid. Use a thymidine kinase (TK)-driven constitutive Renilla luciferase as internal control (pRL-TK, Promega, catalog no. E2241).

5. A GFP or GFP-derivative expressing plasmid for optical transfection control (e.g., mVenus-N1, Addgene, Plasmid #54640).
2.2 **Cells**
HEK293 (ATCC), PC12 cells (PC12 Tet-Off, Clontech), any other heterologous cell line, or primary cells that are amenable to efficient transfection and allow studying protein–protein interactions (see Note 4).

2.3 **Reagents**
1. Medium, depending on the cell line chosen (e.g., for HEK293 cells: DMEM 4.5 g glucose/l without L-glutamine, each with and without phenol red). See complete media formulations for HEK293 and PC12 cells below.
2. Fetal bovine serum (FBS), heat-inactivated, 500 ml (Life Technologies)
3. Horse serum (HS), heat-inactivated, 500 ml (Life Technologies) (only needed for PC12 cells).
4. GlutaMAX, 100×, 100 ml (Life Technologies).
5. Penicillin and Streptomycin, 100×, 100 ml (Life Technologies).
6. Opti-MEM (Life Technologies).
7. Poly-L-lysine, mol wt 70,000–150,000, diluted in a final concentration of 0.02 mg/ml in H2O (Sigma).
8. Lipofectamine 2000 transfection reagent (Life Technologies).

2.4 **Equipment**
1. 96-well plates: white with flat bottom for luciferase-based assays (Falcon).
2. 96-well plates: clear with flat bottom for optical transfection control and fluorescence-based assays (Falcon).
3. 75 cm² flasks (Falcon) or 15 cm dish (Falcon) (for maintenance of cell lines).
4. Cell culture facility including flow hood and incubator set to 5% CO2 and 37 °C.
5. Luciferase reader (e.g., Mithras, Berthold Technologies or Envision, PerkinElmer).
6. Fluorescence microscope (to check for transfection efficiency).

2.5 **Media Formulation**
Prepare for each cell line used the appropriate media for maintenance and assay conditions.
1. HEK293 cells, assay medium: DMEM (4.5 g glucose/l, without L-glutamine and without phenol red), 500 ml, supplemented with 0.5% FBS and 2 mM GlutaMAX.
2. HEK293 cells, medium for maintenance: DMEM (4.5 g glucose/l, without L-glutamine, with phenol red), 500 ml,
supplemented with 10% FBS, 2 mM GlutaMAX, and 100 U/ml of each penicillin and streptomycin.

3. PC12 Tet-Off cells, assay medium: DMEM (1 g glucose/l, without L-glutamine and without phenol red), 500 ml, supplemented with 1% FBS, and 2 mM GlutaMAX.

4. PC12 Tet-Off cells, medium for maintenance: DMEM (1 g glucose/l, without L-glutamine, with phenol red), 500 ml, supplemented with 5% FBS and 10% HS, 2 mM GlutaMAX, and 100 U/ml of each penicillin and streptomycin.

3 Methods

1. Design appropriate ORF(X)-NTEV-tevS-GV and ORF(Y)-CTEV fusion expression plasmids (see guidelines in Subheading 2.1, items 1 and 2, and Note 2). When testing interactions between candidates X and Y, it is strongly advisable to design appropriate interaction controls (see Note 5).

2. Clone and prepare sufficient amount of DNA for the assays, set the concentration for each plasmid to 100 ng/μl (see Note 6).

3. Prepare 96-well plates for split TEV luciferase assays and fluorescence controls.
   (a) Coat a white/clear 96-well plate(s) with a 0.02 mg/ml poly-L lysine (PLL) solution for 30 min at room temperature.
   (b) Wash twice with double deionized water, then let plates air-dry. PLL-coated plates can be stored at 4 °C for up to 6 weeks without losing coating efficiency. This step is optional for HEK293 cells.

4. Seed 20,000 HEK cells resuspended in maintenance medium without antibiotics into each well of the coated 96-well plate. Place the 96-well plate(s) into the incubator. For PC12 Tet-Off cells, seed 40,000 cells per well (see Note 7).

5. Seed cells onto a clear 96-well plate for an optical control of transfection efficiency.

6. On the next day, prepare the plasmid mixes for each condition to be tested. Run 6 replicates per condition to obtain sufficient readings for statistical analyses. For each 96-well, use 15 ng per single DNA. Make a master mix for 6 + 1 replicates, which contains an extra amount needed to allow for pipetting errors (for example, when using 15 ng DNA/well, use 15 ng x 7 = 105 ng in total). For the Renilla luciferase plasmid, use the same DNA amount chosen. Prepare a constitutively expressing GFP plasmid (or variant thereof) (2 ng per well) for control
positions on the clear plate to guarantee an optical control of transfection efficiency (see Note 8).

7. Prepare the transfection mix. Dilute the Lipofectamine 2000 150-fold in Opti-MEM supplemented with 1% GlutaMAX, vortex for 1–2 s, and incubate for 5 min. For each 96-well, we use 0.2 µl Lipofectamine 2000 that is diluted in 30 µl Opti-MEM. We suggest preparing a 7× master mix for all conditions to be tested.

8. Pipette the Lipofectamine/Opti-MEM mix onto the DNA, mix (vortex), and incubate for 20 min at room temperature.

9. Remove medium from each well entirely (be careful not to remove the cells) and transfer the DNA/transfection mix (30 µl per well) onto the cells.

10. After 1 h, add 60 µl medium for maintenance without antibiotics.

11. Depending on whether a constitutive or stimulus-dependent protein–protein interaction assay is run,
    - proceed to step 12 for constitutive, or
    - proceed to step 13 for agonist and agonist/antagonist assays.
    - Common experimental paradigms are schematically displayed in Fig. 4.

12. *Constitutive assay:* Incubate for 12–24 h. Typically, we incubate constitutive assays for 20 h. Before lysing the cells, verify the transfection efficiency using the clear optical control plate. By this time, the transfected plasmids should express considerably, and transfection efficiency, as determined by GFP signals, should be at least 30–50% to obtain stable assay conditions, otherwise start optimizing transfection parameters (see Note 9).
    - Proceed to step 17.

13. *Agonist and agonist/antagonist assays:* Incubate overnight.

14. On the next day, verify transfection efficiency (c.f. step 12, see Note 9) and replace the medium with 80 µl assay medium to induce the starvation of cells.

15. After 16–24 h, add appropriate stimuli in 80 µl assay medium, with the agonists in double concentration. If studying agonist/antagonist effects, add the antagonists 1 h before the agonists (see Note 10).

16. Incubate for 6–24 h, depending on the assay setup. This step strongly depends on the individual interaction measured, the cell type used, or biological question addressed. For a first approach, we incubate for 18–20 h (see Note 11).

17. Aspirate the medium and add 30 µl of 1x Passive Lysis Buffer to each well. Incubate the 96-well plate on a horizontal rocking platform (100–150 rpm) for 10 min at room temperature.
Lysed cells can be stored at $-20 \, ^\circ\text{C}$ for up to 4 weeks without losing significant levels of signal intensity (see Note 12).

18. Perform the Dual luciferase readout. Add 30 μl of Firefly substrate into each well (from commercial Dual Luciferase Kit prepared as indicated by the manufacturer), read the plate using a luciferase reader. Add 30 μl of Renilla substrate into each well, read the plate (see Note 13). Alternatively, self-made substrates may be used (see Note 14).

19. Calculate mean, standard deviations and standard errors of the mean using an Excel spreadsheet. Generate a bar graph to visualize the data. For dose-response analyses, we recommend using the dose response curve package ‘drc’ in R (see Note 15). Alternatively, the GraphPad Prism program can be used, which allows standard visualizations of dose response curves.

20. Advice for troubleshooting addressing critical steps can be found in Table 1.

**Table 1**

<table>
<thead>
<tr>
<th>Step</th>
<th>Problem</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>4, 10</td>
<td>Low signals in general (both firefly and Renilla)</td>
<td>Do not use antibiotics as they might interfere with the transfection reagent</td>
</tr>
</tbody>
</table>
| 12, 16 | Good transfection efficiency, good Renilla signals, but low firefly signals:  
- ORF-NTEV  
- ORF-NTEV-tevS-GV  
- ORF-CTEV fusion protein may not be expressed properly | Although the optical transfection control is fine as indicated by for example EGFP expression, the candidate fusion in question may be poorly expressed  
- Verify protein expression using Western blotting  
- Try to swap candidate proteins for NTEV and CTEV fragments. If using the cytoplasmic assay, you may also place the NTEV moiety to either end. Protein expression, however, also depends on the target protein chosen |
| 16 | Low firefly signals for inducible PPIs | Stimulation time may vary for a given receptor or any other inducible PPI. Therefore, we suggest to run an online luciferase split TEV assay to obtain the best activation window. If no appropriate reader is available, monitor assay activity from 6 to 20 h in an hourly or two-hourly mode after the agonist has been added |
| 18 | Low luminescence signals, in particular for firefly luciferase | Double-check the functionality of the luciferase buffers. This is of particular importance for the firefly luciferase buffer containing ATP, which can decay quickly. To do so, we prepare test lysates from cells that constitutively express both firefly and Renilla luciferases. These lysates can easily be stored at $-80 \, ^\circ\text{C}$ for several months |
4 Notes

1. When opting for the hybrid NTEV-tevS-GV tag, we recommend to first perform control assays that only contain the candidate-NTEV-tevS-GV to exclude high background readings. For example, activated Sterile20-like kinase MST1 (STK4) is cleaved [43] and produces high background readings, thus precluding the use of the hybrid tag. In a next step, the candidate may be tested with a protein that is known not to bind to the candidate to validate assay robustness, and to establish experimental parameters for a reference negative control.

2. Cloning of candidate split TEV fusions. Clone the candidate ORFs including all split TEV assay elements into a regular expression vector, such as pcDNA3.1. The fusions are constructed to contain a 10 amino acid flexible linker (GGGSGGGGS) between the candidate protein and NTEV or CTEV. As expression plasmids, we regularly use pcDNA3-derived (Invitrogen) and pTag-derived (Stratagene) vectors. If required, N- and CTEV fusions can be stably integrated into the cells using the mammalian selection markers present on these plasmids (i.e., neomycin or zeocin resistance). We regularly use Gateway recombination cloning (Life Technologies) to quickly obtain candidate fusions. Gateway-compatible ORFs for more than 16,000 ORFs are available at Plasmid ID (http://plasmid.med.harvard.edu/PLASMID/Home.jsp). The TEV protease mutant S219P can be obtained from Addgene (plasmid no. 8830, pRK792).

3. When opting for fluorescence-based assays, use a UAS-EGFP plasmid (i.e., pJFRC7-20XUAS-IVS-mCD8::GFP, Addgene, plasmid no. 26220). In this case, normalization can be done using nuclear staining or by co-transfecting a constitutively expressing mCherry plasmid. In addition, a fluorescence plate reader is required for quantification.

4. For investigating membrane protein interactions, we start testing assays in PC12 Tet-Off cells (Clontech) and then proceed to HEK293 cells (ATCC). For membrane-associated and soluble proteins, we commonly start with HEK293 cells. Notably, U2OS cells proved to a valuable cell line to monitor GPCR activities. As we observed substantial differences in the activation ratios for some GPCRS, like AVPR1 and DRD2, and RTKs, like ERBB family receptors, we suggest to assess the optimal cell line for each individual assay.

5. Positive (known interacting protein) and negative (known non-interacting protein) controls should always be included in protein–protein interaction assays based on protein fragment complementation [33]. As a starting guideline, we suggest...
using FKBP1A-NTEV or FKBP1A-CTEV as negative controls as FKBP1A represents a small protein with a defined number of interactions described [44]. FKBP1A was used in split TEV assays [40], and also by other split protein approaches. Further, we have successfully used the GCN4-coiled coil domain (GCN4cc) as interaction control [9, 41]. Another strategy may be applying a truncated form of an interacting partner that lacks the defined interaction domain(s). However, it is crucial to validate its correct use first.

6. We suggest to store DNA plasmid aliquots at −20 °C. Avoid unnecessary freeze-thaw cycles as this decreases DNA concentration and quality.

7. Seeding the cells can be done on the previous day or in the morning of the day the experiment is planned. When seeding on the same day, increase the amount of cells by roughly 50% and wait at least 3–4 h to let the cells attach to the surface of the plate.

8. It is also possible to prepare the DNA mix on the previous day. If done so, freeze the DNA as small volumes may evaporate at 4 °C.

9. To optimize transfection conditions, start with adapting plasmid amounts, i.e., increase/decrease DNA amounts, and/or change ratios of the candidate X vs. candidate Y vs. reporter plasmids. Further, we find changing cell densities and/or Lipofectamine 2000 concentrations helpful.

10. Agonists can be added in a 10× or 20× concentration diluted in assay medium. Note that both agonists and antagonists are frequently diluted in DMSO, which must not exceed 1% of the total volume.

11. For RTKs, we regularly incubate for 20 h to obtain stable inducible readings. For GPCRs, we noted, however, that shorter incubation periods of 6–8 h yield better readings. To determine the optimal time window for a given inducible PPI, we also suggest to run an online luciferase split TEV assay before investing into a dose response assay.

12. When opting for a fluorescent reporter, such as UAS_EGFP, do not lyse the cells, and proceed with acquiring the fluorescent signal using a fluorescence reader or an automated microscope.

13. When acquiring luciferase signal intensities using the Mithras device from Berthold Technologies, we apply a time integration protocol, which measures each well for 2 s. Substrates are added by automated injection across the entire plate, followed by an orbital shake for 10 s and an incubation time of 10 min, followed by reading the signals. This process is performed first.
for the firefly luciferase substrates, and then for the *Renilla* luciferase substrates.

14. If using the self-made substrates for the Mithras device from Berthold Technologies, a different protocol for measuring the *Renilla* luciferase signals will be applied, as this substrate has a decreased stability. The substrate is injected, followed by an orbital shake of 2 s, and then immediately measured for 2 s. This Process is repeated well by well. The detailed protocol for making both firefly (a) and *Renilla* (b) substrates is listed below. We regularly order special reagents from p.j.k. GmbH, Kleinblittersdorf, Germany; these include coelenterazine, D-luciferin, co-enzyme A, ATP, DTT (1,4 dithiothreitol). c(0.001, 0.01, 0.1, 1, 10, 100),

(a) Substrate for firefly luciferase

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tricine</td>
<td>20 mM</td>
</tr>
<tr>
<td>(MgCO₃)₄<em>Mg(OH)₂</em>5H₂O (magnesium carbonate-hydroxide-pentahydrate)</td>
<td>1.07 mM</td>
</tr>
<tr>
<td>MgSO₄ * 7 H₂O</td>
<td>2.67 mM</td>
</tr>
<tr>
<td>EDTA</td>
<td>0.1 mM</td>
</tr>
<tr>
<td>DTT</td>
<td>33.3 mM</td>
</tr>
<tr>
<td>Coenzym A</td>
<td>270 µM</td>
</tr>
<tr>
<td>d-Luciferin, free acid</td>
<td>470 µM</td>
</tr>
<tr>
<td>ATP</td>
<td>530 µM</td>
</tr>
</tbody>
</table>

To solve magnesium carbonate, adjust the pH value using HCl (37%, use approximately 1 ml for 1 l) until solution becomes clear. Then, adjust the pH value to 7.8 using 5 M NaOH (use approximately 7.5 ml for 1 l).

(b) Substrate for *Renilla* luciferase

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>1.1 M</td>
</tr>
<tr>
<td>Na₂-EDTA</td>
<td>2.2 mM</td>
</tr>
<tr>
<td>K₂PO₄ (pH 5,1)</td>
<td>0.22 M</td>
</tr>
<tr>
<td>BSA</td>
<td>0.44 mg/ml</td>
</tr>
<tr>
<td>NaN₃</td>
<td>1.3 mM</td>
</tr>
<tr>
<td>Coelenterazine (solved in EtOH)</td>
<td>1.43 mM</td>
</tr>
</tbody>
</table>
$K_2PO_4$ (pH 5.1): 1 M $KH_2PO_4$, adjust pH value to 5.1 using 2 M KOH (use approximately 15 ml for 1 l). Adjust pH value to 5.0. Add coelenterazin afterwards.

For storage, make 50 ml aliquots and store at $-20^\circ C$ in the dark. Aliquots can be stored up to 6 months.

15. The package ‘drc’ is used in R to calculate EC50 and IC50 values and to generate fitted dose response curves. The script uses the drm function and is amended to include error bars based on the standard error of the mean (sem). The assay data is saved as txt file using a tab delimited format. Luciferase activity (activity) is calculated in %, concentration (conc) in $\mu$M, and sem in %.

A sample assay data set is shown below:

<table>
<thead>
<tr>
<th>activity</th>
<th>conc</th>
<th>sem</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.024253731</td>
<td>0.00003</td>
<td>0.969091058</td>
</tr>
<tr>
<td>8.789256841</td>
<td>0.0001</td>
<td>1.875653638</td>
</tr>
<tr>
<td>10.69690609</td>
<td>0.0003</td>
<td>1.271332669</td>
</tr>
<tr>
<td>7.650808458</td>
<td>0.001</td>
<td>0.991970081</td>
</tr>
<tr>
<td>12.68326337</td>
<td>0.003</td>
<td>1.117476168</td>
</tr>
<tr>
<td>9.120219216</td>
<td>0.01</td>
<td>1.537416736</td>
</tr>
<tr>
<td>9.865127488</td>
<td>0.03</td>
<td>1.367412567</td>
</tr>
<tr>
<td>13.02510883</td>
<td>0.1</td>
<td>2.670494192</td>
</tr>
<tr>
<td>26.97566853</td>
<td>0.3</td>
<td>3.160663768</td>
</tr>
<tr>
<td>31.31743626</td>
<td>1</td>
<td>2.324396584</td>
</tr>
<tr>
<td>61.8740283</td>
<td>3</td>
<td>3.084176905</td>
</tr>
<tr>
<td>70.77386505</td>
<td>10</td>
<td>9.107073314</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>5.429347986</td>
</tr>
<tr>
<td>95.9036847</td>
<td>100</td>
<td>13.04036002</td>
</tr>
</tbody>
</table>

The drc script including the addition of error bars reads:

```r
# Loading package
library("drc")

# Loading data
drc <- read.delim("D:/.../DRC_1.txt", header=T, dec=".")

# Returning data
drc

# Fitting the model
DR.1 <- drm(activity~conc, data = drc, fct = LL.4())

summary(DR.1)
```
# Calculating effective dose (ED) values
# first column: the estimates ED50, ED80, and ED90
# second column: the estimated standard errors
ED(DR.1, c(50, 80, 90))
EDout = ED(DR.1, c(50))
# Labelling of x axis, compound name & unit
DRxlab = "EGFId [ng/ml]" #or "Lapatinib [μM] & 10 ng/ml EGFId"
# Plotting graph with y axis from 0 to 100% activity
plot(DR.1, broken = FALSE, type = c("all"), ylab = "activity[%]", xlab = DRxlab, log = "x", xt = c(0.001, 0.01, 0.1, 1, 10, 100), ylim = c(0, 115), legend = FALSE, bty = "l", sub = paste("EC50: ",round(EDout[1,1],4),"μM",sep=" ")) #adjust EC/IC50 and concentration
# Adding error bars as sem
for (i in 1:dim(drc)[1]){
  segments(drc[i,2],drc[i,1]-drc[i,3],drc[i,2],
  drc[i,1]+drc[i,3],col="black")
}

Acknowledgment

This work was supported by the Ludwig Maximilian University of Munich and the DFG grant WE 5683/1-1 to M.C.W.

References

Characterising Dynamic PPIs Using Split TEV


Part VI

Optogenetic Switches
Chapter 15

Development of a Synthetic Switch to Control Protein Stability in Eukaryotic Cells with Light

Christof Taxis

Abstract

In eukaryotic cells, virtually all regulatory processes are influenced by proteolysis. Thus, synthetic control of protein stability is a powerful approach to influence cellular behavior. To achieve this, selected target proteins are modified with a conditional degradation sequence (degron) that responds to a distinct signal. For development of a synthetic degron, an appropriate sensor domain is fused with a degron such that activity of the degron is under control of the sensor. This chapter describes the development of a light-activated, synthetic degron in the model organism *Saccharomyces cerevisiae*. This photosensitive degron module is composed of the light–oxygen–voltage (LOV) 2 photoreceptor domain of *Arabidopsis thaliana* phototropin 1 and a degron derived from murine ornithine decarboxylase (ODC). Excitation of the photoreceptor with blue light induces a conformational change that leads to exposure and activation of the degron. Subsequently, the protein is targeted for degradation by the proteasome. Here, the strategy for degron module development and optimization is described in detail together with experimental aspects, which were pivotal for successful implementation of light-controlled proteolysis. The engineering of the photosensitive degron (psd) module may well serve as a blueprint for future development of sophisticated synthetic switches.

Key words Optogenetics, Protein degradation, Proteasome, Ubiquitin-independent degradation, Protein stability, Synthetic biology, LOV2 domain, Blue light, Degron

1 Introduction

Regulated proteolysis is involved in virtually each and every regulatory process in eukaryotic cells. Continuously, protein levels are adjusted during passage through the cell cycle or as reaction on changing environmental conditions, which results in the degradation of selected proteins by the ubiquitin-proteasome system. Ubiquitin is a small protein that serves among other things as a marker for proteolysis by the proteasome. In general, an enzymatic cascade of ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2) and ubiquitin-protein ligase (E3) modifies a protein substrate with several moieties of ubiquitin in the form of a linear chain, which targets the protein for destruction
by the proteasome [1, 2]. The substrates contain linear motifs that are recognized by a specialized E3, which ensures selectivity of the process. These motifs are called degradation-inducing sequence or shortly degron, if they are necessary and sufficient for recognition by the degradation machinery. Thus, modification of a stable protein by a degron will lead to its destabilization. In rare cases, proteins are also directly recognized by the proteasome and degraded in a ubiquitin-independent way [3].

Due to the importance of proteolysis in eukaryotic cells, synthetic control of protein stability is a powerful tool to interfere with regulatory mechanisms and to influence cellular activities. To implement synthetic regulation, the activity of a degron is switched from an inactive to an active state by a distinct signal. Until now, signals like temperature, small molecules, nutrients, cell cycle stage or light have been used to regulate the stability of conditional degrons in eukaryotic cells [4–9]. To confer synthetic regulation on a selected target protein, one of the degrons is fused to the target gene to create a cell line in which target protein abundance and activity is controlled synthetically.

Development of such a degradation tool requires two protein domains or sequences: a sensor domain that converts the signal into an output, e.g., a conformational change and a degron that is controlled by this switch. For generation of a light-controlled degron, the light–oxygen–voltage (LOV) 2 photoreceptor domain of *Arabidopsis thaliana* phototropin 1 was selected as sensor due to precise knowledge about the molecular changes in the LOV2 domain after light-exposure. The LOV2 domain binds the cofactor flavin-mononucleotide (FMN), which is excited by blue light. Subsequently, the side-chain of a cysteine residue of LOV2 forms a covalent bond with the C4a atom of FMN. This induces structural changes in the LOV2 core resulting in unfolding of the so-called Jα-helix at the carboxy terminus of the photoreceptor [10, 11].

The degradation sequence selected for the conditional degron was derived from the well-studied murine ornithine decarboxylase (ODC). The ODC degron sequences are located at the very carboxy-terminus of the enzyme. It is somewhat unusual, as ubiquitin is not necessary for proteasomal degradation of ODC, whereas the vast majority of proteasome substrates are degraded in an ubiquitin-dependent way [12]. The active ODC degron has two requirements: a cysteine-alanine motif located 19 amino acids upstream of the C-terminus flanked by sequences without secondary structure comprising a length of 37 amino acids in total [13, 14].

To generate the light-activated degron, the *A. thaliana* LOV2 domain was fused to a 23 amino acid long peptide of a synthetic variant of the ODC degron called cODC1 [9, 15]. Several variants of this construct were produced to assess the overall performance of the degron, which was finally named photosensitive degron (psd) module (Fig. 1a). The variants that were used in comparison
Fig. 1 Engineering of the photosensitive degron (psd) module. (a) Design of the psd module and its variations. The psd module is composed of \textit{A. thaliana} LOV2 of phototropin 1 and a degron derived from murine ODC. As photoreceptor, the LOV2 domain (amino acids M460 to P616 indicated by blue color) was chosen; at the C-terminus it carries the so-called J\alpha-helix (orange color). The full degron sequences are given (green color). The control for maximum degron activity was a fusion of LOV2 with ODC36, the control for minimal activity was a LOV2 domain lacking any degron sequence. Degron sequences similar to ODC23 were fused to the LOV2 domain during optimization of the psd module. Variations in the LOV2 sequence comprised point mutations that kept the LOV2 domain in the lit (LOV2^{I608E}) or the dark (LOV2^{C512A}) state and were used to characterize the behavior of the module with minimal and maximal photoreceptor activity. To optimize the construct, site-directed mutagenesis was performed and the fusion point between LOV2 and the degron sequence was varied. Finally, the whole construct was used in a random mutagenesis approach. (b) Mechanism of psd module activation by blue light. Measurement of psd module variant performance was done with the constructs fused to red fluorescence protein (RFP). The constitutive \textit{ADH1} promoter was used to express psd module variants. In vivo, the psd module is inactive in darkness and the protein is stable. Upon blue light excitation of the LOV2 domain, the J\alpha-helix (indicated in yellow) is unfolded and the cODC1 degron is exposed and activated. This induces ubiquitin-independent degradation of the construct by the proteasome. (c) Scheme showing RFP-psd behavior in yeast. During growth in darkness, the RFP-psd fusion protein is highly abundant and the yeast cells show high levels of red fluorescence. Cells grown in the presence of blue light show low fluorescence levels due to depletion of RFP-psd.
with the psd module were constructs with the full cODC1 sequence and without any degron sequence to assess the maximal and minimal degradation activity, respectively. Similarly, LOV2 mutants that abolish signaling and lock the photoreceptor in the dark state or the light state were used as well [9]. The optimization process included the testing of psd module variants with changes in the degron sequence, alterations of the fusion sequence between LOV2 and cODC1 sequence, as well as testing of variants obtained by site-directed mutagenesis and random mutagenesis [9, 16]. For all constructs, red fluorescent protein (RFP) was used as reporter sequence, which allowed facile quantification of psd module variant performance (Fig. 1b, c). Examples for the diverse characteristics that were measured in psd module variants are given in Table 1.

Several conclusions can be drawn from our approach of generating a synthetic degron reactive to light, which might be valid for the generation of other synthetic constructs as well. For successful generation of a synthetic switch, the molecular mechanisms within the sensor and the effector domain should be known. It is essential to take every detail into account that influences the activity states of both domains. Several different fusions of both domains should

### Table 1
Characteristics of selected psd module variants

<table>
<thead>
<tr>
<th>Name</th>
<th>Half-life in darkness (min)</th>
<th>Half-life in 30 μmol m⁻² s⁻¹ blue light (min)</th>
<th>Dark–light ratio of fluorescence measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type psd module</td>
<td>123 ± 21</td>
<td>20 ± 1</td>
<td>11 ± 0.5</td>
</tr>
<tr>
<td>V19I</td>
<td>132 ± 50</td>
<td>28 ± 4</td>
<td>n.d.</td>
</tr>
<tr>
<td>K92R E132A E155G</td>
<td>102 ± 41</td>
<td>12 ± 0.4</td>
<td>22 ± 2.5*</td>
</tr>
<tr>
<td>K92R E132A E139N N148D E155G</td>
<td>66 ± 10</td>
<td>9.8 ± 0.4</td>
<td>16 ± 1.2*</td>
</tr>
<tr>
<td>K121 M N128Y</td>
<td>44 ± 8</td>
<td>8.5 ± 0.3</td>
<td>14 ± 1.4*</td>
</tr>
<tr>
<td>K121 M N128Y G138A</td>
<td>92 ± 28</td>
<td>13 ± 1</td>
<td>22 ± 4*</td>
</tr>
<tr>
<td>G138A</td>
<td>147 ± 52</td>
<td>22 ± 1</td>
<td>n.d.</td>
</tr>
<tr>
<td>G138A N148E</td>
<td>151 ± 53</td>
<td>75 ± 14</td>
<td>n.d.</td>
</tr>
<tr>
<td>ΔL156 ΔP157</td>
<td>187 ± 45</td>
<td>52 ± 9</td>
<td>n.d.</td>
</tr>
<tr>
<td>degODC</td>
<td>86 ± 21</td>
<td>20 ± 3</td>
<td>n.d.</td>
</tr>
<tr>
<td>CACA</td>
<td>87 ± 21</td>
<td>20 ± 3</td>
<td>n.d.</td>
</tr>
<tr>
<td>CACACA</td>
<td>66 ± 13</td>
<td>16 ± 1</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

Half-lives were measured by cycloheximide chase analysis and dark–light ratio by fluorescence intensity measurements. The asterisk (*) marks dark–light ratios that are considered to be statistically different from the ratio of psd (P < 0.05). The data in Table 1 were published previously [16]. Numbering of amino acids refers to the LOV2 domain, which starts at M460 in the full length phototropin
be tested, as receptor and effector domain might influence each other depending on the exact composition of the fusion site. Good controls that mimic maximum and minimum activity for both the receptor and the effector domain are necessary to have a good measure for the performance range of the synthetic switch. Once the switch shows reasonable performance, further optimizations by specific changes or a directed evolution approach may be conducted. In the latter case, it makes sense to allow changes in both, the signal receptor and the effector domain simultaneously, as the best performing variant might require alterations in both parts. Finally, a quick assay that is easily quantifiable is crucial to obtain reliable data about the performance of the different variants. The following part of this chapter contains a detailed description of the engineering steps and the yeast techniques that have been used to obtain and improve the photosensitive degron module.

# Materials

## 2.1 Yeast Strains

For homologous recombination and analysis of the psd module, no specific yeast strain is required; every wild type lab strain that has the necessary auxotrophy markers can be used (e.g., ESM356–1 [17], S288C background).

## 2.2 Cloning Procedures

1. High-Fidelity DNA amplification system (e.g., Phusion, KOD, or Herculase) for amplifying large DNA fragments including polymerase 10x reaction buffer, 10 mM dNTP stock solutions each and oligonucleotide primers.

2. Taq DNA polymerase for error prone PCR including Taq polymerase, 10x reaction buffer, 10 mM MnCl₂ stock solution, separate 10 mM stock solutions for dATP, dGTP, dCTP, and dTTP in order to optimize relative ratios for error prone PCR conditions and oligonucleotide primers.

3. QuickChange Site Directed Mutagenesis kit (Agilent).

4. DpnI restriction enzyme (for removing residual DNA template after PCR).

## 2.3 Yeast Media

Standard media were used for yeast growth [18].

1. Yeast complete medium (YPD): 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose.

2. Low-fluorescence medium [16]: 5 g/l (NH₄)₂SO₄, 1 g/l KH₂PO₄, 0.5 g/l MgSO₄, 0.1 g/l NaCl, 0.1 g/l CaCl₂, 0.5 mg/l H₃BO₄, 0.04 mg/l CuSO₄, 0.1 mg/l KI, 0.2 mg/l FeCl₃, 0.4 mg/l MnSO₄, 0.2 mg/l Na₂MoO₄, 0.4 mg/l ZnSO₄, 2 mg/l biotin, 0.4 mg/l calcium pantothenate, 2 mg/l inositol, 0.4 mg/l niacin, 0.2 mg/l 14-aminobenzoic acid
(PABA), 0.4 mg/l pyridoxine HCl, 0.4 mg/l thiamine, 2% \((w/v)\) glucose, supplemented with appropriate amino acids and/or other nutrients.

3. Solid synthetic complete medium (SC): 0.15\% \((w/v)\) Bacto yeast nitrogen base without amino acids and without ammoniumsulfate, 0.5\% \((w/v)\) ammoniumsulfate, 0.2\% \((w/v)\) drop-out mix lacking the compound used for plasmid selection, 2\% \((w/v)\) glucose, 2\% \((w/v)\) agar.

### 2.4 Yeast Transformation

1. SORB buffer: 100 mM lithium acetate, 10 mM Tris–HCl pH 8.0, 1 mM EDTA–NaOH pH 8.0, 1 M sorbitol, adjusted with acetic acid to pH 8.0 and filter-sterilized.

2. Carrier DNA: 10 mg/ml herring sperm DNA, denatured at 100 °C for 10 min and immediately cooled on ice.

3. Polyethylene glycol (PEG) buffer: 100 mM lithium acetate, 10 mM Tris–HCl pH 8.0, 1 mM EDTA–NaOH pH 8.0, 40\% \((w/v)\) PEG 4000; filter-sterilized.

### 2.5 Plasmid Rescue from Yeast

1. Breaking buffer: 2\% \((v/v)\) Triton X-100, 1\% \((w/v)\) sodium dodecylsulfate (SDS), 100 mM NaCl, 10 mM Tris–HCl pH 8.0, 1 mM EDTA–NaOH pH 8.0.

### 2.6 Cell Lysis and Immunoblotting

1. Alkaline lysis buffer: 1.85 M NaOH, 7.5\% β-mercaptoethanol.

2. High urea buffer: 5\% \((w/v)\) SDS, 8 M urea, 200 mM Na2HPO4/NaH2PO4 pH 6.8, 0.1 mM EDTA, 0.01\% \((w/v)\) bromophenol blue.

3. Protein-precipitation buffer: 55\% trichloroacetic acid (TCA) \((w/v)\)

4. Cycloheximide stock solution: 20 mg/ml.

5. Sodium azide buffer: 100 μM NaN₃.

6. Antibodies recognizing tagRFP are available from evrogen (www.evrogen.com), antibodies against Tub1 (loading control) were obtained from Abcam (www.abcam.com).

### 2.7 Illumination of Yeast with Blue Light

1. Light emitting diode (LED) stripes or clusters for illumination of yeast cells grown on plate or in liquid medium. The LED setup should include a dimmer to achieve a photon flux of 30 \(\mu\)mol m\(^{-2}\) s\(^{-1}\) at the level of the yeast cells. Single wavelength LEDs for blue light illumination (output wavelength 465 nm) or RGB LEDs with an appropriate controller can be used. Homogeneous illumination is advisable, which might be easier to achieve with many LEDs of lower light intensity output than with few high-power LEDs.

2. An optometer (e.g., P2000, equipped with light detector D-9306-2, Gigahertz-Optik, Türkenfeld, Germany) to measure the light-intensity used for illumination of yeast cells.
3. Clear plastic cell culture flasks equipped with a ventilated cap for illumination of yeast cultures growing in liquid medium.


5. A box with LEDs attached to the lid, spaced evenly to ensure uniform illumination. Covering the interior of the box with reflective material optimizes illumination. The box should be of sufficient size to host several Petri dishes or cell culture flasks.

6. Lightproof, heat shrinkable tubing adjusted to a diameter which is slightly bigger than a test tube to protect small yeast cultures from ambient light. To create the tubing, test tubes were covered with aluminum foil and placed in heat shrinkable tubing, which was subsequently adjusted in size with a heat gun. After cooling, the aluminum foil was removed with forceps.

3 Methods

3.1 Cloning by Homologous Recombination in Yeast

Generation of shuttle vectors for yeast and Escherichia coli by homologous recombination has been described previously [19]. The method can be adapted in many ways to generate gene fusions (Fig. 2).

1. Design primers for amplification of the gene of interest. The primers have two parts: for the 3′-end, sequences that allow AtLOV2 amplification are chosen. At the 5′-end, sequences derived from the vector are added, which will be used for the homologous recombination step in yeast that generates the final vector.

2. Perform PCR to generate the DNA fragment with the LOV2 photoreceptor domain (see Note 1). To generate photoreceptor variants, a PCR using mutagenic conditions can be performed (see Note 2).

3. Linearize the target vector by restriction enzymes. The sequences for homologous recombination should flank the gap.

4. Cotransform the fragments into frozen-competent S. cerevisiae cells.

5. Rescue the plasmid from yeast into E. coli cells, isolate it and verify the construct by enzymatic digest and sequencing.

3.2 Design of Primers for Cloning by Homologous Recombination

Target gene-specific primers contain sequences that are homologous to the target gene and sequences homologous to the vector sequence, which are added to the 5′ ends (Fig. 2). The resulting PCR product contains the target gene (e.g., LOV2 photoreceptor) flanked by sequences for homologous recombination. In the same way, the vector backbone can be generated with one or several oligonucleotide pairs (see Note 1).
Fig. 2 Workflow chart of psd module generation. Step 1: Primer design strategy for the cloning of the photoreceptor LOV2 into a yeast vector by homologous recombination. The primers have two parts; the 3′-end contains the sequence necessary to amplify the LOV2 gene, the 5′-end consists of sequence, which is identical to the vector sequence and will be used for homologous recombination in yeast. Step 2: Generation of the PCR product for homologous recombination. For a directed evolution approach, this PCR can be performed with conditions that enhance the error rate of the polymerase, which results in PCR products containing a library of photoreceptor mutants. Step 3: Linearization of the target vector, the sequences for homologous recombination should flank the gap in the plasmid. Step 4: Cotransformation of the linearized vector and the PCR products results in generation of the plasmid by homologous recombination. In case a directed evolution approach is undertaken, an appropriate assay is performed at this stage to identify interesting clones. Step 5: Plasmid rescue from yeast into E. coli, verification and retransformation in yeast. Step 6: Functional characterization of the construct under different conditions.
1. Standard PCR conditions can be used to generate the products used for homologous recombination. For cloning purposes, a kit containing a high-fidelity polymerase (e.g., Phusion, KOD, or Herculase) is advisable to reduce the introduction of errors during the reaction.

2. For a directed evolution approach, error-prone PCR followed by in yeast ligation generates a library of clones [16, 20].

3. PCR is performed with Taq polymerase under conditions that increase the error rate of the enzyme which is achieved by the presence of different concentrations of MnCl₂ (0, 0.62, and 1.25 mM) and a twofold excess of dCTP and dTTP (500 μM each in the final PCR reaction).

4. If necessary, the template vector can be destroyed after the PCR step by enzymatic digest with DpnI, which cleaves methylated DNA. Products obtained in this way are used without further manipulations for homologous recombination in yeast.

The QuickChange protocol can be followed to change single positions in the psd module by site-directed mutagenesis, which may result in psd modules with distinct characteristics. Mutations should be selected based on previous studies examining the same photoreceptor domain or a homologous one. Furthermore, changes that have been found in different variants obtained by random mutagenesis can be combined in a single construct by this method.

1. The oligonucleotides are designed as follows: the changed codon is flanked by 15 bases matching the vector sequence. Both forward and reverse primers are the reverse complement of each other. At each end, a cytosine or guanidine is preferred due to increased binding stability.

2. A whole vector PCR is performed with the template plasmid using standard conditions. A low number of cycles is recommended (<20).

3. The template plasmid is digested with DpnI that cleaves specifically methylated DNA.

4. The product of the PCR is transformed into E. coli by a standard method [21].

The protocol for yeast transformation is based on the lithium acetate method [22] and has been described previously [23–26].

1. Yeast cells are inoculated from an overnight preculture (approx. 1:50 dilution) and grown to an optical density (A₆₀₀) of 0.8–1.0 at 30 °C in 50 ml of YPD medium.
2. Yeast cells are harvested by centrifugation (3 min, 500 × g), washed once with sterile water (0.1–0.5 volumes) followed by a washing step with sterile SORB buffer (0.1–0.2 volumes).

3. Cells are suspended in 450 μl SORB buffer and 50 μl of carrier DNA is added. Cells are divided into appropriate aliquots and placed at −80 °C (no shock freezing).

4. Usually, 50 μl of competent cells are used for the transformation with a PCR product. Thawed competent cells are mixed with PCR product (5–15 μl of DNA for 50 μl of cells) and six volumes of PEG buffer are added.

5. Cells are incubated at room temperature for approximately 30 min.

6. Cells are incubated at 42 °C for 5–20 min (15 min works well with most strains).

7. Cells are sedimented (3 min, 500 × g) and washed once with YPD.

8. The pellet is resuspended in YPD and spread onto an appropriate selection medium. Plates with yeast cells harboring a light-sensitive construct should be protected from ambient light during growth and storage.

### 3.6 Plasmid Rescue from Yeast

A modified standard method is used to transfer plasmids from yeast to *E. coli* for further usage [21].

1. An amount of yeast cells corresponding to the size of a match head is scraped off a plate and dissolved in 500 μl of breaking buffer.

2. 200 μl of phenol–chloroform–isoamyl alcohol mixture (25/24/1, buffered with TE, pH 7.5–8) and 0.3 g of glass beads (~200 μl) are added.

3. Cells are disrupted by vortexing (5 min, highest speed) and phases are separated in a microcentrifuge (10 min, 16,000 × g).

4. 400 μl of the aqueous layer are transferred to a new tube and 1 ml of ethanol is added to this tube.

5. The tube is incubated at −20 °C for 10 min and then subjected to centrifugation (10 min, 16,000 × g).

6. The supernatant is removed and the pellet is washed with 70% ethanol.

7. The tube is centrifuged again (3 min, 16,000 × g), the supernatant removed and the pellet air-dried for 3 min.

8. The pellet is dissolved in 20 μl of desalted water.

9. 1 μl of the solution is used for transformation of *E. coli* cells by electroporation.
3.7 Detection of the Target Protein RFP by Immunoblotting

The target protein can be detected by immunoblotting using antibodies directed against the tester protein tagRFP. For immunoblotting, crude cell extracts can be prepared by alkaline lysis [27].

1. 1 ml of logarithmically growing cells (A600 = 1) is treated with 150 μl of alkaline lysis buffer and kept for 10 min on ice.

2. Protein precipitation is induced by addition of 150 μl 55% (w/v) trichloroacetic acid (TCA) followed by 10 min incubation on ice.

3. The samples are subjected to centrifugation (10 min, 16,000 × g) and the supernatant is removed. The pellet is dissolved in 60 μl of high urea buffer by mixing the sample vigorously at 65 °C.

4. The extracts are cleared from cell debris by centrifugation (10 min, 16,000 × g), 10–20 μl of sample is loaded onto an SDS-PAGE gel. Standard procedures can be used for SDS-PAGE and blotting [28, 29].

3.8 Measurement of Protein Degradation by Cycloheximide Chase

Cycloheximide is a translational elongation inhibitor that can be used to stop protein biosynthesis in eukaryotic cells. Thus, the stability of a protein can be inferred from the decay rate after addition of cycloheximide to logarithmically growing yeast cells. Target protein levels can be detected by an appropriate assay, e.g., immunoblotting.

1. Grow yeast cells in darkness until logarithmic growth phase is reached.

2. Take first sample (t = 0 h) and subject to alkaline lysis.

3. Add cycloheximide to a final concentration of 200 μg/ml to the yeast culture to stop protein synthesis.

4. Expose cells to specific illumination conditions, e.g., darkness or a distinct amount of blue light for the rest of the experiment.

5. Collect equal amounts of sample at several time points that match the protein decay rate (e.g., every 30 min) and subject to alkaline lysis.

6. Perform immunoblotting to measure protein abundance over time.

3.9 Quantification of Target Protein Levels by Fluorescence

The target protein RFP can directly detected by fluorescence measurements with a fluorimeter. For the interpretation of the results, the maturation time of the used fluorescent protein has to be taken into account.

1. 1 ml of logarithmically growing cells (A600 = 1) is treated with 100 μl of sodium azide buffer.

2. Samples are subjected to centrifugation (3 min, 500 × g) and 1 ml of the supernatant is removed. The pellet is dissolved in the remaining supernatant.
3. The samples are transferred in a lightproof, flat-bottom microtiter plate and transferred to a fluorimeter.

4. The fluorescence intensity of each sample is measured according to the protocol given by the manufacturer of the fluorimeter.

The photosensitive degron module is activated by blue light (465 nm, 30 μmol m$^{-2}$ s$^{-1}$); cells are grown under continuous illumination (see Note 3). The illumination regimen can be easily adapted for yeast cells growing on solid medium (see Note 4).

1. Yeast cells are grown in low fluorescence medium supplemented with 2% glucose in darkness or under continuous illumination until mid-log phase is reached (see Note 5).

2. Degradation of the psd module is initiated by exposure of the cells to blue light (465 nm, 30 μmol m$^{-2}$ s$^{-1}$). The time frame until depletion of the target protein is achieved depends on the psd module variant, but it can be expected that 2–4 h are sufficient for short-lived psd module variants (see Note 3). Immunoblotting or fluorescence measurements can be used to quantify RFP-psd levels. Measuring the abundance of the construct in cells kept under restrictive conditions (blue light-exposed) and permissive conditions (darkness) gives a measure for the depletion efficiency and the overall performance of the psd module variant. Ongoing target protein synthesis will lead to minimal amounts of target protein even after prolonged exposure to blue light (see Note 6). The overall performance of a psd module variant can be assessed with mutants, in which the degron or the photoreceptor is inactivated (see Note 7).

4 Notes

1. Homologous recombination offers high flexibility for vector design. The illustration shows the most simplified case, in yeast ligation of a single gene or gene fragment into an existing vector (Fig. 2). However, with the same approach, several fragments can be ligated together in one step. To do so, it is necessary to generate DNA fragments with flanking homologous sequences. The same strategy as depicted for AtLOV2 can be used to combine several genes or gene fragments, e.g., one-step-fusion of a photoreceptor with a degron sequence. Also the vector backbone can be produced by homologous recombination of PCR fragments [19].

2. To generate photoreceptor variants, the PCR to amplify the AtLOV2 domain can be performed with a polymerase lacking proofreading activity like the Taq polymerase. Furthermore,
addition of manganese ions and unbalanced amounts of nucleotides will increase the error rate. Standard protocols can be found in the literature [16, 20, 21].

3. The FMN cofactor of the LOV2 domain is excited by blue light (400–500 nm) with a maximum at 450 nm. Thus, the wavelength used for illumination can be varied without affecting the protein destabilization activity of the psd module. At a light-flux of 30 µmol m$^{-2}$ s$^{-1}$ (465 nm), strong activation of the psd module has been observed, whereas yeast cell growth is not affected [9, 16]. A much higher light flux is not recommended, as yeast cells are sensitive to high-intensity blue light [30]. To characterize the behavior of the psd module at different light conditions, the light flux can be varied between 0 and 30 µmol m$^{-2}$ s$^{-1}$. It is advisable to measure the actual light-flux before the experiment with an optometer. The illumination strength influences the psd module degradation rate; reduced illumination will prolong the time that is necessary to reduce the target protein amounts in most cases [9, 16].

4. For growth in the presence of blue light on solid medium, yeast cells bearing a plasmid can be grown on synthetic medium lacking the appropriate auxotrophy compound. An upright orientation of the Petri dishes is recommended to expose the cells directly to the light, which avoids unwanted shading. The same LED-equipped boxes that are used for growth in liquid culture can be used to grow yeast on solid medium.

5. Low fluorescence medium is recommended to grow cells exposed to defined amounts of light. The carbon source can be chosen freely. Other media can be used as well, as long as the blue light is not completely absorbed by the medium. If light absorption takes place, the illumination conditions have to be adjusted accordingly. The main difference between low fluorescence medium and synthetic complete medium is the lack of flavin derivatives in low fluorescent medium. If a yeast strain is used that does not produce FMN in appropriate amounts, it can be added to the medium or synthetic complete medium can be used. In both cases, the blue light absorption of the flavin derivative in the medium has to be taken into account and the illumination conditions have to be adjusted accordingly.

   Depending on the experiment, different illumination regimens are applied. For cycloheximide chase experiments, cells are kept in darkness during the initial growth phase and illumination is started after the first sample has been collected. For protein level quantifications, cells are grown under constant conditions, either exposed to a certain light flux or kept in darkness.
6. The psd module constructs are expressed by a constitutive promoter, which leads to constant synthesis of the psd module. Under blue light illumination, a steady-state level is reached, which depends on production and degradation rates of the protein. Usage of a regulatable promoter may further decrease the depletion time and lower target protein abundance at restrictive conditions.

7. A suitable control to assess the influence of a certain mutation on the degradation mechanism of the construct under consideration is a mutation that abolishes cODC1 degron activity. For constructs with differences in light-regulated activation of the degron, a similar stability is expected after inactivation of the degron. If a mutation in the photoreceptor domain influences the stability of the construct without increasing the degron presentation, the increased degradation rate persists after mutational degron inactivation. Such a mutation is less valuable, as the contribution of the mutation is not controlled by light and cannot be influenced by further modifications of the photoreceptor. Ideally, all modifications influence the degradation rate solely by impacting on degron presentation. In this case, the signaling state of the photoreceptor determines protein stability.

8. Another set of controls are mutations that abolish light-dependent signaling of the photoreceptor. For the Arabidopsis thaliana phototropin 1 LOV2 domain examples are I608E and C512A, which lock the photoreceptor in the light and the dark state, respectively. Please note that for mutations similar to C512A some residual light activation has been observed in homologous LOV domains [31].

Acknowledgments

I thank D Störmer for her excellent technical assistance as well as S Hepp, C Renicke, and J Trauth for helpful comments on the manuscript. A Batschauer is acknowledged for the opportunity to use his photobiology equipment. This work was supported by the DFG grant TA320/3-1 and the DFG-funded graduate school GRK1216.

References


Chapter 16

Light-Regulated Protein Kinases Based on the CRY2-CIB1 System

Wignand W.D. Mühlhäuser, Maximilian Hörner, Wilfried Weber, and Gerald Radziwill

Abstract

Optogenetic approaches enable the control of biological processes in a time- and space-resolved manner. These light-based methods are noninvasive and by using light as sole activator minimize side effects in contrast to chemical inducers. Here, we provide a protocol for the targeted control of the activity of protein kinases in mammalian cells based on the photoreceptor cryptochrome 2 (CRY2) of Arabidopsis thaliana and its interaction partner CIB1. Blue light (450 nm)-induced binding of CRY2 to CIB1 allows the recruitment of a chimeric cytosolic protein kinase AKT1 to the plasma membrane accompanied with stimulation of its kinase activity. This protocol comprises the transient and stable implementation of the light-regulated system into mammalian cells and its stimulation by blue light-emitting diodes (450 nm) irradiation as well as analysis of the light-activated AKT1.

Key words Optogenetics, Signal transduction, Protein kinases, Membrane recruitment, AKT, CRY2

1 Introduction

Protein kinases play a key role in orchestrating the complex cellular signal transduction machinery. By phosphorylation they can influence activity, localization, and/or the interactome of these signaling complexes. Thereby, protein kinases significantly influence cell fate decisions (e.g., proliferation vs. differentiation; survival vs. apoptosis). For this reason, it is not surprising that protein kinases themselves underlie a tight regulation. Reoccurring mechanisms comprise regulation by other kinases, by local recruitment to specific subcellular compartments or clustering (e.g., dimerization) of kinases [1].

Conventional approaches, which try to mimic activation mechanisms, often employ chemical inducers or mutants. Major drawbacks of those approaches lie in their poor or altogether missing spatiotemporal resolution and, in certain cases, unwanted or unknown side effects in the case of chemical inducers. Therefore,
the use of constantly improving light-inducible dimerization/oligomerization systems (i.e., optogenetic tools) becomes more and more popular. Using light as the sole activator allows unprecedented spatiotemporal resolution with minimal side effects. This allows the intervention of signal transduction in a most sophisticated manner [2, 3].

One optogenetic tool that has been employed in the activation of various kinases is the blue light-responsive cryptochrome 2/cryptochrome-interacting basic-helix-loop-helix 1 (CRY2/CIB1) system [4–6]. Cryptochromes are flavoprotein photoreceptors found in bacteria, higher plants, and animals. They were first identified in Arabidopsis thaliana, where they regulate growth and developmental processes [7]. Cryptochromes contain a C-terminal cryptochrome extension and an N-terminal photolyase homology region (PHR) that noncovalently binds flavin adenine dinucleotide (FAD). Irradiation with blue light of wavelengths ranging from 390 nm to 480 nm leads to FAD reduction, which converts cryptochrome into its biological active conformation [8]. The activated Arabidopsis thaliana CRY2 can interact with CIB1, which is involved in the mechanism of floral initiation. Besides the interaction with CIB1, photoexcited CRY2 can also form homomers. Light-activated CRY2 forms homo- and heteromers within seconds and dissociates in the dark within minutes in a repeatable manner. Optogenetic systems have been generated based on the light-induced oligomerization of the PHR region of CRY2 (aa 1—489) and the N-terminus of CIB1 (CIBN; aa 1—170). The characteristics of the CRY2-based optogenetic are especially suited for the control of signaling events in mammalian cells [3, 9] (Fig. 1). For instance, blue light-induced clustering of CRY2PHR fused to the protein kinase RAF as well as the recruitment of chimeric RAF to membrane-bound CIBN activates the mitogen-activated protein kinase cascade (MAPK) signaling pathway [4, 5, 10]. To gain control of the PI3K/AKT pathway, CRY2PHR either fused to the inter-SRC-homology 2 (iSH2) domain of the regulatory p85α subunit of PI3K or to the protein kinase AKT were recruited to a membrane-anchored CIBN to activate signaling upon blue light irradiation [6, 11]. Due to its high relevance constant efforts to further improve this system are undertaken (e.g., enhanced oligomerization) [12, 13].

In this chapter, we describe a detailed approach of how to use the CRY2PHR/CIBN optogenetic tool to control AKT1 (optoAKT) to activate the PI3K/AKT signaling pathway. A detailed description of the experimental setup, as well as a step-by-step instruction of the experimental procedure, will be given. By following the instructions provided, the experienced scientist will be able to generate, either transiently or stably, optoAKT-expressing cells. Upon irradiation with blue light AKT1, fused to the PHR domain of CRY2, is recruited toward a membrane-bound CIBN. The light-dependent
recruitment toward the membrane leads to the activation of AKT1. The validation of kinase activation is shown by Western blot analysis using antibodies against AKT, AKT<sub>pT308</sub>, and AKT<sub>pS473</sub>. The used constructs further contain fluorescence proteins, which further allow the visualization of light-dependent membrane recruitment via fluorescence microscopy. The described methods can be transferred to other protein kinases depending on oligomerization or recruitment to defined subcellular regions for activation.

### 2 Materials

#### 2.1 Cell Culture

1. Culture medium: Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% (v/v) fetal calf serum (FCS) and penicillin (100 U/ml)/streptomycin (100 μg/ml). Stored at 4 °C.

2. Phosphate buffered saline (PBS): 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, 137 mM NaCl, and 2.7 mM KCl, pH 7.4.

3. Trypsin-EDTA solution: 0.05% Trypsin, 0.02% EDTA in PBS without Ca<sup>2+</sup> and Mg<sup>2+</sup>. Stored at 4 °C.

4. CASY model TT cell counting device, 150 μm (Roche, cat. no. 05651697001) with CASY cups (Roche, cat. no. 05651794001) and CASY ton buffer (Roche, cat. no. 05651808001). Alternatively, a standard hemocytometer can be used.
Cell culture-certified disposable plasticware: 10-cm Petri dishes, 6-well and 24-well plates, serological pipettes, and 15- and 50 ml conical centrifuge tubes.

Centrifuge with rotor which allows centrifugation of 15 ml conical centrifuge tubes.

CO₂ incubator at 37 °C and 5% CO₂.

2.2 Transient Transfection

1. Opti-MEM (Thermo Scientific, cat. no. 31985-062). Store at 4 °C.

2. Polyethylenimine (PEI) transfection solution: 1 mg/ml PEI in H₂O, adjusted to pH 7.0 with HCl. After sterile filtration (Ø 0.22 μm) 500 μl aliquots are stored at −80 °C. Avoid repeated freeze-thaw cycles.

3. Silica-based anion-exchange DNA purification kits.

2.3 Transduction

1. Advanced DMEM (Thermo Scientific, cat. no. 12391-015).

2. Cholesterol: 5 mM diluted in PBS.

3. Egg lecithin (Serva Electrophoresis GmbH): 5 mM diluted in Ethanol.

4. 1× chemical defined lipid concentrate (Gibco, cat. no. 11905031).

5. Hexadimethrine bromide (polybrene): Dilute in deionized water to a final concentration of 8 mg/ml.

6. Centrifuge that allows centrifugation of cell culture plates.

7. Syringe and 0.45 μm filter (low protein binding).

2.4 Irradiation and Readout

1. Light boxes: Build out of opaque PVC material. Panels of LEDs emitting light with a wavelength of 450 nm should be placed on the top of the box (Roithner LaserTechnik, LED450-series). The intensity of the LEDs has to be adjustable. Further, it is advantageous to include a mean of further programming the LEDs ON/OFF time to address temporal aspects of signaling. A ventilation of the light box should be included to ensure that the temperature as well as atmosphere in the box corresponds to the surroundings (see Note 1).

2. LED safe lights with wavelength longer than 500 nm.

2.5 Western Blot

1. 5× SDS-loading buffer: 10% (w/v) SDS, 0.3 M Tris–HCl (pH 6.8), 50% (v/v) Glycerin, 12.5% (v/v) β-mercaptoethanol, 0.05% (w/v) bromphenol blue in fully desalinated H₂O.

2. ECL solutions.

3. TENT cell lysis buffer (modified): 20 mM Tris–HCl (pH 8.0), 1 mM EDTA, 100 mM NaCl, 0.5% (v/v) Triton X-100, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate,
50 mM sodium fluoride, 10 mM β-glycerophosphate, 1× protease inhibitors (Roche, cat. no. 04693132001), 0.1% SDS (w/v). Store at 4 °C.

4. Running buffer: 192 mM glycine, 25 mM Tris, 0.1% (w/v) SDS in fully desalinated H₂O.

5. PVDF membrane.

6. Transfer buffer: 125 mM glycine, 25 mM Tris, 10% (v/v) methanol in fully desalinated H₂O.

7. Tris-buffered saline, 0.025% (v/v) Tween-20 (TBS-T): 50 mM Tris–HCl (pH 7.5), 150 mM NaCl in distilled H₂O. Adjust pH to 7.5 with HCl before adding 0.025% (v/v) Tween-20.

8. TBS-T supplemented with 5% BSA (v/v).

9. Primary antibodies (Cell Signaling Technology): AKT (pan) (C67E7) (cat. no. 4691), Phospho-AKT (Ser473) (cat. no. 9271), Phospho-AKT (Thr308) (cat. no. 4056).

10. Secondary antibodies (Santa Cruz Biotechnology): goat anti-rabbit IgG-HRP (cat. no. sc-2030).

### 2.6 Fluorescence Microscopy

1. 1,4-diazabicyclo[2,2,2]octane (DABCO).

2. Dulbecco’s phosphate buffered saline (DPBS).

3. Immersion oil.

4. Mowiol mounting solution: 2.4 g mowiol, 6 g glycerol in 133 mM Tris–HCl, pH 8.5. Store at 4 °C.

5. Paraformaldehyde (PFA) cell fixation solution: 4% (w/v) PFA in PBS. Store at −20 °C.

6. 12 mm round cover slips. Autoclave and store in a sealed container prior to use.

7. Fluorescence microscope with filter for green and red fluorescent proteins.

8. Microscope slides.

### 2.7 Cell Lines


2. HEK293T: DSMZ, Braunschweig, Germany, cat. no. ACC 635.


4. MCF7: DSMZ, Braunschweig, Germany, cat. no. ACC 115.

HEK293T, Platinum E, and MCF7 cells are cultured in DMEM supplemented with 10% (v/v) FCS and penicillin (100 U/ml)/streptomycin (100 μg/ml). C2C12 cells are cultured in DMEM supplemented with 15% (v/v) FCS and 1% (v/v) penicillin/streptomycin. Starvation was performed with DMEM supplemented with penicillin (100 U/ml)/streptomycin (100 μg/ml) (see Note 2).
See Table 1 for the list of constructs. For the transfection, it is important to use high-purity plasmid DNA, which can be obtained for example with Jetstar 2.0 Midiprep kit.

2.8 DNA Constructs

3 Methods
Day 1: Begin with a 10-cm plate with ~80% confluent HEK293T (transfection and lentiviral vector transduction) or Platinum E (retroviral vector transduction) cells. Pre-warm culture medium, PBS, and trypsin-EDTA solution in a 37 °C water bath. When preparing samples for fluorescence microscopy, place autoclaved coverslips into wells of 24-well plate.

2. Aspirate culture medium and wash once with PBS to remove residual culture medium before adding 1 ml of trypsin-EDTA solution. Place cells back into incubator for ~3 min. After incubation use a serological pipette to flush the plate with 5 ml culture medium and transfer the cell suspension into a 15 ml conical centrifuge tube. Centrifuge the conical centrifuge tube containing the cell suspension for 3 min at 300 × g. Afterward, discard the supernatant and resuspend the cell pellet in 10 ml culture medium. Dilute appropriate amount of cell suspension (dependent on CASY setup) in 10 ml CASY ton buffer within a CASY cup and determine cell number using CASY cell counter (see Note 3).

1. Seed HEK293T cells with a density of 0.75 × 10⁵ cells/cm². Either diluted in 2 ml culture medium for 6-well plates (for Western blot analysis) or in 0.5 ml culture medium for 24-well plates (for fluorescence microscopy). It is important that for each experimental condition at least one replicate on a separate plate is prepared, which is going to serve as dark control later.

2. Day 2: Cells should have reached ~60–70% confluence for a good transfection efficiency. For the transfection of a 6-well use a total of 3 μg DNA and for 24-well plate a total of 0.75 μg DNA. Transfect pGR427 and pGR464 in a ratio of 1:1 (see Note 4). For sole transfection of pGR427 use an empty vector instead of pGR464 to guarantee comparable transfection conditions. Also include an empty vector control to ensure that irradiation itself is causing the observed effects.

3. Dilute the DNA in 100 μl or 50 μl opti-MEM for 6-well and 24-well, respectively. Next add PEI solution (9 μl for 6 well; 2.25 μl for 24 well) and vigorously vortex PEI/-opti-MEM/-DNA solution for 5 s (see Note 5). Let the solution incubate for 15 min at room temperature. Prior to applying the solution pipette it up and down a couple of times to ensure homogeneity then add it dropwise onto cells. Afterward, gently rock the plate to evenly distribute the PEI/-opti-MEM/-DNA mixture and put the cells back into the incubator (see Note 6).

4. After 5 h replace the transfection medium with fresh culture medium and return the cells back into the incubator. After an incubation period of ~24 h proceed with the experiment (continue with Subheading 3.2).
1. **Day 1**: For lentiviral vector production complement 10 ml of Advanced DMEM with 2% (v/v) FCS and penicillin (100 U/ml)/streptomycin (100 μg/ml). Further add 20 μl cholesterol, 20 μl egg lecithin, and 1× chemical defined lipid concentrate. Dilute HEK293T cells in 10 ml with a density of 0.75 × 10^6 cells/cm² and seed them in a 10-cm plate.

   For retroviral vector production dilute Platinum E cells with a density of 0.75 × 10^6 cells/cm² in 10 ml culture medium and seed them in a 10-cm plate (see Note 8).

2. **Day 2**: Cells should have reached ~60–70% confluence before continuing. For lentiviral vector production prepare PEI transfection mix by adding 5 μg pLTR-G, 5 μg pCD/NL BH*DDD, 10 μg pWM043 (or pWM023), and 1 ml opti-MEM. Afterward, add 60 μl PEI and vigorously vortex the solution for 5 s. For retroviral vector production dilute 8 μg PA-AKT (or pWM029) in 1 ml opti-MEM. Then add 24 μl PEI and vortex vigorously. In both cases, incubate the mixture for 15 min at room temperature before adding it dropwise on the cells. Gently rock the plate to ensure an even distribution of transfection mix. Place the cells back into the incubator.

3. After 5 h replace the transfection medium with fresh Advanced DMEM including the above-listed additives and return the cells back into the incubator. At this time point the medium may already contain viral particles. BSL2 precautionary measures have to be undertaken when working with lentiviral vectors.

4. **Day 3**: As target cells seed MCF7 cells, for lentiviral vector transduction, or C2C12 cells, for retroviral vector transduction, in 6-well plate with a density of 10^5 and 0.3 × 10^5 cells/cm², respectively, according to the procedure described in Subheading 3.1.

5. **Day 4**: After 48 h incubation the transfected HEK293T/Platinum E cells have produced a sufficient amount of viral particles. Collect the supernatant, which contains either the retro- or lentiviral particles, using a serological pipette and filter it through a 0.45 μm sterile filter into a 50 ml conical centrifuge tube (see Note 9). Add polybrene solution to a final concentration of 8 μg/ml to the viral particle-suspension.

6. Replace the medium of the target cells with a total of 2 ml viral particle-suspension. For a double transduction use 1 ml of each viral particle-suspension (e.g., 1 ml of pWM043 and pWM023 harboring vectors). Next place cells back into the incubator for 15 min. After incubation centrifuge plates for 30 min at 1100 × g (see Note 10). Subsequently to centrifugation discard viral particle containing medium and add fresh culture medium before returning cells into the incubator again.
7. **Day 5**: The transduced cells should have reached a confluence of ~80% before continuing. Switch off ambient light and turn on the safe lights. Aspirate the medium and wash the cells once with PBS before adding 150 μl trypsin-EDTA solution. Place cells back into incubator for ~3 min or until they detach.

8. Wash cells off the plate with 2 ml culture medium, transfer them to a new 10-cm plate, and add culture medium up to a total of 10 ml (see Note 11). Keep cells in incubator until they have reached ~80% density.

9. The cells now stably express the transduced constructs (see Note 12). Seed the cells (10^6 and 0.3 × 10^5 cells/cm² for MCF7 and C2C12 cells, respectively) in either 6-well plates (for Western blot analysis) or on coverslips in 24-well plates (for fluorescence microscopy). Again, include a replicate on a separate plate for each condition as dark control.

### 3.2 Light Experiment and Readout

#### 3.2.1 Western Blot Analysis

1. **Day 1**: Start with optoAKT expressing cells on 6-well or 24-well plate transfected/seeded the day before (see Subheadings 3.1.1 and 3.1.2, respectively), which should have a confluence of ~80% in the evening. Replace culture medium with starvation medium (see Note 13) and place plates back into incubator. It is important that the cells are not exposed to any source of irradiation (390–480 nm) that would inadvertently excite the optogenetic system during starvation.

2. **Day 2**: In the morning (~12–15 h post starvation) cells are ready for the experiment. The exact conditions of irradiation vary depending on the desired readout.

1. For analysis of optoAKT by Western blotting set the intensity of the LEDs in the light box to 150 μmol/(m² s) and irradiate the cells 15 times for 5 s/min or constantly with a lower intensity (see Note 14). Intensities higher than 50 μmol/(m² s) may induce photo-cytotoxic effects when applied for extended time periods. Before further processing the cells ensure that ambient light is turned off and safe light is turned on (see Note 15) and that there are no additional sources of light which accidently could activate the optogenetic system.

2. Aspirate medium and wash cells once with PBS before adding 250 μl ice-cold modified TENT cell lysis buffer. Gently shake the plates to evenly distribute the lysis buffer. Incubate the plates for 10 min on ice.

3. Use a cell scraper to detach the cells and transfer the cell lysate into a precooled 1.5 ml Eppendorf tube. Sonicate the lysate for 15 min (30 s/min pulses) at 4 °C to fragment DNA and lower viscosity. In the next step, centrifuge lysates for 15 min at 10,000 × g and 4 °C to pellet the insoluble components.
Transfer 200 μl of supernatant into a new Eppendorf tube and add 50 μl 5× SDS-loading buffer (see Note 16), next boil the sample for 5 min at 95 °C (see Note 17). The lysates can be stored at −20 °C when not directly used.

4. For SDS-PAGE, load protein ladder and 10 μl of each sample onto a 9% SDS-gel with 15 pockets (8.3 × 6.4 cm). Conduct gel electrophoresis in running buffer at 90 V for approximately 2 h.

5. For the semidry Western blot using a PVDF membrane, the membrane has to be activated in the first step. Therefore, incubate the membrane 1 min in methanol and then incubate for another minute in transfer buffer. At the same time, incubate two Whatman paper per gel in transfer buffer. Break open the SDS-gel chamber and assemble a “sandwich” consisting out of Whatman paper -- PVDF membrane -- SDS gel -- Whatman paper (anode to cathode) in a semidry blotter. Perform Western blot for 1 h at 0.35 A/SDS-Gel.

6. Afterward block the membrane in TBS-T supplemented with 5% BSA (v/v) for 1 h. Dilute primary antibodies 1:1000 in TBS-T supplemented with 5% BSA (v/v) and add them to the membrane. Incubate overnight at 4 °C.

7. Day 3: Wash the membrane thrice with TBS-T for 10 min each time. Dilute secondary antibody coupled to horse radish peroxidase 1:10,000 in TBS-T supplemented with 5% BSA (v/v) and incubate membrane for 1 h at room temperature. Again wash the membrane thrice with TBS-T for 10 min each time.

8. Apply ECL I/ECL II mixture (1:1) onto the membrane incubate for some seconds and detect chemiluminescence with luminescent image analyzer (Fig. 2).

3.2.2 Fluorescence Microscopy

1. Day 1: Place one of the 24-well plates you transfected/seeded (see Subheadings 3.1.1 and 3.1.2) the day before into a light box while the control plate remains in the dark. Afterward irradiate cells with 1.5 μmol/(m² s) for 5 min.

2. Switch on safe lights before taking cells out of the incubator to prevent uncontrolled activation of the optoKinase system. Subsequently, wash them once with DPBS before adding 200 μl 4% paraformaldehyde cell fixation solution per well (see Note 18).

3. Let the cells incubate for 10 min on ice and subsequently for additional 10 min at room temperature. Afterward, the samples do not have to be handled under safe light anymore.

4. Apply 7 μl mowiol (with DABCO) onto a microscope slide, try to avoid generating air bubbles.

5. Use pincers to pick up cover slip and dip it a couple of times into water to remove excess of paraformaldehyde. Use a cos-
metic tissue to absorb the water and place the coverslip with the cells first on the drop of mowiol. Let the slides dry over-night in the dark to prevent bleaching.

6. Day 2: Use nail polish to seal the samples before looking at them under the fluorescence microscope (Fig. 3).

---

**Fig. 2** Western blot analysis of cells expressing optoAKT. (a) HEK293T cells were transiently co-transfected with CRY2-EGFP-AKT1 (pGR427) and either membrane anchored CIBN-mCherry-CaaX (pGR464) (+) or junk DNA (−). After overnight serum-starvation those cells were irradiated with blue light (450 nm, 40 μmol/(m² s)) for 15 min (+) or kept in the dark (−). Whole cell lysates were prepared and subjected to Western blot analysis using the designated antibodies. (b) MCF7 cells co-transduced with CRY2-EGFP-HA-AKT1ΔPH (pWM043) and membrane anchored m/p-mCherry-CIBN (pWM023) were serum-starved overnight. Afterward, those cells were either irradiated with blue light (450 nm) with an intensity of 150 μmol/(m² s) for 5 s/min 15 times (Pulse), with a constant intensity of 5 μmol/(m² s) for 15 min (Const.) or kept in the dark (−). Whole cell lysates were prepared and equal amounts of lysate were subjected to Western blot analysis using the designated antibodies. *endog. AKT* endogenous AKT

**Fig. 3** Fluorescence microscopic analysis of light-induced plasma membrane recruitment. Transduced cells, co-expressing photoactivatable CRY2-YFP-AKT1 [6] and membrane anchored m/p-mCherry-CIBN (pWM029), were seeded on cover slips and serum-starved overnight. Serum-starved cells were then either irradiated with blue light (450 nm, 1.5 μmol/(m² s)) for 5 min or kept in the dark. Afterward, cells fixated with paraformaldehyde were analyzed by fluorescence microscopy. Scale bar: 10 μm
4 Notes

1. For a detailed description of the light boxes, see refs. 16, 17.

2. It is also possible to use other cell lines. However, be aware that signal transduction can differ between cell lines and that the procedure described here might not be directly transferable to other cell lines.

3. When no CASY cell counter is available, a hemocytometer in combination with a bright field microscope can be used to determine the cell number. If there is also no hemocytometer available, the amount of cells to be seeded can be vaguely estimated. One dense 10-cm plate contains approximately $1.5 \times 10^7$ cells.

4. When rocking the plate try to form a lying eight or perform a north-south followed by an east-west movement. Avoid circular movement, since this is causing an accumulation of the cells in the middle of the plate. Once placed in the incubator do not move the plate for at least a couple of hours to ensure attachment of the cells.

5. To optimize the optoSystem the amount and/or the ratio of CRY2 and CIBN constructs can be adjusted.

6. Perform same movement as for seeding to prevent local accumulation of transfection mix.

7. The generation of a stable cell line takes time to begin with but is going to save time later. Furthermore, the overall expression level of the recombinant proteins can be further refined by cell sorting.

8. Instead of using a packaging cell line (i.e., Platinum E) the generation of retroviral vectors can be conducted by transfecting HEK293T cells with the appropriate helper plasmids.

9. When filtrating viral particles use a filter with low protein-binding properties. PVDF and nitrocellulose filter will bind viral particles and strongly reduce transduction efficiency.

10. Wrap plates carefully with parafilm to prevent accidental splashing of viral particles and to delay equilibration of CO$_2$ with surrounding.

11. In case trypsin-sensitive cells are used it may be advisable to centrifuge the cells and discard the supernatant to remove all trypsin; alternatively, an enzyme-free detachment solution can be used.

12. It is possible (and a good idea) to expand the stable cell line generated at this point or after a first experiment, ensuring proper functionality of the optoSystem. This allows skipping future work under BSL2 conditions and also provides a backup of cells.
13. The residual amount of culture medium after aspiration gains increasing significance with decreasing surface area. Therefore, washing the wells with PBS in case of small well format is recommended (i.e., 48- /96-well format).

14. Light intensities of 5 μmol/(m² s) or lower are strongly recommended, since they sufficiently trigger the optogenetic System, while showing negligible phototoxicity.

15. Illuminate the cell culture lab by safelight LEDs (i.e., wavelength longer than 500 nm). Measure the wavelength of LEDs to ensure that they will not excite the blue light-responsive CRY2/CIBN system.

16. After the centrifugation the almost clear pellet may be hard to detect. Make sure not to transfer it, since it will lower the quality of the Western blot.

17. When boiling samples use Eppendorf Safe-Lock tubes. This will prevent accidental opening of the tubes during boiling, which leads to uneven protein amounts between samples.

18. Do not process too many samples at once. In case of the CRY2/CIBN system, the light-induced interaction of the fusion proteins is limited when working under safe light.

Acknowledgments

This work was supported by the Excellence Initiative of the German Federal and State Governments (EXC-294, BIOSS – Centre for Biological Signaling). We thank J. Schmidt, D. Schächtele, and J. Meßmer (University of Freiburg) for designing and constructing the illumination boxes.

References


8. Cui Y, Choudhury SR, Irudayaraj J (2014) Quantitative real-time kinetics of optogenetic proteins CRY2 and CIB1/N using single-


Yeast-Based Screening System for the Selection of Functional Light-Driven K+ Channels

Cristian Cosentino, Laura Alberio, Gerhard Thiel, and Anna Moroni

Abstract

Ion channels control the electrical properties of cells by opening and closing (gating) in response to a wide palette of environmental and physiological stimuli. Endowing ion channels with the possibility to be gated by remotely applied stimuli, such as light, provides a tool for in vivo control of cellular functions in behaving animals. We have engineered a synthetic light-gated potassium (K+) channel by connecting an exogenous plant photoreceptor LOV2 domain to the K+ channel pore Kcv. Here, we describe the experimental strategy that we have used to evolve the properties of the channel toward full control of light on pore gating. Our method combines rational and random mutagenesis of the channel followed by a yeast-based screening system for light-activated K+ conductance.

Key words Functional complementation, Protein evolution, Rational and random mutagenesis, Light, Screening, S. cerevisiae, Optogenetics, Ion channels, Potassium (K+), Gating

1 Introduction

Potassium (K+) channels are membrane proteins ubiquitous in living organisms. They drive the flux of K+ ions across the ion-impermeant lipid membrane. In higher organisms, they play a fundamental role in controlling cell excitability and intercellular communication, particularly between neurons. Potassium channels are tetrameric proteins in which each monomer contributes to the formation of a central ion conductive pore surrounded by regulatory or sensor domains. All K+ channels operate according to a common principle: upon perception of the stimuli, the peripheral sensors convey the information to the pore that opens or closes in response (gating). Given their common modular architecture, K+ channels tolerate swapping of pore and sensor modules between members of different families without losing functionality [1, 2]. More surprising is the evidence that exogenous sensors found in proteins unrelated to the K+ channel superfamily can be grafted on the pore to control gating. In the past, we have engineered a
functional voltage-gated K⁺ channel by connecting the voltage sensor of CiVSP, a phosphatase found in the ascidian *Ciona intestinalis*, to the pore of a Kcv, a K⁺ channel found in the algae virus PBCV-1 [3]. An exciting perspective is the expansion of signals perceived by K⁺ channels beyond the realm of those already sensed by native channels. Regulation of K⁺ channels by infrared light, magnetic fields or ultrasounds that freely penetrate mammalian tissues, will offer the unprecedented possibility to control cellular functions in freely moving and behaving organisms.

We recently succeeded in engineering BLINK1, a blue light (475 nm)-gated K⁺ channel that can be used to inhibit the electrical activity of neurons in optogenetics [4]. BLINK1 was engineered by connecting the light sensor module LOV2 of *Avena sativa* phototropin [5] to the N-terminus of Kcv (Fig. 1a). Crucial for the project was the development of a yeast-based screening system that allowed fast screening for functional K⁺ channels. We had previously shown that Kcv expression rescues growth of *S. cerevisiae* Δtrak1 Δtrak2 mutants in low external K⁺ [6, 7]. In this project, we were further screening for the ability of the engineered channels to promote differential growth in light and dark. A prototype channel retrieved from the first round of screening was further improved through a random mutagenesis approach that generated a library of mutants. In a second round of screening, about 30,000 clones from the library were transformed in yeast and the colonies

![Fig. 1](image)

**Fig. 1** Schematic representation of the engineering principles of the light-activated K⁺ channel and of the yeast-based screening assay used to identify functional light-driven K⁺ channels. (a) The prototype light-gated K⁺ channel (LK) was engineered by fusing the LOV2-Jα domain (L) to the viral K⁺ channel Kcv (K). K and LK are inserted in the lipid membrane while L is cytosolic. N and C indicate the first and the last aminoacid of each construct. (b) Yeast complementation assay for the detection of functional light-regulated K⁺ channels: the mutant strain SGY1528 does not grow in low external K⁺ concentration (4 mM) when it is transformed with L, neither in the dark nor in light. Expression of the functional K⁺ channel Kcv (K) restores yeast growth, independently of light and dark growth conditions. Expression of LK promotes differential yeast growth in light and dark, indicating light regulation of channel gating.
tested for differential yeast growth in light/dark. Channel variants showing improved complementation ability, compared to the parental channel, were selected for further electrophysiological characterization of the channel currents.

2 Materials

2.1 Coding Sequences for the LOV2 Domain and the Kcv Channel

1. LOV2 domain: amino acids 404–546 of *Avena sativa* Phototropin 1 (NPH1-1) (GenBank: AAC05083.1).

2. Kcv: amino acids 2–94 of *Paramecium bursaria* Chlorella virus 1 potassium ion channel protein (PBCV-1-Kcv) (NP_048599.1).

2.2 Expression Vector

1. The pYES2-Met25 expression vector [6] was used for protein expression in *Saccharomyces cerevisiae*.

2.3 Cloning and Expression Organisms

1. *Escherichia coli* (DH5α) for cloning procedures and plasmid DNA amplification.

2. *Saccharomyces cerevisiae Δtrk1 Δtrk2* mutant (SGY1528) for functional complementation by light-driven K⁺ ion channels.

2.4 *E. coli* Growth Media

1. LB medium: add 10 g Tryptone, 5 g Yeast Extract, 10 g NaCl to 800 mL ultrapure water. Mix and adjust pH to 7.0 with 1 M NaOH. Make up to 1 L with water. Autoclave and store at RT. Before use, add selective antibiotics. For solid LB medium proceed as previous step and add 15 g/L agar. After sterilization allow the medium to cool down, add selective antibiotics, and then pour medium into petri dishes. Seal solid plates with parafilm and store in the dark at +4 °C (see Note 1).

2. Ampicillin: 50 mg/mL stock solution. Weigh 500 mg Ampicillin sodium salt and dissolve in 10 mL ultrapure water. Make aliquots and store in the dark at −20 °C (see Note 2).

3. Gentamycin: 20 mg/mL stock solution. Weigh 200 mg Gentamycin sulfate; prepare a 10 mL solution and aliquot as in the previous step (see Note 2).

2.5 Yeast Growth Media

1. Complete Yeast Medium supplemented with 100 mM KCl (YPDA + 100 KCl): weigh 20 g Peptone, 20 g D-Glucose, 10 g Yeast Extract, 50 mg Adenine hemisulfate, 7.4 g KCl and add to 800 mL ultrapure water. Mix and adjust pH to 5.6 with 1 M HCl; make up to 1 L with water and autoclave. To obtain solid medium, add 20 g Agar (Plant cell culture tested, Sigma-Aldrich) before autoclaving. Pour agar-containing medium into petri dishes while still molten and seal with parafilm when cooled. Store liquid medium at RT and plates at +4 °C.

2. Minimal SD medium without Uracil and supplemented with Adenine and 100 mM KCl (SD-U + 100 KCl): Weigh 26.7 g
Minimal SD base (Clontech), 765 mg CSM dropout Uracil (CSM-Ura; Formedium), 100 mg Adenine hemisulfate, 7.4 g KCl and dissolve in water as described in the previous step. Mix and adjust pH to 5.5 with 1 M KOH. Make up to 1 L with water. To obtain solid medium, add 20 g Agar (Plant cell culture tested, Sigma-Aldrich). Autoclave and store liquid medium at RT, pour agar plates as previously described, and store at +4 °C.

3. Selective medium without Uracil and Methionine, supplemented with Adenine and KCl (SEL-U-M+KCl): weigh 725 mg L-Arginine, 100 mg Adenine hemisulfate, 1.5 g CSM dropout Uracil and Methionine (CSM-Ura-Met; Formedium), 10 g D-Glucose and add to 800 mL of water. Add 2 mL of 500 mM MgSO₄ and 200 μL of 500 mM CaCl₂. Supplement with KCl at the desired final concentration diluting a 2 M stock solution (see Note 3). Mix and adjust pH to 6.0 with Phosphoric acid. Make up to 1 L with water. Autoclave and store liquid medium at RT. Before use, add 1 μL Vitamin solution and 1 μL Trace Elements per each mL of media. For solid medium, add 15 g Low Potassium Agar (Sigma-Aldrich #P9338). After sterilization, allow the medium to cool down and add Trace Elements and Vitamins as previously described; pour medium into petri dishes. Seal solid plates and store at +4 °C in the dark (see Note 4).

4. Vitamin solution: weigh and add in a 50 mL tube 10 mg Biotin, 20 mg D-Panthotene acid calcium salt, 20 mg Nicotinic acid, 20 mg Pyridoxine HCl, 20 mg Thiamine, and 10 mg Myo-inositol. Add ultrapure grade water up to 50 mL, mix and filter the solution through 0.2 μm syringe filter. Store at +4 °C, in the dark (see Note 5).

5. Trace elements: weigh 25 mg Boric acid, 2 mg CuSO₄, 5 mg KI, 25 mg FeCl₃, 20 mg MnSO₄, 45 mg Molybdic acid, 20 mg ZnSO₄, 500 μL HCl 37% and dissolve in 50 mL of water. Sterilize the solution as described in the previous step. Store at +4 °C, in the dark (see Note 5).

2.6 Mutation Strategies

2.6.1 Enzymes and Reagents

1. High-fidelity PCR Amplification System (e.g., Pfu DNA polymerase).
2. QuikChange II XL Site-Directed Mutagenesis Kit.
4. Restriction Enzymes.
5. Commercial PCR purification kit (e.g., QIAquik PCR Purification Kit).
6. Commercial DNA gel extraction kit (e.g., Wizard SV Gel and PCR Purification Kit).
7. Commercial DNA plasmid purification kit (e.g., QIAprep Spin Miniprep Kit).
8. Commercial yeast transformation kit: e.g., the Frozen-EZ yeast transformation II kit (Zymo Research).

2.6.2 **PCR Primers**

Primer characteristics for each application:

1. Overlap extension PCR (OE-PCR): presence of G or C at 5′ end; \( T_m \) between 64 and 66 °C; the overlapping region with the target sequence has to be 15–22 bp long.
2. Terminal Extension PCR (TE-PCR): 5′ extension can be up to 35 bp long; \( T_m \) of region annealing to target sequence has to be between 64 and 66 °C.
3. Site directed mutagenesis PCR (SDM-PCR): primer length should be between 42 and 45 bp; the desired mutation has to be close to the center of the primer sequence; the \( T_m \) (calculated as per QuikChange II XL Site-Directed Mutagenesis Kit) has to be greater than 78 °C.
4. Error-prone PCR (EP-PCR): the primer binding sites are located on the vector sequence at a distance greater than 70 bp from desired restriction sites. \( T_m \) is between 64 and 66 °C. In Error-Prone PCR reactions, the required primers stocks concentration is 250 ng/μL.

2.7 **Light Irradiation**

1. Functionally effective Blue light: Tri-Star Rebel LED (Luxeon #MR-R0500-20T) containing three Royal-Blue LEDs, mounted on 20 mm Tri-Star Saber base: 2730 mW at recommended operating current 700 mA; \( \lambda = 447.5 \) nm (440–460 nm); beam angle 125°.
2. LED Heat Sink: Alpha Heat Sink (Luxeon #N80-20B) with convection thermal resistance rating of 2.85 °C/W and 80 × 80 × 20 mm footprint.
3. LED adhesive tape: thermally conductive, electrically isolating and adhesive Bond-Ply 100 pad (Luxeon #LXT-T-12) cut to fit the 20 mm Luxeon LED’s Tri-Star base. Operational temperature from −30 °C to 120 °C; thermal performance 4.5 °C/W.
4. Functionally ineffective Red light: Tri-Start Rebel LED (Luxeon #MR-H2060-20T) containing three Red–Orange LEDs, mounted on 20 mm Tri-Star Saber base: 366 mW at recommended operating current 700 mA; \( \lambda = 617 \) nm (610–620 nm); beam angle 125°.
5. Safety light: Deep-Orange co-extruded polycarbonate film (Rosco Supergel #22) applied to main room light source: RGB 240:80:0; transmittance ~0% at \( \lambda < 540 \) nm.
2.8 Light/Dark Differential Screening Chamber

The screening is performed in a temperature-controlled chamber where culture plates can be either exposed to homogeneous light irradiation or kept in the dark. Figure 2 shows a schematic drawing of the chamber, from the side view (Fig. 2a) and from the top (Fig. 2b).

1. Homogenous light irradiation is provided as follows: a metal scaffold is built around a 1 m² surface, in order to place 9 LED assemblies 70 cm above cellular cultures (see Note 6). Each LED assembly is composed of a heat sink mounted on the top of a Tri-star LED, through thermal conductive adhesive tape. The thermal sink allows an adequate passive cooling system that ensures long-time LED functionality (Fig. 2a). LED assemblies are mounted on three metal bars to allow

Fig. 2 Growth chamber for yeast-based light/dark complementation assay: (a) side view of the incubation chamber; (b) top view of the workbench. Detailed scheme of LED assembly and LED components is shown at the right side of each panel.
homogenous radiation distribution over the table surface. LED assemblies are grouped in series and groups are in parallel with each other, according to power supply specifications. This connection system allows easy removal of a single LED in case of technical maintenance (Fig. 2b). White panels reflecting irradiated light are placed around the table, in order to homogeneously distribute light over the plate surface. A space between the panels and the table was left to allow air circulation. Moreover, the open conformation of the scaffold ensures homogenous temperature conditions all over the chamber.

2. Dark condition is obtained by means of a dark cardboard and a black cloth covering the plates and allowing air circulation.

### 3 Methods

The selection assay is based on a *S. cerevisiae* mutant Δtrk1 Δtrk2 (SGY1528) that cannot grow in low external K⁺. Expression of heterologous K⁺ transport proteins restores growth in low K⁺. Functional synthetic channels were identified by their ability to complement the yeast growth phenotype in low external K⁺.

#### 3.1 Preparation of *S. cerevisiae* Competent Cells

1. Grow SGY1528 cells on YPDA + 100 mM KCl solid medium up to 3 days at 30 °C.
2. Inoculate a single colony in 3 mL YPDA + 100 mM KCl broth and grow overnight in 30 °C shaking incubator (*see Note 7*).
3. Dilute cell culture 1:50 in YPDA + 100 mM KCl broth and grow in 30 °C shaking incubator until OD₆₀₀ is between 0.8 and 1.0 (*see Note 8*).
4. Centrifuge cells at 500 × g for 4 min, and then discard supernatant.
5. Commercially available kits such as the Frozen-EZ yeast transformation II provide reagents and protocols to obtain competent yeast cells with high transformation efficiency.
6. Aliquot cells in small volumes and store at −80 °C (*see Note 9*).

#### 3.2 Transformation of *S. cerevisiae* Cells

Commercially available kits such as the Frozen-EZ yeast transformation II allow for a high yield in yeast transformation procedures.

1. Follow manufacturer’s instruction to transform cells (*see Note 10*).
2. Plate up to 5 μL cells over 1/3 of a minimal SD medium plate.
3. Incubate plates up to 3–4 days at 30 °C.
Two main strategies were followed to build a synthetic light-driven ion channel: first, a rational approach was pursued, which consisted in connecting the sensor to the pore module in different positions guided by what is known on the gating mechanism of the channel. This was followed by a random mutagenesis approach aiming to improve the properties of a promising candidate retrieved from the previous approach. Rational mutagenesis was performed using a combination of different PCR amplification methods. The standard thermal cycler program in combination with Pfu DNA polymerase is as follows: step 1, 30 s at 95 °C; step 2 (15–35 cycles), 10 s at 95 °C, 20 s at $T_m$, 1 min per kb of amplicon at 72 °C; step 3, 5–7 min at 72 °C; step 4, 7 °C for unlimited time.

**3.3 Rational and Random Mutagenesis Strategies**

**3.3.1 Rational Mutagenesis: Terminal Extension PCR (TE-PCR)**

TE-PCR extends both termini of a target sequence for further sequence manipulation (e.g., restriction enzyme cloning; OE-PCR; motif addiction; see Fig. 3a). To set up PCR parameters, only the primer region annealing to the target sequence has to be taken into account. The reaction is based on a standard PCR protocol based on Pfu DNA polymerase, following the standard reagent concentration and the standard thermal cycler program as per manufacturer’s instructions.

3.1. Generate partially overlapping amplicons through TE-PCR (see Subheading 3.3.1) using a combination of one TE-PCR primer and one OE-PCR primer: the resulting fragments will have a common annealing region.

2. Isolate amplicons running the PCR reaction on 1% Agarose gel and purify the desired fragment with any commercially available DNA gel extraction kit.

3. First OE-PCR reaction: mix within a PCR tube 0.75 U Pfu DNA polymerase, 1× Pfu polymerase buffer, 200 μM dNTPs, 40 ng of the longest and partially overlapping amplicon; 1:1 molar ratio of the shorter overlapping amplicon; make final volume up to 50 μL with ultrapure water.

4. Set up standard PCR protocol, with following modifications: Step 2, 15 cycles and $T_m$ at 55 °C; skip step 3, move directly to step 4 (7 °C).

5. Second OE-PCR reaction: add 0.5 μL of 10 μM primers (forward and reverse) to allow full amplification of the newly generated construct.
6. Set up standard PCR protocol, with following modifications:
   $T_m$ 55 °C.
7. Isolate amplicons running the reaction on 1% agarose gel and purify the desired fragment with a commercially available DNA gel extraction kit.

This procedure introduces amino acid substitutions, insertions, or deletions into a target sequence (see Fig. 3c). For this purpose, commercial products as QuikChange site-directed mutagenesis kit are available.

1. Perform SMD-PCR as per manufacturer’s guide, with minor adaptations: $T_m$ value and elongation time regulation, based on the target sequence.
2. Transform PCR product into *E. coli* cells, extract and purify the obtained plasmid.
3. Verify the presence of the desired mutation by sequencing.

EP-PCR randomly mutates a target region by using an error-prone DNA polymerase. Commercially available kits such as GeneMorph II random mutagenesis are specifically designed for this purpose. EP-PCR primers used in this reaction were designed to perfectly
match the target plasmid sequence, in order to allow subsequent recombination in gap repair (Fig. 4).

1. The following setup was designed to generate 10–15 random point mutations per 1 Kb of target DNA as per instructions of the GeneMorph II random mutagenesis kit. Use 50 ng target DNA (e.g., 510 ng of 5.5 Kb plasmid containing a 560 bp long target sequence) and run the amplification reaction as follows: step 1, 95 °C for 2 min; step 2, 95 °C for 30 s, 64 °C for 30 s, 72 °C for 1 per 1 Kb amplified, repeated for 30 cycles; step 3, 72 °C for 10 min.

2. Use 2 μL of the PCR product as a template for another EP-PCR. Repeat this step twice.

3. Isolate amplicons running the reaction on 1% agarose gel and purify the desired fragment with a commercially available DNA gel extraction kit.

4. Proceed to yeast transformation and gap repair.

3.3.5 Random Mutagenesis: Gap Repair

Gap repair allows the generation of libraries of a randomly mutated clone within the same non-mutagenized vector background (see Fig. 4). High efficiency yeast transformation is recommended. Gap repair transformation requires an open plasmid, with terminal regions annealing to the terminal regions of the mutagenized fragment that has to be inserted (see Note 11).

Fig. 4 Gap repair-based library preparation in yeast. RE stands for restriction site; P stands for EP-PCR primer
1. To prepare the vector for a gap repair reaction, prepare a restriction digest of the plasmid by mixing in a tube 10 μg of circular pYES2-Met25 expression plasmid, 50 U of each restriction enzyme, 1× reaction buffer according to restriction enzyme manufacturer’s instructions. Make final volume up to 50 μL with water (see Note 12).

2. Incubate at the appropriate temperature following restriction enzyme manufacturer’s instructions for more than 1 h 30 min.

3. Isolate digested fragments running the reaction on 1% Agarose gel and purify the linearized plasmid with any commercially available DNA gel extraction kit.

4. To perform a gap repair reaction, prepare into a 2 mL tube a mixture of 500 ng EP-PCR randomly mutated amplicons with 1 μg of cut yeast expression vector.

5. Transform the mixture into 50 μL SGY1528 competent cells using any commercially available yeast transformation kit, providing high transforming rate. When transforming yeast cells with Frozen-EZ yeast transformation II kit, eight transformation reactions are enough to plate cells onto 36 Ø 150 mm petri dishes (see Note 13).

6. Dilute transformed cells with MilliQ water up to 6.5 mL and mix.

7. Plate 180 μL of diluted cells on 36 Ø 150 mm petri dishes containing SD-U + 100 mM KCl

8. Number the plates and provide each with a mark reporting the orientation for the following replica plating step.

9. Incubate at 30 °C in the dark up to 4 days.

### 3.4 Light-Driven Functional Complementation

The yeast mutant strain SGY1528 (see Note 14) is suitable for functional complementation tests of light-gated synthetic potassium channels inserted in pYES2-Met25 vector. Functional complementation is performed on SEL-U-M medium supplemented with KCl at a selective concentration. The differential growth between light and dark exposed colonies determines which clones are functional. Light-driven functional complementation can either be performed in low throughput by means of a drop-test assay or in high-throughput.

#### 3.4.1 Low-Throughput Screening by Drop-Test

The drop test method allows fine screening of tens of colonies on a single plate by functional complementation. Tenfold serial dilutions of liquid cell culture are spotted onto solid selective medium and grown under light irradiation or in the dark. The differential growth rate of the same colony, at the same dilution, at the two different conditions, reflects a different ability in complementing the defective yeast phenotype.
1. Pick up the desired colony from solid nonselective medium (SD-U + A + 100 mM KCl) and inoculate it in 3 mL of liquid nonselective medium. Grow cells overnight in 30 °C shaking incubator in the dark.

2. Centrifuge 2 mL of the overnight cell culture at $2700 \times g$ for 2 min and discard supernatant.

3. Resuspend pellet with 2 mL sterile water to wash cells.

4. Centrifuge cells at $2700 \times g$ for 2 min and discard the supernatant.

5. Add 1 mL sterile water and resuspend cells.

6. In a different tube, transfer 50 μL resuspended cells and add 950 μL water (1:20 dilution).

7. Measure OD$_{600}$ of the 1:20 dilution and calculate the OD$_{600}$ of the resuspended cells.

8. Dilute the resuspended cells to OD$_{600} = 0.8$ with water.

9. Prepare tenfold serial dilutions in 2 mL tubes: transfer 100 μL cell from OD$_{600} = 0.8$ preparation in a new tube and add 900 μL water (1:10 dilution). After mixing, repeat the same procedure starting from 1:10 dilution to obtain 1:100 dilution and from the 1:100 dilution to obtain 1:1000 dilution.

10. Spot 7 μL drops of each dilution onto plates of SEL-U-M medium, supplemented with KCl at the desired selective concentration. Prepare each plate in duplicate to test simultaneously two functional conditions.

11. Seal plates with parafilm and incubate inverted for 3 days at 30 °C under light or dark conditions in the chamber described in Subheading 2.8.

12. To purify the plasmid from selected colonies, pick the desired colony from solid nonselective medium and inoculate in 3 mL of liquid nonselective medium.


14. Purify yeast plasmid with any commercially available plasmid extraction kit (see Note 15).

15. In a new 2 mL tube transform 10 μL DH5α E. coli competent cells with 1 μL of the yeast extracted plasmid.

16. Plate 90 μL transformed cells on LB agar plate supplemented with selective antibiotics and incubate overnight at 37 °C.

17. Inoculate colonies in a 50 mL tube containing 7 mL LB broth, supplemented with selective antibiotics; grow cells overnight at 37 °C in a shaking incubator.

18. Proceed as per any commercially available DNA plasmid purification kit workflow.
This step allows the replication of the same colony over two plates with selective medium. The replica plates are then incubated under different growing conditions and differentially growing colonies identified and selected for further analysis.

1. With a sterile velvet cloth, replicate colonies grown onto each nonselective plate onto two different Ø150 mm plates containing SEL-U-M+A supplemented with the desired selective KCl concentration.

2. Seal and number the two selective plates according to mother plate number and selection condition and mark each to report the orientation as for the nonselective plate (see Note 16).

3. Incubate plates at 30 °C in the growth chamber. One subset of plates must be placed under effective light irradiation, the second subset must be grown in the dark (see Note 17).

4. After 3 days, light-irradiated and dark grown plates are compared to find differentially grown yeast colonies (see Note 18). Colonies that have been identified by differential growth in replica plating must be validated to eliminate false positives.

5. To validate the phenotype, identify the colonies that show differential growth between light and dark conditions following replica plating.

6. Inoculate the selected colonies in liquid nonselective medium (SD-U+A + 100 mM KCl) and purify plasmid from the liquid growth.

7. Transform S. cerevisiae competent cells and confirm the phenotype (dark/light differential growth) by drop test.

4 Notes

1. Cover plates with aluminum foil to keep LB agar supplemented with ampicillin in the dark. Ampicillin is light sensitive and degrades under high temperature and light exposure. Moreover, storing plates in an inverted position (with their lid down) allows the condensed water to accumulate on the plastic lid instead of on the medium.

2. Aliquot 1 mL of antibiotics stock solution into 1.5 mL Eppendorf tubes. When preparing aliquots of Ampicillin, cover tube with aluminum foil to keep solution in the dark. Stock solutions are 1000×, when supplemented to LB media add 1 μL antibiotic to each mL of medium.

3. Dissolve 14.9 g KCl in 100 mL ultrapure water to have a 2 M stock solution. Add 250 μL, 500 μL, or 2 μL of the obtained stock solution to supplement 1 L medium with 0.5 mM,
1 mM, or 4 mM KCl, respectively. Weigh 7.4 g of KCl powder to prepare 1 L of 100 mM KCl containing medium.

4. Low Potassium Agar from Sigma-Aldrich (#P9338) allows the preparation of media supplemented with very low potassium concentrations (e.g., 0.5 mM).

5. Filtering vitamins and trace elements through 0.2 μm syringe filters ensure sterilization of the solution. Filter under laminar airflow hood directly into sterile 50 mL falcon tubes. Cover the tubes with aluminum foil to protect from light irradiation. Open tubes only in sterile environment.

6. Nine assemblies of Royal-blue Tri-Star Rebel LED, at 2730 mW each, are able to irradiate 1 sq. m table at an average light distribution (λ = 447 nm) of 160–230 μW/cm².

7. Inoculate 3 mL of cell culture in a 15 mL tube.

8. This process typically requires about 4 h when applying 1:50 dilution and about 3 h when working with 1:25 dilution.

9. Allow cells to cool down slowly by placing tubes into a polystyrene box, within the −80 °C refrigerator. This procedure allows gradual freezing of the cells and improves their transformation efficiency.

10. Frozen-EZ Yeast Transformation II kit workflow adaptation: prepare in a 2 mL tube a mixture of 1 μL (>200 ng/ul) expression vector, 5 μL SGY1528 competent cells, 50 μL EZ3 solution. Incubate cells at 30 °C for 45 min and mix the solution every 15 min.

11. The product of three consecutive EP-PCR reactions is typically enough to perform a gap repair transformation plated on up to 18 petri dishes (Ø150 mm).

12. Cut the plasmid by restriction digestion leaving sticky and blunt end opposite each other to avoid plasmid re-circularization. Lacking restriction sites can be inserted by SDM-PCR.

13. This number of plates occupies roughly 1m² of an irradiated area during functional complementation tests. Depending on the cell density per plate, it can provide about 20,000 growing colonies per functional screening.

14. SGY1528 is a K⁺ uptake defective yeast strain: Mata Ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 trk1::HIS3 trk2::TRP1 [8].

15. When using Zymoprep Yeast plasmid Miniprep II kit a typical purification from 2 mL cell culture has an average yield between 20 and 120 ng/μL. However, the purity of this extraction is not suitable for sequencing, for this reason a further cloning step in E. coli cells and subsequent plasmid purification are necessary.
16. Seal plates with parafilm to protect cell culture from contamination and drying. Orientation mark will help to identify the original colony on the mother plate.

17. Effective irradiation to allow yeast functional complementation of BLINK light-driven potassium channel [5]: $210 \pm 20 \mu W/cm^2$ using Royal Blue Light ($\lambda = 447$ nm).

18. The average yield of differentially growing colonies obtained is up to 20 per plate.

Acknowledgments

This work was supported by MAE (Ministero Affari Esteri) grant n. 01467532013-06-27 and ERC 2015 AdG n. 695078 noMAGIC to A.M.

References


Primer-Aided Truncation for the Creation of Hybrid Proteins

Robert Stabel, Birthe Stüven, Robert Ohlendorf, and Andreas Möglich

Abstract

Proteins frequently display modular architecture with several domains and segments connected by linkers. Proper protein functionality hinges on finely orchestrated interactions among these constituent elements. The underlying modularity lends itself to the engineering of hybrid proteins via modular rewiring; novel properties can thus be obtained, provided the linkers connecting the individual elements are conducive to productive interactions. As a corollary, the process of protein engineering often encompasses the generation and screening of multiple linker variants. To aid these steps, we devised the PATCHY method (primer-aided truncation for the creation of hybrid proteins) to readily generate hybrid gene libraries of predefined composition. We applied PATCHY to the mechanistic characterization of hybrid receptors that possess blue-light-regulated histidine kinase activity. Comprehensive sampling of linker composition revealed that catalytic activity and response to light are primarily functions of linker length. Variants with linkers of $7n$ residues mostly have light-repressed activity but those with $7n + 1$ residues mostly have inverted, light-induced activity. We further probed linker length in the context of single residue exchanges that also lead to an inversion of the signal response. As in the original context, activity is only observed for certain periodic linker lengths. Taken together, these results provide mechanistic insight into signaling strategies employed by sensory photoreceptors and sensor histidine kinases. PATCHY represents an adequate and facile method to efficiently generate and probe hybrid gene libraries and to thereby identify key determinants for proper function.

Key words DNA library, Hybrid gene, Light–oxygen–voltage, Protein engineering, Sensor histidine kinase, Sensory photoreceptor, Signal transduction

1 Introduction

Proteins of diverse biological roles consist of multiple modules, often corresponding to distinct protein domains, that are connected through linker segments [1]. Proper function generally depends on productive and precisely calibrated interactions among these modules, as facilitated by the intervening linkers [2]. Depending upon protein context, linkers differ in a number of parameters including length, sequence, surface charge, structure and flexibility. Given their eminent role in mediating interactions
between protein modules, linkers are also decisive in the engineering of proteins with novel function. In particular, desired traits can often be obtained by covalently connecting different modules via suitable linkers and by thus generating hybrid (or, chimeric) proteins. Some cases call for flexible linkers that bring modules into spatial proximity, but allow reorientation and relative movements; in other cases, rigid connectors are required to fix modules at discrete distances and defined angular orientations. Additional considerations in linker design include proper folding, intracellular trafficking, chemical and biological stability of resultant hybrid proteins [3]. Structural information, where available, and multiple sequence alignments often provide valuable clues as to which linker suffices for a specific engineering purpose. Nonetheless, a priori it is difficult to select among many possible linker variants the one(s) best suited for meeting the above criteria. As a corollary, often multiple linkers are constructed and empirically tested for best performance, which incurs considerable expenditure of time and effort.

These concepts are exemplified in signal receptors and their engineering. In particular, sensory photoreceptors constitute the group of signal receptors that impart sensation of light [4]. Photosensor modules absorb light of appropriate quality and in response modulate the biological activity of effector (or, output) modules. Across different photoreceptor classes, the linkers connecting these modules are often of α-helical or coiled-coil conformation; modifications to the linker as confined as exchange, addition or deletion of single residues can profoundly affect receptor activity and regulation [5–7]. As a case in point, we constructed the photoreceptor YF1 by recombining the blue-light-responsive light–oxygen–voltage (LOV) photosensor module of Bacillus subtilis YtvA (BsYtvA) with the Bradyrhizobium japonicum FixL (BjFixL) histidine kinase effector [8] (Fig. 1). The crystal structure of homodimeric YF1 in its dark-adapted state showed its two LOV photosensor modules to be connected to the effector module via a parallel coiled-coil linker, denoted Jα [9]. Length variations of Jα revealed heptad (i.e., seven-residue) periodicities of catalytic activity and regulation by light [8]. The original YF1 construct derived almost its entire linker from the parental protein BjFixL, but equally one could have used the corresponding linker of the other parental protein BsYtvA, or hybrids of both linkers. Given that the parental linkers are 23 and 27 residues long, there are \((23 + 1) \cdot (27 + 1) = 672\) possible combinations for connecting BsYtvA and BjFixL if one restricts hybrid fusions to these linker segments (Fig. 1). Although comprehensive interrogation of all linker combinations could provide invaluable insight into signaling mechanisms and engineering principles, manual construction and separate testing of each individual variant is prohibitively cumbersome. We hence sought to assess all possible linker variants in parallel. To this
end, we established the PATCHY (primer-aided truncation for the creation of hybrid proteins) method for the efficient generation of hybrid gene libraries with defined composition [10].

Prior to this, alternate strategies for creating hybrid gene libraries had been proposed and successfully applied for the construction of novel signal receptors, in particular the methods SHIPREC (sequence homology-independent protein recombination) and ITCHY (incremental truncation for the creation of hybrid proteins) [11–14]. However, for fragmentation of the parental genes, these methods rely on endonucleolytic or exonucleolytic DNA cleavage, both of which are challenging to titrate. Moreover, it is inherently difficult to confine hybrid fusions to specific regions of the parental genes such as their linker regions. Taken together, these limitations have hampered the wider application of either method [15]. In contrast, PATCHY circumvents problems arising from nucleolytic DNA fragmentation by means of PCR amplification with sets of staggered primers to truncate the parental genes and thus produce linker libraries of defined composition (Fig. 2). As detailed in this chapter (see Subheading 3.1), PATCHY starts from a template construct in which two parental gene fragments are concatenated such that they are separated by a nucleotide stretch that introduces a frameshift and a unique restriction site. The template is then amplified in a one-pot PCR reaction with sets of forward and reverse primers. Both primer sets are staggered in

![Fig. 1](image_url)

(a) Domain architecture of *BsYtvA*, *BjFixL*, and YF1. The hybrid YF1 consists of the *BsYtvA* LOV domain fused to the *BjFixL* histidine kinase, where almost the entire linker derives from *BjFixL*. (b) The crystal structure (PDB entry 4GCZ [9]) of homodimeric YF1 in its dark-adapted state shows the coiled-coil linker between the LOV photosensor and histidine kinase effector modules. (c) Schematic of the possible linker combinations for hybrids between the *BsYtvA* photosensor and the *BjFixL* effector modules.

Generation of Hybrid Gene Libraries
increments of three nucleotides such as to truncate the template construct in the register of nucleotide triplets which corresponds to single amino acids at the protein level. The resultant linearized plasmid fragments are phosphorylated and religated to produce a library of circular plasmids. Optionally, remnants of the template construct can be depleted from the library by restriction digest. Plasmid libraries are then transformed and screened for favorable phenotypes (see Subheading 3.2).

We applied PATCHY to the fusion of the $B_{YtvA}$ LOV photoreceptor and the $B_{jFixL}$ effector module that previously yielded the blue-light-repressed histidine kinase YF1 [8], see above. To facilitate library screening, hybrid gene libraries were constructed in the context of the pDusk-DsRed reporter plasmid [16] which encodes YF1, the cognate response regulator $B_{jFixJ}$ and the DsRed fluorescent reporter under control of the $B_{jFixK2}$ promotor. In the dark, YF1 phosphorylates $B_{jFixJ}$ which in turn binds to the $B_{jFixK2}$ promotor and thereby ramps up the expression of the reporter gene. Under blue light, YF1 acts as a net phosphatase on $B_{jFixJ}$ which in turn results in about 10–15 fold decreased DsRed expression compared to in the dark (Fig. 3). Fluorescence-based screening (see Subheading 3.3.1) of the PATCHY library revealed that photoreceptor activity and regulation by light are by and large determined by linker length with a pronounced heptad (seven-residue) periodicity. Insertion of single residues sufficed to invert the response to blue light. An inversion of the light response of YF1 can not only be effected via linker-length modifications but
also via the introduction of the single amino-acid exchanges D21V or H22P within the LOV photosensor module [9, 17]. To obtain a better mechanistic understanding of signal transduction and modulation of the signal response, we applied PATCHY to the systematic interrogation of linker determinants in the background of either residue exchange (see Subheading 3.3.2). Similar to the results for YF1, activity was only observed in discrete heptad registers, i.e., 7n for H22P, as well as 7n and 7n + 1 for D21V. In one variant, elongation of the linker by three residues counteracted the inversion of the signal response originally caused by the D21V exchange.

2 Materials

2.1 Chemicals and Consumables

1. 5× HF Buffer for PCR.
2. 2 U/μL Phusion High-Fidelity DNA Polymerase.
3. 10 mM dNTPs each.
4. 10 μM total forward primer pool for PATCHY.
5. 10 μM total reverse primer pool for PATCHY.
6. 10 U/μL Nhel restriction enzyme.
7. 10 U/μL DpnI restriction enzyme.
8. 10 U/μL T4 Polynucleotide Kinase.
9. 50% PEG-4000 (w/v).
10. 0.5 mM ATP.
11. 30 U/μL T4 DNA Ligase.
12. 50× TAE buffer: 2 M Tris-Acetate, 50 mM EDTA, pH 8.5.
13. 1% Agarose in 1× TAE Buffer.
14. LB/Kanamycin medium: 10 g peptone, 5 g yeast extract and 10 g NaCl in 1 L dH2O supplemented with 50 mg/L Kanamycin.

Fig. 3 The pDusk-DsRed reporter plasmid. In the absence of blue light, YF1 phosphorylates the response regulator BjFixJ which binds to the BjFixK2 promotor and thereby upregulates reporter gene expression. Blue light inhibits expression of DsRed by around 10–15-fold as it converts YF1 to a net phosphatase [16].
15. LB/Kanamycin agar plates: LB/Kanamycin medium supplemented with 20 g/L agar.
16. Chemically or electrocompetent *Escherichia coli* cells.
17. 50% glycerol (w/v).
18. NucleoSpin Extract Kit for purification of PCR products (Macherey Nagel).
19. 96-deep-well microtiter plates for growth of bacterial clones (e.g., Axygen).
20. Clear 96-well microtiter plates for absorption measurements (e.g., Nunc).
21. Black 96-well microtiter plates for fluorescence measurements (e.g., Nunc).

### 2.2 Lab Equipment

1. Gradient thermal cycler for PCR amplification (e.g., Thermal Cycler S1000, Bio-Rad).
2. Electrophoresis chamber (e.g., Wide Mini-Sub Cell GT Cell, Bio-Rad).
3. Nanodrop spectrophotometer (e.g., Spark 10 M with Nanoquant plate, Tecan).
4. Microplate reader with absorption and fluorescence optics (e.g., Infinite M200 pro, Tecan).
5. Two incubators (e.g., Incu Line IL10, VWR).
6. Two shakers for microtiter plates (e.g., PMS-1000i, Grant).
7. Blue-light LED array, custom built, 10 × 8 LEDs of 470 ± 10 nm (Winger Electronics).
8. Lamp power meter (model 842-PE, Newport) with silicon photo detector (model 918D–UV-OD3, Newport).
10. *Optional*: Flow cytometer with sort functionality (e.g., S3e, Bio-Rad).

### 2.3 Software

1. Python script for design of staggered primers: [https://github.com/vrylr/PATCHY.git](https://github.com/vrylr/PATCHY.git)

### 3 Methods

The generation and analysis of PATCHY hybrid gene libraries are described in Subheadings 3.1 and 3.2, respectively. To illustrate the general method and individual steps, by way of example we repeatedly refer to a recent study in which we applied PATCHY to recombine the *B*YtVA photosensor with the *B*xFxL effector module (see Subheading 1) [10]. Key results from these experiments are
summarized as a case study in Subheading 3.3.1. The application of PATCHY to the same two modules but with either the residue exchange D21V or H22P within the photosensor is covered as a second case study in Subheading 3.3.2.

3.1 Generation of PATCHY Libraries

1. Principal considerations: In a one-pot reaction, PATCHY generates defined libraries of hybrid genes with single fusion sites between an upstream gene A and a downstream gene B. Individual library members differ in which fragments of the parental genes A and B they comprise. As a first step, the desired composition of the library is specified by deciding which set of gene fragments of A and B are to be recombined. In the case study, we constructed hybrid variants that connect the entire BsYtvA LOV photosensor module including a variable number \( i \) of residues of its C-terminal linker (\( i \in \{0, 1, ..., 23\} \)) to the entire BjFixL effector module including a variable number \( j \) of residues of its N-terminal linker (\( j \in \{0, 1, ..., 27\} \)) (see Figs. 1 and 2).

2. PATCHY template construct: PATCHY libraries are obtained by PCR amplification of a template construct with staggered primer sets. The template construct contains a hybrid fusion between the longest desired fragments of each of genes A and B. A spacer sequence that harbors a unique restriction site and deliberately introduces a frameshift is inserted between the two gene fragments. In the case study, the BsYtvA LOV photosensor including its entire C-terminal linker (\( i = 23 \)) was connected to the BjFixL effector including its entire N-terminal linker (\( j = 27 \)). The intervening spacer sequence encoded a frameshift and a unique NheI restriction site. To facilitate subsequent analysis and screening of PATCHY libraries (see Subheading 3.2), the template construct was assembled in the background of the pDusk-DsRed reporter plasmid [16] which affords facile fluorescence readout of receptor activity, see Subheading 1 and Fig. 3.

3. Primer design: Sets of forward and reverse oligonucleotide primers are devised such that during the PATCHY PCR reaction they lead to incremental truncations of the template construct at the 3’ end of the upstream gene A and at the 5’ end of the downstream gene B (see Fig. 2). For this purpose, both sets are staggered in increments of base triplets, corresponding to single amino acids at the protein level. To facilitate the PATCHY PCR reaction, the primers should be designed with largely uniform melting temperatures (\( T_m \)), ideally all within \( \pm 1^\circ \text{C} \). Primers can either be designed manually or in automated manner with a Python script (see Note 1). All forward primers are pooled at equimolar ratios, and the total concentration is adjusted to 10 \( \mu \text{M} \); likewise, a pool of all reverse primers at 10 \( \mu \text{M} \) total concentration is prepared (see Note 2).
4. PATCHY PCR reaction: The PCR reaction mixture is prepared according to Table 1, and PCR amplification is conducted as described in Table 2 where the annealing temperature is set at $T_m - 5 \, ^\circ C$. Analysis of the reaction products by agarose gel electrophoresis should show one dominant or even a single DNA band of the expected size. For $n$ forward and $m$ reverse primers, the PCR product should theoretically comprise $n \cdot m$ different, incrementally truncated, linear PCR fragments.

5. Workup of PATCHY PCR reaction: The product of the PCR reaction is purified using standard molecular biology kits (e.g., NucleoSpin Gel and PCR clean up, Macherey-Nagel). To achieve high DNA concentration for subsequent steps, the PCR product should be eluted with 30 μL ddH$_2$O or elution buffer. The concentration is determined spectrophotometrically or by analysis via gel electrophoresis and comparison to a standard of known concentration. Ideally, one should have a total of at least 1–2 μg DNA at this stage (see Note 3).
6. **Optional**: To deplete the template construct from the reaction mixture, the PCR product is alternatively purified via agarose gel extraction. Depending upon the plasmid size, the circular, negatively supercoiled template can show different electrophoretic mobility on agarose gels from the desired linear PCR products. It is hence often possible to separate the original template construct from the PATCHY library (see Note 4).

7. Phosphorylation of PCR product: The PCR product is phosphorylated at the 5′ end by polynucleotide kinase according to Table 3. The reaction mixture is incubated at 37 °C for 30 min.

8. Circularization of linear PCR product: Ligation of the phosphorylated, linear PCR products yields circular plasmids and is carried out in the same reaction solution as phosphorylation. The reaction mix is allowed to cool to 22 °C before T4 DNA ligase, PEG-4000, and ATP (optional) are added (Table 4). The reaction mix is then incubated at 22 °C for 1 h (see Note 5).

9. **Optional**—Depletion of template construct from PATCHY library: The original template construct is selectively depleted from the PATCHY library via restriction digest with the enzyme that cuts at the deliberately introduced unique site within the spacer sequence of the template (see step 2). For the example case of the PATCHY library between the *BbYtvA* LOV sensor and the *BjFixL* effector, 1.5 μL *NheI* was added, followed by incubation at 37 °C for 30 min. To prevent recircularization, 1.5 μL of the phosphatase FastAP was added and

---

**Table 3**

**Phosphorylation of linear DNA fragments**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10× T4 DNA ligase buffer</td>
<td>3.5 μL</td>
</tr>
<tr>
<td>Linear PCR fragment</td>
<td>30 μL (50 pg–1 μg)</td>
</tr>
<tr>
<td>10 U/μL T4 polynucleotide kinase</td>
<td>2 μL</td>
</tr>
</tbody>
</table>

**Table 4**

**Ligation of phosphorylated linear DNA fragments**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphorylation reaction mix</td>
<td>35 μL</td>
</tr>
<tr>
<td>50% PEG-4000 (w/v)</td>
<td>4 μL</td>
</tr>
<tr>
<td>0.5 mM ATP (optional)</td>
<td>0.5 μL</td>
</tr>
<tr>
<td>30 U/μL T4 DNA ligase</td>
<td>1 μL</td>
</tr>
</tbody>
</table>
incubation continued for 15 min at 37 °C. Enzymes were inactivated by incubation at 75 °C for 15 min (see Note 4).

10. Optional – Depletion of template construct from PATCHY library: Alternatively, the template construct can selectively be digested with DpnI which cleaves DNA in methylation-dependent manner (see Note 4).

11. Transformation of PATCHY library: The PATCHY library can directly be transformed into *Escherichia coli* cells (e.g., DH10b). For example, 5–10 μL of the reaction mix were transformed into chemically competent cells, or up to 1.5 μL were transformed into electrocompetent cells. After plating on LB agar and overnight incubation at 37 °C, the library can either be processed right away (see Subheading 3.2) or stored at −80 °C as a glycerol stock.

### 3.2 Analysis of PATCHY Libraries

1. Sequence analysis of the naïve PATCHY library: To assess the quality of the PATCHY library, a number of individual clones can be analyzed by Sanger sequencing. In case the original template construct is found with significant frequency, extra efforts should be taken to deplete it from the library (see Subheading 3.1, steps 6, 9, and 10 as well as Note 4).

2. Optional: Alternatively, the library can be analyzed by next-generation sequencing (NGS). In the case study [10], the entire PATCHY plasmid library was fragmented by ultrasound and sequenced on an Illumina platform with paired-end 150-base-pair reads. Although most reads covered the invariant plasmid backbone, the low cost of NGS still allowed to get good coverage of the variant parts of the plasmids. We obtained approximately 5500 paired-end reads covering the linker region, corresponding to an ~8-fold oversampling of the expected 672 different hybrid genes (Fig. 4). Out of these 672 variants, 578 could be detected by NGS. The data also indicated that each forward and reverse primer had been used in the PATCHY PCR reaction albeit to varying extents.

3. Optional: In case the sequence analysis indicates a nonrandom, biased distribution of expected constructs in the PATCHY library, one may rerun the PATCHY PCR reaction with adjusted relative primer concentrations (see Subheading 3.1, steps 3 and 4 as well as Note 2).

4. Optional: To further diversify the PATCHY library, it can be subjected to an error-prone PCR reaction (see Note 6).

5. Isolation of hybrid variants from the PATCHY library: The identification of hybrid variants with desirable traits in the PATCHY library is obviously governed by the assays available to read out activity. In principle, individual clones can be isolated and analyzed separately, but ideally functionality can
directly be assessed with high throughput and in parallel in vivo, i.e., without laborious intermediate steps such as protein purification. Efficient in vivo assays largely fall into two clades: first, screening assays are based on a readily measurable, often colorimetric readout such as fluorescence; second, selection systems link functionality of the hybrid variants to cell survival or proliferation. In the example case of the chimerae between BsYtvA LOV and BjFixL histidine kinase, we resorted to the pDusk-DsRed screening system which links hybrid receptor activity to expression of a DsRed fluorescent reporter gene \[10, 16\]. Transformed cells are cultured at 37 °C on plate in darkness or under constant blue light illumination (470 nm, 40 μW/cm²). Clones harboring functional hybrid variants that expressed DsRed under the given incubation condition are visually identified via observation of fluorescence through a 550-nm long-pass filter goggle. As detailed in Subheading 3.1 step 1, the template construct is incapacitated by a deliberate frameshift between the two gene fragments recombined via PATCHY. Screening for function is thus well suited to further discriminate against the template construct (see Note 4).

6. Optional: As an alternative screening method, fluorescence-activated cell sorting can be used, provided function of hybrids is linked to a fluorescent readout \[16\] (see Note 7).

7. Storage of isolated hybrid variants: Clones with desirable phenotypes identified in the previous two steps are grown at 37 °C to stationary phase. To this end, single wells of a 96-deep-well microtiter plate are filled with 600 μL LB medium supplemented with the required antibiotic and are inoculated with a
single clone. For long-term storage, 200 μL of the saturated cultures is mixed with 200 μL 50% (w/v) glycerol and frozen at −80 °C.

8. Quantitative characterization of isolated hybrid variants: Using suitable functional assays, the biological activity of the hybrid variants isolated in Subheading 3.2 step 5 should be characterized in more detail. In the example case, replicates of each clone were inoculated in 5 mL LB/Kanamycin and grown at 37 °C for 16 h in darkness or constant blue light (470 nm, 100 μW/cm²) (see Note 8). Saturated cultures were diluted tenfold in ddH₂O, and OD₆₀₀ was measured in clear microtiter plates with a microplate reader. Samples were further diluted fivefold in ddH₂O, and DsRed fluorescence was measured in black microtiter plates with settings of 554 ± 9 nm for excitation and 591 ± 20 nm for emission, respectively. Fluorescence readings were normalized to the OD₆₀₀ of the corresponding sample.

9. Sequence analysis of isolated hybrid variants: Clones that showed desirable properties in the previous step are analyzed by Sanger DNA sequencing (see Note 9).

3.3 PATCHY Case Studies

3.3.1 Linker Libraries of YF1

We applied PATCHY to better characterize the underlying design principles and signal-transduction mechanisms in the chimeric photoreceptor YF1 [8]. The original YF1 receptor originated from the fusion of the blue light-sensitive light–oxygen–voltage photosensor module of BșYtvA to the effector module of the BļFixL histidine kinase. Notably, in YF1 the linker between these modules essentially derived from the BļFixL parental protein, i.e., i = 3 and j = 25 (see Subheading 3.1 step 1). Sparse sampling of linker composition had earlier identified linker length as the main determinant for activity and regulation of the hybrid receptors [8]. A seven-residue periodicity of the dependence of activity and regulatory properties on linker length resulted from the coiled-coil conformation of the linker, as later evidenced in the high-resolution structure of dark-adapted YF1 [9]. Notably, only a tiny fraction of many conceivable hybrid variants were studied at this point [8].

Despite these biochemical and structural data, the mechanism by which signals are transduced from the LOV photosensor to the histidine kinase effector remained unclear. We reasoned that comprehensive interrogation of linker sequence space could yield additional mechanistic insight. The linkers between the respective sensor and effector modules in the parental receptors BșYtvA and BļFixL comprise 23 and 27 residues, respectively. Provided that fusions between the parental proteins are restricted to these linker regions, there are (23 + 1) • (27 + 1) = 672 different ways to recombine the BșYtvA LOV photosensor with the BļFixL effector (see Figs. 1 and 2). Using PATCHY as described in Subheading 3.1, we generated a construct library that theoretically contains
these 672 linker variants. Library construction was performed in the context of the pDusk-DsRed reporter plasmid [16] that affords rapid screening for functional hybrid variants as described in Subheading 3.2 (see Fig. 3).

Key results from the analysis of the PATCHY library are summarized in the following and in Fig. 5a. Strikingly, light-regulated activity in \( BsYtvA-Bj\)FixL hybrids is supported by quite different linker lengths ranging from 4 to 50 residues, whereas the original YF1 possesses a linker of 28 residues. Assuming a continuous \( \alpha \)-helical coiled-coil structure of the linker elements, this translates into distances of separation of \( \sim 6-75 \) Å between LOV photosensor and effector modules. In addition to hybrids with light-repressed activity as in YF1, several variants showed an inverse response to light, i.e., they had light-induced activity. Hybrid variants with alike responses to light usually displayed a pronounced

---

**Fig. 5** Linker-length analysis of light-regulated \( BsYtvA-Bj\)FixL hybrid variants. (a) For YF1, light-repressed variants (filled circles) showed periodic linker lengths of \( 7n \) or \( 7n + 5 \) residues but light-activated variants (open circles) mostly occurred at linker lengths of \( 7n + 1 \) residues. (b) Light-activated D21V variants had linkers of \( 7n \) or \( 7n + 1 \) residues. Two variants with linkers belonging to the \( 7n \) and \( 7n + 3 \) classes showed inverted, light-repressed response. (c) For H22P, only light-activated variants were identified that predominantly possessed linkers of \( 7n \) residues.
heptad periodicity in their linker lengths. Whereas light-repressed hybrid variants predominantly possessed linkers of \(7n\) and \(7n + 5\) residues, those with light-induced activity mostly had linkers of \(7n + 1\) residues \((n \in \{1, \ldots, 7\})\). The heptad dependence on linker length can be accounted for by the coiled-coil nature of the linker. Sequence analysis of a large group of related, natural sensor histidine kinases revealed their sensor and effector modules to be mainly connected by coiled-coil linkers of length \(7n\) and \(7n + 5\) residues [10].

The difference in the preferred linker lengths for light-repressed and light-activated hybrid variants provides clues as to the mechanistic bases for signal transduction in YF1. Notably, the registers \(7n\) and \(7n + 1\) differ by insertion of a single residue; within a canonical coiled coil such an insertion induces an angular shift of around 100° [18]. We thus propose that signal transduction in YF1 is predicated on angular reorientation of the LOV photosensor relative to the histidine kinase effector modules [8, 9, 19, 20]. This type of angular reorientation could also be brought about by a change in the degree of supercoiling of the linker. We recently investigated light-dependent signal transduction in YF1 by electron paramagnetic resonance (EPR) spectroscopy and X-ray solution scattering [19, 20]. Indeed, the two independent and complementary approaches identify light-induced left-handed supercoiling as the mode for signal transduction and thereby provide a structural rationale for the PATCHY data.

The inversion of the response to light in YF1 cannot only be achieved by linker variations but also by the replacement of certain residues within the LOV photosensor units [9, 17]. To investigate the basis for this profound effect on signal transduction, we repeated the PATCHY analysis of the \(BsYtvA-BjFixL\) chimerae in the background of either the D21V or H22P exchange, both of which induce inversion of the light response in YF1.

In case of D21V, hybrid variants with the same qualitative signal response as the original YF1 D21V, i.e., with an increase in activity upon blue-light exposure, fell into two distinct classes with linkers of \(7n\) and \(7n + 1\) residues, respectively (Fig. 5b). The preference for these classes is indicative of the coiled-coil structure of the linker in the D21V variants. Interestingly, in the YF1 context the clades \(7n\) and \(7n + 1\) were associated with light-repressed and light-induced activity, respectively, whereas in case of D21V they both predominantly gave rise to light activation. However, two D21V variants with linkers of 31 \((= 7 \cdot 4 + 3)\) and 42 \((= 7 \cdot 6)\) residues displayed light-repressed activity; in particular, in the former construct, \(DsRed\) reporter fluorescence was enhanced by around 9-fold in darkness versus blue light. Intriguingly, the signal inversion introduced by the D21V exchange could thus be reverted in a
construct with a linker that is elongated by three residues. Currently, only two D21V variants with distinctly light-repressed activity could be identified, and a more detailed analysis of the underlying linker lengths is therefore precluded. Exhaustive sampling of linker variants in the D21V context by PATCHY might yield additional light-repressed variants and allow this type of analysis.

For the H22P background, only hybrid variants with light-induced activity as the original YF1 H22P were found but none with light-repressed activity (Fig. 5c). Consistent with a coiled-coil conformation, the linkers of these variants predominantly comprised $7n$ residues. A recent EPR investigation [19] revealed that introduction of the H22P exchange leads to a complete rearrangement of the dimer interface; two N-terminal $\alpha$ helices that in YF1 are tucked in between the LOV photosensor domains (see Fig. 1) are displaced and probably unfolded in H22P. Given this drastic change of the dimer interface, it is perplexing that the H22P variant transduces signals, let alone in inverted manner. Despite these differences, signal transduction in H22P apparently employs a highly similar structural mode as YF1, albeit with inverted sign [19]. In support of this view, we now find that the linkers of variants with signal response corresponding to that of YF1 H22P mostly conformed to $7n$ residues. By that token, one would also expect to find H22P linker variants with inverted, i.e., light-repressed activity. Our inability to do so may stem from insufficient sampling of linker sequence space or from an inherent difference in the signal mechanism of H22P. The above finding that the light response of the D21V variant could be inverted for certain linker lengths implies that the former is true and that signal inversion by linker-length variations also applies to the H22P context.

The PATCHY method offers an efficient route toward libraries of hybrid genes with a single fusion between defined fragments of two parental genes A and B. In contrast to related approaches for the construction of hybrid gene libraries [11–15], PATCHY uses a simpler protocol and obviates incremental nucleolytic digest of the parental genes which is difficult to precisely adjust. Rather, PCR amplification with sets of staggered primers exactly delineates which fragments of A and B are generated and recombined. The analysis of PATCHY hybrid libraries benefits from efficient activity assays, in particular cell-based screening and selection approaches.

We demonstrate the utility of PATCHY for the test case of hybrid blue light photoreceptors. Catalytic activity and response to light are largely governed by the length of the linker intervening the constituent photosensor and effector modules of the receptors. A striking seven-residue dependence of receptor function on linker length is explained by the continuous coiled-coil
structure of said linker. Insertion of single residues in the linker can lead to inversion of the receptor response to blue light [10]. Recent mechanistic investigations [19, 20] reveal light-induced left-handed supercoiling of the linker and thus provide the structural rationale for signal inversion via linker-length variations. In the context of the residue exchanges D21V and H22P, each of which induces signal inversion in YF1, light-regulated function is only sustained by linkers of discrete lengths.

As a method, PATCHY is generally suitable for the construction and mechanistic interrogation of hybrid gene libraries. In addition to the engineering of signal receptors, PATCHY also appears well suited to the construction of fluorescent reporters.

4 Notes

1. The computer script is available from Github at https://github.com/vrylr/PATCHY.git. To execute it, a Python interpreter is required which can be obtained free of charge from http://www.python.org or https://winpython.github.io.

2. We recommend to initially pool all forward primers at equimolar ratio and use them at a total concentration of 10 μM. The reverse primers should be prepared accordingly. Given that the individual primers differ in their sequences, they may well possess different annealing and amplification efficiencies which would result in a biased distribution of hybrid genes. Adjustment of the relative primer concentrations may be a remedy for this problem. In principle, one can employ non-equal concentrations within each primer pool to deliberately predispose the PATCHY PCR reaction toward the generation of certain sets of hybrid variants. Alternatively, the PATCHY protocol can be rerun with subsets of the forward and reverse pools that entirely lack certain primers.

3. One can carry on with the subsequent steps, even if less DNA has been obtained at this stage.

4. We found it useful to implement several steps for removal of the initial template construct from the PATCHY library. (a) In many cases, the linear PATCHY PCR products can effectively be separated from the circular template via gel extraction (see Subheading 3.1 step 6). (b) Following phosphorylation and circularization of the PATCHY PCR products, the template can selectively be degraded via restriction digest with DpnI or the specific endonuclease chosen in the template design (see Subheading 3.1 steps 9 and 10). (c) Lastly, a deliberate frameshift can be introduced into the template construct to disable it. Functional analysis of the PATCHY library, in particular by
efficient screening or selection procedures, provides a ready means for discrimination against the template construct (see Subheading 3.2 step 5). In case the resultant PATCHY library still contains an exceedingly high fraction of the initial template construct, steps (a) and (b) can be repeated.

5. The ligation reaction can alternatively be carried out at 16 °C over night.

6. A particular strength of PATCHY is the generation of hybrid genes of defined composition. Although this usually is the wanted outcome, in certain scenarios the desired traits may not manifest in any of the hybrid gene variants within the PATCHY libraries. In these cases, the PATCHY libraries can further be diversified via error-prone PCR [21]. Introduction of random residue exchanges, in particular within the linker region between protein modules, may well yield variants with the desired functional properties.

7. The screening of PATCHY libraries can be expedited with fluorescence-activated cell sorting. This is especially useful in case of protein switches that are supposed to show differential activity in response to a signal. PATCHY hybrid libraries can alternately be screened in the absence and presence of this signal, and stringently responding variants can thus be enriched, identified and isolated.

8. Alternatively, the activity of the PATCHY variants can be assessed in microtiter format. Within a 96-deep-well plate, aliquots of 600 μL LB/Kanamycin medium are inoculated with individual variants and incubated for 16 h in darkness or under constant blue light (470 nm, 40 μW/cm²). All subsequent steps are conducted as described in Subheading 3.2 steps 8 and 9.

9. Many companies offer plasmid preparation and Sanger sequencing in microtiter format. For larger number of hybrid variants, this approach usually is more economic in terms of time and money.

Acknowledgments

Discussions with C. Engelhard and R. Bittl (Freie Universität Berlin) are appreciated. Financial support through Deutsche Forschungsgemeinschaft grant MO2192/4-1 (A.M.), through a Sofja-Kovalevskaya Award by the Alexander-von-Humboldt Foundation (A.M.), and through Boehringer-Ingelheim Fonds (R.O.) is gratefully acknowledged.
References


Part VII

Cellular Signaling Switches
Chapter 19

Engineering Small Molecule Responsive Split Protein Kinases

Javier Castillo-Montoya and Indraneel Ghosh

Abstract

The over 500 human protein kinases are estimated to phosphorylate at least one-third of the proteome. This posttranslational modification is of paramount importance to intracellular signaling and its deregulation is linked to numerous diseases. Deciphering the specific cellular role of a protein kinase of interest remains challenging given their structural similarity and potentially overlapping activity. In order to exert control over the activity of user-defined kinases and allow for understanding and engineering of complex signal transduction pathways, we have designed ligand inducible split protein kinases. In this approach, protein kinases are dissected into two fragments that cannot spontaneously assemble and are thus inactive. The two kinase fragments are attached to chemical inducers of dimerization (CIDs) that allow for ligand induced heterodimerization and concomitant activation of kinase activity.

Key words Protein kinase, Split kinase, Split protein, Ligand gating, Chemical inducer of dimerization

1 Introduction

The activity and function of many, if not all, proteins are regulated by a multitude of chemical perturbations, collectively termed posttranslational modifications (PTMs) [1]. Of these, protein phosphorylation may be the most common [2], where some estimate that three-quarters of the proteome can be phosphorylated [3, 4]. Protein phosphorylation is catalyzed by a class of enzymes known as protein kinases. Protein kinases are involved in almost all pathways, from cell division to cell death [5], and their aberrant function is implicated in a plethora of diseases such as cancer [6], metabolic disorders [7], inflammation [8], and neurological disorders [9]. This renders kinases very attractive targets for therapeutic intervention [10–12], and protein kinase inhibitors are currently the second largest group of therapeutics. Kinase research has advanced tremendously since the discovery of the first serine [13] and tyrosine kinases [14], respectively. However, deciphering the specific cellular role of a
specific kinase remains challenging. Though powerful, genetic methods such as siRNA-based genetic knockdown and CRISPR-Cas gene editing, which provide insight regarding the function of a specific protein kinase, presently lack spatial and temporal control and often fall prey to compensatory cellular mechanisms. Toward the goal of establishing temporal control of a specific protein kinase, three elegant posttranslational control methods have been developed that build on structural studies [15–17] and seek to either turn a specific kinase on or off.

Shokat and coworkers pioneered a pharmacological knockdown approach by designing inhibitor-sensitive kinase alleles, where mutation of the gatekeeper residue adjacent to the kinase ATP-binding cleft allowed for inhibition with uniquely complementary inhibitors [18–20]. A more recent protein engineering approach described by Hahn and coworkers provides allosteric control over kinase activity by using a modified FK506 binding protein (FKBP) as an insertion motif in a highly conserved region of the catalytic domain of a protein kinase, which renders the kinase inactive. The subsequent addition of the small molecule rapamycin as well as the FKBP12 – rapamycin binding protein (FRB), likely restores catalytic activity through the reestablishment of the native tertiary structure of the disrupted catalytic domain [21, 22]. A third approach, pioneered by Chin and coworkers, incorporated a genetically encoded nonnatural photocaged lysine implicated in catalysis that renders the kinase inactive [23]. Irradiation of cells expressing the mutant kinase uncages lysine, leading to a gain of kinase activity and downstream signaling. Each of these innovative approaches, focusing on rendering a specific kinase inactive or active, has merits for interrogating kinase-dependent biological pathways. However, these approaches do not currently allow for the simultaneous control and study of multiple kinases.

The approach discussed in this chapter, split protein kinases (split kinases), complements the three previously described methods, providing ligand-induced gating of the activity of not only one, but multiple kinases, which can allow for understanding and engineering of more complex signal transduction pathways [24–26]. A split kinase consists of a protein kinase dissected into two fragments, each of which is attached to an interacting protein pair from a system commonly called chemical inducers of dimerization (CID) (Fig. 1). A CID system is composed of two proteins that do not interact with each other until an input, typically a small molecule, is added, which causes the proteins to interact or dimerize [27]. Thus in our design, each fragment of a split kinase attached to ligand responsive proteins is inactive until reassembled through the addition of the appropriate chemical input with attendant restoration of catalytic activity. Conceptually, the control of multiple kinases is possible by using orthogonal sets of CIDs. We have
successfully designed and constructed split kinases using three orthogonal CIDs: (1) rapamycin (Rap), which heterodimerizes FKBP and FRB; (2) abscisic acid (ABA), which heterodimerizes modified versions of interacting proteins, pyrabactin resistance 1-like protein (PYLcs, amino acids 33 to 209 of PYL1), and type 2C protein phosphatase, abscisic acid-insensitive 1 protein (ABIcs*, amino acids 126 to 423, inactive D143A mutant of ABI) [28, 29]; and (3) gibberellic acid (GA3), which heterodimerizes interacting proteins gibberellin insensitive dwarf1 receptor (GID1) and a truncated version of the protein gibberellin insensitive (GAI(92), amino acids 1 to 92) [30, 31]. Rapamycin is a widely utilized CID, but interferes with the mTOR pathway, and designed analogs (known as rapalogs) can be used in its place in studies where mTOR activity is relevant. ABA and GA3 are both plant hormone-based CIDs and can be potentially used in mammalian cells without perturbing native pathways.

In order to develop split kinases, the two main challenges are the identification of appropriate dissection sites and the development of robust methods for identifying functionally reassembled proteins. We identified potentially general sites for constructing ligand-gated split kinases by identifying sequences in the catalytic domain of protein kinases that harbor significant dissimilarities [25]. These sites correspond to positions 268, 311, and 393 of tyrosine kinase Lyn. With regard to a general and practical method to interrogate for split kinase activity, we developed a rapid in vitro or in cellulo system for production of the split kinase, with subsequent purification and testing of kinase activity through a direct
radioactivity-based assay, which remains a gold standard for activity measurements. The in vitro procedure can be completed in 9–10 h, which allows for rapid and practical iterative optimization of a particular split kinase design. We have successfully used this approach to construct split kinases for the catalytic domains of tyrosine kinases Src, Lyn, Fyn, Abl, Fak, Hck, the catalytic domain of serine/threonine kinase PKA, and full-length tyrosine kinases Src, Lyn, and Abl. We envision that other desired split enzymes can be similarly designed, interrogated, and optimized.

The following protocol describes strategies used to construct a split protein sensor for specific kinases, and to test the design by expressing the fragments either using in vitro or in cellulo approaches, and subsequently measuring kinase activity in its on and off states (Fig. 2).

**Fig. 2** General scheme showing the (a) in vitro and (b) in cellulo expression of split protein kinase sensors, followed by (c) His6-tag based protein purification and (d) radioactivity-based kinase activity assay. This methodology allows for the rapid interrogation of activity and optimization of split kinases.
2 Materials

2.1 Plasmid Construction and Cloning

1. Plasmids encoding for the kinase of interest and the interacting proteins from the desired CID system (Addgene is a good source for plasmids). We have almost exclusively focused on protein tyrosine kinases and the protocols reflect our bias.

2. Plasmid vector for cloning, such as pRSFduet-1 and pcDNA3.1(+).

3. Appropriate oligonucleotide primers to generate inserts.

4. Enzymes for cloning procedures with the necessary reaction buffers and dNTPs. Thermostable DNA polymerase (such as Taq, e.g., from New England BioLabs or KAPA from KAPA Biosystems), Klenow fragment (e.g., New England BioLabs), appropriate restriction enzymes, calf intestinal alkaline phosphatase (CIP, e.g., from New England BioLabs), DNA ligase (such as T4 DNA ligase, e.g., from New England BioLabs).

5. Competent cells, such as *E. coli* XL1-Blue, and appropriate growth media, such as Luria–Bertani agar and broth.

6. Plasmid DNA and PCR product purification kits. We typically use Macherey-Nagel NucleoSpin Plasmid, and NucleoSpin Gel and PCR Clean-up kits.

2.2 Split Kinase Protein Expression

2.2.1 In Vitro Expression in Rabbit Reticulocyte Lysate

1. DNA polymerase such as Taq or KAPA with provided reaction buffer and dNTPs.

2. RNA production system (such as Promega T7 RiboMAX Large Scale RNA production system), with provided reaction buffer and rNTPs.

3. PCR-product purification kit (e.g., from Macherey-Nagel) and G50-microcolumns for mRNA purification (e.g., from GE Healthcare).

4. Rabbit reticulocyte lysate (RRL) system (ours was donated by Luceome Biotechnologies), which contains the cellular components necessary for protein synthesis (tRNAs, amino acids, ribosomes, initiation, elongation, and termination factors). Commercial systems are also optimized to include an energy-generating system consisting of prequalified phosphocreatine and phosphocreatine kinase, a mixture of tRNAs to expand the range of mRNAs that can be translated, hemin to prevent inhibition of initiation, and potassium acetate and magnesium acetate.

5. CID: 6.25 μM Rapamycin (Rap, Sigma-Aldrich).

6. CID: 2.5 mM Abscisic acid (ABA, AG Scientific).

7. CID: 2.5 mM Gibberellic acid (GA₃, Sigma-Aldrich).

8. Dimethylsulfoxide (DMSO)
1. Dulbecco’s Modified Eagle’s Medium (DMEM) with 4 mM L-glutamine, 4500 mg/L glucose, 1 mM sodium pyruvate, and 1500 mg/L sodium bicarbonate.
2. Fetal bovine serum (FBS).
3. 100× penicillin/streptomycin solution (10,000 U/mL penicillin, 10 mg/mL streptomycin) (P/S).
4. 100× amphotericin B solution (250 μg/mL) (AmpB).
5. Phosphate-buffered saline (PBS): 5.6 mM Na₂HPO₄, 1.1 mM Na₂HPO₄ and 154 mM NaCl.
6. 0.25% trypsin/0.53 mM EDTA solution.
8. CID: 6.25 μM Rapamycin (Rap, Sigma-Aldrich).
9. CID: 2.5 mM Abscisic acid (ABA, AG Scientific).
10. CID: 2.5 mM Gibberellic acid (GA₃, Sigma-Aldrich).
11. Dimethylsulfoxide (DMSO).
12. Lysis buffer such as Thermo Scientific mPER Mammalian Protein Extraction Reagent.
13. 100× Protease inhibitor cocktail from Sigma-Aldrich.
14. 100× Phosphatase inhibitor cocktail from Santa Cruz, cocktails A and B.
15. Bicinchoninic acid protein quantification kit such as Thermo Scientific Micro BCA Protein Assay Kit.

2.2.2 In Cellulo Expression in Mammalian Cells (HEK293T)

2.3 Split Kinase Protein Purification

1. Ni-NTA agarose resin for His6-tag based purification such as Macherey-Nagel.
2. Buffer A: 20 mM Tris, 250 mM NaCl, 20 mM imidazole, pH 8.
3. Buffer B: 20 mM Tris, 250 mM NaCl, 20 mM imidazole, pH 7.
4. Kinase Assay Buffer: 20 mM MOPS, 1 mM EDTA, 10 mM MgCl₂, pH 7.

2.4 Radioactivity-Based Kinase Activity Assay

5. Kinase Assay Buffer: 20 mM MOPS, 1 mM EDTA, 10 mM MgCl₂, pH 7.
6. Appropriate kinase peptide substrate such as Srctide, Lyntide, Kemptide, and PKAtide. These can be bought from suppliers such as SignalChem or synthesized.
7. Adenosine triphosphate (Cold ATP, Sigma-Aldrich).
8. ³²P-labeled ATP (Hot ATP, PerkinElmer).
10. 0.85% Phosphoric acid.
11. Acetone.
12. Cocktail for liquid scintillation counting such as Budget Solve Complete Count Cocktail.

### 2.5 Instruments and Equipment

1. Microcentrifuge.
5. Microscope for cell counting.

### 3 Methods

#### 3.1 Plasmid Construction/Cloning

1. Choose an appropriate site to split the kinase of interest. We have utilized a dissimilarity based approach using sequence alignment of several members of a particular kinase group. *(see Note 1).*

2. Choose desired CID system for your split kinase, either under the control of rapamycin, abscisic acid, or gibberellic acid, which heterodimerize the protein pairs FRB/FKBP, ABIs*/PYLcs, and GAI(92)/GID1, respectively.

3. Using standard cloning techniques *(see Note 2)*, attach the N-Terminal portion of the kinase (NTerm) to one of the CID interacting proteins (CID1), and the C-Terminal portion of the kinase (CTerm) to the other CID interacting protein (CID2), to get the final NTerm-Linker-CID1 and CID2-Linker-CTerm constructs *(see Note 3).* Make sure you use the appropriate vector for your intended purpose *(see Note 4).*

4. Design one of your split kinase halves to include a His6-tag for protein purification purposes. We commonly incorporate this tag at the end of the kinase CTerm (CID2-Linker-CTerm-His6) *(see Note 5).*

#### 3.2 Split Kinase Protein Expression

1. In order to use the in vitro rabbit reticulocyte lysate system, in vitro translation (IVT) PCR linear products are generated using the appropriate plasmid as templates, and using a forward primer that incorporates a RNA polymerase promoter, such as T7, and a mammalian Kozak sequence, and a reverse primer that incorporates a stabilizing RNA stem loop.

2. With the appropriate IVT PCR linear products in hand, generate mRNA products using a commercial RNA production system, such as Promega’s T7 RiboMAX Large Scale RNA production system. Transcribe 3 μg of IVT-PCR products using a RNA polymerase for 4 h at 30 °C *(see Note 6)* in a final volume of 25 μL, with the provided reaction buffer and rNTPs. mRNA is further purified using G50-microcolumns.
3. Using the generated mRNA products, express proteins using a RRL system: Mix 1 pmol of mRNA with 0.25 μL of 1 mM amino acid mixture, 17.5 μL of RLL, 1 μL of 25-fold solution of the desired CID (either 6.25 μM Rap, 2.5 mM ABA or 2.5 mM GA₃ in 1% DMSO), and nuclease-free water to a final volume of 25 μL. The final DMSO concentration in the mix will be 0.04%. Incubate at 30 °C for 90 min (see Note 7).

4. Once protein expression is complete, proceed to Subheading 3.3 for protein purification and subsequent testing of kinase activity. If desired (optional), the expression of the split kinase proteins can be simultaneously performed using an in vitro protein expression reaction using 35S-methionine to label your protein of interest and subsequently running the samples on a SDS-PAGE gel and detection by autoradiography. The procedure for this step is out of the scope of this chapter, and can be found elsewhere [32].

1. Maintain HEK293T cells in DMEM media supplemented with 10% FBS, 1× P/S solution, and 1× AmpB in a 5% CO₂ incubator, at 37 °C.

2. Plate HEK293T cells at a density of 1 × 10⁶ cells per well in 6-well-plates, and incubate at 37 °C for 18–24 h prior to transfection.

3. Co-transfect cells with 1 μg of total plasmid DNA. Typically, for split constructs, 0.5 μg of each plasmid is co-transfected and compared to transfections with 0.5 μg of wild-type kinase and 0.5 μg of empty vector using a PolyJet transfection reagent or equivalent according to manufacturer protocols (see Note 8). In brief, replenish media in cells plated in 6-well-plates, 1 h before transfection; mix DNA plasmids in 50 μL of DMEM media, and separately mix 3 μL of PolyJet reagent in another 50 μL of DMEM media; rapidly add the PolyJet solution to the plasmid DNA solution, mix 3–4 times, and incubate at room temperature for 15 min. Add the resulting 100 μL of transfection mix to each well in the 6-well plates; gently mix the media in the 6-well plates and incubate at 37 °C for 10 h.

4. 10 h post-transfection, remove transfection complexes, and replenish wells with DMEM/0.1% FBS/1× P/S/AmpB media (see Note 9) containing either 80 nM rapamycin/0.1% DMSO or 100 μM ABA/0.1% DMSO or 100 μM GA₃/0.1% DMSO and 0.1% DMSO as a negative control. Incubate cells with CID for 24–36 h at 37 °C (see Note 10).

5. 34–46 h post-transfection, wash each well with 500 μL of PBS, and lyse cells with 100 μL of lysis buffer (mPER supplemented with 1× protease and phosphatase inhibitors, and with the appropriate CID in activated samples, typically, 250 nM Rap, 100 μM ABA or 100 μM GA₃. Incubate cells with mPER solu-
tion for 20 min at 4 °C. Cell lysate is then cleared at 14,000 rcf for 10 min at 4 °C.

6. Quantitate total protein concentration in each sample using BCA reagent, according to protocol: Mix 5 μL of each sample with 245 μL of 10% mPER in water, and 250 μL of BCA reagent. Incubate samples at 60 °C for 30 min and then mix 200 μL of each sample with 800 μL of water in a plastic spectrometer cuvette. Read absorbance at 562 nm and compare with a standard curve of 0–0.05 μg BSA/μL.

7. Take 150 μg of total protein from each sample and make up the volume to 100 μL using mPER solution. Proceed to Subheading 3.3 for protein purification and subsequent testing of kinase activity.

### 3.3 Split Kinase Protein Purification

1. All of the buffers used in this procedure should be supplemented with the appropriate CID (250 nM Rap, 100 μM ABA, or 100 μM GA3) in the activated samples.

2. For in vitro translated samples, dilute the 25 μL of RLL after translation with 75 μL of Buffer A, for a final volume of 100 μL. The in cellulo translated samples are already at a volume of 100 μL.

3. For each sample, add 5 μL of Ni-NTA agarose resin (Macherey-Nagel) and 100 μL of Buffer A in a microcentrifuge tube (see Note 11), and equilibrate for 30 min at 4 °C. Spin down the resin at 1530 rcf for 1 min and remove 75 μL of supernatant (see Note 12).

4. Add the 100 μL of cell lysate either from in vitro or in cellulo translation and bind to the resin for 1 h at 4 °C. Gently spin down the resin (1530 rcf/1 min) and remove 100 μL of supernatant.

5. Wash the resin for 4 min with 100 μL each time with the following buffers: 1× Buffer A, 3× Buffer B, 1× Kinase Assay Buffer (see Note 13).

6. After the last wash, remove a total of 120 μL of supernatant, leaving a total volume of 10 μL in the microcentrifuge tube (see Note 14).

### 3.4 Kinase Assay

1. Add 10 μL of an appropriate peptide substrate solution (see Note 15) and incubate at RT for 30 min.

2. Add 10 μL of a 300 μM ATP solution (83.3 nCi/μL, equivalent to adding $\frac{1}{24}$ of hot $^{32}$P–ATP PerkinElmer BLU02250UC) and incubate at room temperature for 4 h (see Note 16).

3. Spot the reaction mixture after incubation on P81 paper, and then wash the paper for 3 min 3× with 500 mL of 0.85% phosphoric acid and 1× with 500 mL of acetone.

4. Immerse the P81 paper in 10 mL of scintillation cocktail in a scintillation vial and measure counts using a scintillation counter.
5. Compare your activated split kinase samples (+CID), non-activated ones (−CID), as well as the negative control, for example, RRL or cell lysate only, to confirm that the split kinase is small molecule dependent.

4 Notes

1. We have found three general fragmentation sites for tyrosine kinases, equivalent to Lyn kinase 298/299, 311/312, and 393/394 sites (using the numbering system in UniProt P07948) [25]. These are a good starting point to create a working split kinase system for protein tyrosine kinases.

2. In brief, a standard cloning procedure consists of generating desired inserts by PCR amplification using appropriate oligonucleotide primers, followed by digestion with appropriate restriction enzymes. A selected vector is digested using the same restriction enzymes, and the insert and vector are subsequently ligated using a DNA ligase. The ligated product is then used to transform competent cells, grow colonies, purify plasmid DNA in those colonies, and finally confirm the intended construct by DNA sequencing.

3. We most commonly use the NTerm-CID1 and CID2-CTerm arrangement, but different arrangements and orientations should be tried to optimize a particular split kinase system. The most practical way to generate these constructs is by creating a vector with two multiple cloning sites (MCS) separated by a linker (MCS-Linker-MCS). The CID interacting proteins can then be cloned to generate general CID vectors, such as MCS-Linker-CID1 and CID2-Linker-MCS. We have used different linker lengths (13, 17, 21 and 25aa, with GGGS repeats), and several should be tried to optimize a particular split kinase system.

4. We most commonly use the pRSFDuet-1 vector from Novagen and the pcDNA3.1(+) vector from Invitrogen. Both of them can be used to produce protein using in vitro methods, while only the pcDNA3.1(+) vector is used to express proteins in mammalian cells.

5. We have previously placed the His6-tag at the N-termini of the constructs, with equally successful results. Other purification tags, such as FLAG-tag, can also be used in place of the His6-tag.

6. If mRNA yields are low at 4 h and 30 °C, yields may improve with 4 h at 37 °C.

7. A ribonuclease inhibitor, such as RNasin, can be added to the RLL reaction to prevent mRNA degradation. If the RLL sys-
tem does not express your split kinase system efficiently, a wheat germ extract system can also be used.

8. If this transfection conditions do not work for your particular split kinase sensor, the so-called hard-transfection conditions using 2 μg of total plasmid, and trypsinizing and pelleting cells before incubation with transfection mix can be tried. Specific procedures for hard transfection can be found for different commercial suppliers.

9. FBS can sometimes interfere with small molecule permeability. Thus, a 0.1% FBS media instead of standard 10% FBS is recommended at the step where the CID is added.

10. Incubation of cells with CID should be optimized as a function of time. Maximum expression levels of functional split kinases are usually reached at 24 h after addition of CID.

11. Using smaller microcentrifuge tubes (such as 0.2 or 0.6 mL) makes it convenient for removal of supernatant without disturbing the Ni-NTA agarose resin.

12. The 25 μL or supernatant left in the microcentrifuge tube is necessary to perform subsequent washes without disturbing Ni-NTA agarose resin. The excess supernatant is removed at the end of the washes, prior to the kinase activity assay.

13. For washes, removing supernatant, adding fresh buffer, doing a quick gentle vortex, and then incubating for 4 min before spinning down is recommended.

14. Ni-NTA agarose resin is hygroscopic and absorb buffer and thus at the end of the washes there is usually less than the 25 μL extra volume added at the beginning. Make sure to remove the same amount of supernatant in all samples without disturbing the resin, since the results for the kinase activity assay are relative to how each sample is handled.

15. Different kinases phosphorylate different commercial peptide substrates. Find an appropriate substrate for your split kinase of interest and validate all protocols with the full-length or catalytic domain only and conduct experiments at the reported $K_m$ or independently verify $K_m$ of the parent kinase prior to studies with the split kinase.

Acknowledgment

We thank members of the Ghosh lab, Dr. Reena Zutshi, and Luceome Biotechnologies for helpful comments and reagents. Indraneel Ghosh is the CSO at Luceome Biotechnologies, Tucson, AZ. We thank the NIH (1R01GM115595-01) and NSF (CHE-1506091) for supporting this research.
References


Chapter 20

Directed Evolution Methods to Rewire Signaling Networks

Raphaël B. Di Roberto, Benjamin M. Scott, and Sergio G. Peisajovich

Abstract

The ability to sense and process cues about changing environments is fundamental to life. Cells have evolved elaborate signaling pathways in order to respond to both internal and external stimuli appropriately. These pathways combine protein receptors, signal transducers, and effector genes in highly connected networks. The numerous interactions found between signaling proteins are essential to maintain strict regulation and produce a suitable cellular response. As a result, a signaling protein’s activity in isolation can differ greatly from its activity in a native context. This is an important consideration when studying or engineering signaling pathways. Fortunately, the difficulty of studying network interactions is fading thanks to advances in library construction and cell sorting. In this chapter, we describe two methods for generating libraries of mutant proteins that exhibit altered network interactions: whole-gene point mutagenesis and domain shuffling. We then provide a protocol for using fluorescence-activated cell sorting to isolate interesting variants in live cells by focusing on the unicellular eukaryotic model organism Saccharomyces cerevisiae, using as an example recent work that we have done on its G protein-coupled receptor Ste2.

Key words Signaling pathway, Protein network, Directed evolution, Domain shuffling, Random mutagenesis, Error-prone PCR, Fluorescence-activated cell sorting

1 Introduction

Cells must constantly respond to changes in their environment to fulfill their role or simply to survive. It is the function of signaling pathways to detect these changes and to decide on an appropriate response. These pathways are composed of a variety of proteins and small molecules interacting together and forming highly connected networks. Signaling proteins include extracellular or intracellular receptors, small molecules, modifiers (kinases, methylases, proteases, etc.) and transcription factors, among others.

Due to their complex roles in signal transmission and regulation, signaling networks are especially rich in protein–protein interactions. These interactions encompass simple binding, or scaffolding, which ensures that effectors and their targets are in close proximity, as well as more complex regulatory interactions that can
control protein activity [1]. Importantly, protein interactions are highly context-dependent. Each interaction partner must be present in threshold amounts and in the right state or conformation, and this may depend on the presence of a third partner at an allosteric site. As such, a signaling protein’s activity in its natural context can dramatically differ from its activity in vitro or even in a heterologous expression system. These possible discrepancies can lead us astray when studying signaling networks as the sum of individual parts rather than as a complex system of functional interdependence. Conversely, changes in network interactions, or “rewiring,” can alter network output in myriad ways, from simple changes in signal dynamics to more complex changes in input/output relationships [2].

Despite their significance, network interactions are often overlooked when studying protein function and evolution; furthermore, they are also under-utilized in protein engineering. Functional changes during protein evolution are typically studied by either comparing proteins from related organisms, or by introducing mutations in a gene and observing the consequences of those changes in the immediate vicinity of the function being analyzed. For example, when studying the evolution of a signaling kinase, one could analyze how changes in the amino acid sequence of that kinase affect binding to upstream activators, downstream substrates, scaffolds, or other regulators. A conceptually different approach is to investigate how those changes affect the function of the network, rather than the function of an individual gene. For that, it is necessary to look beyond the direct interactions between the protein being analyzed and its close network neighbours. A particularly powerful approach for studying how changes in an individual gene affect network function consists in combining random mutagenesis with selection, a procedure known as directed evolution [3]. To take full advantage of this approach, the method of mutagenesis, as well as the context of selection, should be carefully considered. As such, we provide here some guidelines for the directed evolution of signaling proteins in a way that emphasizes how changes in individual components affect overall network function. We then follow with detailed protocols for mutant library creation and selections, focusing on how different types of mutagenesis will allow us to explore very different evolutionary trajectories. These are written with the yeast mating pathway in mind, an example of a highly connected signaling pathway, but they can be used for other contexts as well.

In any evolutionary process, a source of variability is needed, and so directed evolution begins with gene mutagenesis. This process has improved considerably since the use of chemical and physical mutagens. Today, these methods have been supplanted by a wide range of approaches, ranging from error prone PCR [4], in which random mutations are introduced in a specific gene, to more
complex strategies based on DNA recombination [5], in which large DNA fragments (e.g., whole genes, specific protein domains, or other DNA regulatory regions) are swapped among different library variants. While these approaches are capable of generating billions of unique mutant genes (or gene combinations), this potential is much reduced by the necessary vector cloning and cell transformation/transfection steps that follow. Both steps have limited efficiencies that significantly reduce the size of the mutant library that can be generated. Furthermore, the introduction of random point mutations throughout an entire gene inevitably results in a large pool of inactive or unchanged mutants [6], which can be seen as wasted potential. This problem has often been remediated at least in part by restricting the region of mutagenesis. In this way, known structural features are avoided and the region of interest is mutated more thoroughly. However, this solution comes with bias, as catalytic sites are typically targeted because they are often well defined and perceived as key to changing function [7, 8]. This neglects other possible functional changes, including network rewiring events which are rarely well characterized.

To ensure that network interactions are not excluded, it is thus recommended that targeting specific gene regions be kept to a minimum, especially in the absence of knowledge on the protein’s interaction motifs. Instead, it is better to increase the efficiency of cloning and transformation. As transformation efficiency is strongly cell- and vector-dependent, we will not discuss it further. Instead, we will focus on the mutagenesis and cloning steps. In our first protocol, we will describe the construction of a library of random point mutants through a combination of error-prone DNA polymerase, high-fidelity polymerase and type IIS restriction enzymes. The latter enzymes, reviewed in [9], are notable for their ability to cut DNA downstream of a recognition sequence. Though they are useful for scar-free, one-directional cloning, perhaps their greatest advantage comes from the possibility to design custom non-palindromic overhangs. These can prevent self-ligation between inserts (illustrated in Fig. 1) and can therefore dramatically increase the efficiency of proper ligation.

Alternatively, one can avoid point mutagenesis altogether by focusing instead on protein domains. Many proteins consist of functional units, called domains, which can fold independently of each other and accomplish self-contained functions. This modularity allows different domains to be joined into new proteins and is thought to be an essential source of novelty in natural evolution [10]. For instance, in the yeast mating pathway, protein kinases and scaffolds involved in signal transduction are composed of one or more interaction domains [11]. When these domains are manually “shuffled” with each other, new functional proteins are generated and these can confer a novel response to the overall signaling pathway [11–13]. By promoting novel interactions between
Fig. 1 AarI-based multi-insert cloning and domain shuffling. (a) AarI recognizes a 7 bp sequence but cuts 4 and 8 bases away from it. Because AarI is not sensitive to the specific sequence at its cutting sites, it is possible to design non-palindromic overhangs that will reduce the chances of insert (or plasmid) self-ligation, thereby...
unchanged catalytic units, domain shuffling is a powerful way to rewire a protein network and gain understanding of how those interactions contribute to signaling. Of course, this approach is dependent on the presence of clear domain boundaries, and public databases are helpful for identifying these. Here, we provide a protocol for the construction of a library of shuffled domains. This protocol also takes advantage of the unique properties of type IIS restriction enzymes in order to efficiently assemble multiple protein domains in a defined order.

Once a mutant DNA library has been generated, it must be expressed and screened in order to analyze the resulting diversity of phenotypes. Because this screening step is often conducted in vitro or in a nonnative context (via heterologous expression in \textit{E. coli} or yeast), mutations affecting network interactions can be missed due to the lack of native protein partners. As such, it is preferable to conduct the assay in a native or near-native context or in an environment where the relevant protein partners are exogenously supplied. Furthermore, if a specific function is desired, a selection step is necessary to isolate promising mutants. Importantly, the phenotype selected should reflect pathway output rather than an individual protein’s activity. Our third protocol describes how to select interesting mating pathway mutants using fluorescence-activated cell sorting (FACS) in yeast. Together with efficient library cloning and protein domain shuffling, this strategy can reveal the contribution of network rewiring in protein and pathway function, or unlock its potential for applied protein engineering.

## 2 Materials

### 2.1 Error-Prone PCR

1. DNA primers for error-prone PCR and amplification PCR reactions (see \textbf{Note 1}).
2. Target DNA (on plasmid).
3. 40 mM dNTP mix (10 mM each dNTP).
5. 10× Mutazyme II reaction buffer (Agilent).
6. DpnI enzyme.

\textbf{Fig. 1} (continued) significantly increasing the efficiency of the multi-insert cloning step. \textbf{(b)} The identification of domain boundaries via protein databases is followed by a PCR amplification step that adds AarI recognition sequence and specific overhangs designed for sequential ligation. The domain library is then ligated into a vector for which the overhangs match those of the first and last domains that will form shuffled proteins, resulting in the creation of a large domain-shuffled gene library. Note that an individual domain could be included at multiple positions in the final multi-domain protein, simply by altering the identity of the flanking overhangs.
7. QIAquick PCR Purification Kit (Qiagen).
8. Pfu DNA polymerase.
9. Acceptor plasmid DNA.
10. Antarctic phosphatase.
11. T4 DNA ligase.
12. DH5α competent *E. coli* cells (High Efficiency).
13. LB liquid and agar media containing appropriate antibiotic.
14. QIAprep Spin Miniprep Kit (Qiagen).

### 2.2 AarI Mediated Domain Shuffling

1. Acceptor plasmid DNA.
2. Donors’ DNA: either genomic DNA containing targeted genes, or plasmids carrying the cloned targeted genes.
3. DNA primers for PCR reactions.
4. Pfu DNA polymerase.
5. QIAquick PCR Purification Kit (Qiagen).
6. AarI restriction enzyme.
7. 50× (0.025 mM) oligonucleotide containing AarI recognition sequence.
8. AarI 10× reaction buffer.
10. 10× Antarctic phosphatase buffer.
11. T4 DNA ligase.
12. 10× T4 DNA ligase buffer.
13. DH5α competent *E. coli* cells (high efficiency).

### 2.3 Yeast-Based Sorting of Rewired Pathway Interactions

1. Synthetic drop-out medium (liquid and solid).
2. Mutant plasmid library.
3. Round-bottom tubes with 35 μm cell strainer caps (BD).
4. 10 mM *K. lactis* α-factor pheromone.
5. 10 mg/mL cycloheximide.
6. Zymoprep™ Yeast Plasmid Miniprep II (Zymo Research).

### 3 Methods

#### 3.1 Error-Prone PCR

1. Prepare a stock solution of Target DNA in water. The amount of Target DNA depends on desired mutation frequency (*see Table 1*). As an example we use the yeast G protein-coupled receptor Ste2.
2. Prepare a 50 μL reaction as follows:
3. Place reaction in thermocycler, using the program depicted in Table 2.

4. To digest original plasmid containing target DNA, add 1 μL DpnI enzyme to the error-prone PCR mixture. Incubate at 37 °C for 3 h.

5. Purify the error-prone PCR reaction using a PCR purification kit. Elute with no more than 30 μL water or the provided elution buffer.

6. To amplify the Error-Prone PCR product, prepare two 50 μL reactions as follows:

   - 5 μL of 10× Mutazyme II reaction buffer.
   - 1 μL of 40 mM dNTP mix.
   - 0.5 μL of PCR primer mix (250 ng/μL of each PCR primer).
   - 1 μL of Mutazyme II DNA polymerase (2.5 U/μL).
   - 1–15 μL of target DNA (see Table 1).
   - Water to a final total volume of 50 μL.
7. Place reactions in thermocycler, using the program depicted in Table 3:

8. Combine amplification PCR reactions and purify using a PCR purification kit.

9. Digest >4 μg of purified amplification PCR product, and >4 μg of desired acceptor plasmid with appropriate restriction endonuclease(s). See Subheading 3.2 step 7 for recommended AarI digestion protocol.

10. Dephosphorylate the ends of the linearized acceptor plasmid by incubation with Antarctic phosphatase at 37 °C for 1 h to reduce self-ligation background.

11. Purify digested amplification PCR product and acceptor plasmid with a PCR purification kit. Elute with no more than 30 μL water or provided elution buffer.

12. Ligate purified digested products using T4 DNA ligase. The number of ligation reactions to perform depends on the desired size of the library (see Table 4). Use 150 ng digested acceptor plasmid and a 3 molar excess of digested amplification PCR product per 20 μL reaction. Incubate at 16 °C overnight.

---

<table>
<thead>
<tr>
<th>Segment</th>
<th>Number of cycles</th>
<th>Temperature</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>95 °C</td>
<td>2 min</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>95 °C</td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td>Primer Tm—5 °C</td>
<td></td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td>72 °C</td>
<td></td>
<td>1 min (&lt;1 kb targets) or 1 min/kb (&gt;1 kb targets)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>72 °C</td>
<td>10 min</td>
</tr>
</tbody>
</table>

---

Table 3
PCR cycling protocol for amplifying error-prone PCRs with Pfu DNA polymerase

<table>
<thead>
<tr>
<th>Segment</th>
<th>Number of cycles</th>
<th>Temperature</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>95 °C</td>
<td>2 min</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>95 °C</td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td>Primer Tm—5 °C</td>
<td></td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td>72 °C</td>
<td></td>
<td>2 min/kb</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>72 °C</td>
<td>10 min</td>
</tr>
</tbody>
</table>

---

Table 2
PCR cycling protocol for error-prone PCR with mutazyme

<table>
<thead>
<tr>
<th>Segment</th>
<th>Number of cycles</th>
<th>Temperature</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>95 °C</td>
<td>2 min</td>
</tr>
<tr>
<td>2</td>
<td>20–30, see Table 1</td>
<td>95 °C</td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td>Primer Tm—5 °C</td>
<td></td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td>72 °C</td>
<td></td>
<td>1 min (&lt;1 kb targets) or 1 min/kb (&gt;1 kb targets)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>72 °C</td>
<td>10 min</td>
</tr>
</tbody>
</table>
13. Transform 50 μL competent *E. coli* cells with 5 μL ligation reaction. Plate on LB agar media containing the appropriate antibiotic, and incubate at 37 °C overnight.

14. Randomly select at least ten colonies, and culture each overnight in liquid LB media containing the appropriate antibiotic.

15. Prepare plasmid DNA mini-preparations from overnight cultures, and sequence the inserted mutant gene to determine the mean mutation frequency.

16. If the desired mutation frequency has been established, transform the remaining ligation reaction volume into competent *E. coli* cells, with 5 μL ligation reaction per 50 μL cells (see Table 4).

17. The total number of plated *E. coli* colonies is the approximate diversity of your plasmid library. Add 1.5 mL liquid LB media to each plate of transformed *E. coli* cells, steriley scrape cells, and pipette into the same tube.

18. Prepare plasmid DNA mini-preparations, using one QIAprep Spin Column per plate of *E. coli* cells. Elute half the columns each with 35 μL elution buffer. Then pass this elutate over the remaining columns. Combine all mini-preparations to create the final plasmid library (see Note 3).

### 3.2 AarI Mediated Domain Shuffling

AarI is a Type IIIS restriction enzyme that recognizes a 7 bp sequence (CACCTGC), cutting 4 and 8 bp away from the recognition site, leaving a 4 base overhang, as show in Fig. 1. The fact that AarI is agnostic to the sequence of the overhangs means that users can “design” those sequences, thus enabling scarless cloning or, more importantly in this case, allowing the use of NON-PALINDROMIC overhangs. This property is key for multi-insert cloning, as ligation of multiple fragments flanked by non-palindromic overhangs is orders of magnitude more efficient than the ligation of fragments with palindromic ends (as palindromic overhangs can self-ligate, decreasing the ligation efficiency dramatically). The steps required to generate libraries of shuffled domains are outlined below.

1. Identify domain boundaries for the genes of interest from databases (e.g., UniProtKB/Swiss-Prot).

---

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Setup for ligation reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligation reaction volume (5 × 20 μL)</td>
<td>100 μL</td>
</tr>
<tr>
<td>Competent <em>E. coli</em> cells volume (20 × 50 μL)</td>
<td>1000 μL</td>
</tr>
<tr>
<td>Expected number of <em>E. coli</em> colonies (i.e., library size)</td>
<td>70,000 colonies</td>
</tr>
</tbody>
</table>
2. Design PCR primers to amplify protein domains with appropriate “overhang” sequences and AarI recognition sequences that will result in the multi-part assembly of a domain-shuffled gene. In Fig. 1, we show the overhang design strategy for a three-domain shuffling library. In this case, we have 4 distinct overhangs, of the following sequences as summarized in Table 5:

As an example, here we show in detail the design of four primers, one pair amplifying a domain flanked by “A” and “B” overhangs, and one domain flanked by “B” and “D” overhangs. These two fragments could then be joined as an “AB” and “BD” pair.

Table 5
Summary of 4 bp overhangs

<table>
<thead>
<tr>
<th>Overhang name</th>
<th>DNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>GGAG</td>
</tr>
<tr>
<td>B</td>
<td>CCCT</td>
</tr>
<tr>
<td>C</td>
<td>GCGA</td>
</tr>
<tr>
<td>D</td>
<td>TGCG</td>
</tr>
</tbody>
</table>

Primers for “AB” flanked domain:
Forward primer: 5’-8 random bases—CACCTGCACACAAGGAG—18–20 bases annealing to the 3′ end of your gene fragment of interest-3′ (note that for the first fragment, the primer should contain an ATG start codon).
Reverse primer: 5’-8 random bases—CACCTGCGTTCAGGG—18–20 bases annealing to the 3′ end of your gene fragment of interest-3′ (note that for multi-part shuffling it is fundamental to ensure that all three parts be in frame).

Primers for “BD” flanked domain:
Forward primer: 5’-8 random bases—CACCTGCACACAACCCT—18–20 bases annealing to the 5′ end of your gene fragment of interest-3′ (note that as this will be an internal domain, there is no need to include a START codon).
Reverse primer: 5’-8 random bases—CACCTGCGTTCGCCCA—18–20 bases annealing to the 5′ end of your gene fragment of interest-3′ (note that, as above, additional bases will be needed to ensure that all fragments are in frame and since this will be the last domain in the shuffled protein, you will need to include a STOP codon as well).

3. Perform the PCR reactions using a high fidelity polymerase, such as Pfu, as indicated by the manufacturer. Note that the
The protocol listed here can be used to clone three consecutive domains leading to a single three-domain protein, or to clone libraries in which each “AB,” “BC,” or “CD” fragment is actually a large collection of different domains sharing the same overhangs. In the latter case, the result is not a single three-domain protein, but rather a library of multiple different combinations of three domains arranged in the desired orders. This order is determined by the presence of consecutive matching overhangs (e.g., “AB” will ligate upstream of “BC”).

4. Clean the PCR reactions using the QIAquick PCR cleanup kit as indicated by the manufacturer.

5. Determine the concentration of each amplified fragment library by measuring absorbance at 260 nm using a Nanodrop spectrophotometer (or equivalent instrument). This step is key, as AarI digestion is sensitive to DNA concentration.

6. For AarI-mediated multi-insert cloning of domain shuffled libraries, it is necessary to use an acceptor vector flanked by AarI recognition sequences with the first and last overhangs used (“A” and “D” in our case), as show below:

```
5′…PROMOTER…GGAGCAAGGCAGGTG…(~20 bp intermediate sequence)…CACCTGCAACATGCG…TERMINATOR…3′
```

Note the flanking overhangs in bold, and the AarI recognition sequences (in opposite directions) underlined.

7. Set up individual AarI digestion reactions for each PCR-generated DNA fragment library, as well as for the acceptor plasmid DNA. For each reaction, set up the components as depicted in Table 6, mix gently and incubate at 37 °C for 3 h:

8. Inactivate AarI by incubation at 65 °C for 20 min.

9. Purify digested DNA (both plasmid and PCR fragments) directly with QIAquick PCR clean up kit. Note that as long as the fragment released from the acceptor vector is small (up to 20 bp) there is no need for gel purification.

### Table 6
**Setup for AarI restriction enzyme reaction**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA (either acceptor plasmid OR PCR fragment)</td>
<td>5 μg</td>
</tr>
<tr>
<td>10× AarI reaction buffer</td>
<td>6 μL</td>
</tr>
<tr>
<td>50× AarI oligonucleotide (0.025 mM)</td>
<td>1 μL</td>
</tr>
<tr>
<td>AarI enzyme</td>
<td>2.5 μL</td>
</tr>
<tr>
<td>Water</td>
<td>To complete 60 μL</td>
</tr>
</tbody>
</table>
10. Dephosphorylate the ends of the linearized acceptor plasmid by incubation with Antarctic phosphatase at 37 °C for 1 h to reduce self-ligation background.

11. Determine the concentration of each digested DNA fragment by measuring absorbance at 260 nm using a Nanodrop spectrophotometer (or equivalent instrument). Once again, this step is key, as multi-insert ligations are very sensitive to the relative concentration of all fragments being ligated. We have found that a molar ratio of 2:1:1:1 (where “2” is the acceptor plasmid) is efficient for 3-insert ligations.

12. Set up ligation reactions as depicted in Table 7 (for a 20 μL reaction) and incubate the reaction for 4 h at 16 °C, followed by overnight incubation at 4 °C.

13. Transform the ligations in highly competent cells (protocols vary depending on the particular bacterial cell strain used). Normally, chemically competent cells with efficiencies of ~10^8 CFU are suitable. The target number of colonies depends on the theoretical size of the library being generated. For example, while in principle a single colony will suffice for a single 3-insert ligation (though several more would obviously be preferred), a library in which “N1” number of inserts are shuffled at position “AB,” “N2” number of inserts are shuffled at position “BC,” and “N3” number of inserts are shuffled at position “CD,” will have a theoretical size equal to “N1 × N2 × N3”. In this case, it is good to have at least ten times more colonies than the expected library size to ensure that all variants are represented.

14. Transformants can be screened by colony PCR, using standard methods. Clone identities should be determined by DNA sequencing.

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Setup for ligation reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Component</strong></td>
<td><strong>Amount</strong></td>
</tr>
<tr>
<td>Digested/dephosphorylated acceptor plasmid DNA</td>
<td>150 ng</td>
</tr>
<tr>
<td>Digested insert “AB” (or mix of multiple “AB” inserts in desired ratios)</td>
<td>As required (based on desired molar ratio)</td>
</tr>
<tr>
<td>Digested insert “BC” (or mix of multiple “BC” inserts in desired ratios)</td>
<td>As required (based on desired molar ratio)</td>
</tr>
<tr>
<td>Digested insert “CD” (or mix of multiple “CD” inserts in desired ratios)</td>
<td>As required (based on desired molar ratio)</td>
</tr>
<tr>
<td>T4 DNA ligase</td>
<td>1 μL</td>
</tr>
<tr>
<td>10x T4 DNA ligase buffer</td>
<td>2 μL</td>
</tr>
<tr>
<td>Water</td>
<td>To complete 20 μL</td>
</tr>
</tbody>
</table>
The following protocol is designed for the identification of mating pathway gene variants that enable a strong response to the peptide pheromone α-factor. For instance, the gene STE2, which encodes the pheromone receptor, can be mutated to identify changes that lead to mating pathway induction in the presence of a pheromone from the related species Kluyveromyces lactis [14]. In this system, a ste2Δ yeast strain which expresses green fluorescent protein (GFP) from a mating-inducible promoter is transformed with a mutant STE2 library and used to select active Ste2 variants.

1. Transform yeast cells with the plasmid library of mutant DNA. We recommend using the high-efficiency lithium acetate transformation method [15]. This method can be expected to yield 15,000 colonies from an initial 5 mL of log-phase yeast culture and 1 μg of a 5–10 kbp DNA plasmid.

2. Pick 100 colonies or more and assay the mating pathway response of each. This pre-sorting screen will reveal the ratio of active to inactive mutants as well as the diversity of the active mutants’ phenotypes (Fig. 2a).

3. Place individual mutant colonies in 2 mL of drop-out medium. Also inoculate 2 mL volumes with a negative control, such as cells transformed with an empty vector, and a positive control, such as cells expressing wild-type STE2. Grow overnight at 30 °C in a 225 RPM shaker incubator.

4. Transfer 40 μL of the overnight culture to 2 mL of drop-out medium. Grow this dilution to an optical density at 600 nm (OD600) of 0.4 to 0.6 (early log-phase).

5. Add α-factor pheromone to each culture to a final concentration of 100 nM. Cultures can also be split into two to measure the pathway response in the absence of pheromone. Grow for 2 h.

6. Add cycloheximide to a final concentration of 10 μg/mL to each culture to arrest protein expression, including GFP.

7. Sonicate each log-phase culture briefly to break large cell aggregates. This typically requires two sonication pulses at the lowest setting.

8. Run each culture in a flow cytometer to measure GFP fluorescence. This requires a 488 nm laser and a 525/50 nm filter.

9. To proceed with cell sorting, combine the mutant library into a single liquid culture. For this, dispense 5 mL of drop-out medium onto each plate of transformed yeast cells and scrape off the colonies using a plating stick. Aspirate the mixed colonies and add them to a tube on ice.

10. Vortex the colony mixture on a low setting for 30 s.

11. Inoculate 50 mL of drop-out medium with 50 μL of the colony mixture. Also inoculate 2 mL volumes with the negative and positive controls. Grow overnight.
Fig. 2 Fluorescence-activated cell sorting of yeast mutant libraries. In this example, we show sorting data for yeast carrying a Ste2 receptor library, with mutants conferring a mating response to *K. lactis* pheromone. (a) Gating strategy needed to ensure that most events collected correspond to single cells. The rectangular gates capture cells in the diagonal, which mostly correspond to single cells (cell aggregates display higher forward
12. Perform **steps 4–7** to induce the mating pathway, but do not add cycloheximide.

13. Transfer 2 mL of each culture to an empty 5 mL round bottom tube with a 35 μm cell-strainer cap. Dispense the liquid culture through this cap to further reduce aggregation.

14. Run the filtered cultures in a cell sorter:

15. Vortex a 5 mL round bottom tube containing 500 μL of sterile drop-out medium and place it in the cell sorter’s collection chamber.

16. Set up the gating strategy outlined in Fig. 2a to target live, single cells. Most cell types (including bacteria, yeast, and animal cells) will form cell aggregates when suspended in solution. Thus, there is a risk that a simplistic gating strategy (e.g., selecting cells that display high GFP fluorescence) will be compromised by the presence of cell aggregates that may include cells with low fluorescence levels considered to be a single “event” by the flow cytometer. To minimize this risk, it is important to devise a multigate selection strategy that includes at least two gates based on forward and side scattering signals capable of distinguishing between single cells and aggregates. In the case of yeast cells, two gates are sufficient: one plotting the width of the forward scattering signal vs. the height of the forward scattering signal, and a second gate plotting the width of the side scattering signal vs. the height of the side scattering signal, as shown in Fig. 2a. Similar strategies can be devised for other cell types, though it is important to be familiar with the specific aggregation propensities and light scattering properties of your cell of choice.

17. Use your negative control to establish the cells’ baseline fluorescence and your positive control to visualize the fluorescence of active mutants. Draw a gate that includes the latter but does not overlap with the former. Note that in separate experiments you may want to select cells that activate the pathway in the absence of stimulus, or only in the presence of stimulus (Fig. 2b).

18. Sort the desired number of mutants and collect them in a 5 mL tube containing 500 μL of drop-out medium. Once the sort has ended, lightly vortex the collection tube to wash the walls.

19. Plate the content of the collection tube onto solid medium (100 μL per plate). Approximately 40% of the collected events can be expected to be recovered this way.

---

**Fig. 2** (continued) and side scattering heights. (b) Gating strategies needed to select cells that activate the pathway response upon pheromone treatment. Note the exclusion of inactive mutants (left). (c) Flowchart of a cell sorting experiment. A plasmid library of STE2 mutants is transformed in yeast. An initial phenotypic screen shows that most mutants cannot sense the pheromone of *K. lactis* as well as wild type. Following a first round of cell sorting, almost all mutants selected can sense the pheromone and about half can do so better than wild type. A second cell-sorting step yields an even greater proportion of strong *K. lactis*-responsive mutants.
20. Run a screen as done in steps 2–8 in order to confirm that the sorted mutants have the desired phenotype (see Fig. 2c for an example with our data on *S. cerevisiae* Ste2 mutants capable of responding to a pheromone from a distantly related yeast species). Interesting mutants can be isolated and their plasmid extracted using the Zymoprep™ Yeast Plasmid Miniprep II to be sequenced.

21. Optional: Perform iterative rounds of sorting on the sorted library to isolate an ever-greater proportion of desired mutants.

### 4 Notes

1. DNA primers must contain AarI restriction sites, or alternative endonuclease sites appropriate for inserting the product into the desired acceptor plasmid, followed by 18–20 bp annealing to the Target DNA. This annealing sequence will not be mutated.

2. Listed values are similar to those suggested by the manufacturer, although several conditions can be tested simultaneously to achieve the desired mutation frequency. In general, using less Template DNA and more cycles will increase the mutation frequency.

3. If the plasmid library will be transformed into yeast, a plasmid DNA concentration of 1 μg/μL is recommended. The DNA concentration can be increased by standard precipitation with sodium acetate and ethanol.

### Acknowledgments

Supporting Grant Information: This work was funded by an NSERC (National Science and Engineering Research Council, Canada) Discovery Grant 418467-2012 (S.G.P), a CFI-ORF (S.G.P.), an Early Research Award from the Province of Ontario (S.G.P.), a Boehringer-Ingelheim Young Researcher Award (S.G.P.), an NSERC Canadian Graduate Scholarship (R.B.D.), and an Ontario Graduate Scholarship (B.S.).

### References

Rewiring Signaling Networks

Natl Acad Sci USA 91(22):10747–10751. doi:10.1073/pnas.91.22.10747


Index

A

Actuator ........................................................................................................ 3, 4, 6, 8–11, 13, 14, 198
Affinity clamp .......................................................................................... 14, 200, 204
Allosteric derepression ............................................................................ 167–176
Allosteric Protein Switches ..................................................................... 28–40
Allosteric regulation ................................................................................. 151
Alternative frame folding ....................................................................... 6
Amyloid fibrils .......................................................................................... 59–66
Ancestral Gene Resurrection (AGR) ......................................................... 9
Antibody-like scaffolds
  fibronectin ............................................................................................. 14
  Fn3 ......................................................................................................... 28
Antibody scaffolds
  V_{H} .................................................................................................... 150, 155–156, 160
  V_{L} .................................................................................................... 150, 155, 156, 160

B

Beta-lactamase .......................................................................................... 31
Beta-lactamase inhibitory protein (BLIP) .................................................. 169, 171, 
  172, 174–176, 180, 182, 185–187, 189–193
Beta-lactamase sensor ............................................................................. 149, 167, 179
Binding receptors ..................................................................................... 4, 6, 9, 14, 198–200
Bioluminescent enzyme assays .............................................................. 137
Bioluminescent sensors ......................................................................... 101, 119, 131
Biomarker detection ................................................................................. 149
Biophysics ................................................................................................. 4, 6–10, 14, 72, 200
Biosensors
  amino acids .......................................................................................... 185, 221
  antibodies ........................................................................................... 15, 171, 180
  cell surface .......................................................................................... 103, 110–111, 219
  cellular .................................................................................................. 219, 221
  DNA ..................................................................................................... 103, 179–193
  drugs ................................................................................................... 101, 181, 220
  in vitro .................................................................................................. 110, 179
  peptides ............................................................................................... 107, 151
  phosphorylation ................................................................................... 131, 220
  protein–protein interactions ................................................................. 219–236
  proteins ................................................................................................. 28, 72, 101, 105, 171, 186

C

Cell biology ................................................................................................. 17, 104
Cell-free expression
  rabbit reticulocyte lysate .................................................................. 311, 313–314

Chemical denaturation curves
  in GdnHCl ........................................................................................... 33
  in urea ................................................................................................. 33
Chemical inducers of dimerization
  abscisic acid ....................................................................................... 309, 311–313
  gibberellic acid ................................................................................... 309, 311–313
  rapamycin .......................................................................................... 309, 311–313
Clampbody ............................................................................................. 150–162
Computational modelling ...................................................................... 89–99

D

Design-build-test ..................................................................................... 5, 16
Design-build-test and learn ................................................................... 5
Diagnostic assays .................................................................................... 119
Directed evolution .................................................................................. 43–54, 245, 322
DNA cloning
  in-fusion .............................................................................................. 12, 153, 170
  overlap extension ................................................................................ 95, 156
  restriction ligation .............................................................................. 11
  TOPO .................................................................................................. 153, 156
  USER Enzyme .................................................................................. 12, 202, 206
DNA mutagenesis
  error prone PCR ................................................................................ 245, 275, 279,
  323, 325–329
  gap repair ........................................................................................... 280, 281
  site-directed ....................................................................................... 81, 122, 128, 134, 274
DNA nuclease digest
  DNase I ............................................................................................... 44, 45, 47, 51, 54
  S1, 44, 45, 54
DNA recombination
  homology-dependent ......................................................................... 10, 11
  ligation-dependent ............................................................................. 5, 11–12
  modular ............................................................................................... 10, 12, 180
  nuclease-dependent ......................................................................... 11
Domain shuffling .................................................................................... 324–326, 329–333
Dose response curves ............................................................................... 161, 231, 235
Drug screening ......................................................................................... 119, 169

E

Enzyme substrate channeling .................................................................. 119
Escherichia coli
  DNA cloning ....................................................................................... 202
  electroporation .................................................................................. 74, 80
  protein expression ............................................................................. 17, 172, 202
Fluorescence microscopy ........................................ 115, 228, 259, 261, 263, 265–268
Fluorescence spectroscopy ..................................... 79, 205
Fluorescent protein (FP) ........................................... 4, 14, 17, 30, 71, 82, 85, 89–99, 101–103, 109, 137, 180, 244, 333
Fluorescent protein sensors ...................................... 13
FREX (protein/fragment exchange). See Protein/fragment exchange (FREX)

Genetic assays......................................................... 17, 18, 308
Genetic complementation ........................................ 18
Genetic screening .................................................... 16, 17
Genetic selection ..................................................... 17, 18
Green fluorescent protein (GFP) .............................. 6, 10, 13, 17, 221, 227, 229, 230, 232, 333, 335

Ion channel engineering ........................................... 273, 278

Mammalian cell culture
lentiviral transduction ............................................ 134, 263, 264
protein expression ................................................... 17
transfection .......................................................... 114, 263, 264, 312
Membrane recruitment .......................................... 259, 267
Mouse .................................................................... 97, 132, 134–136, 139, 140, 142–144

Optogenetics .......................................................... 258, 265, 269, 272

Peptide fibrils .......................................................... 59–66
Peptide switches ....................................................... 59
Periplasmic-binding proteins (PBPs) ......................... 9, 17
Phosphorylation ....................................................... 257
Photoreceptor engineering
Cry2-CIB1 .............................................................. 257–269
Lov2 .................................................................... 18, 243, 247, 248, 254, 272, 273

Photoreceptors ......................................................... 242–244, 247–249, 252, 254, 258, 288, 290, 298, 301
Photosensitive degron .............................................. 242, 243, 245, 252
Phylogenetic analysis .............................................. 73, 75–77, 82–84
Polymerase chain reaction (PCR)
multiplex inverse PCR ........................................... 44, 45, 48, 54
overlap extension PCR ........................................... 10, 11, 35, 155, 156, 275, 278, 279
PATCHY ................................................................. 289, 290, 293, 294, 296–302
Proteases
receptors ............................................................... 198, 199, 201–205, 210–212, 216
sensors ................................................................. 169, 173, 176, 197–217
split-TEV sensors ................................................... 219
switches ............................................................... 198, 199, 203, 204, 209–214, 216, 217
transducers ......................................................... 198–201, 203, 206, 209–213, 216, 217
Protein analysis
SDS-PAGE ............................................................. 112, 159, 171, 266, 314
semi-native SDS-PAGE ........................................... 183, 190
western blotting ...................................................... 133, 137, 139, 144, 150, 153–154, 159–160, 231, 259–261, 263, 265–267, 269
Protein circular permutation ..................................... 31, 44, 52, 73, 85, 107
Protein complementation assay ............................... 121, 141, 151, 220
Protein conjugation ................................................ 185, 186, 188, 189
Protein degradation ................................................ 251
Protein-DNA conjugation ....................................... 179
Protein engineering
chemical conjugation ............................................. 38, 91, 182, 185
computational ....................................................... 7, 18, 105
domain insertion .................................................... 6, 10
high-throughput screening and selection .................. 10, 13, 18, 200, 208, 209
modular ............................................................... 10
rational ................................................................ 13, 200
Protein evolution ..................................................... 76, 77, 322
Protein/fragment exchange (FREX) ......................... 28–30, 34, 35, 37, 39, 40
Protein kinases
AKT kinase ............................................................ 258
ATM kinase sensor ................................................ 131–144
histidine kinase sensor .......................................... 290, 298, 300
Protein linker engineering
glycine-serine ....................................................... 11
polyproline .......................................................... 108
Protein phosphorylation ......................................... 198, 307
Protein purification
His/TALON-tag affinity .......................................... 111, 181
Strep-tag affinity .................................................... 111, 181
Protein receptors ................................................... 322, 326
Protein refolding .................................................... 150, 159
Protein sensors ....................................................... 7, 8, 310
Synthetic Protein Switches: Methods and Protocols

Index

Protein stability ........................................... 241–254
Protein thermostability ................................ 72, 73, 75, 79

R

Radioactivity-based kinase activity assays ........ 310, 312–313
Reaction kinetics
  chemical ................................................... 63
  enzymatic ............................................... 123, 204, 216
Rewiring signal transduction networks.......... 321–336

S

Saccharomyces cerevisiae
  chemical transformation ............................. 249–250
  drop test ............................................... 283
  functional complementation ....................... 273, 281
  genetic complementation ............................ 18
  replica plating ....................................... 283
Self-labelling proteins
  CLIP-tag ................................................. 103, 110, 113
  SNAP-tag ............................................... 102, 103, 105, 108, 110, 113
Semi-synthetic sensors
  bioluminescent ....................................... 101
  fluorescent .......................................... 89, 101
  LUCIDs ............................................... 101–115
  SNIFITs ............................................... 101–115
  Solute binding proteins (SBP) ..................... 72, 73, 75, 76
  79–82, 84, 85, 92, 93
Spectrophotometry
  absorbance ............................................ 113, 331, 332
  circular dichroism .................................. 33
  fluorescence ......................................... 33
  luminescence ....................................... 101, 119, 131
  UV/Vis .................................................... 62, 182, 187
  Split-proteases ....................................... 219
  Split protein kinase ................................ 307–317
  Split-protein sensor ................................ 310
  Split-tyrosine kinase ............................... 310, 316
  Statistical sequence analysis ...................... 5, 8–9
  Synthetic biology .................................. 5, 179, 198
  Synthetic dye labelling ............................ 90, 91, 98
  Synthetic protein switches ...................... 3, 179, 191, 201
  Synthetic signal transduction ................. 4

T

Therapeutic drug monitoring ..................... 101

V

Viral potassium channel engineering ........... 271