



Arnaud Gautier
Marlon J. Hinner *Editors*

Site-Specific Protein Labeling

Methods and Protocols

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
<http://www.springer.com/series/7651>

Site-Specific Protein Labeling

Methods and Protocols

Edited by

Arnaud Gautier

Department of Chemistry, École Normale Supérieure, Paris, France

Marlon J. Hinner

Sensitive Farbstoffe GbR, Munich, Germany



Editors

Arnaud Gautier
Department of Chemistry
École Normale Supérieure
Paris, France

Marlon J. Hinner
Sensitive Farbstoffe GbR
Munich, Germany

ISSN 1064-3745
ISBN 978-1-4939-2271-0
DOI 10.1007/978-1-4939-2272-7
Springer New York Heidelberg Dordrecht London

ISSN 1940-6029 (electronic)
ISBN 978-1-4939-2272-7 (eBook)

Library of Congress Control Number: 2014958475

© Springer Science+Business Media New York 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Cover Image Description: Confocal and STED microscopy image of methanol fixed U2OS cell expressing microtubule binding fusion protein SNAP-Cep41. SNAP-tagged protein is stained with cell permeable SiR-SNAP substrate before fixation.

Cover Courtesy: The image is courtesy of Gražvydas Lukinavičius and Kai Johnsson, École Polytechnique Fédérale de Lausanne

Printed on acid-free paper

Humana Press is a brand of Springer
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Recombinant DNA technologies have revolutionized the way biologists study and manipulate proteins. The ability to produce chimeric proteins by inserting a peptide sequence before, after, or within a protein through genetic manipulation has led to the development of a multitude of techniques that render a protein of interest unique merely by adding an encoded label. Prominent examples are the introduction of small epitopes for immunolabeling, the use of affinity tags for protein purification, and the fusion to fluorescent proteins for imaging. The power of those approaches lies in the simplicity and absolute specificity of genetic encoding. However, the genetically encodable tags are *a priori* limited by the 20 proteogenic amino acids, which cover a very limited part of the chemical space.

This limitation is overcome by techniques that allow the covalent functionalization of a protein of interest with a synthetic probe, which includes fluorescent dyes, radiolabels, chemical cross-linkers, photoactivatable molecules, pharmacologically active compounds, toxins, synthetic biosensors, or nanoparticles [1, 2]. The application of such artificial synthetic objects in living cells or living organisms opens new avenues for studying and manipulating protein function in living systems. The issue of labeling specificity becomes critical for labeling *in situ* in a physiological context or in the cases where well-defined chemically modified biomolecules are desired. Classical reactive labeling techniques, however, are usually not selective enough for this purpose. This problem has been overcome over the last 15 years based on the pioneering work of Roger Y. Tsien and his group, and today various covalent labeling techniques are available that are perfectly site-specific and can be applied in the context of cells and organisms.

Today, the field as a whole is at an exciting stage: while some site-specific labeling approaches are now fully mature and well adopted by the molecular and cell biology community, new approaches and ingenious ways of applying existing approaches continue to emerge. The creative application of site-specific protein labeling techniques in cell biology beyond simple fluorescent labeling requires both a biologist's knowledge of biological problems and an organic chemist's understanding of the opportunities and problems involved in generating a custom label for the problem in question. *Methods of Site-Specific Protein Labeling* is directed at scientists from all fields that want to get a better understanding of labeling techniques. In particular, it aims at providing researchers interested in such techniques with advice on how to choose the most appropriate labeling method for their biological question and information on general considerations and problems involved in the design, the generation, and the application of the corresponding organic molecules used for the labeling step.

The first chapters deal with the background and basic considerations of site-specific protein labeling. As often, the historical perspective is insightful: In Chapter 1, B. Albert Griffin, Stephen R. Adams, and Roger Y. Tsien provide a highly interesting recollection of why and how they came to invent the FlAsH-tag. Chapter 2, written from the industrial perspective by Lukas Leder from Novartis, provides an overview of applications of labeled proteins in assays that are common in the industry, and Lukas Leder shares experiences that his laboratory made with adopting site-specific protein labeling. Chapter 3 was motivated

by a recurring issue in the site-specific labeling of intracellular proteins: whether the compound used for labeling can at all cross the cell membrane in sufficient amounts to enable intracellular reaction. A lack of permeability can render the most creative labeling molecule useless, which can be painful if it is realized only after the synthesis has been performed. The chapter, written by Nicole Yang and Marlon J. Hinner, provides a comprehensive overview of the factors that govern membrane translocation not only for small molecules and peptides but also for proteins. As the last of the overview articles, Chapter 4 by Ivan Correa provides a broad overview of general considerations for the design of labeling molecules, exemplified by SNAP-tag and CLIP-tag technology. His chapter includes a number of protocols that should be of high interest for chemists and nonchemists alike.

The chapters that follow cover the most relevant methods of site-specific protein labeling with selected applications. The techniques described include tag-based methods (which can be further subdivided), methods that rely on the incorporation of unnatural amino acids during protein translation, and methods that work specifically on native, untagged proteins.

In tag-based methods, a protein of interest is fused to a peptide sequence that acts as a genetic anchor for the attachment of the probe. This peptide sequence can contain just a few residues or be a full protein. Depending on the size of the tag and whether the tag requires an added enzyme to be linked to the label of interest, tag-based methods can be grouped into self-labeling tags, self-labeling proteins, and enzyme-mediated labeling of tags [2].

Developed by Roger Y. Tsien and coworkers, the archetype of a self-labeling tag is the tetracysteine tag which can specifically react with biarsenical compounds [3]. A recently developed self-labeling tag is described in the contribution of Lina Cui and Jianghong Rao (Chapter 5), which presents how a single terminal cysteine can be exploited for site-specific labeling with cyanobenzothiazole derivatives. The contribution of Thomas K. Berger and Ehud Y. Isacoff (Chapter 6) demonstrates additionally how well positioned cysteines within a cell-membrane receptor can be functionalized with thiol-linked environment-sensitive dyes to measure protein motion in ion channels in real time.

Relying on an uncatalyzed chemical reaction can limit the kinetics of the labeling step, and using short peptides as a recognition sequence may also lead to a less-than-perfect selectivity of labeling. These limits can be overcome with self-labeling protein tags that rely on a rapid and selective, catalyzed labeling reaction. The contributions from Gražvydas Lukinavičius, Luc Reymond, and Kai Johnsson (Chapter 7) and from Hélène A. Benink and Marjeta Urh (Chapter 8) describe aspects of two self-labeling proteins that are commercially available, SNAP-tag and HaloTag. Lukinavičius et al. show in particular how the SNAP-tag technology can be exploited in the context of super-resolution microscopy. Split inteins are another example for a catalyzed reaction that can be exploited for site-specific protein labeling. In two chapters from the group of Henning Mootz, Julian Matern et al. (Chapter 9) and Anne-Lena Bachmann et al. (Chapter 10) present two different approaches that exploit split inteins for attaching a small peptide functionalized with a chemical probe to a protein of interest.

The size of the added tag sometimes being a concern, strategies combining the small size of a short peptide sequence with the speed and high specificity of protein-catalyzed labeling have also been designed. In these methods, the labeling reaction is trimolecular and involves a transferase enzyme, the molecule used for labeling, and the recognition (acceptor) peptide sequence. Here, the transferase enzyme can be added in medium or needs to be coexpressed if intracellular labeling is required. The enzyme-mediated labeling of tags is described for Sfp-mediated labeling—applied in phage display—by Bo Zhao et al. (Chapter 11), for BirA-mediated labeling by Michael Fairhead and Mark Howarth (Chapter 12), and for Sortase-mediated labeling by Max Popp (Chapter 13).

Fusing a peptide or protein tag to the protein of interest is not required in techniques relying on unnatural amino acid incorporation during protein synthesis. The inserted unnatural amino acid plays the role of the molecular anchor in this case. Since the size of the side chain of the unnatural amino acid can be limited by the cell's protein translation machinery, often a small chemical functionality is introduced to which a chemical probe can be tethered in a second step using various bioorthogonal chemical "click" reactions. Using this methodology, the contribution of Peter Landgraf, Elmer R. Antileo, Erin M. Schuman, and Daniela C. Dieterich (Chapter 14) illustrates how metabolic labeling can be used to mark newly synthesized proteomes. The contribution of Kathrin Lang, Lloyd Davis, and Jason W. Chin (Chapter 15) describes the recent development of methods to fully genetically encode these unnatural "anchor" amino acids in order to be able to selectively label a single protein at a specific residue in living mammalian cells.

The "Holy Grail" in protein labeling is to be able to specifically target any native, non-tagged protein with a chemical probe in a physiological context. The two final chapters are reserved for this topic and are written by Itaru Hamachi with coworkers Tomonori Tamura (Chapter 16) and Shinya Tsukiji (Chapter 17), respectively. They describe two related approaches to how native protein labeling can be achieved by relying on labeling probes made of three parts, (1) a recognition moiety, binding selectively to the native protein of interest, (2) the probe to be attached, and (3) a reactive group, which can react with nucleophilic residues on the protein surface. While this reactive group is in principle capable of labeling any protein in a mixture, selectivity is achieved due to close proximity of the reactive group to the protein of interest, enforced by the recognition moiety.

In putting together this edition, we have attempted to include what we perceive as the currently most relevant and best established labeling methods across the different general methodologies. A number of important techniques are not presented, however, because detailed reviews and protocols have been recently published elsewhere. This includes the tetracysteine tag [3], lipoic acid-mediated labeling [4], labeling based on the genetically encoded aldehyde tag [5], and transglutaminase-based labeling [6]. While we have not attempted to include examples for every possible application of site-specific protein labeling, the chapters are nonetheless designed to provide guidance on the limits and possibilities of each technique and references to applications that have been described in the literature. For more information on applications and a comparative analysis of the various techniques, as well as introductions to other labeling methods not included here, we invite the readers to consult recent reviews on site-specific labeling [1, 2].

Finally, we thank all the authors that have contributed to this edition of *Methods in Molecular Biology*. We hope that both authors and readers will find this compendium useful and that it will support the further development of creative ideas in the field and facilitate making site-specific protein labeling a standard, widely used lab technique.

Paris, France
Munich, Germany

Arnaud Gautier
Marlon J. Hinner

References

- O'Hare H, Johnsson K, Gautier A (2007) Chemical probes shed light on protein function. *Curr Opin Struct Biol* 17:488–494. doi: [10.1016/j.sbi.2007.07.005](https://doi.org/10.1016/j.sbi.2007.07.005)
- Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21:766–776. doi: [10.1016/j.copbio.2010.09.011](https://doi.org/10.1016/j.copbio.2010.09.011)

3. Hoffmann C, Gaietta G, Zürn A, Adams SR, Terrillon S, Ellisman MH, Tsien RY, Lohse MJ (2010) Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat Protoc* 5:1666–1677. doi: [10.1038/nprot.2010.129](https://doi.org/10.1038/nprot.2010.129)
4. Uttamapinant C, Sanchez MI, Liu DS, Yao JZ, Ting AY (2013) Site-specific protein labeling using PRIME and chelation-assisted click chemistry. *Nat Protoc* 8:1620–1634. doi: [10.1038/nprot.2013.096](https://doi.org/10.1038/nprot.2013.096)
5. Rabuka D, Rush JS, deHart GW, Wu P, Bertozzi CR (2012) Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nat Protoc* 7:1052–1067. doi: [10.1038/nprot.2012.045](https://doi.org/10.1038/nprot.2012.045)
6. Dennler P, Schibli R, Fischer E (2013) Enzymatic antibody modification by bacterial transglutaminase. *Methods Mol Biol* 1045:205–215. doi: [10.1007/978-1-62703-541-5_12](https://doi.org/10.1007/978-1-62703-541-5_12)

Contents

Preface	v
Contributors	xi
1 How FlAsH Got Its Sparkle: Historical Recollections of the Biarsenical-Tetracysteine Tag	1
<i>B. Albert Griffin, Stephen R. Adams, and Roger Y. Tsien</i>	
2 Site-Specific Protein Labeling in the Pharmaceutical Industry: Experiences from Novartis Drug Discovery	7
<i>Lukas Leder</i>	
3 Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins	29
<i>Nicole J. Yang and Marlon J. Hinner</i>	
4 Considerations and Protocols for the Synthesis of Custom Protein Labeling Probes	55
<i>Ivan R. Corrêa Jr.</i>	
5 2-Cyanobenzothiazole (CBT) Condensation for Site-Specific Labeling of Proteins at the Terminal Cysteine Residues	81
<i>Lina Cui and Jianghong Rao</i>	
6 Fluorescent Labeling for Patch-Clamp Fluorometry (PCF) Measurements of Real-Time Protein Motion in Ion Channels	93
<i>Thomas K. Berger and Ehud Y. Isacoff</i>	
7 Fluorescent Labeling of SNAP-Tagged Proteins in Cells	107
<i>Gražvydas Lukinavičius, Luc Reymond, and Kai Johnsson</i>	
8 HaloTag Technology for Specific and Covalent Labeling of Fusion Proteins	119
<i>Hélène A. Benink and Marjeta Urb</i>	
9 Ligation of Synthetic Peptides to Proteins Using Semisynthetic Protein <i>trans</i> -Splicing	129
<i>Julian C.J. Matern, Anne-Lena Bachmann, Ilka V. Thiel, Gerrit Volkmann, Alexandra Wasmuth, Jens Binschik, and Henning D. Mootz</i>	
10 Chemical-Tag Labeling of Proteins Using Fully Recombinant Split Inteins	145
<i>Anne-Lena Bachmann, Julian C.J. Matern, Vivien Schütz, and Henning D. Mootz</i>	
11 Phage Selection Assisted by Sfp Phosphopantetheinyl Transferase-Catalyzed Site-Specific Protein Labeling	161
<i>Bo Zhao, Keya Zhang, Karan Bhuripanyo, Yiyang Wang, Han Zhou, Mengnan Zhang, and Jun Yin</i>	
12 Site-Specific Biotinylation of Purified Proteins Using BirA	171
<i>Michael Fairhead and Mark Howarth</i>	

13	Site-Specific Labeling of Proteins via Sortase: Protocols for the Molecular Biologist	185
	<i>Maximilian Wei-Lin Popp</i>	
14	BONCAT: Metabolic Labeling, Click Chemistry, and Affinity Purification of Newly Synthesized Proteomes	199
	<i>Peter Landgraf, Elmer R. Antileo, Erin M. Schuman, and Daniela C. Dieterich</i>	
15	Genetic Encoding of Unnatural Amino Acids for Labeling Proteins	217
	<i>Kathrin Lang, Lloyd Davis, and Jason W. Chin</i>	
16	Labeling Proteins by Affinity-Guided DMAP Chemistry	229
	<i>Tomonori Tamura and Itaru Hamachi</i>	
17	Ligand-Directed Tosyl Chemistry for Selective Native Protein Labeling In Vitro, In Cells, and In Vivo	243
	<i>Shinya Tsukiji and Itaru Hamachi</i>	
	<i>Index</i>	265

Contributors

STEPHEN R. ADAMS • *Department of Pharmacology, University of California San Diego, La Jolla, CA, USA*

ELMER R. ANTILEO • *Institute of Pharmacology and Toxicology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany; Emmy Noether Group Neuralomics, Leibniz Institute for Neurobiology, Magdeburg, Germany*

ANNE-LENA BACHMANN • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*

HÉLÈNE A. BENINK • *Promega Corporation, Madison, WI, USA*

THOMAS K. BERGER • *Research Center Caesar, Bonn, Germany*

KARAN BHURIPANYO • *Department of Chemistry, University of Chicago, Chicago, IL, USA*

JENS BINSCHIK • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*

JASON W. CHIN • *Medical Research Council—Laboratory of Molecular Biology, Department of Protein and Nucleic Acid Chemistry, Centre for Chemical and Synthetic Biology, Cambridge, UK*

IVAN R. CORRÉA JR. • *New England Biolabs Inc., Ipswich, MA, USA*

LINA CUI • *Molecular Imaging Program at Stanford, Department of Radiology, School of Medicine, Stanford University, Stanford, CA, USA; Department of Chemistry, School of Medicine, Stanford University, Stanford, CA, USA*

LLOYD DAVIS • *Medical Research Council—Laboratory of Molecular Biology, Department of Protein and Nucleic Acid Chemistry, Centre for Chemical and Synthetic Biology, Cambridge, UK*

DANIELA C. DIETERICH • *Institute of Pharmacology and Toxicology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany; Emmy Noether Group Neuralomics, Leibniz Institute for Neurobiology, Magdeburg, Germany*

MICHAEL FAIRHEAD • *Department of Biochemistry, University of Oxford, Oxford, UK*

ARNAUD GAUTIER • *Department of Chemistry, École Normale Supérieure, Paris, France*

B. ALBERT GRIFFIN • *Independent Researcher, San Diego, CA, USA*

ITARU HAMACHI • *Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto, Japan; Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Tokyo, Japan*

MARLON J. HINNER • *Sensitive Farbstoffe GbR, Munich, Germany*

MARK HOWARTH • *Department of Biochemistry, University of Oxford, Oxford, UK*

EHUD Y. ISACOFF • *Department of Molecular and Cell Biology and Helen Wills Neuroscience Institute, 142 Life Sciences Addition, University of California Berkeley, Berkeley, CA, USA; Physical Bioscience Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

KAI JOHNSSON • *Ecole Polytechnique Fédérale de Lausanne, Institute of Chemical Sciences and Engineering (ISIC), National Centre of Competence in Research (NCCR) in Chemical Biology, Lausanne, Switzerland*

PETER LANDGRAF • *Institute of Pharmacology and Toxicology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany; Emmy Noether Group Neuralomics, Leibniz Institute for Neurobiology, Magdeburg, Germany*

- KATHRIN LANG • *Medical Research Council—Laboratory of Molecular Biology, Department of Protein and Nucleic Acid Chemistry, Centre for Chemical and Synthetic Biology, Cambridge, UK*
- LUKAS LEDER • *Center for Proteomic Chemistry, Novartis Campus, Basel, Switzerland*
- GRAŽVYDAS LUKINAVIČIUS • *Ecole Polytechnique Fédérale de Lausanne, Institute of Chemical Sciences and Engineering (ISIC), National Centre of Competence in Research (NCCR) in Chemical Biology, Lausanne, Switzerland*
- JULIAN C.J. MATERN • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*
- HENNING D. MOOTZ • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*
- MAXIMILIAN WEI-LIN POPP • *Department of Biochemistry and Biophysics, School of Medicine and Dentistry, University of Rochester, Rochester, NY, USA*
- JIANGHONG RAO • *Molecular Imaging Program at Stanford, Departments of Radiology and Chemistry, School of Medicine, Stanford University, Stanford, CA, USA*
- LUC REYMOND • *Ecole Polytechnique Fédérale de Lausanne, Institute of Chemical Sciences and Engineering (ISIC), National Centre of Competence in Research (NCCR) in Chemical Biology, Lausanne, Switzerland*
- ERIN M. SCHUMAN • *Max Planck Institute for Brain Research, Frankfurt am Main, Germany*
- VIVIEN SCHÜTZ • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*
- TOMONORI TAMURA • *Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan*
- ILKA V. THIEL • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*
- ROGER Y. TSIEN • *Department of Pharmacology, Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA, USA; Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA, USA*
- SHINYA TSUKIJI • *Top Runner Incubation Center for Academia-Industry Fusion, Nagaoka University of Technology, Nagaoka, Niigata, Japan*
- MARJETA URH • *Promega Corporation, Madison, WI, USA*
- GERRIT VOLKMANN • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*
- YIYANG WANG • *Department of Chemistry, Georgia State University, Atlanta, GA, USA*
- ALEXANDRA WASMUTH • *Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Muenster, Münster, Germany*
- NICOLE J. YANG • *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*
- JUN YIN • *Department of Chemistry, Georgia State University, Atlanta, GA, USA*
- KEYA ZHANG • *Department of Chemistry, University of Chicago, Chicago, IL, USA*
- MENG NAN ZHANG • *Department of Chemistry, Georgia State University, Atlanta, GA, USA*
- BO ZHAO • *Department of Chemistry, University of Chicago, Chicago, IL, USA*
- HAN ZHOU • *Department of Chemistry, Georgia State University, Atlanta, GA, USA*

Chapter 1

How FlAsH Got Its Sparkle: Historical Recollections of the Biarsenical-Tetracysteine Tag

B. Albert Griffin, Stephen R. Adams, and Roger Y. Tsien

Abstract

The biarsenical-tetracysteine tagging system was the first of (and inspiration for) the now numerous methods for site-specifically labeling proteins in living cells with small molecules such as fluorophores. This historical recollection describes its conception and the trial-and-error chemical development required to become a versatile technique.

Key words FlAsH, ReAsH, Biarsenical, Tetracysteine, Protein labeling, Genetically targeted

Prior to the introduction of GFP, the ability to label and image specific proteins in living cells was severely limited and time-consuming. The protein(s) of interest were isolated from tissue or, as a recombinant protein in bacteria, labeled with a fluorescent dye after often considerable trial and error [1, 2], and then microinjected into the cytoplasm of suitably large and robust tissue culture cells. Even after the introduction of GFP [3, 4], labeling with different colors or spectroscopic phenotypes not limited to fluorescence was an unsolved problem. A simple method for modifying an expressed protein in any cell was desirable, but how could one protein be made more chemically reactive than the many others in a cell? The biarsenical-tetracysteine tag was inspired by a talk given by Jack Dixon at a FASEB meeting in 1994 on protein tyrosine phosphatases, in which he explained that their surprisingly specific inhibition by phenyl arsenoxide resulted from the requirement for vicinal cysteines and more importantly, the rarity of such sequences in proteins. Only up to a few tens of cysteine pairs were labeled by phenyl arsenoxide. This remark suggested (to R.Y.T.) that if two vicinal cysteines were better than one, then two pairs of cysteine (four total) might be still better. In other words, the desired chemical specificity might be achievable by incorporating tandem vicinal cysteines in expressed recombinant proteins in cells and labeling

with a membrane-permeable fluorophore bearing two appropriately spaced arsenoxides. To prevent biarsenicals from binding to endogenous vicinal cysteines and thiols such as lipoic acid cofactors, one of the known 1,2-dithiol arsenic antidotes would be added at a concentration that could not compete with the expected higher tetracysteine-biarsenical affinity.

The peptide portion of the pair should be as small as possible to minimize potential disruption of the labeled protein's function. At the same time, it would be advantageous for the peptide to have some secondary structure, to mitigate the entropy cost of two arsenic atoms constraining four cysteines. The obvious candidate for a small peptide with secondary structure was an alpha-helix. Based on recently published studies on the role of natural amino acids in helix formation [5–7] we designed the following alanine-rich peptide: Acetyl-WEAAAREACCRECCARA-amide. Three arginine-glutamate pairs at i and $i+4$ positions form salt bridges, each across one turn of the helix. Glutamates toward the N-terminus and arginines toward the C-terminus interact favorably with the helix dipole. Cysteines are not strong alpha-helix formers by themselves, but induce alpha-helix formation in the i and $i+4$ positions when complexed with divalent cations [8]. Binding of the first arsenic to a pair of cysteines should nucleate an alpha-helix, positioning the other two cysteines favorably to bind the other arsenic. An even shorter peptide, without the N-terminal turn, was also prepared but proved to be difficult to maintain in the reduced state for purification and binding studies.

To validate the design, two dicysteine peptides were prepared, modeling two ways in which a single arsenic could bind to the tetracysteine binding site: i and $i+1$ (Acetyl-WEAAARECCARA-amide) and i and $i+4$ (Acetyl-WEACARECAARA-amide). Phenylarsine oxide (PAO) bound to both peptides as demonstrated by HPLC and MS. As expected, a single equivalent of the dithiol antidote 2,3-dimercaptopropanol (British anti-Lewisite or BAL) completely reversed binding. CD spectra confirmed dose-dependent alpha-helix formation when the i and $i+4$ peptide was titrated with 4-arsenophenylaminosulfonic acid, a more water-soluble analogue of PAO. Interestingly, in equilibrium experiments, PAO showed a preference for binding to the i and $i+1$ peptide. We reasoned that the arsenic atom might be too small to comfortably bridge the helix turn and perhaps a larger atom would be advantageous.

Phenylstibonoxide [9] was prepared to test this hypothesis. Antimony is larger (ionic radius of 0.76 Å vs. 0.58 Å) and shares much chemistry with arsenic. The antimony compound bound to both model peptides. In equilibrium experiments, phenylstibonoxide showed a preference for binding across alpha-helix turns (i and $i+4$) instead of to adjacent cysteines (i and $i+1$), opposite of the behavior of PAO, supporting the hypothesis that arsenic might not be sterically ideal. An organobismuth compound also bound to the i and $i+4$ peptide but was not further characterized. Since toxicity and biological stability of organoantimony and organobismuth

compounds were not as well studied as those of arsenic, the heavier group 15 elements would be pursued only if arsenic compounds failed to deliver the expected high tetracysteine affinity. We then set about the task of synthesis and testing of biarsenicals.

At the time of this research, reliable software was not available for modeling arsenic and peptides together, so we relied on Corey–Pauling–Koltun models to estimate that the arsenics should be spaced between 5 and 6 Å apart to fit well to the peptide. The first biarsenical compound tested for binding to the tetracysteine peptide was m-phenylenebisarsenoxide [10–12], in which the arsenics are 5.8 Å apart. There are no degrees of rotational freedom and a protruding C–H bond in the spacer between them. MS of the reaction mixture showed some of the expected mass, but also a peak that could only arise from a 2:2 or larger complex. After a few hours, no peaks were seen above noise. Our interpretation was that the early peaks were temporary kinetic products that went on to form larger polymers which precipitated. In designing the biarsenical probe, we had to be either incredibly lucky to get the spacing and orbital orientations correct, or include some degree of flexibility between the arsenics so that the molecule might adjust to the tetracysteine binding site. The first attempt was a shot in the dark that missed. Subsequent compounds included at least one degree of rotational freedom.

In a second molecule, 1,1'-bisarsenosoferroocene [13], an arsenic was attached to each cyclopentadienyl ring (Fig. 1a) via a mercurial intermediate. Rotation of the rings around the sandwiched iron ion would allow the probe to adjust to the tetracysteine binding site. The ferrocene compound induced a clean shift in HPLC retention time of the tetracysteine peptide when mixed stoichiometrically in neutral aqueous buffer. A mass spectrum of the collected fraction gave the expected molecular weight, clean of any peaks that would indicate larger complexes. However, two equivalents of BAL, one for each arsenic, rapidly returned the peptide to its original HPLC retention time. A quantitative, stable 1:1 bisarsenical/tetracysteine peptide complex had been achieved, but the hypothesized enhanced affinity that would make the system

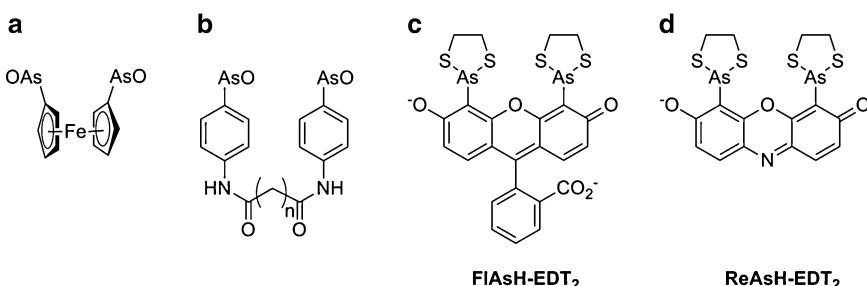


Fig. 1 The evolution of biarsenicals designed to bind tetracysteine (CCXXCC)-containing peptides. (a) 1,1'-bisarsenosoferroocene, an unsuccessful biarsenical. (b) Amides formed from 4-arsenosaniline (also known as 4-aminophenylarsenoxide) and dicarboxylic acids, also unsuccessful. (c) FIAsh-EDT₂, the first successful biarsenical. (d) ReAsH-EDT₂, another successful biarsenical

useful in live cells was lacking. It would only be a matter of some tweaking of the spacer to achieve our goal.

Eleven compounds later the goal was no closer. Biarsenicals, some aliphatic, some aromatic, with a variety of spacers in a range of flexibilities were synthesized. Four of these bound the tetracysteine peptide, but all were easily dissociated by stoichiometric concentrations of small dithiols. These compounds were prepared by linking 4-arsenosaniline via amide bonds (Fig. 1b) or by arsenic trioxide action on aliphatic bromides. A fresh approach was needed. We looked for commercially available compounds containing mercuric acetate or chloride groups, intending to employ the synthetic methodology used to prepare the ferrocene compound. One such compound from Aldrich was fluorescein mercuric acetate (FMA), which was sold with the compound name 2',7'-bis(acetoxymercuri) fluorescein. In fluorescein nomenclature, substituents on the 2' and 7' positions protrude from opposite sides of the molecule, as displayed in the Aldrich catalog. This was of no use since the arsenics needed to be closer together. The Aldrich Library of ¹H FT-NMR Spectra was sitting on the benchtop and for unknown reasons we flipped it open and turned to the FMA page. It was immediately apparent from proton coupling that the hydrogens in the fluorophore region of the molecule were ortho to each other and not para as in the structure as drawn, which meant that the mercury atoms were in the 4' and 5' positions, close and aligned parallel to each other on the same side of the molecule. Although unlikely that a compound with no flexibility between the arsenics would succeed where our first rigid compound had failed, the chemistry was straightforward and the compound was attractive since a useful chromophore was built in the biarsenical (as a footnote, Aldrich was notified of the error in their catalog and it was subsequently corrected). Isolated yield from early preparations was quite low (<1 %). Separation of the desired compound from the complex mixture was accomplished by complexing the reaction products with aliphatic dithiols, 1,3-propanedithiol (PDT) or 1,2-ethanedithiol (EDT). Mono-substituted arsenoxides migrate cleanly on a thin-layer chromatogram when complexed with small dithiols that form a tight S-As-S containing ring, shielding the arsenic from silica oxygens. BAL was not used because it was thought that its chirality might give rise to multiple isomers, complicating product characterization. To our great delight, the compound not only bound the tetracysteine peptide, but persisted in the presence of a moderate excess of BAL. We promptly named it *Fluorescein Arsenic Helix-binder* (FlAsH, Fig. 1c). To our even greater surprise FlAsH was totally nonfluorescent as the bis-1,2-ethanedithiol (EDT) adduct. But when EDTs were replaced by the tetracysteine peptide, it becomes brightly fluorescent. Our explanation, buttressed by molecular models, was that the thiol groups could become coplanar with the fluorescein in the ethanedithiol complex, allowing quenching in the excited state,

but were constrained to twist out of the fluorescein plane in the peptide complex.

Once the remarkable properties of the compound were confirmed, we set about improving the synthetic yield. Two alternate methods were explored, arsenic trichloride action on 4',5'-bis-diazoniumfluorescein (Bart conditions) and direct arsonation of fluorescein using arsenic trioxide [14–16]. While both techniques had worked in our hands on other substrates, neither produced detectable quantities of the desired product. Revisiting the bis-mercuric compound, Prof. Jay Siegel suggested that a catalytic amount of palladium(II) might facilitate substitution of arsenic for mercury. This added reagent did indeed greatly improve the yield and is the method still used today to synthesize FlAsH-EDT₂. Overall, our first paper on the remarkable partnership between FlAsH and tetracysteine peptides was (surprisingly) accepted in *Science* [17] and launched the field of site-specific small-molecule labeling of proteins in living cells.

For final confirmation that the labeling system had worked in vitro as originally envisioned, CD spectra of the pair were collected to confirm the presence of an alpha-helix. When there was no evidence of helix in the spectra, we were at first perplexed. But then we concluded that the helix must be present but its CD signature was obscured by fluorescein absorbance. The PAO-dicysteine peptide adduct was helical, so we believed FlAsH-tetracysteine peptide should be as well! Later, after peptides containing a helix-breaking proline-glycine (PG) between the cysteine pairs showed a higher affinity for FlAsH [18], and sufficient material with the resorufin-based biarsenical ReAsH (Fig. 1d) was available for NMR spectroscopy [19], those peptides were found to have adopted a beta turn conformation rather than an alpha-helix. Fortunately “helix-binder” and “hairpin-binder” both start with “h”, so the acronym did not have to be revised.

Studies from our group [17–29] and others have described many applications of this biarsenical-tetracysteine tagging system, summarized in a recent review [30], which is why this article is a historical reminiscence.

References

1. Taylor DL, Amato P, Luby-Phelps K et al (1984) Fluorescent analog cytochemistry. *Trends Biochem Sci* 9:88–91
2. Adams SR, Bacska BJ, Taylor SS, Tsien RY (1993) Optical probes for cyclic AMP. In: Mason WT, Relf G (eds) *Fluorescent probes for biological activity of living cells*. Academic, London, pp 133–149
3. Chalfie M, Tu Y, Euskirchen G et al (1994) Green fluorescent protein as a marker for gene expression. *Science* 263:802–805
4. Heim R, Prasher DC, Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* 91:12501–12504
5. O’Neil KT, DeGrado WF (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:646–651
6. Merutka G, Shalongo W, Stellwagen E (1991) A model peptide with enhanced helicity. *Biochemistry* 30:4245–4248

7. Merutka G, Stellwagen E (1991) Effect of amino acid ion pairs on peptide helicity. *Biochemistry* 3:1591–1594
8. Ghadiri MR, Cho C (1990) Secondary structure nucleation in peptides. Transition metal ion stabilized alpha-helices. *J Am Chem Soc* 112:1630–1632
9. Doak GO, Steinman HG (1946) The preparation of stibonic acids by the scheller reaction. *J Am Chem Soc* 68:1987–1989
10. Steinkopf W, Schmidt S, Penz H (1934) Zur Kenntnis organischer Arsenverbindungen, XVII. Über ein polymeres Phenylarsinoxid und über Reaktionen der m-phenylen-diarsinsäure. *J Prakt Chem* 141:301–305
11. Lieb H (1921) Aromatische Diarsinsäuren und deren reduktionsprodukte (I. Mitteilung). *Berichte d D Chem Gesellschaft* 54:1511–1519
12. Doak GO (1940) A modified Bart reaction. *J Am Chem Soc* 62:167–168
13. Spang C, Edelmann FT, Noltemeyer M, Roesky HW (1989) Anorganische Ringsysteme mit ferrocenyl-substituenten. *Chem Ber* 122:1247–1254
14. Michaelis A (1908) Über die p-dimethylaminophenyl-arsinsäure (dimethylatoxyl). *Berichte d D Chem Gesellschaft* 41:1514–1516
15. Jacobs WA, Heidelberger M (1919) The isomeric hydroxyphenylarsonic acids and the direct arsenation of phenol. *J Am Chem Soc* 41:1440–1450
16. Benda L (1909) Über o-amino-arylarsinsäuren. *Berichte d D Chem Gesellschaft* 42:3619–3622
17. Griffin BA, Adams SR, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281:269–272
18. Adams SR, Campbell RE, Gross LA et al (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J Am Chem Soc* 124:6063–6076
19. Madani F, Lind J, Damberg P et al (2009) Hairpin structure of a biarsenical-tetracysteine motif determined by NMR spectroscopy. *J Am Chem Soc* 131:4613–4615
20. Griffin BA, Adams SR, Jones J, Tsien RY (2000) Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Methods Enzymol* 327:565–578
21. Gaietta G, Deerinck TJ, Adams SR et al (2002) Multicolor and electron microscopic imaging of connexin trafficking. *Science* 296:503–507
22. Tour O, Meijer RM, Zacharias DA et al (2003) Genetically targeted chromophore-assisted light inactivation. *Nat Biotechnol* 21:1505–1508
23. Ju W, Morishita W, Tsui J et al (2004) Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nat Neurosci* 7:244–253
24. Hoffmann C, Gaietta G, Bünenmann M et al (2005) A FlAsH-based FRET approach to determine G protein-coupled receptor activation in living cells. *Nat Methods* 2:171–176
25. Martin BR, Giepmans BN, Adams SR, Tsien RY (2005) Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* 23:1308–1314
26. Gaietta GM, Giepmans BN, Deerinck TJ et al (2006) Golgi twins in late mitosis revealed by genetically encoded tags for live cell imaging and correlated electron microscopy. *Proc Natl Acad Sci U S A* 103:17777–17782
27. Tour O, Adams SR, Kerr RA et al (2007) Calcium green FlAsH as a genetically targeted small-molecule calcium indicator. *Nat Chem Biol* 3:423–431
28. Martin BR, Deerinck TJ, Ellisman MH et al (2007) Isoform-specific PKA dynamics revealed by dye-triggered aggregation and DAKAP1alpha-mediated localization in living cells. *Chem Biol* 14:1031–1042
29. Adams SR, Tsien RY (2008) Preparation of the membrane-permeant biarsenicals FlAsH-EDT₂ and ReAsH-EDT₂ for fluorescent labeling of tetracysteine-tagged proteins. *Nat Protoc* 3:1527–1534
30. Hoffmann C, Gaietta G, Zürn A et al (2010) Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat Protoc* 5:1666–1677

Chapter 2

Site-Specific Protein Labeling in the Pharmaceutical Industry: Experiences from Novartis Drug Discovery

Lukas Leder

Abstract

Chemically modified proteins play an important role in several fields of pharmaceutical R&D, starting from various activities in drug discovery all the way down to biopharmaceuticals with improved properties such as antibody–drug conjugates. In the first part of the present chapter the significance and use of labeled proteins in biophysical methods, biochemical and cellular assays, *in vivo* imaging, and biopharmaceuticals is reviewed in general. In this context, the most relevant methods for site-specific modification of proteins and their application are also described. In the second part of the chapter, in-house (Novartis) results and experience with different techniques for selective protein labeling are discussed, with a focus on chemical or enzymatic (Avi-tag) biotinylation of proteins and their application in biophysical and biochemical assays. It can be concluded that while modern methods of site-specific protein labeling offer new possibilities for pharmaceutical R&D, classical methods are still the mainstay mainly due to being well established. However, site-specific protein labeling is expected to increase in importance, in particular for antibody–drug conjugates and other chemically modified biopharmaceuticals.

Key words Biophysical methods, Biochemical assay, Cellular assays, *In vivo* imaging, Biopharmaceutical, Antibody–drug conjugates, Biotin ligase, Avi-tag, SNAP-tag, Transglutaminase, Lipoic acid ligase, Click chemistry, Sortase, Phosphopantetheinyl transferase

1 Introduction

Modern drug discovery in pharmaceutical research is a highly diverse, protracted, and intricate process encompassing many activities such as target identification/validation, development of *in vitro* assays, screening for active compounds, structural studies, biophysical methods, medicinal chemistry, and *in vivo* pharmacology. Several of these disciplines are absolutely dependent on the supply of purified proteins in order to deliver meaningful results. For many applications like enzymatic assays or structure determination, nonmodified, native proteins are perfectly suitable. On the contrary, chemically modified proteins are needed or at least preferred for various experimental techniques such as certain

types of fluorescence-based assays, biophysical methods which rely on immobilization of proteins, *in vivo* imaging approaches, or chemically modified biopharmaceuticals with improved properties like antibody–drug conjugates. Therefore pharmaceutical research labs have made use of labeled proteins for quite a long time. Originally, the classical nonspecific approach by attachment of small molecules to reactive groups (mainly amino and thiol groups) on proteins was used exclusively for the production of labeled proteins. Because of the intrinsic disadvantages of the chemical labeling approach such as inhomogeneous incorporation of the chemical label and a potential impairment of the function and/or stability of the protein, the need for more specific and controllable methods to modify proteins was recognized. The last decade has witnessed the invention of several new techniques for site-specific protein labeling as reviewed generally in [1, 2] and more specifically for enzyme-catalyzed approaches in [3]. These latter methods include for instance the already well-established Avi-tag approach enabling site-specific attachment of biotin to a 15 aa long peptide tag catalyzed by the enzyme biotin ligase BirA as described in [4]. In analogy to the Avi-tag approach, other methods relying on short peptide tags for enzyme-catalyzed modification were developed as well. For example, recognition sequences have been designed and optimized for enzymes such as lipoic acid ligase [5, 6] or phosphopantetheinyl-transferases [7, 8] enabling the selective conjugation of lipoic acid analogues or coenzyme A derivatives to a specified Lys or Ser residue, respectively. Other enzyme-based labeling methods with short recognition sequences rely on enzymes such as transglutaminase [9, 10] or sortase [11, 12]. Finally, if no additional amino acids are tolerated or desired at all on a certain protein, incorporation of nonnatural amino acids with specific linking chemistries [13, 14] can be considered as well.

These novel technologies are increasingly being explored by the industry, and may become a part of standard procedures within drug discovery and development. However, the well-established classical reactive chemistry remains the dominant labeling method in many areas. In the following sections, I discuss several applications in which chemically modified proteins play an important role, and to what extent site-specific methods may provide an advantage over classical labeling techniques in the respective fields. In the final section, experiences made with several selected labeling techniques performed in several labs mainly in Novartis drug discovery will be described, with the aim of providing the reader with an idea whether we considered a given technique easy to implement and successful for our purposes. I am aware, of course, that this is my personal view and experience, which may therefore be biased and not congruent with the experiences of others.

2 Biophysical Methods

Biophysical methods are used to determine a direct binding of a pure target protein to a ligand which can either be another protein, a peptide, or a low molecular weight (LMW) compound. The most frequently used methods are for instance differential scanning fluorimetry, isothermal calorimetry, affinity-based mass spectroscopy, ligand or protein observed NMR spectroscopy, and techniques based on surface plasmon resonance (SPR). While the first methods are label-free as they work with proteins that are not chemically modified at all, the SPR-related approaches need a covalent or non-covalent immobilization of the target protein onto a surface as indicated in Fig. 1 and reviewed in [15, 16]. Covalent immobilization is mostly performed by nonspecific coupling of amino groups on the protein to carboxyl groups present on the dextran surface of the chip which is very similar the nonspecific protein labeling with *N*-hydroxy-succinimide (NHS) derived reagents. The nonspecific immobilization approach leads to a random orientation of the immobilized proteins which might result in a population of molecules where the binding site is not accessible or functional anymore. In addition, an efficient immobilization of a protein requires enrichment at the negatively charged SPR chip surface, which can be challenging for proteins with a low iso-electrical point. In order to circumvent these problems a strong non-covalent and/or directed covalent immobilization approaches can be a useful alternative. Ideally a strong non-covalent interaction should have a very low dissociation constant K_d in order to be reasonably stable over time. This requirement is perfectly fulfilled by the interaction between biotin and avidin/streptavidin with a K_d of 10^{-15} M, making biotinylated proteins highly important for SPR applications as described in [17, 18]. Besides classical modification of amino or

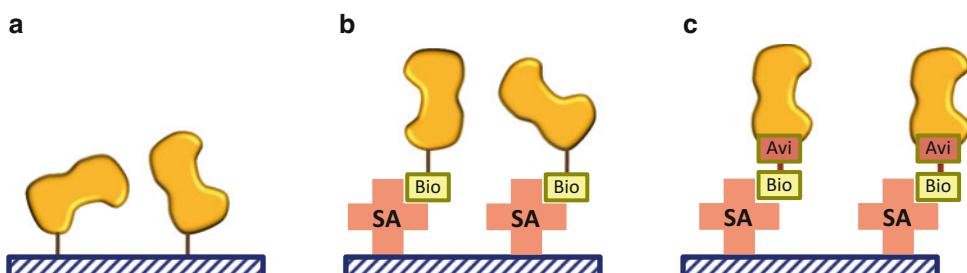


Fig. 1 Different approaches for immobilization of protein onto solid surfaces: Scheme (a) represents the direct immobilization of a protein through coupling of amino groups, resulting in random orientation. Scheme (b) shows the so-called minimal biotinylation technique, in which chemically biotinylated proteins are bound to a streptavidin surface. The biotinylated proteins are thus also oriented in a random manner but in contrast to the first method the immobilization procedure is less harsh. In scheme (c) specific biotinylated proteins (Avi-tagged) are immobilized onto streptavidin in a directed manner

thiol groups with biotin, site-specific biotinylation has become more and more important and mostly relies on the Avi-tag technology resulting usually in a complete incorporation of biotin [4]. Incomplete labeling is not problematic, as nonmodified protein molecules do not bind to the streptavidin surface and will be washed away. More recent approaches for attachment of proteins to surfaces such as click chemistry or binding of poly-histidine-tagged proteins on specific metal chelates can be considered as interesting alternatives as described in [19–22].

3 Biochemical Assays

A prominent example requiring modified proteins is the study of protein–protein interactions (PPI), where low molecular weight compounds or larger biomolecules are employed that disrupt or enhance the binding between two proteins. For measuring PPIs, proximity-based assays relying on fluorescence energy transfer (FRET, [23]) or time-resolved (TR-FRET or HTRF [24]) are widespread assay formats [25]. As depicted in Fig. 2, there are several possibilities to build up a proximity assay for a PPI, with the simplest setup being that two proteins are labeled directly with either donor or acceptor fluorophores. On the other hand, more complex setups exist in which one or more of the proteins contain “recognition handles” for helper molecules such as antibodies (epitope tags) or streptavidin (biotin). In this context, biotinylation is widely used for protein modification and can be performed either in

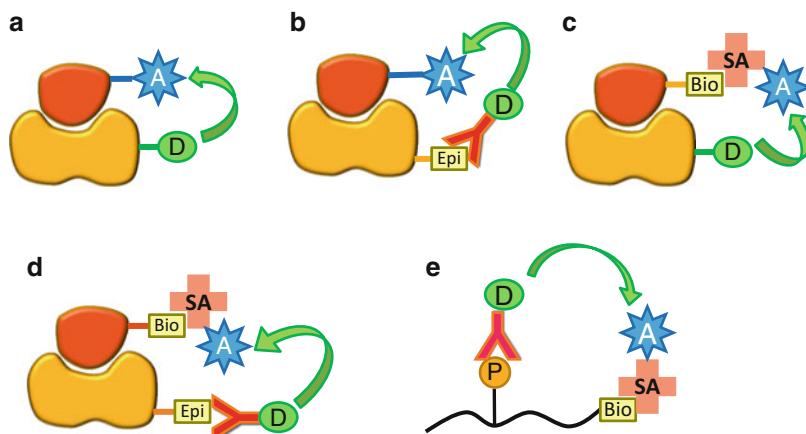


Fig. 2 Various layouts for TR-FRET assay formats in biochemical screening. Pictures (a–d) represent the mainly used formats for protein–protein or protein–peptide interaction assays, starting from directly labeled components (A) to most complex layout with two accessory tags/molecules (D). Picture (e) represents a format mainly used in protein kinase assays in which a biotinylated substrate peptide becomes phosphorylated and is recognized by a phospho-specific antibody

a specific manner (Avi-tag) or through reactive chemistry [26, 27]. The same considerations apply to helper antibodies and streptavidin with attached acceptor or donor molecules. However, this modification is usually done with chemical modification of amino groups. Since antibodies and streptavidin are large and stable proteins with many reactive amino groups (about 80–100 in the case of an mAb), classical chemical labeling remains the main method, since interference with the binding site and/or destabilization are of less relevance. In a typical layout for a PPI assay, the target protein (against which the inhibitors are searched for) is labeled with the donor (lanthanide chelate in case of the widely used TR-FRET) in order to obtain better signal/noise ratios in the read-out. Until recently these reagents were available only with limited reactive groups (as isothiocyanate derivatives for amino group labeling and iodoacetamide derivatives for thiol group labeling). However, lanthanide chelate reagents have been also developed for site-specific attachment with proteins containing a SNAP, CLIP, or Halo-tag from CisBio (<http://www.htrf.com/tag-lite-toolbox>). These so-called self-labeling protein tags allow the covalent and irreversible attachment of a large set of labels containing either benzyl-guanine/cytosine moieties reacting with alkyl-guanine/cytosine transferase (SNAP/CLIP tag [28, 29]) or chloro-alkane moieties reacting with haloalkane-dehalogenase (Halo-tag [30]). On the other side, there are many acceptor fluorophores such as Cy5 or various Alexa dyes available with a large variety of reactive chemistries making them amenable for newer site-specific labeling techniques such as for the aforementioned self-labeling tags. In cases where these large fusion tag encompassing around 200 aa might pose a problem for biochemical assays alternative approaches with shorter tags such as the aldehyde tag [31], a trans-glutaminase acceptor tag [10], or non-natural amino acids [32, 33] have been published.

4 Structural Biology

Another cornerstone in industrial drug discovery is structural biology for solving the 3D-structure of isolated drug targets and in-depth elucidation of the binding/active site of targets in complex with chemical compounds or biopharmaceuticals. Especially the latter has become a very important tool for systematic exploration of structure–activity relationship (SAR) in order to optimize hit and lead compounds in medicinal chemistry based on rational and structure-guided design. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the two state-of-the-art technologies for structural investigation of protein–drug interactions. The incorporation of NMR-active isotopes into proteins for 2D-NMR is an interesting field of application for site-specific labeling, in particular using methods of unnatural amino acid

incorporation. The standard method for incorporation is isotope labeling with ^{13}C and/or ^{15}N labeled precursor compounds such as the 20 proteinogenic amino acids in the case of eukaryotic cell culture, or more simple molecules like NH_4Cl and glucose for the metabolically more competent *E. coli* bacteria [34]. Besides labeling all amino acids uniformly, isotope labeling is also performed with selected amino acid residues having distinct spectroscopic properties such as methyl groups of the aliphatic amino acids like Met, Ile, Leu, Val, the aromatic rings of Phe, Tyr or nitrogen atoms of side chains in Trp, His, Lys, and Arg residues [35, 36]. This more selective approach allows monitoring specifically the binding of ligands if one or several of selectively labeled amino acids are located in sufficient proximity of the interaction site on the protein. As an alternative to metabolic labeling, chemical probes with NMR active moieties such as spin labels (unpaired electron in a stabilized radical), ^{19}F bearing molecules, or other chemically distinct amino acids can be incorporated into the protein [37]. Since it is advantageous to attach such NMR active labels in the proximity of the active/binding site, modification at selected residues can for instance be achieved by using incorporation of nonnatural amino acids as described in [38], mutation of suitable residues to cysteine for specific thiol modification [39], or by using the trans-glutaminase reaction [40].

5 Cellular Assays

Besides looking at isolated and purified protein targets in structural biology and biochemical assays, it is equally important to study them within a more natural environment which is by definition either the whole cell, an isolated organ, or even the whole organism. Additionally, certain important classes of targets such as GPCRs, ion channels, or large protein complexes are very challenging to be produced as isolated proteins and therefore need to be investigated within a cellular context. Visualization and tracking of proteins in response to a stimulus is performed by (immuno-)fluorescence microscopy. In the classical approach, the proteins of interest are detected with specific antibodies and visualized by fluorescent dyes that are attached directly to these antibodies or to secondary antibodies used for detection. With new techniques in microscopy and the introduction of automated acquisition and analysis of images, the so-called high-content screening (HCS) technology was developed. HCS allows the investigation of cellular imaging at medium to high-throughput and is now well established within pharmaceutical R&D as an important tool in profiling and optimization of compounds, secondary screening, and in some cases even for primary screening. In a typical HCS experiment not only the entire appearance of cells in terms of size, morphology, and organelle

distribution, but also the individual fate (expression, degradation, and translocation) of distinct proteins can be investigated at once. For the observation of individual proteins the detection with fluorescently labeled antibodies is still a widely used approach; however, the direct labeling of the target protein is gaining importance as reviewed in [41]. Especially the development of fluorescent proteins based on green fluorescent protein (GFP) with various improved and different properties such as wavelengths for excitation/emission, stability, and reactivity has enabled a lot of new possibilities in cellular imaging [42]. In addition to the genetically encoded auto-fluorescent proteins, other methods for protein labeling in a cellular environment were introduced like the self-labeling enzymes tags (SNAP-tag, Halo-tag) or small cysteine-rich peptide tags such as the FlAsH-tag [41, 43]. More recently it was also shown that a mutated version of lipoic acid ligase is able to attach a coumarin-based fluorescent dye on intracellular proteins that contain the corresponding lplA acceptor peptide sequence [44]. Especially the SNAP-tag technology has now become a quite popular alternative to the fluorescent proteins, since it has some advantages such as a greater variety of fluorescent labels with improved properties (sensitivity, stability, wavelengths) or the possibility of labeling of cell surface proteins with non-cell permeable dyes [45]. As a newer development the so-called cooper-free or strain-promoted click chemistry approach becomes interesting for cell-based applications as it utilizes reagent that do not need cooper ions for efficient coupling between azide and alkyne moieties [46] or even more recently trans-cyclooctene and tetra-azine derived labels [47, 48].

6 In Vivo Imaging

Besides imaging of single cells, whole-body imaging (animal or even human) has a very high importance not only in medical diagnostics but also within pharmaceutical research and development. One of the main goals of whole organism imaging is to determine pharmacodynamics and pharmacokinetics aspects of a given drug. This involves answers to the questions where drugs act in the body, how they reach their target, what organs are affected, at which doses side or toxicological effect becomes relevant, how long the drug stays on the target organ, and how fast it is eliminated from the body. Generally molecular imaging techniques are based on radioactive nuclides, like positron emission tomography (PET) with ¹¹C, ¹⁸F, or ¹²⁴I as the most common isotopes and single photon emission computerized tomography (SPECT) using gamma radiation emitting isotopes such as ⁹⁹Tc or ¹¹¹In. Nonradioactive methods like optical imaging with fluorescent dyes in the long wavelength range and magnetic resonance imaging (MRI) with contrast enhancing metals (e.g., Gd) are also established [49, 50].

If the drug is a biopharmaceutical, the specific label must be linked stably to the protein during the whole residence time within the organism to avoid unspecific and high background signals caused by dissociated label. The radioactive or nonradioactive metals are normally present as cations and therefore need to be integrated into a stable complex with an organic chelator such as EDTA or similar molecules. For covalent attachment of the metal-chelator complex or the fluorescent dye to the protein of interest usually classical nonspecific modification techniques are used, and as mentioned previously concerns regarding negative effects caused by unspecific labeling are of lower importance. Several radionuclide labeled mAbs currently are commercialized and used especially in diagnostic oncology [51, 52]. However, if the protein of interest is a smaller biomolecule such as an antibody fragment, protein hormone, cytokine, or growth factor, a more selective and controllable conjugation could be desirable for protein preparation with improved stability and activity. Therefore site-specific labeling approaches have also emerged for applications around *in vivo* imaging as reviewed [53]. For instance selective labeling has been achieved by reaction with engineered thiols [54, 55] or selenocysteine [56]. Further, *in vivo* imaging has been performed with proteins fused both to SNAP-tag [57] and Halo-tag [58].

7 Biopharmaceuticals

A lot of recombinant endogenous proteins and specific antibodies acting on extracellular targets became available to patients in the last decades. In particular, monoclonal antibodies in the field of oncology and autoimmune diseases such as Herceptin, Avastin, and Humira generate multibillion revenues. While in the beginning most of the proteins were produced in their native form with only cell-derived posttranslational modifications like glycosylation, there was a growing need to obtain chemically modified biopharmaceuticals with improved properties. For instance, rather small proteins with a mass below 30–50 kDa suffer fast clearance from the circulation by excretion through the kidneys. Proteolytic degradation and decreased activity by denaturation or aggregation can also be problematic for proteins used as therapeutics. As a remedy to overcome these shortcomings, it was found that the covalent attachment of hydrophilic polymers such as poly-ethylene-glycol (PEG) to proteins not only resulted in larger molecules with a lower clearance rate but also improved properties in terms of stability, solubility, and overall bioavailability [59, 60]. This modification strategy was also applied to several biopharmaceuticals such as the cytokines interferon alpha 2a or 2b (Pegasys or PEG-Intron) or erythropoietin (Mircera), leading to commercial success. However, these modifications were and are still performed with

classical nonspecific conjugation techniques using reactive amino, thiol, or carbohydrate groups on the surface of the protein, which usually yield an inhomogeneous product in terms of location and number of the attached PEG chains, unless there are only single reactive groups such as thiols present on the protein. Such randomly modified proteins might be problematic in terms of characterization, batch to batch variations, and a decreased potency caused by masked binding sites. Therefore, strategies to incorporate PEG chains in a site-specific manner to biopharmaceutical protein are of a great interest. For instance, a PEG molecule can be conjugated only to the N-terminal amino group of a protein based on its lower pKa compared to ε-amino groups of lysine residues; selectivity is achieved by a careful adaptation of the reaction conditions. This method was successfully established with the commercial product Neulasta through reductive alkylation of the N-terminal amino group with PEG-aldehyde under acidic conditions [61]. Other approaches include the site-directed exchange of lysine residues to arginine residues which preserve the charge and functionality but do not react with PEG reagents as demonstrated for TNF-α [62], or introduction of unique cysteine residues at selected sites of the protein such as interferon alpha 2 [63] or thyroid-stimulating hormone [64]. While these procedures still use classical conjugation chemistry, newer technologies relying on truly targeted modification were also evaluated: For instance PEG was attached to glutamine residues catalyzed by the enzyme transglutaminase [65] or by using the sortase technology [66]. In another example nonnatural amino acids such as azido-homoalanine were incorporated into the polypeptide chain and coupled with alkyne labels by using click chemistry [67].

Another emerging and highly interesting topic of protein modification related to biopharmaceuticals is the so-called antibody-drug conjugates (ADCs) in which cytotoxic drugs are covalently attached to specific antibodies. The purpose of this approach is the selective delivery of cytotoxic compounds to tumor cells without affecting noncancerous cells. Since the antibody part of an ADC typically binds to extracellular proteins nearly exclusively expressed on tumor cells, the ADCs are selectively internalized, followed by release of the attached cytotoxic agent into the cytoplasm and cell killing. The marketed products Adcetris (conjugate between the anti-CD30 mAb brentuximab and monomethyl-auristatin) [68] and Kadcyla (anti-HER2 mAb trastuzumab coupled to a derivative of maytansine) [69] have demonstrated that this strategy is highly promising and many other ADCs are now in several phases of clinical studies [70, 71]. Adcetris and Kadcyla are generated based on traditional reactive chemistry: in the case of Kadcyla, the antibody and cytotoxic drug are linked through a bifunctional reagent with an NHS moiety forming amide bonds with NH₂ groups of the antibody and a maleimide group reacting with a thiol group on the

maytansine derivative. These linkers are non-cleavable and the drug gets released in the cell only by lysosomal digestion of the whole antibody. Other linkers contain cleavable parts: the linker of the ADC Adcetris contains a specific protease cleavage site that is cleaved by a lysosomal protease thus enhancing the liberation of the drug [72]. Other approaches rely on disulfide or hydrazone moieties which are labile against reducing or acidic conditions which occur within the cell [72]. At the present the generation of ADC with classical chemical modification is the most widespread approach since the large size of the antibody might mitigate the potential negative impact of random labeling. Nevertheless, the first ADC (Gemtuzumab ozogamicin; Mylotarg) was removed from the market due to safety and efficacy issues which may have also been caused by the conjugation technique [73]. In order to improve the properties of ADCs approaches for a more directed conjugation such as introduction of additional specifically reactive cysteine residues [74], nonnatural amino acids [75, 76], or the conjugation with the help of transglutaminase [77, 78] have been described, highlighting the potential of site-specific modification in this highly competitive and commercially attractive field.

Another important application for site-specific labeling in the field of biopharmaceuticals is not the direct modification of biopharmaceutical proteins but rather the generation of monoclonal antibodies against pharmacologically relevant targets. Besides the classical hybridoma approach with immunization of mice, the mainly used technology is the selection of specific antibodies with the phage display technology [79]. In this method usually the antigen against which antibodies need to be selected are immobilized onto a solid surface, enabling the binding of well-binding phages and washing away of weakly binding phages in several rounds. Since this immobilization step should preserve the conformation and accessibility of the antigen, there are similar requirements and challenges as in the SPR technologies described in the section above. Hence, either chemical or site-specific biotinylation of the antigen and immobilization on streptavidin-coated surfaces or beads are widely used in phage display selection technique [80, 81].

8 Efforts and Experiences from In-house (Novartis)

8.1 Example 1: Biotinylation and SPR

Site-specific biotinylation with the Avi-tag technology was evaluated quite early in Novartis Research labs [4], and since then its use has been continuously expanded and it is now established as a standard technology for various applications like SPR, immobilization for phage display, and biochemical assays. In our department close to 20 proteins from various families and different lengths (full-length vs. single or multiple domains) were used for site-specific labeling using the Avi-tag approach. In our experience we found

this method easy to implement and straightforward in its routine usage. In the beginning we mostly relied on the *in vitro* labeling approach in which the isolated target protein is incubated with biotin, ATP, and purified biotin ligase (BirA). We have seen, however, that this method was detrimental to some of our target proteins as even these rather mild conditions (incubation at room temperature for several hours and buffers with rather low salt concentration) were too harsh for some proteins. Therefore we invested considerable effort to optimize biotinylation during expression in *E. coli* as described before [4], especially by varying location of Avi-tag (N-terminal vs. C-terminal), concentrations of biotin and arabinose (induction for co-expression of BirA) in the cultivation medium, and optimizing expression times and temperatures during expression. With this optimization a vast majority of different proteins as listed in Table 1 could be biotinylated with an incorporation >95 % as determined by LC-MS. Interestingly the

Table 1
Examples of proteins expressed in *E. coli* and modified by *in vivo* biotinylation with the Avi-tag method

Protein	Location of Avi-tag	Expression conditions	Biotin incorporation (%)
Bromodomain 1 (15 kDa)	C-terminal	200 µM biotin, 8 mg/ml arabinose, TB, 20 °C ON	>95
	N-terminal		40
Bromodomain 2 (15 kDa)	C-terminal	400 µM biotin, 8 mg/ml arabinose, TB, 20 °C ON	65
	N-terminal		>95
Bromodomain 2 (16 kDa)	C-terminal	400 µM biotin, 8 mg/ml arabinose, TB, 20 °C ON	>95
	N-terminal		>95
Catalytic domain of histone-methyl-transferase 1 (28 kDa)	C-terminal	100 µM biotin, 4 mg/ml arabinose, TB,	>95
	N-terminal		>95
Internal fragment of histone-methyl-transferase 2 (7 kDa)	C-terminal	200 µM biotin, 4 mg/ml arabinose, TB, 20 °C ON	80
	N-terminal		50
Catalytic domain of protein deacetylase 1 (50 kDa)	C-terminal	50 µM biotin, 2 mg/ml arabinose, TB, 20 °C ON	>95
	N-terminal		>95
Hydrolase (65 kDa)	C-terminal	100 µM biotin, 4 mg/ml arabinose, TB, 20 °C ON	>95
	N-terminal		>95
Ligand binding domain of nuclear receptor 1 (28 kDa)	N-terminal	100 µM biotin, 4 mg/ml arabinose, TB, 20 °C ON	90
N-terminal part of E3 ligase (21 kDa)	C-terminal	200 µM biotin, 2 mg/ml arabinose, LB, 25 °C, 5 h	>95
Ser/Thr kinase (35 kDa)	C-terminal	100 µM biotin, 4 mg/ml arabinose, LB, 20 °C ON	>95

The expression conditions are described in terms of amount of biotin added to the cultivation medium, arabinose for co-expression of biotin ligase (BirA). TB and LB refer to the cultivation media (terrific broth or Luria broth)

only three cases in which the maximal biotin incorporation reached less than 90 % were observed with small domains or stretches within a protein, whereas it worked very well with larger domains or full-length proteins.

A successful biotinylation via Avi-tag technology is no guarantee for a successful application in an SPR experiment, however. We therefore routinely apply at least two of the three different protein immobilization approaches (amine coupling, minimal chemical biotinylation, and site-specific biotinylation with Avi-tag) with the proteins under study. Examples are provided in Table 2. In our experience, the outcome with the three methods can be quite different, and there is no clear favorite single approach. Sufficient and stable immobilization was achieved in all examples with the selected methods and proteins. However, in terms of binding efficiency

Table 2

Outcome of SPR-based assays with different proteins and immobilization methods, immobilization is referring whether sufficient amounts of the protein could be bound on the surface; outcome of the assay describes whether a useful assay with detection of low molecular weight compounds could be established

Protein	Immobilization approach	Outcome assay
Bromodomain 1 (15 kDa)	Amino-coupling, Avi-tag	Worked well with amino coupling, low binding efficiency with all Avi-tagged variants
Bromodomain 2 (15 kDa)	Amino-coupling, Avi-tag	Results better with Avi-tag than amino-coupling, assay implemented
Bromodomain 2 (16 kDa)	Amino-coupling, Avi-tag	Results better with Avi-tag than amino-coupling, assay implemented
Ligand binding domain of nuclear receptor 2 (27 kDa)	Avi-tag (in vitro)	Assay implemented and used in SPR pilot studies
Catalytic domain of protein deacetylase 2 (50 kDa)	Avi-tag (in vitro)	Low binding efficiency, no assay implemented
Protease (25 kDa)	Amino coupling, minimal biot	Low binding efficiency, no assay implemented
Ser/Thr kinase 2 (75 kDa)	Amino coupling, minimal biot	Assay implemented and used
Hydrolase (65 kDa)	Avi-tag	Low binding efficiency, no assay implemented
Catalytic domain of Ser/ Thr kinase 3 (37 kDa)	Avi-tag (in vitro)	Assay implemented and used
N-terminal part of E3 ligase (21 kDa)	Avi-tag	Assay implemented and used
Catalytic domain of histone-methyl-transferase 1 (28 kDa)	Avi-tag, minimal biotinylation	No assay implemented yet, rather low signals
Hydroxylase (45 kDa)	Minimal biotinylation	Assay implemented and used

(percentage of the protein molecules which are still able to bind the ligand after immobilization) and sensitivity, the picture is much more mixed. For instance, with two different bromodomains SPR measurements could only be performed when they were immobilized through random amino coupling, whereas with a third bromodomain or the catalytic domain of Ser/Thr kinase the Avi-tag approach worked perfectly well. In other cases, different immobilization techniques all enabled a good assay, while with still other proteins no useful assay could be developed regardless of the biotin attachment method. The quintessence of these various experiments is that no single “gold standard” immobilization technique exists and the result seems to depend on various properties of the proteins, such as size, charge distribution, accessibility of the binding state, and thermal stability.

8.2 Example 2: Protein Labeling and Protein–Protein Interactions

In our company many protein–protein interactions were subjected to assay development and HTS in order to find modulating compounds and some of them are listed in Table 3. We usually employ pairs of donors and acceptors where one contains a directly attached label, while the other contains one or even two accessory detection partners such as streptavidin or a specific antibody (cf. Fig. 2). In a majority of the examples one of the proteins could be replaced by a short peptide without compromising the binding properties, thus facilitating the assay by the straightforward synthesis and labeling of the peptide during synthesis.

Table 3
Examples of assays that assessed protein–protein interactions in a proximity assay format such as TR-FRET

Protein 1	Label/tag (donor)	Protein/peptide 2	Label/tag (acceptor)
E3-ligase complex	Eu-chelate (NH ₂ groups)	Peptide	Cy5 synthetic
Ubiquitin	Biotin (single Cys)	C-terminal part of E3 ligase 2	Cy5 (NH ₂ groups)
N-terminal part of E3 ligase 1	Biotin (Avi-tag)	Peptide	Cy5 synthetic
Immunoglobulin	Biotin (NH ₂ groups)	Extracellular part of Ig receptor	Alexa647 (NH ₂ groups or carbohydrate)
Ligand binding domains of nuclear receptors	His ₆ -tag	Co-activator derived peptides	Cy5 synthetic
Cytokine	His ₆ -tag	Cognate receptor	Cy5 (NH ₂ groups)
Various bromodomains	His ₆ -tag	Histone-derived peptides	Biotin synthetic

Labeling/detecting whether the donor/acceptor moiety was either conjugated directly to the protein or through an accessory molecule such as an antibody. In case of protein labeling on amino groups either *N*-hydroxy-succinimide-derived fluorescent dyes or isothiocyanate-derived Eu-chelate was used. For peptides the label was attached during synthesis at specific site (usually N-terminal NH₂ group)

Labeling proteins using classical chemical modifications (via lysine, cysteines, or sugar side chains) usually works sufficiently well. Although an optimization of reaction conditions is required in many cases, we have as of now not had any examples in which random chemical labeling did not result in proteins that were functionally active in the TR-FRET assay. Due to a lack of necessity, the use of site-specifically modified proteins in biochemical assays is therefore all in all still quite limited in our in-house drug discovery.

Nonetheless, we have investigated the use of modern labeling methods: In one instance, we biotinylated the N-terminal domain of an E3 ligase with the Avi-tag technology and used it in a setup as shown in Fig. 2b with Eu-chelate streptavidin as the helper molecule. Additionally, a specifically biotinylated protein served as the substrate for a protein kinase in a functional assay. The format of this assay was based on a TR-FRET readout between a Eu-chelate labeled antibody directed against a phosphorylated residue within the protein and streptavidin binding to the biotin on the Avi-tag, and worked.

We have also tested selective labeling of several chemokines with fluorescent dyes with the help of transglutaminase. These specifically labeled chemokines were intended for binding experiments to their receptors (GPCR) present either on intact cells or on membrane preparations. In the case of the chemokine MCP1 the approach worked quite well since just one distinct glutamine residue near the C-terminus of the protein was modified with the label tetra-methyl-rhodamine-cadaverine when transglutaminase from liver extracts was used. In consequence, this modified chemokine showed similar affinities in the binding assay when compared to a scintillation proximity assay format. However, in other chemokines such as IL-8 and SDF1, no glutamine residue was reactive or accessible enough to enable a specific attachment of label to a sufficient extent. We tried to overcome this problem with microbial transglutaminase. However, this enzyme led to highly cross-linked proteins in which several glutamine and lysine residues reacted with each other. As an alternative, we tried to fuse small peptide sequences containing reactive glutamine residues such as the first seven amino acids of substance P to the C-terminus of the chemokine. After extensive optimization of labeling conditions using liver transglutaminase, specific modification was achieved, but cross-linked side products were still obtained. Due to the rather poor predictability and variability of optimal reaction conditions, we abandoned the transglutaminase-catalyzed approach for site-specific modification and pursued other approaches.

One of the alternative methods was the SNAP-tag technology as described earlier [4] with an E2 ubiquitin conjugating enzyme as the target protein. Expression of the fusion protein and site-specific modification on the SNAP-tag moiety worked well, and in

consequence we performed some preliminary TR-FRET experiments with labeled ubiquitin, showing rather low FRET signals. Even though the cause of these modest results could not be explained with the properties of the SNAP-tag we did not see an obvious advantage of this approach in the context of biochemical assays and did not perform further studies.

In the course for exploration of the ubiquitin pathway we attached fluorescent peptides through an isopeptide linkage specifically to the C-terminus of ubiquitin with the help of ubiquitin activating and conjugating enzymes [82]. These specifically modified ubiquitins served then as substrates for deubiquitinating enzymes in fluorescence-based assays. Even though this approach enabled successful assay development and screening it is clearly restricted to ubiquitin and related molecules and cannot be expanded to other classes of proteins.

Most recently we also started to explore the short ACP and LAP tags for selective labeling. These two tags are comparable in size to the well-characterized Avi-tag; thus based on the experiences with Avi-tagged proteins we assume they are not problematic in terms of interference with the properties of the protein under study. Internal efforts to use these tags for specific modification of proteins with fluorescent dyes have been initiated and some first preliminary results show a potential for these novel techniques, especially for the LAP2 technology in conjunction with copper-free click chemistry.

In conclusion, Avi- and SNAP-tag technology worked well in our hands and LAP2 labeling shows promise, while transglutaminase-mediated labeling required excessive optimization due to either nonreactiveness or the formation of covalent protein–protein aggregates. Nonetheless, classical reactive labeling remains the main method applied in biochemical assays at Novartis so far. The main reason for the reluctant use of site-specific labeling methods is that so far the established approaches using labeled accessory molecules have worked well and reliably, reducing the pressure for switching to different methods.

8.3 Example 3: Cellular Imaging

In the field of cellular imaging, the established methods rely on detection through labeled protein-specific antibodies and fusion to fluorescent proteins. These approaches are very reproducible, robust, and also reliable in higher throughput applications, and are therefore also the main approach used by Novartis. In one exploratory study a key member of a signaling pathway was fused to GFP or three self-labeling tags (SNAP, Halo, and FlAsH-tags) in either N- or C-terminal position and compared in terms of labeling efficiency and biological functionality of the fused protein. In our experience the labeling worked equally well for both SNAP and Halo-tags with low background staining, and fluorescence signals were comparable to eGFP. In contrast, fusion proteins with the

short FlaSH tag resulted in significant background signals, most likely caused by nonspecific interaction between Cys-rich sequences of cellular proteins and the biarsenic fluorescent dye. In terms of biological activity as determined by a luciferase-based reporter gene assay there were some differences between the different fusion proteins: Generally if the fluorescent partners such as GFP and SNAP-tag were fused to the N-terminus of the target protein, the signal of the reporter gene assay was significantly higher when compared to an attachment at the C-terminus. Interestingly we observed the opposite behavior with the Halo-tag showing a higher signal when fused to the C-terminus.

8.4 Example 4: Antibodies for Diagnostic Imaging and ADCs

For whole-body imaging the specific labeling of a therapeutic antibody or of antibody fragments like Fab or single-chain Fv was explored with incorporation of a nonnatural amino acid using the PCL approach [14]. For that purpose, different labels such as long wavelength fluorescent dyes and metal chelates were selected. The goal of this study was to evaluate whether specifically labeled antibody and especially the smaller fragments behave more stably and robustly within a whole organism compared to random labeled proteins; however no results are yet available.

In another very recent publication from the Novartis Institute for Functional Genomics, the successful production and preclinical studies of an ADC using a site-specific modification approach have been published [83]. In these experiments, selected peptide sequences serving as substrate for the enzyme phosphopantetheinyl transferase were inserted at defined locations within the Fc part of the well-known mAb Trastuzumab (Herceptin). The most promising Trastuzumab variants were then conjugated with various derivatives of coenzyme A, including some coupled to the cytotoxic drug mono-methyl-auristatin. Well-behaving ADCs with a drug to antibody ratio of about 2 were then tested in cellular assays and in animal xenograft tumor models where they proved to be efficacious. These site-specifically modified ADCs might be a truly valuable alternative to ADCs generated with a nonspecific modification approach.

9 Concluding Remarks

The different previous sections demonstrated that modified proteins play an important role in pharmaceutical research and development in many different areas. Starting from early discovery activities with biochemical assays and biophysical methods, through imaging in cells or whole organisms, to marketed products such as labeled antibodies for diagnostics and antibody-drug conjugates, chemically labeled proteins have become indispensable. At this point in time, it appears that the vast majority of applications using

modified proteins still relies on classical chemical chemistry. Various methods for site-specific labeling have been explored at Novartis, but are not yet broadly applied. The lack of publications from pharmaceutical companies in this area in general indicates that the situation is similar in the industry as a whole. This is most likely due to the time-pressure and result-driven environment of the pharmaceutical industry, where there is a certain reluctance to alter processes unless established methods fail to deliver results. Nonetheless, site-specific labeling is expected to become more important in different applications as methods and applications mature, and new areas are explored where there is a proven advantage of using specifically labeled proteins over randomly labeled protein. In the case of Novartis drug discovery, Avi-tagged based site-specific biotinylation has now become a standard method, as it works very well and has proven superior over chemical biotinylation in various examples. Other, more recent approaches for site-specific labeling of proteins with other moieties than biotin are also being applied, but only occasionally and on an opportunistic basis, as the traditional techniques work reasonably fine. Probably the greatest potential for site-specific modification lies in chemically modified biopharmaceuticals such as PEGylated proteins, antibodies for diagnostic imaging, and especially ADCs, as several examples described above already point in this direction. But also in this commercially highly attractive field, the adaptation of the labeling procedure to a more site-directed approach needs to overcome some hurdles, and a clear benefit over the existing methods in terms of pharmacological properties such as efficacy safety and/or cost still needs to be demonstrated.

Acknowledgements

I would like to thank the following colleagues: Jutta Blank, Marjo, Goette, Christian Bergsdorf, and Rainer Kneuer who gave valuable input and discussion in terms of performed experiments and general approaches in their respective field of expertise.

References

1. Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21(6):766–776
2. Foley TL, Burkart MD (2007) Site-specific protein modification: advances and applications. *Curr Opin Chem Biol* 11(1):12–19
3. Rashdian M, Dozier JK, Distefano MK (2013) Enzymatic labeling of proteins: techniques and approaches. *Bioconjug Chem* 24(8): 1277–1294
4. Tirat A, Freuler F, Stettler T, Mayr LM, Leder L (2006) Evaluation of two novel tag-based labelling technologies for site-specific modification of proteins. *Int J Biol Macromol* 39(1–3):66–76
5. Fernandez-Suarez M, Baruah H, Martinez-Hernandez L, Xie KT, Baskin JM, Bertozzi CR, Ting AY (2007) Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* 25(12):1483–1487

6. Puthenveetil S, Liu DS, White KA, Thompson S, Ting AY (2009) Yeast display evolution of a kinetically efficient 13-amino acid substrate for lipoic acid ligase. *J Am Chem Soc* 131(45): 16430–16438
7. Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* 102(44):15815–15820
8. Zhou Z, Cironi P, Lin AJ, Xu Y, Hrvatin S, Golan DE, Silver PA, Walsh CT, Yin J (2007) Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem Biol* 2(5):337–346
9. Lin CW, Ting AY (2006) Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* 128(14):4542–4543
10. Jager M, Nir E, Weiss S (2006) Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification. *Protein Sci* 15(3):640–646
11. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3(11):707–708
12. Antos JM, Chew GL, Guimaraes CP, Yoder NC, Grotenbreg GM, Popp MW, Ploegh HL (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* 131:10800–10801
13. Young TS, Schultz PG (2010) Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* 285(15):11039–11044
14. Ou WJ, Uno T, Chiu HP, Grunewald J, Cellitti SE, Crossgrove T, Hao XS, Fan Q, Quinn LL, Patterson P et al (2011) Site-specific protein modifications through pyrroline-carboxylysine residues. *Proc Natl Acad Sci U S A* 108(26):10437–10442
15. Danielson UH (2009) Integrating surface plasmon resonance biosensor-based interaction kinetic analyses into the lead discovery and optimization process. *Future Med Chem* 1(8):1399–1414
16. Huber W (2011) SPR-based direct binding assays in drug discovery. In: Cooper M, Mayr LM (eds) Label-free technologies for drug discovery. John Wiley & Sons, Ltd., Chichester
17. Syguda A, Kerstan A, Ladnorg T, Stuben F, Woll C, Herrmann C (2012) Immobilization of biotinylated hGBP1 in a defined orientation on surfaces is crucial for uniform interaction with analyte proteins and catalytic activity. *Langmuir* 28(15):6411–6418
18. Hutsell SQ, Kimple RJ, Siderovski DP, Willard FS, Kimple AJ (2010) High-affinity immobilization of proteins using biotin- and GST-based coupling strategies. *Methods Mol Biol* 627:75–90
19. Huang ZH, Hwang P, Watson DS, Cao LM, Szoka FC (2009) Tris-nitrilotriacetic acids of subnanomolar affinity toward hexahistidine tagged molecules. *Bioconjug Chem* 20(8): 1667–1672
20. Chen YX, Triola G, Waldmann H (2011) Bioorthogonal chemistry for site-specific labeling and surface immobilization of proteins. *Acc Chem Res* 44(9):762–773
21. Wong LS, Khan F, Micklefield J (2009) Selective covalent protein immobilization: strategies and applications. *Chem Rev* 109(9): 4025–4053
22. Wammes AE, Fischer MJ, de Mol NJ, van Eldijk MB, Rutjes FP, van Hest JC, van Delft FL (2013) Site-specific peptide and protein immobilization on surface plasmon resonance chips via strain-promoted cycloaddition. *Lab Chip* 13(10):1863–1867
23. Joo C, Balci H, Ishitsuka Y, Buranachai C, Ha T (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu Rev Biochem* 77:51–76
24. Degorce F, Card A, Soh S, Trinquet E, Knapik GP, Xie B (2009) HTRF: a technology tailored for drug discovery—a review of theoretical aspects and recent applications. *Curr Chem Genomics* 3:22–32
25. Gotoh Y, Nagata H, Kase H, Shimonishi M, Ido M (2010) A homogeneous time-resolved fluorescence-based high-throughput screening system for discovery of inhibitors of IKKbeta-NEMO interaction. *Anal Biochem* 405(1): 19–27
26. Nakamura K, Zawistowski JS, Hughes MA, Sexton JZ, Yeh LA, Johnson GL, Scott JE (2008) Homogeneous time-resolved fluorescence resonance energy transfer assay for measurement of Phox/Bem1p (PB1) domain heterodimerization. *J Biomol Screen* 13(5): 396–405
27. Kim B, Tarchevskaya SS, Eggel A, Vogel M, Jardetzky TS (2012) A time-resolved fluorescence resonance energy transfer assay suitable for high-throughput screening for inhibitors of immunoglobulin E-receptor interactions. *Anal Biochem* 431(2):84–89
28. Mollwitz B, Brunk E, Schmitt S, Pojer F, Bannwarth M, Schiltz M, Rothlisberger U, Johnsson K (2012) Directed evolution of the suicide protein O6-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* 51(5):986–994

29. Gautier A, Juillerat A, Heinis C, Correa IR, Kindermann M, Beaufils F, Johnsson K (2008) An engineered protein tag for multiprotein labeling in living cell. *Chem Biol* 15(2):128–136
30. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M et al (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3(6):373–382
31. Shi X, Jung Y, Lin LJ, Liu C, Wu C, Cann IK, Ha T (2012) Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat Methods* 9(5):499–503
32. Lemke EA (2013) Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. In: Mark SS (ed) *Bioconjugation protocols*, vol 751, *Methods in molecular biology*. Humana Press, New York, NY, pp 3–15
33. Kim J, Seo MH, Lee S, Cho K, Yang A, Woo K, Kim HS, Park HS (2013) Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis. *Anal Chem* 85(3):1468–1474
34. Lian LY, Middleton DA (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog Nucl Mag Res Sp* 39(3):171–190
35. Hoogstraten CG, Johnson JE (2008) Metabolic labeling: taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. *Concept Magn Reson* 32A(1):34–55
36. Weigelt J, Wikstrom M, Schultz J, van Dongen MJ (2002) Site-selective labeling strategies for screening by NMR. *Comb Chem High Throughput Screen* 5(8):623–630
37. Jahnke W, Rudisser S, Zurini M (2001) Spin label enhanced NMR screening. *J Am Chem Soc* 123(13):3149–3150
38. Jones DH, Cellitti SE, Hao X, Zhang Q, Jahnz M, Summerer D, Schultz PG, Uno T, Geierstanger BH (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. *J Biomol NMR* 46(1):89–100
39. Yang Y, Li QF, Cao C, Huang F, Su XC (2013) Site-specific labeling of proteins with a chemically stable, high-affinity tag for protein study. *Chemistry* 19(3):1097–1103
40. Shimba N, Yamada N, Yokoyama K, Suzuki E (2002) Enzymatic labeling of arbitrary proteins. *Anal Biochem* 301(1):123–127
41. Crivat G, Taraska JW (2012) Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol* 30(1):8–16
42. Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* 90(3):1103–1163
43. Jing C, Cornish VW (2011) Chemical tags for labeling proteins inside living cells. *Acc Chem Res* 44(9):784–792
44. Uttamapinant C, White KA, Baruah H, Thompson S, Fernandez-Suarez M, Puthenveetil S, Ting AY (2010) A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* 107(24):10914–10919
45. Reymond L, Lukinavicius G, Umezawa K, Maurel D, Brun MA, Masharina A, Bojkowska K, Mollwitz B, Schena A, Griss R et al (2011) Visualizing biochemical activities in living cells through chemistry. *Chimia* 65(11):868–871
46. Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, Lo A, Codelli JA, Bertozzi CR (2007) Copper-free click chemistry for dynamic *in vivo* imaging. *Proc Natl Acad Sci U S A* 104(43):16793–16797
47. Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* 4(4):298–304
48. Seitchik JL, Peeler JC, Taylor MT, Blackman ML, Rhoads TW, Cooley RB, Refakis C, Fox JM, Mehl RA (2012) Genetically encoded tetrazine amino acid directs rapid site-specific *in vivo* bioorthogonal ligation with trans-cyclooctene. *J Am Chem Soc* 134:2898–2901
49. Ripoll J, Ntziachristos V, Cannet C, Babin AL, Kneuer R, Gremlich HU, Beckmann N (2008) Investigating pharmacology *in vivo* using magnetic resonance and optical imaging. *Drugs R&D* 9(5):277–306
50. Kneuer R, Gremlich HU, Beckmann N, Jetzfellner T, Ntziachristos V (2012) *In vivo* fluorescence optical and multi-modal imaging in pharmacological research: from chemistry to therapy monitoring, *RSC Drug Discovery Series*. RSC Publishing, Cambridge, pp 343–370
51. Wu AM, Olafsen T (2008) Antibodies for molecular imaging of cancer. *Cancer J* 14(3):191–197
52. van Dongen GA, Visser GW, Lub-de Hooge MN, de Vries EG, Perk LR (2007) Immuno-PET: a navigator in monoclonal antibody development and applications. *Oncologist* 12(12):1379–1389
53. Wang H, Chen X (2008) Site-specifically modified fusion proteins for molecular imaging. *Front Biosci* 13:1716–1732
54. Blankenberg FG, Backer MV, Levashova Z, Patel V, Backer JM (2006) *In vivo* tumor angiogenesis imaging with site-specific labeled

- (99 m)Tc-HYNIC-VEGF. *Eur J Nucl Med Mol Imaging* 33(7):841–848
55. Backer MV, Levashova Z, Levenson R, Blankenberg FG, Backer JM (2008) Cysteine-containing fusion tag for site-specific conjugation of therapeutic and imaging agents to targeting protein. *Methods Mol Biol* 494: 275–294
 56. Wällberg H, Grafström J, Cheng Q, Lu L, Martinsson Ahlzen HS, Samén E, Thorell JO, Johansson K, Dunás F, Olofsson MH, Stone-Elander S et al (2012) HER2-positive tumors imaged within 1 hour using a site-specifically ¹¹C-labeled Sel-tagged affibody molecule. *J Nucl Med* 53(9):1446–1453
 57. Pardo A, Stöcker M, Kampmeier F, Melmer G, Fischer R, Thepen T, Barth S (2012) In vivo imaging of immunotoxin treatment using Katushka-transfected A-431 cells in a murine xenograft tumour model. *Cancer Immunol Immunother* 61(10):1617–1625
 58. Kosaka N, Ogawa M, Choyke PL, Karassina N, Corona C, McDougall M, Lynch DT, Hoyt CC, Levenson RM, Los GV et al (2009) In vivo stable tumor-specific painting in various colors using dehalogenase-based protein-tag fluorescent ligand. *Bioconjug Chem* 20(7): 1367–1374
 59. Molineux G (2003) Pegylation: engineering improved biopharmaceuticals for oncology. *Pharmacotherapy* 23(8 Pt 2):3S–8S
 60. Veronese FM, Mero A (2008) The impact of PEGylation on biological therapies. *BioDrugs* 22(5):315–329
 61. Molineux G (2004) The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta). *Curr Pharm Des* 10(11):1235–1244
 62. Yamamoto Y, Tsutsumi Y, Yoshioka Y, Nishibata T, Kobayashi K, Okamoto T, Mukai Y, Shimizu T, Nakagawa S, Nagata S et al (2003) Site-specific PEGylation of a lysine-deficient TNF-alpha with full bioactivity. *Nat Biotechnol* 21(5):546–552
 63. Rosendahl MS, Doherty DH, Smith DJ, Carlson SJ, Chlipala EA, Cox GN (2005) A long-acting, highly potent interferon alpha-2 conjugate created using site-specific PEGylation. *Bioconjug Chem* 16(1):200–207
 64. Qiu H, Boudanova E, Park A, Bird JJ, Honey DM, Zarazinski C, Greene B, Kingsbury JS, Boucher S, Pollock J et al (2013) Site-specific PEGylation of human thyroid stimulating hormone to prolong duration of action. *Bioconjug Chem* 24(3):408–418
 65. Sato H, Hayashi E, Yamada N, Yatagai M, Takahara Y (2001) Further studies on the site-specific protein modification by microbial transglutaminase. *Bioconjug Chem* 12(5): 701–710
 66. Popp MW, Dougan SK, Chuang TY, Spoonera E, Ploegh HL (2010) Sortase-catalyzed transformations that improve the properties of cytokines. *Proc Natl Acad Sci U S A* 108:369–374
 67. Kochendoerfer GG (2005) Site-specific polymer modification of therapeutic proteins. *Curr Opin Chem Biol* 9(6):555–560
 68. Katz J, Janik JE, Younes A (2011) Brentuximab Vedotin (SGN-35). *Clin Cancer Res* 17(20): 6428–6436
 69. Lewis Phillips GD, Li G, Dugger DL, Crocker LM, Parsons KL, Mai E, Blattler WA, Lambert JM, Chari RV, Lutz RJ et al (2008) Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Res* 68(22):9280–9290
 70. Adair JR, Howard PW, Hartley JA, Williams DG, Chester KA (2012) Antibody-drug conjugates - a perfect synergy. *Expert Opin Biol Ther* 12(9):1191–1206
 71. Zolot RS, Basu S, Million RP (2013) Antibody-drug conjugates. *Nat Rev Drug Discov* 12(4): 259–260
 72. Flygare JA, Pillow TH, Aristoff P (2013) Antibody-drug conjugates for the treatment of cancer. *Chem Biol Drug Des* 81(1):113–121
 73. Ricart AD (2011) Antibody-drug conjugates of calicheamicin derivative: gemtuzumab ozogamicin and inotuzumab ozogamicin. *Clin Cancer Res* 17(20):6417–6427
 74. Junutula JR, Raab H, Clark S, Bhakta S, Leipold DD, Weir S, Chen Y, Simpson M, Tsai SP, Dennis MS et al (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat Biotechnol* 26(8):925–932
 75. Axup JY, Bajjuri KM, Ritland M, Hutchins BM, Kim CH, Kazane SA, Halder R, Forsyth JS, Santidrian AF, Stafin K et al (2012) Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc Natl Acad Sci U S A* 109(40):16101–16106
 76. Tian F, Lu Y, Manibusan A, Sellers A, Tran H, Sun Y, Phuong T, Barnett R, Hehli B, Song F et al (2014) A general approach to site-specific antibody drug conjugates. *Proc Natl Acad Sci U S A* 111(5):1766–1771
 77. Strop P, Liu SH, Dorywalska M, Delaria K, Dushin RG, Tran TT, Ho WH, Farias S, Casas MG, Abdiche Y et al (2013) Location matters: site of conjugation modulates stability and pharmacokinetics of antibody drug conjugates. *Chem Biol* 20(2):161–167
 78. Jeger S, Zimmermann K, Blanc A, Grunberg J, Honer M, Hunziker P, Struthers H, Schibli R (2010) Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. *Angew Chem Int Ed Engl* 49(51): 9995–9997

79. Matz J, Chames P (2012) Phage display and selections on purified antigens. *Methods Mol Biol* 907:213–224
80. Kruszynski M, Tsui P, Stowell N, Luo J, Nemeth JF, Das AM, Sweet R, Heavner GA (2006) Synthetic, site-specific biotinylated analogs of human MCP-1. *J Pept Sci* 12(5):354–360
81. Scholle MD, Kriplani U, Pabon A, Sishtla K, Glucksman MJ, Kay BK (2006) Mapping protease substrates by using a biotinylated phage substrate library. *Chembiochem* 7(5):834–838
82. Tirat A, Schilb A, Riou V, Leder L, Gerhartz B, Zimmermann J, Worpenberg S, Eidhoff U, Freuler F, Stettler T et al (2005) Synthesis and characterization of fluorescent ubiquitin derivatives as highly sensitive substrates for the deubiquitinating enzymes UCH-L3 and USP-2. *Anal Biochem* 343(2):244–255
83. Geierstanger BH, Grunewald J, Bursulaya B (2013) Use of phosphopantetheinyl transferase substrate peptides for site-specific labeling of immunoglobulins. *PCT Int Appl.* PCT/US2013/043684

Chapter 3

Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins

Nicole J. Yang and Marlon J. Hinner

Abstract

The ability to efficiently access cytosolic proteins is desired in both biological research and medicine. However, targeting intracellular proteins is often challenging, because to reach the cytosol, exogenous molecules must first traverse the cell membrane. This review provides a broad overview of how certain molecules are thought to cross this barrier, and what kinds of approaches are being made to enhance the intracellular delivery of those that are impermeable. We first discuss rules that govern the passive permeability of small molecules across the lipid membrane, and mechanisms of membrane transport that have evolved in nature for certain metabolites, peptides, and proteins. Then, we introduce design strategies that have emerged in the development of small molecules and peptides with improved permeability. Finally, intracellular delivery systems that have been engineered for protein payloads are surveyed. Viewpoints from varying disciplines have been brought together to provide a cohesive overview of how the membrane barrier is being overcome.

Key words Cell membrane, Permeability, Translocation, Intracellular delivery, Cytosolic delivery, Fluorescent probe, Passive diffusion, Membrane transporter, Endosomal escape

1 Introduction

Molecules that can readily cross cell membranes are frequently needed in biological research and medicine. Permeable molecules that are useful for biological research include indicators of ion concentrations and pH, fluorescent dyes, crosslinking molecules, fluorogenic enzyme substrates, and various protein inhibitors. In medicine, numerous drugs are small molecules acting on intracellular targets, such as statins that inhibit cholesterol production, and reverse transcriptase inhibitors used for the treatment of HIV. Given the high level of interest across multiple areas of study in modulating intracellular targets, a broad overview of cytosolic delivery strategies could contribute to orienting researchers newly entering the field, and bringing together the solutions that have been proposed for various cargo.

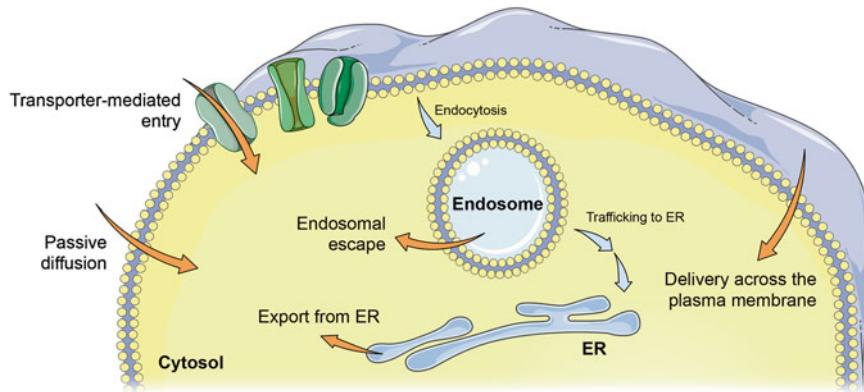


Fig. 1 Possible routes of cytosolic entry. Molecules may passively diffuse across the cell membrane, or be shuttled in via natural or artificial delivery mechanisms. Membrane transporters allow the passage of various ions and metabolites. Protein toxins and viruses have evolved complex translocation mechanisms, hijacking the host's ER transporters in some instances. Engineering approaches to improve a payload's permeability may involve physically disrupting the membrane, chemically modifying the payload, or attaching the payload—covalently or non-covalently—to an intracellular delivery system that can disrupt cell membranes. In any case, the translocation event can occur across the plasma membrane, or across internal cellular membranes following endocytosis (termed "endosomal escape"). Images were adapted from Servier Medical Art

This review examines how varying types of molecules—namely, small molecules, peptides, and proteins—are thought to cross the mammalian plasma membrane, how such permeation events are measured experimentally, and what kinds of technologies are being developed to deliver impermeable molecules to the cytoplasm (Fig. 1). Inspired by Stein et al. [1], we first discuss the structure and organization of the cell membrane, and the endocytic and secretory pathways responsible for the uptake and discharge of material. Then, we outline the experimental methods that have been used to determine, qualitatively or quantitatively, whether a molecule has successfully traversed the cell membrane. Next, we survey how various molecules, including ions, small solutes and metabolites, along with bacterial toxins and viruses, are thought to traverse the cell membrane. Lastly, we introduce engineering strategies that have been proposed to improve the permeability of small molecules, peptides, and proteins. Although we do not discuss the delivery of nucleic acids explicitly, certain approaches that have been proposed for introducing peptides and proteins into the cytoplasm have been (or could be) applied to nucleic acids, and vice versa. Overall, significant advances have been made in the prediction and design of permeable small-molecule compounds, and the repertoire of intracellular delivery technologies is steadily expanding for peptide and protein payloads.

Understandably, much effort has been dedicated towards developing therapeutic entities that can modulate intracellular proteins, and thus we have drawn heavily from the drug development and delivery literature. However, engineering considerations regarding the optimization of *in vivo* properties were considered beyond the scope of this review and are not discussed. Due to the breadth of subjects we attempted to cover in a small space, we have preferentially referenced pertinent review articles when possible and strongly encourage readers to further explore the primary literature.

2 Cellular Organization

Before discussing membrane transport mechanisms, we will briefly describe the structure of the mammalian plasma membrane and components of the endocytic and secretory pathways. For the purpose of intracellular delivery, it is important to note that the interior of endocytic vesicles and the lumen of the organelles involved in the secretory pathway topologically correspond to the extracellular space.

2.1 The Structure and Organization of the Cell Membrane

The plasma membrane is a complex composite of multiple lipid species and membrane proteins [2]. Three major classes of lipids, including glycerophospholipids, sphingolipids and cholesterol, form a bilayer approximately 5nm in width. Spatially, these lipids are distributed asymmetrically across the bilayer. Additionally, according to the lipid raft hypothesis, the membrane is thought to contain lateral organizations enriched in sphingolipids, cholesterol, and glycosylphosphatidylinositol (GPI)-anchored proteins.

The ratio of protein to lipid in cellular membranes has been approximated to be 1:40 by number [3], suggesting that the membrane may in fact be crowded with proteins [4, 5]. This ratio can vary substantially by cell type, where metabolically active membranes are richer in protein [1, 6]. Membrane proteins can actively influence the organization of the membrane by forming specific and nonspecific interactions with lipids in the immediate boundary [7, 8].

Finally, the plasma membrane is in continuous motion [1], creating a highly dynamic structure. In addition to lateral diffusion, phospholipid flip-flop for some lipids is thought to occur on the order of minutes [9], and faster so for cholesterol [10, 11]. Also, cells constantly internalize and recycle their membranes, as discussed below.

2.2 The Secretory Pathway

In mammalian cells, secretory proteins are translocated across the Endoplasmic Reticulum (ER) membrane into the lumen cotranslationally via the translocon complex. Misfolded proteins in

the ER are transported back to the cytosol and degraded by the proteasome, a process termed ER-associated degradation (ERAD). Correctly folded proteins are transported across the Golgi network and released into the extracellular space via secretory vesicles. Specialized vesicles also mediate retrograde transport from the Golgi to the ER, and from older to newer Golgi [2].

2.3 Endocytic Pathways

Multiple endocytic pathways facilitate the internalization of exogenous cargo, creating a complex web of intracellular traffic. The choice of which endocytic pathway is utilized may depend on the cargo [12]. Nonspecific internalization of large volumes of fluid—pinocytosis—occurs in all cells, typically triggered by external stimuli such as growth factors [13]. Clathrin-dependent and independent routes of endocytosis generate primary endocytic vesicles that subsequently fuse with early endosomes, a major sorting station. Traveling down tracks of microtubules towards the perinuclear space, the early endosomes mature into multivesicular bodies (MVB), late endosomes and lysosomes. Endocytosed material that has not been recycled to the plasma membrane or exchanged with the *trans* Golgi network is proteolyzed by hydrolytic enzymes in the lysosome [2].

3 Methods to Measure Membrane Permeation

The permeability of a given molecule can be quantitatively represented by its permeability coefficient (typically in units of cm/s) (Table 1), which is a measure of how fast it can cross a membrane [14]. High-throughput methods to measure the permeability coefficients of small molecules are routinely performed in drug development [15]. Artificial lipid bilayers [16] of various compositions or cell monolayers (typically the colorectal Caco-2 or renal MDCK cell lines) are widely used as model barriers [17]. While the former allows passive permeation only, the latter also allows transporter-mediated permeation. To disentangle these two modes of transport, cell lines that lack certain transporters, such as P-glycoprotein, have been developed [18].

In contrast to small molecules, the permeability of peptides or proteins across model membranes is rarely reported, reflecting in part the difficulty of translocating such molecules across a lipid membrane. It is technically challenging to accurately quantify the number of functional peptides or proteins that have successfully entered the cytoplasm. Selective isolation of the cytosol (and not endosomal compartments) using cellular fractionation [19, 20] or digitonin-mediated permeabilization of the plasma membrane [21] have been reported. Immunoprecipitation demonstrating the intended disruption of intracellular protein–protein interactions has also been presented as evidence of permeation [22, 23].

Table 1
Permeability coefficients of select molecules

Species	Molecule	Permeability coefficient (cm/s)	Membrane type	Reference
Ions	Na ⁺	5.0×10^{-14}	Artificial membrane	Papahadjopoulos et al. [40]
	K ⁺	4.7×10^{-14}		
Small molecules	O ₂	2.3×10^1	Artificial membrane	Subczynski et al. [36]
	CO ₂	3.5×10^{-1}	Artificial membrane	Gutknecht et al. [37]
	H ₂ O	3.4×10^{-3}	Artificial membrane	Walter and Gutknecht [38]
	EtOH	2.1×10^{-3}	Erythrocyte membrane	Stein and Lieb [1]
	Steroids	10^{-3} to 10^{-4}	Cell monolayer	Giorgi and Stein [42]
	Urea	4.0×10^{-6}	Artificial membrane	Finkelstein [35]
	Glycerol	5.4×10^{-6}	Artificial membrane	Orbach and Finkelstein [39]
	Small molecule drugs	10^{-5} to 10^{-6}	Artificial membrane	Dobson et al. [112]
Peptides	Cyclosporin A	2.5×10^{-7}	Artificial membrane	Rezai et al. [45]
	TAT	2.7×10^{-9}	Artificial membrane	Jones and Howl [81]

Fluorescence microscopy-based methods or biological assays measuring the activity of the payload in the cytoplasm are also employed. In microscopy, diffuse cytosolic staining (indicating endosomal release) is contrasted with punctate signal (indicating endosomal entrapment) to provide a qualitative assessment of permeation. However, it should be noted that payloads in the cytoplasm may also aggregate or associate with subcellular organelles to produce punctate patterns. In some cases, automated image analyses have been reported to identify endosomal release events [24]. In any case, the presence of labeled payload in the cytoplasm does not guarantee that it has retained its function, and the label itself or fixation steps may cause artifacts in cellular distribution [25, 26]. All observations are subject to the detection threshold of the instrument. Flow cytometry may be used as an alternative to microscopy if the fluorescence spectra are distinct in the endosomal compartment and the cytoplasm [27].

Alternatively, cytosolic uptake can be confirmed by measuring a biological effect that is generated only when the payload is in the cytoplasm. For example, peptides have been conjugated to dexamethasone (Dex) derivatives, which bind to transiently expressed glucocorticoid receptor (GR)-fusion proteins in the cytosol to induce a reporter [28] or alter its localization [29]. It should be noted that reporter gene expression inherently amplifies the signal through multiple rounds of transcription and translation [29]. In cases where the biological activity of the payload is reported, certain payloads can generate the measured

macroscopic effect with fewer numbers. This is particularly true for catalytic proteins. For example, approximately 50 molecules of β -lactamase in a single cell have been reported to generate a detectable signal from catalyzing a fluorogenic substrate, albeit over a long period of time (16 h) [30]. Similarly, in theory, four molecules of Cre recombinase can repeatedly catalyze multiple recombination events to promote recombined gene expression [31]. Single molecules of toxins such as diphtheria and ricin have been estimated to kill a cell [32, 33].

4 Natural Membrane Transport Mechanisms

Small, moderately polar molecules are able to passively diffuse across the cell membrane. To transport larger, more polar compounds such as most sugars, amino acids, peptides, and nucleosides, membrane transporters are utilized. Interestingly, bacteria and viruses have developed sophisticated mechanisms to transport whole organisms, protein toxins, or genetic material into the mammalian cytoplasm.

4.1 Passive Diffusion

Passive diffusion across a cellular membrane is driven by the concentration and electric gradient of the solute and does not require the use of energy. In the simplest terms, passive diffusion is considered a three-step process, where the permeant first partitions into the membrane, diffuses across, and is released into the cytosol (known as the homogeneous solubility-diffusion model) [34, 35].

The most important parameters that govern transmembrane diffusion are polarity and size. For example, small nonpolar gases such as O₂, CO₂, and N₂, and small polar molecules such as ethanol cross lipid membranes rapidly. High permeability coefficients have been reported for such molecules across artificial lipid membranes, such as 2.3×10^1 cm/s for O₂ [36] and 3.5×10^{-1} cm/s for CO₂ [37]. The small, but highly polar water molecule is still able to diffuse across artificial membranes rapidly with a permeability coefficient of 3.4×10^{-3} cm/s [38].

In comparison, even slightly larger polar metabolites such as urea and glycerol have lower permeability across artificial membranes (approximately 10^{-6} cm/s) [35, 39]. The plasma membrane is virtually impermeable against larger, uncharged polar molecules and all charged molecules including ions. Indeed, despite their small size, Na⁺ and K⁺ have extremely low permeability coefficients (approximately 10^{-14} cm/s) [40]. Apart from small solutes of moderate polarity, the number of natural molecules known to passively diffuse across the cell membrane is surprisingly limited. Steroid hormones have been assumed to do so [41], although direct experimental evidence is scarce. Permeability coefficients on the order of 10^{-4} cm/s have been reported for a number of steroids across cell monolayers [42].

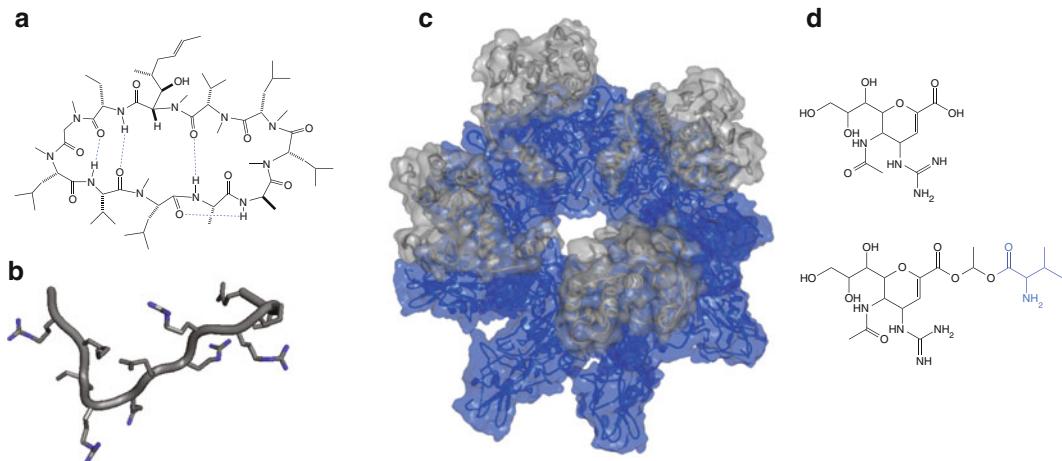


Fig. 2 (a) Cyclosporin A (CsA) in its closed conformation in nonpolar solvent [186]. The four intramolecular hydrogen bonds (*dotted lines in blue*) are thought to shield the polarity of the molecule. (b) The TAT peptide segment excerpted from the NMR structure of HIV-1 TAT protein (adapted from PDB 1TIV) [187]. The guanidinium nitrogens (*blue*) are thought to enhance the interaction between TAT and the cell membrane. (c) A slanted top-down view of the pre-pore formed by anthrax toxin protective antigens (PAs) (*blue*) in complex with lethal factors (LFs) (*gray*), which are translocated across the full pore. Shown in the figure are eight molecules of PA bound to four molecules of LF (PA₈(LF_N)₄) (PDB 3KWV) [188]. (d) The neuraminidase inhibitor Zanamivir (*top*) and Zanamivir-L-Val (*bottom*) [189]. The conjugated valine (*blue*) has been proposed to render Zanamivir into a substrate for amino acid transporters

Interestingly, some non-endogenous natural products have been proposed to passively diffuse across the cell membrane despite their relatively higher polarity and size, such as the cyclic peptide Cyclosporin A (CsA) (Fig. 2a). Prescribed as an immunosuppressant, its intracellular mode of action and low EC₅₀ in cells (7–10 nM) [43] suggests that CsA is capable of passively permeating the cell membrane [43, 44]. Still, the reported permeability coefficient of CsA— 2.5×10^{-7} cm/s across artificial membranes [45]—is relatively low compared to those of small molecules that are considered highly permeable (on the order of 10^{-5} cm/s or higher [46]).

4.2 Transporter-Mediated Entry

To facilitate the entry or export of molecules that are insufficiently permeable, cells utilize membrane transporters, the expression of which may depend on cell type. Active transporters use energy to translocate substrates against their concentration gradients, whereas passive transporters allow transmembrane diffusion without additional energy. Approximately 10 % of all human genes are transporter related, emphasizing their functional significance [47]. In the following, a selection of transporters is described, ordered according to the size of the substrate. Please refer to the Transporter Classification Database (www.tcdb.org) [48] and the Solute Carrier (SLC) Tables (www.bioparadigms.org) [49] for comprehensive reviews, and detailed information regarding substrate specificity and tissue/cellular distribution.

Ion channels allow the passive diffusion of inorganic ions with high specificity, often in response to stimuli such as changes in transmembrane potential, ligands, light, or mechanical stress [50, 51]. Alternatively, ions can also be actively transported by ion pumps, such as the sodium/potassium pump (the Na^+ , K^+ -ATPase), which transports three Na^+ ions extracellularly and two K^+ ions intracellularly for every molecule of ATP hydrolyzed [52]. Microbe-synthesized ionophores, such as valinomycin, facilitate the diffusion of ions across the cell membrane by complexing and shuttling ions [53]. Other ionophores such as gramicidin A form channels [54].

Small molecules are also often transported. Water is transported across the membrane by the aquaporin (AQP) family of passive channels. Aquaporins have been reported to transport other gases and solutes as well, such as CO_2 , NO , H_2O_2 , arsenite, ammonia (in addition to the Rh proteins [55, 56]), urea (in addition to the urea transporters [57]) and glycerol [58]. (This is an abbreviated list excerpted from Bienert et al. [59]).

Sugars, including glucose, galactose, and fructose, are molecules of high polarity and intermediate size, and are imported into the cell by the glucose transporter (GLUT) family of facilitated transporters [60]. Most amino acids are transported across the cell membrane by secondary active transporters that utilize the energy stored in the electrochemical gradient of another solute [61]. Nucleobases and nucleosides also have associated secondary transporters on the plasma membrane [62]. Di- and tri-peptides of various chemical character are transported by the oligopeptide transporter PepT1, which has been reported to transport neutral tripeptide-like β -lactam antibiotics and peptide-like drugs as well [63]. Alternatively, α -Amanitin, a cyclic octapeptide that inhibits eukaryotic RNA polymerase II, has been reported to enter cells via an organic anion transporting polypeptide (OATP) transporter [64].

To note, transporters may mediate the efflux of molecules as well. A variety of structurally unrelated compounds, including small-molecule drugs, are substrates of efflux pumps in the ATP-binding cassette (ABC) transporter family such as the multidrug resistance protein (MRP) family [65], the P-glycoprotein pump (P-gp, also known as multidrug resistance protein 1(MDR1)) [66], and the breast cancer resistance protein (BCRP) [67].

4.3 Other Methods of Cytosolic Entry

A majority of the examples discussed in the following first involve the cargo being internalized into the cell via various endocytic pathways. Reiterating an earlier point, endocytosed cargo are topologically still in an extracellular space separated from the cytoplasm by a lipid membrane. Thus, an additional “endosomal escape” (or “endosomal release”) step is required where the cargo

is transported across the membrane to access the cytoplasm. Some peptidic, viral, or bacterial components are thought to accomplish this step, not through passive diffusion or active transport, but by disrupting cellular membranes, allowing the passage of large and charged compounds. The mechanisms of most such processes are not yet fully elucidated and subjects of active research.

4.3.1 Peptides

Cell-penetrating peptides (CPP), also known as peptide transduction domains (PTD), are a diverse class of peptides that have been reported to traverse the cell membrane [68]. Representative members of this family such as the Trans-Activator of Transcription (TAT) peptide (Fig. 2b) and penetratin were initially identified as segments within naturally occurring proteins with proposed membrane permeability [26], such as homeoproteins [69].

The mechanisms of how these peptides cross the cell membrane is still unclear and likely differs amongst peptides. The TAT peptide for example (GRKKRRQRRRPSQ) is rich in arginines, and the delocalized positive charge in their guanidinium moieties has been proposed to allow extensive interactions with negatively charged cell membranes [70, 71]. TAT is thought to bind to the glycosaminoglycans (GAG) on the cell surface such as heparin sulfate [72] or adsorb into the glycerol backbone region of the lipid bilayer [73], eventually being macropinocytosed [74, 75]. Various models of TAT-mediated perturbations of the cell membrane have been proposed, including the formation of transient pores [76–80].

The reported permeability coefficient of TAT across artificial membranes is very low at 2.7×10^{-9} cm/s [81], which may reflect its need for structural features specific to the cell membrane to be able to translocate. Typically, relatively high (μM) concentrations of TAT are required *in vitro* to observe translocation, and the efficiency of such may depend on the cell line [82, 83]. Also, as mentioned earlier, fixation of cells treated with fluorescently labeled TAT may lead to artifacts in cellular distribution [25], and thus reported results need to be interpreted with caution.

4.3.2 Protein Toxins

A number of plant and bacterial toxins are potent inhibitors of central cellular functions such as protein synthesis. However, before they can have their effect, they must gain access to their cytosolic targets [84]. Typically, a separate domain (typically denoted the B domain, translocation domain, or translocation complex (when an oligomer)) is responsible for binding to cellular receptors and translocating the catalytic domain (the A domain) into the cytoplasm (*see* [85] for an illustration).

Some toxins form their own pores, such as the diphtheria and anthrax toxins. The translocation domains of anthrax toxin, known as protective antigen (PA), oligomerizes into a pre-pore complex

following proteolytic activation (Fig. 2c). Subsequent internalization and endosomal acidification is thought to trigger its conversion into a full pore, through which catalytic domains escape into the cytosol [86].

A number of other toxins, such as the plant toxin ricin, *Pseudomonas* exotoxin A, and cholera toxin, take advantage of the ERAD machinery to enter the cell [87–89]. Following binding to gangliosides via its B domains, cholera toxin is internalized and trafficked to the ER where the A domain is reduced and unfolded. This domain is subsequently refolded in the cytoplasm [90, 91].

4.3.3 Viruses

Some viruses enter the cell through the plasma membrane, but more commonly from endocytic compartments after binding to cellular receptors and triggering various endocytic pathways [92]. Viruses can be classified into enveloped viruses, which are encased in a lipid membrane containing glycoproteins, or non-enveloped viruses, which lack a membrane.

In general, enveloped viruses are thought to orchestrate the fusion of host and viral membranes using viral fusion proteins, which expose hydrophobic peptides upon environmental triggers such as receptor binding, low pH, or proteolytic cleavage. For example, influenza A exposes a hydrophobic segment of hemagglutinin (HA) upon endosomal acidification [93]. With this mechanism, there is no need to translocate across the cell membrane.

Non-enveloped viruses, in contrast, have to cross the membrane in order to access the cytoplasm. In general, these viruses are thought to mediate the disruption of cellular membranes by exposing or releasing lytic peptides that are amphipathic or hydrophobic [94–96]. Alternatively, members of the polyomavirus family such as the simian virus (SV40) use a strategy similar to the aforementioned cholera toxin, and hijack the ERAD machinery [97, 98].

5 Approaches to Design and Improve Membrane Permeability

5.1 Small Molecule Cargo

Decades of pharmaceutical research have provided design principles that maximize the chance of obtaining a drug able to efficiently distribute within an organism and permeate through cell membranes. While bioavailability is often the reported parameter of interest, efficient membrane permeation is likely necessary for bioavailability [99]. Therefore, rules that have been devised in medicinal chemistry to achieve favorable bioavailability are a reasonable guide for the design of membrane-permeating small molecules.

5.1.1 Predicting Passive Permeation

Lipinski's "Rule of 5" has been the most influential framework correlating the physicochemical properties of a given compound with its membrane permeability and bioavailability in the context of small-molecule drug development [100]. It postulates that poor

absorption or permeation is more likely when: (1) the calculated lipophilicity ($c\log P$) is over 5; (2) the molecular weight is over 500; (3) there are more than five hydrogen bond donors (well represented by the sum of OH and NH bonds); and (4) there are more than ten hydrogen bond acceptors (represented roughly, by the sum of Ns and Os).

The Rule of 5 has been generally successful at predicting membrane permeability, but not all compounds that comply with the rules are permeable, and permeable compounds that deviate from the rules are not uncommon [46, 101]. Nonetheless, as suggested by Guimarães et al. [46], the Rule of 5 does identify key physicochemical parameters, namely the polarity, size, and lipophilicity of the permeant, that are important for passive diffusion. These interrelated factors can affect the partitioning, diffusion, or both, of the molecule into and across the membrane.

Alternative metrics for these parameters have also been proposed. Regarding polarity, the polar surface area (PSA) of a compound has been used in addition to the number of hydrogen bond donors and acceptors [99, 102]. For molecular size, studies have inversely-correlated the permeability of small solutes with molecular volume [38] or cross-sectional area [103]. A different but related parameter, the number of rotatable bonds, has been suggested as well, where molecules with fewer rotatable bonds and lower PSA were reported to have better permeability across artificial membranes [99]. Additionally, it has also been proposed that conformationally flexible molecules that are able to form intramolecular hydrogen bonds in a low dielectric environment may adaptively reduce their surface polarity for improved permeation [104]. Unsurprisingly, even if the hydrogen bond counts or PSA is low, localized charge or highly polar groups can significantly decrease the permeability of an otherwise permeable parent compound by orders of magnitude [105, 106].

Beyond empirical correlations, molecular dynamics (MD) simulations are increasingly applied to calculate the energetic barrier of transmembrane diffusion, from which permeability coefficients can be derived [15, 107, 108]. Improved computational power and coarse-grained modeling have reduced computing time. However, although these methods are invaluable in estimating permeabilities that are difficult to obtain experimentally, utilizing them on a routine basis is yet hampered by the computational cost and the effort involved in building a suitable representation of the molecule of interest. Estimates for large molecules may be particularly prone to inaccuracy due to insufficient sampling of their conformational space during the simulations.

5.1.2 Predicting Transporter-Mediated Permeation

Designing compounds to be substrates of a specific transporter is currently difficult [109], although indirect approaches have been proposed to identify metabolites that are structurally similar to a

given compound [110]. Alternatively, conjugating compounds to known transporter substrates such as amino acids has been reported to improve permeation and oral adsorption by engaging PepT1 [111] (Fig. 2d). In such “prodrug” approaches, the conjugated substrates are designed to be cleaved intracellularly or during circulation to release the free drug [112]. Although designing specific transporter substrates is infeasible at the moment, it should be kept in mind that transporters can affect a compound’s permeation.

5.1.3 Comparing Theory and Empirical Data for Molecular Probes

Empirical permeability data from molecular probes and labeling molecules roughly agree with the theoretical expectations discussed above. Generally, small and uncharged fluorophores, and those whose charge is delocalized over the fluorophore (e.g., TAMRA), are sufficiently membrane-permeable to be used in intracellular protein labeling applications [113]. However, fluorescent dyes carrying localized charges (e.g., the sulfonic acid derivatives of Cy3 or Cy5) display low membrane permeability [113]. Esterification of charged groups is one strategy to mask the effects of charge [114].

An example of the size-dependence of membrane translocation is provided by fluorescent dyes modified with long and hydrophobic lipid-like tails. For the voltage-sensitive dyes Di-4-ANEPPS and Di-8-ANEPPS (equipped with two octyl and butyl chains, respectively), a strong decrease in membrane flip-flop was observed across planar black lipid membranes for the long-chain variant [115]. A similar result was obtained for the dyes DiI-C12 and DiI-C18 [116]. The counterintuitive result where increasing the overall hydrophobicity of the molecules strongly reduced the rate of flip-flop is likely due to the concomitant increase in molecular size. A similar result has been reported with anthroyl fatty acids in liposomes, where the rate of flip-flop was observed to be 200-fold faster for a C11-fatty acid compared to a C18-fatty acid [117].

As mentioned earlier, these molecular probes may also be substrates of cellular transporters. For example, acetoxyethyl ester (AM) derivatives of various fluorescent indicators were observed to be actively exported from cells by multidrug transporters [118]. Of note, passive diffusion and active transport may occur concomitantly. Chidley et al. studied the intracellular access of various organic molecules used for protein labeling via the SNAP-tag system in yeast strains that were either wild-type or had three efflux transporters deleted [119]. The study showed a strong decrease in uptake with increasing size and polarity of the labeling molecule, suggesting entry by passive diffusion. Additionally, it showed that labeling efficiency increased in the modified yeast strain, presumably due to reduced active export.

5.2 Peptide Cargo

It is unlikely that peptides will passively diffuse across the cell membrane, but altering their physical properties (such as conformational flexibility and polarity) has been proposed to improve their permeability. Despite interesting findings—a selection of which is

discussed in the following—conflicting experimental results have been reported. A straightforward method for converting a non-permeable peptide into an efficiently permeating entity is thus not available so far.

5.2.1 Addressing Conformation and Polarity

Macrocyclic drugs—those with a ring architecture of 12 or more atoms, including cyclic peptides—tend to be larger and more polar than most small-molecule drugs, falling outside the Rule of 5 [120, 121]. Yet some are administered orally [121], suggesting that they may be membrane permeable [43, 99, 122]. In the case of cyclosporin A (Fig. 2a), this is believed to occur by passive diffusion.

Following such examples, cyclizing a given peptide and methylating select amide bond nitrogens have been proposed to improve its membrane permeation and/or bioavailability. Such modifications, when made judiciously [123], are thought to facilitate the formation of intramolecular hydrogen bonds in response to the low dielectric environment of the membrane interior [43, 45, 124]. Passive permeability values ranging from 6.3×10^{-7} cm/s [45] to approximately 7.7×10^{-6} cm/s [124] (estimated from [125]) have been reported for certain hydrophobic cyclic peptides.

Alternatively, cyclization and amidation may alter a compound's specificity towards membrane transporters. In a study of 54 cyclic alanine hexapeptides containing various degrees of N-methylation, Ovadia et al. reported that none of the tested peptides showed permeation across artificial membranes. However, some peptides were found to be highly permeable across Caco-2 cell monolayers (on the order of 10^{-5} cm/s), suggesting that transporters may be involved [126].

In some instances, cyclization by covalently linking internal residues has been proposed to increase permeability by changing the peptide's α -helical content [127]. Such modifications include hydrocarbon “staples” linking the side chains of nonnatural amino acids inserted into the peptide [128], and “hydrogen bond surrogates” replacing a main chain hydrogen bond with a carbon–carbon [129] or disulfide bond [130]. Such modifications have lead to the development of peptide inhibitors against intracellular targets such as the ICN1/CSL complex (involved in the NOTCH signaling pathway) [22], Ras [131] and MDM2/MDMX [23].

However, introduction of a staple alone does not guarantee an improvement in permeability [132–134]. Extensive optimization may still be required for multiple factors such as the position, length, and stereochemistry of the staple [135], as well as the charge and amino acid sequence of the peptide [136].

5.2.2 Designing Cell-Penetrating Peptides

Extensive effort in discovering novel membrane-permeable peptides has generated significant diversity in the physicochemical character of reported CPPs [137]. Methods have been proposed to synthetically design permeable peptides or predict such segments from a given protein sequence [138].

Introducing arginine residues within α -helices has been proposed to improve permeability [139]. In a study of the avian pancreatic polypeptide (aPP), a 36-residue peptide/miniature protein, and CP1, a 28-residue zinc finger, substituting five residues within the α -helix with arginine increased the permeability to that comparable with TAT [139]. The authors estimated that approximately 1–5 % of the internalized peptides were being released into the cytosol.

5.3 Protein Cargo

Proteins cannot passively diffuse across the cell membrane due to their size and polarity. Thus, a delivery system or technique is always required, similar to nucleic acid transfection. However, while nucleic acid transfection reagents are now routinely used in the laboratory, there are no equivalent standards for the delivery of proteins. In the following, we survey strategies that have been proposed to deliver proteins across the cell membrane. Given the physicochemical diversity of proteins and their delicate nature, it is challenging to design a system or method that is readily generalizable to multiple proteins while maintaining the cargo's respective function and stability. For more comprehensive reviews, please see [140–143], as well as those cited below.

5.3.1 Mechanical Disruption of the Membrane

Varying physical methods of disrupting the cell membrane, such as microinjection and electroporation [144], have been proposed for delivering compounds ranging from small molecules to proteins. Sharei et al. developed a microfluidic device that transiently disrupts the plasma membrane through physical constriction [145]. Silicon “nanowires” that pierce the cell membrane have also been reported [146, 147].

5.3.2 Peptide-Based Strategies

CPPs have been reported to enhance the permeability of various macromolecules, including proteins [148–150]. Early studies showed that the TAT peptide can mediate the translocation of covalently coupled proteins [151, 152]. In later studies, an amphiphilic CPP Pep-1 was reported to noncovalently complex and translocate peptide and protein cargos [153].

Substance P (SP), an 11-residue neuropeptide implicated in cancer progression [154], has been proposed to mediate the cytosolic delivery of synthetic antibody fragments [155] and nucleic acids [156] following covalent conjugation. Its natural GPCR partner, the neurokinin-1 receptor (NK1R), has been suggested to play a role in mediating uptake. The mechanisms by which such peptides mediate translocation remains to be clarified.

5.3.3 Protein-Based Strategies

Various pore- or channel-forming proteins of bacterial origin have been utilized to translocate exogenous proteins. Highly sophisticated secretion systems, which transport proteins directly from the bacterial cytoplasm to the eukaryotic host's [157], have been

reported to deliver proteins to the cytosol of antigen-presenting cells [158]. Doerner et al. reported the functional expression of an engineered bacterial channel (MscL) in mammalian cells, the opening and closing of which could be controlled chemically [159]. Alternatively, the cholesterol-dependent cytolysin (CDC) family of pore-forming toxins, which are capable of forming macropores up to 30 nm in diameter [160], have been proposed as “reversible permeabilization” reagents for delivering exogenous proteins [161, 162].

In addition to pore- or channel-forming proteins, the membrane-translocating domains of bacterial toxins have been proposed as a modular tool that can be fused to, and enhance the intracellular delivery of, other proteins [163, 164]. In instances where the receptor-binding domain of the toxin is physically distinct from the translocation domain, the former has been replaced with alternative targeting moieties to generate immunotoxins. Immunotoxins retain the cytotoxicity of the parent toxin but are directed at specific cell types [165, 166].

Additionally, “supercharged” GFP, a variant engineered to have high net positive charge (+36) [167], and certain human proteins with naturally high positive charge [168, 169] have been reported to translocate across the cell membrane. Curiously, 3E10, an autoantibody proposed to bind to dsDNA [170], has been proposed to penetrate into the nucleus and impair DNA repair [171], or translocate an exogenous phosphatase across the cell membrane [172].

5.3.4 Virus-Based Strategies

Packaging proteins in virus-like particles [173] or attaching them to an engineered bacteriophage T4 head [174] has been reported to enhance cytosolic delivery. In addition, although not yet utilized as a delivery system, it has been reported that virus-bound antibodies co-internalize into the cytoplasm along with the virus [175].

5.3.5 Lipid- and Polymer-Based Strategies

With lipid-based materials, the protein cargo is either encapsulated in liposomes [176] or complexed with lipids. Regarding the latter strategy, lipid formulations that have been successful in the transfection of DNA have been attempted for the protein delivery. For example, a formulation based on a mixture of cationic and neutral lipids was reported to translocate negatively charged proteins [177].

Similarly, polymer-based formulations that have been successfully used for nucleic acid transfactions have also been examined for their ability to “transfect” proteins. The “proton sponge effect” is an influential hypothesis still undergoing debate [178], which states that materials such as polyethylenimine (PEI) that are rich in protonatable amines, will cause a significant buffering of protons and subsequent osmotic swelling in endosomes. The endosome is then proposed to stall its maturation and eventually rupture [179]. Poly-β-amino esters (PBAEs), successfully developed for the transfection of nucleic acids [180], are thought to take advantage of this proton sponge effect. Su et al. reported that biodegradable PBAE

nanoparticles enhance the endosomal escape of various cargos, including proteins, when co-administered [181].

Alternatively, Yan et al. reported a technology to encapsulate single proteins in a polymeric shell (termed “nanocapsule”) after attaching the monomeric building blocks of the polymer directly to the protein [182, 183]. Such nanocapsules, designed to be degraded in response to environmental stimuli such as protease activity or changes in pH or redox potential, were reported to deliver proteins including transcription factors [184].

5.3.6 Inorganic Material-Based Strategies

A variety of inorganic materials have also been proposed to translocate protein cargo, including silica, carbon nanotubes, quantum dots, and gold nanoparticles (*see* [142, 185]).

6 Conclusions

The plasma membrane of a mammalian cell is an intricate composite of multiple lipid and protein species continually undergoing endocytosis and exocytosis. Whereas small molecules with moderate polarity are able to diffuse through the cell membrane passively, most metabolites and short peptides require specialized membrane transporters for translocation. Proteins are generally unable to cross the cell membrane, with protein toxins being exceptions where sophisticated (and yet to be fully elucidated) mechanisms have been evolved for translocation.

In this review, we have surveyed proposed strategies on how to obtain membrane-permeable molecules that utilize passive diffusion, membrane transporters, or engineered delivery systems (Fig. 1). For passive diffusion, the Rule of 5 and its derivatives provide a rough guide for design. Minimizing the size of the desired permeant and its effective polarity is recommended. The latter can be achieved by minimizing the number of polar groups and localized charges in the molecule, or by cyclization, amidation and esterification strategies that shield polar groups in the interior of the molecule or mask a charge. It is currently unfeasible to explicitly design molecules as transporter substrates, although conjugation strategies have been proposed to known transporter substrates. When interpreting experimental results showing the entry or export of molecules, the possible contribution from membrane transporters should be kept in mind. Finally, various methods and delivery systems have been proposed to transport proteins across the cell membrane, from mechanical disruption to utilizing delivery systems that are either covalently attached or noncovalently complexed to the protein of interest.

In addition to providing an overview of engineering approaches, we have strived to provide a framework for evaluating the effectiveness of such strategies. Here, it cannot be understated that the

experimental assay employed to demonstrate cytosolic delivery, depending on its sensitivity and mode of detection, could greatly impact the perceived efficacy of the delivery method. Particularly with protein delivery, an easy-to-implement, standardized assay that can accurately quantify delivery performance would be invaluable in objectively comparing different platforms. Also, while frequently overlooked, cytotoxicity caused by the delivery vehicle, if any, should be explicitly addressed. Finally, the mechanism(s) of action allowing delivery should be thoroughly investigated to avoid experimental artifacts.

In summary, we have summarized strategies employed by nature or devised by man to transport small molecules, peptides, and proteins across cell membranes. We hope this review will provide scientists interested in designing cell-permeable probes or effector molecules with a starting point to approach the task. Second, we hope it will aid in bringing together the concepts and solutions generated for diverse payloads. Although overcoming the membrane barrier remains a challenging and incompletely solved problem, significant progress continues to be made towards enabling potentially powerful applications in biological research and medicine.

Acknowledgements

The authors thank Bradley Pentelute, Alessandro Angelini, Sandrine Sagan, Alexander H. de Vries, and Christopher Chidley for helpful discussions and critical reading of the manuscript.

References

1. Stein WD, Lieb WR (1986) Transport and diffusion across cell membranes, 1st edn. Academic, Orlando, FL
2. Alberts B, Johnson A, Lewis J et al (2007) Molecular biology of the cell, 5th edn. Garland Science, New York
3. Di L, Artursson P, Avdeef A et al (2012) Evidence-based approach to assess passive diffusion and carrier-mediated drug transport. *Drug Discov Today* 17:905–912. doi:[10.1016/j.drudis.2012.03.015](https://doi.org/10.1016/j.drudis.2012.03.015)
4. Engelman DM (2005) Membranes are more mosaic than fluid. *Nature* 438:578–580. doi:[10.1038/nature04394](https://doi.org/10.1038/nature04394)
5. Jacobson K, Mouritsen OG, Anderson RGW (2007) Lipid rafts: at a crossroad between cell biology and physics. *Nat Cell Biol* 9:7–14. doi:[10.1038/ncb0107-7](https://doi.org/10.1038/ncb0107-7)
6. Koichi K, Michiya F, Makoto N (1974) Lipid components of two different regions of an intestinal epithelial cell membrane of mouse. *Biochim Biophys Acta* 369:222–233. doi:[10.1016/0005-2760\(74\)90253-7](https://doi.org/10.1016/0005-2760(74)90253-7)
7. Marsh D, Horváth LI (1998) Structure, dynamics and composition of the lipid-protein interface. Perspectives from spin-labelling. *Biochim Biophys Acta* 1376:267–296. doi:[10.1016/S0304-4157\(98\)00009-4](https://doi.org/10.1016/S0304-4157(98)00009-4)
8. Lee AG (2003) Lipid–protein interactions in biological membranes: a structural perspective. *Biochim Biophys Acta* 1612:1–40. doi:[10.1016/S0005-2736\(03\)00056-7](https://doi.org/10.1016/S0005-2736(03)00056-7)
9. Zachowski A (1993) Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement. *Biochem J* 294:1–14
10. Leventis R, Silvius JR (2001) Use of cyclodextrins to monitor transbilayer movement and differential lipid affinities of cholesterol. *Biophys J* 81:2257–2267. doi:[10.1016/S0006-3495\(01\)75873-0](https://doi.org/10.1016/S0006-3495(01)75873-0)
11. Steck TL, Ye J, Lange Y (2002) Probing red cell membrane cholesterol movement with

- cyclodextrin. *Biophys J* 83:2118–2125. doi:[10.1016/S0006-3495\(02\)73972-6](https://doi.org/10.1016/S0006-3495(02)73972-6)
12. Conner SD, Schmid SL (2003) Regulated portals of entry into the cell. *Nature* 422:37–44. doi:[10.1038/nature01451](https://doi.org/10.1038/nature01451)
 13. Mercer J, Helenius A (2009) Virus entry by macropinocytosis. *Nat Cell Biol* 11:510–520. doi:[10.1038/ncb0509-510](https://doi.org/10.1038/ncb0509-510)
 14. Alberts B, Johnson A, Lewis J et al (2002) Molecular biology of the cell. <http://www.ncbi.nlm.nih.gov/books/NBK21054/>. Accessed 27 Feb 2014
 15. Orsi M, Essex JW (2010) Passive permeation across lipid bilayers: a literature review. In: Molecular simulations and biomembranes: from biophysics to function, p 76–90
 16. Kansy M, Senner F, Gubernator K (1998) Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J Med Chem* 41:1007–1010. doi:[10.1021/jm970530c](https://doi.org/10.1021/jm970530c)
 17. Sugano K, Kansy M, Artursson P et al (2010) Coexistence of passive and carrier-mediated processes in drug transport. *Nat Rev Drug Discov* 9:597–614. doi:[10.1038/nrd3187](https://doi.org/10.1038/nrd3187)
 18. Di L, Whitney-Pickett C, Umland JP et al (2011) Development of a new permeability assay using low-efflux MDCKII cells. *J Pharm Sci* 100:4974–4985. doi:[10.1002/jps.22674](https://doi.org/10.1002/jps.22674)
 19. Shamu CE, Story CM, Rapoport TA, Ploegh HL (1999) The pathway of Us11-dependent degradation of Mhc class I heavy chains involves a ubiquitin-conjugated intermediate. *J Cell Biol* 147:45–58. doi:[10.1083/jcb.147.1.45](https://doi.org/10.1083/jcb.147.1.45)
 20. Bartz R, Fan H, Zhang J et al (2011) Effective siRNA delivery and target mRNA degradation using an amphipathic peptide to facilitate pH-dependent endosomal escape. *Biochem J* 435:475–487. doi:[10.1042/BJ20101021](https://doi.org/10.1042/BJ20101021)
 21. Bittner MA, Holz RW (1988) Effects of tetanus toxin on catecholamine release from intact and digitonin-permeabilized chromaffin cells. *J Neurochem* 51:451–456. doi:[10.1111/j.1471-4159.1988.tb01059.x](https://doi.org/10.1111/j.1471-4159.1988.tb01059.x)
 22. Moellering RE, Cornejo M, Davis TN et al (2009) Direct inhibition of the NOTCH transcription factor complex. *Nature* 462:182–188. doi:[10.1038/nature08543](https://doi.org/10.1038/nature08543)
 23. Chang YS, Graves B, Guerlavais V et al (2013) Stapled α -helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc Natl Acad Sci* 110:E3445–E3454. doi:[10.1073/pnas.1303002110](https://doi.org/10.1073/pnas.1303002110)
 24. Bonner DK, Leung C, Chen-Liang J et al (2011) Intracellular trafficking of polyamidoamine-poly(ethylene glycol) block copolymers in DNA delivery. *Bioconjug Chem* 22:1519–1525. doi:[10.1021/bc200059v](https://doi.org/10.1021/bc200059v)
 25. Richard JP, Melikov K, Vives E et al (2003) Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. *J Biol Chem* 278:585–590. doi:[10.1074/jbc.M209548200](https://doi.org/10.1074/jbc.M209548200)
 26. Bechara C, Sagan S (2013) Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett* 587:1693–1702. doi:[10.1016/j.febslet.2013.04.031](https://doi.org/10.1016/j.febslet.2013.04.031)
 27. Cebrian I, Visentin G, Blanchard N et al (2011) Sec22b regulates phagosomal maturation and antigen crosspresentation by dendritic cells. *Cell* 147:1355–1368. doi:[10.1016/j.cell.2011.11.021](https://doi.org/10.1016/j.cell.2011.11.021)
 28. Yu P, Liu B, Kodadek T (2005) A high-throughput assay for assessing the cell permeability of combinatorial libraries. *Nat Biotechnol* 23:746–751. doi:[10.1038/nbt1099](https://doi.org/10.1038/nbt1099)
 29. Holub JM, LaRochelle JR, Appelbaum JS, Scheetz A (2013) Improved assays for determining the cytosolic access of peptides, proteins, and their mimetics. *Biochemistry (Mosc)* 52:9036–9046. doi:[10.1021/bi401069g](https://doi.org/10.1021/bi401069g)
 30. Zlokarnik G, Negulescu PA, Knapp TE et al (1998) Quantitation of transcription and clonal selection of single living cells with β -lactamase as reporter. *Science* 279:84–88. doi:[10.1126/science.279.5347.84](https://doi.org/10.1126/science.279.5347.84)
 31. Bordonaro M (2009) Modular Cre/lox system and genetic therapeutics for colorectal cancer. *J Biomed Biotechnol*. doi:[10.1155/2009/358230](https://doi.org/10.1155/2009/358230)
 32. Yamaizumi M, Mekada E, Uchida T, Okada Y (1978) One molecule of diphtheria toxin fragment a introduced into a cell can kill the cell. *Cell* 15:245–250. doi:[10.1016/0092-8674\(78\)90099-5](https://doi.org/10.1016/0092-8674(78)90099-5)
 33. Eiklid K, Olsnes S, Pihl A (1980) Entry of lethal doses of abrin, ricin and modeccin into the cytosol of HeLa cells. *Exp Cell Res* 126:321–326. doi:[10.1016/0014-4827\(80\)90270-0](https://doi.org/10.1016/0014-4827(80)90270-0)
 34. Diamond JM, Katz Y (1974) Interpretation of nonelectrolyte partition coefficients between dimyristoyl lecithin and water. *J Membr Biol* 17:121–154. doi:[10.1007/BF01870176](https://doi.org/10.1007/BF01870176)
 35. Finkelstein A (1976) Water and nonelectrolyte permeability of lipid bilayer membranes. *J Gen Physiol* 68:127–135. doi:[10.1085/jgp.68.2.127](https://doi.org/10.1085/jgp.68.2.127)

36. Subczynski WK, Hyde JS, Kusumi A (1989) Oxygen permeability of phosphatidylcholine-cholesterol membranes. *Proc Natl Acad Sci* 86:4474–4478
37. Gutknecht J, Bisson MA, Tosteson FC (1977) Diffusion of carbon dioxide through lipid bilayer membranes: effects of carbonic anhydrase, bicarbonate, and unstirred layers. *J Gen Physiol* 69:779–794. doi:[10.1085/jgp.69.6.779](https://doi.org/10.1085/jgp.69.6.779)
38. Walter A, Gutknecht J (1986) Permeability of small nonelectrolytes through lipid bilayer membranes. *J Membr Biol* 90:207–217. doi:[10.1007/BF01870127](https://doi.org/10.1007/BF01870127)
39. Orbach E, Finkelstein A (1980) The nonelectrolyte permeability of planar lipid bilayer membranes. *J Gen Physiol* 75:427–436. doi:[10.1085/jgp.75.4.427](https://doi.org/10.1085/jgp.75.4.427)
40. Papahadjopoulos D, Nir S, Oki S (1972) Permeability properties of phospholipid membranes: effect of cholesterol and temperature. *Biochim Biophys Acta* 266:561–583
41. Mendel CM (1989) The free hormone hypothesis: a physiologically based mathematical model. *Endocr Rev* 10:232–274. doi:[10.1210/edrv-10-3-232](https://doi.org/10.1210/edrv-10-3-232)
42. Giorgi EP, Stein WD (1981) The transport of steroids into animal cells in culture. *Endocrinology* 108:688–697. doi:[10.1210/endo-108-2-688](https://doi.org/10.1210/endo-108-2-688)
43. Bockus AT, McEwen CM, Lokey RS (2013) Form and function in cyclic peptide natural products: a pharmacokinetic perspective. *Curr Top Med Chem* 13:821–836
44. Augustijns PF, Bradshaw TP, Gan LSL et al (1993) Evidence for a polarized efflux system in Caco-2 cells capable of modulating cyclosporine A transport. *Biochem Biophys Res Commun* 197:360–365. doi:[10.1006/bbrc.1993.2487](https://doi.org/10.1006/bbrc.1993.2487)
45. Rezai T, Bock JE, Zhou MV et al (2006) Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J Am Chem Soc* 128:14073–14080. doi:[10.1021/ja063076p](https://doi.org/10.1021/ja063076p)
46. Guimarães CRW, Mathiowitz AM, Shalaeva M et al (2012) Use of 3D properties to characterize beyond rule-of-5 property space for passive permeation. *J Chem Inf Model* 52:882–890. doi:[10.1021/ci300010y](https://doi.org/10.1021/ci300010y)
47. Hediger MA, Cléménçon B, Burrier RE, Bruford EA (2013) The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med* 34:95–107. doi:[10.1016/j.mam.2012.12.009](https://doi.org/10.1016/j.mam.2012.12.009)
48. Saier MH, Reddy VS, Tamang DG, Vastermark A (2013) The transporter classification database. *Nucleic Acids Res* 42: D251–D258. doi:[10.1093/nar/gkt1097](https://doi.org/10.1093/nar/gkt1097)
49. Hediger MA (2013) The ABCs of membrane transporters in health and disease (SLC series). *Mol Aspects Med* 34(2–3):95–752
50. Kew JNC, Davies CH (2010) Ion channels: from structure to function. Oxford University Press, Oxford
51. Enyedi P, Czirják G (2010) Molecular background of leak K⁺ currents: two-pore domain potassium channels. *Physiol Rev* 90:559–605. doi:[10.1152/physrev.00029.2009](https://doi.org/10.1152/physrev.00029.2009)
52. Toyoshima C, Kanai R, Cornelius F (2011) First crystal structures of Na⁺, K⁺-ATPase: new light on the oldest ion pump. *Structure* 19: 1732–1738. doi:[10.1016/j.str.2011.10.016](https://doi.org/10.1016/j.str.2011.10.016)
53. Duax WL, Griffin JF, Langs DA et al (1996) Molecular structure and mechanisms of action of cyclic and linear ion transport antibiotics. *Pept Sci* 40:141–155. doi:[10.1002/\(SICI\)1097-0282\(1996\)40:1<141::AID-BIP6>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0282(1996)40:1<141::AID-BIP6>3.0.CO;2-W)
54. Wallace BA (1998) Recent advances in the high resolution structures of bacterial channels: gramicidin A. *J Struct Biol* 121:123–141. doi:[10.1006/jbsb.1997.3948](https://doi.org/10.1006/jbsb.1997.3948)
55. Zheng L, Kostrewa D, Bernèche S et al (2004) The mechanism of ammonia transport based on the crystal structure of AmtB of Escherichia coli. *Proc Natl Acad Sci U S A* 101:17090–17095. doi:[10.1073/pnas.0406475101](https://doi.org/10.1073/pnas.0406475101)
56. Andrade SLA, Einsle O (2007) The Amt/Mep/Rh family of ammonium transport proteins. *Mol Membr Biol* 24:357–365. doi:[10.1080/09687680701388423](https://doi.org/10.1080/09687680701388423)
57. Shayakul C, Cléménçon B, Hediger MA (2013) The urea transporter family (SLC14): physiological, pathological and structural aspects. *Mol Aspects Med* 34:313–322. doi:[10.1016/j.mam.2012.12.003](https://doi.org/10.1016/j.mam.2012.12.003)
58. Ishibashi K, Hara S, Kondo S (2009) Aquaporin water channels in mammals. *Clin Exp Nephrol* 13:107–117. doi:[10.1007/s10157-008-0118-6](https://doi.org/10.1007/s10157-008-0118-6)
59. Bienert GP, Chaumont F (2013) Aquaporin-facilitated transmembrane diffusion of hydrogen peroxide. *Biochim Biophys Acta*. doi:[10.1016/j.bbagen.2013.09.017](https://doi.org/10.1016/j.bbagen.2013.09.017)
60. Mueckler M, Thorens B (2013) The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med* 34:121–138. doi:[10.1016/j.mam.2012.07.001](https://doi.org/10.1016/j.mam.2012.07.001)
61. Schweikhard ES, Ziegler CM (2012) Amino acid secondary transporters: toward a common transport mechanism. *Curr Top Membr*

- 70:1–28. doi:[10.1016/B978-0-12-394316-3.00001-6](https://doi.org/10.1016/B978-0-12-394316-3.00001-6)
62. Young JD, Yao SYM, Baldwin JM et al (2013) The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29. *Mol Aspects Med* 34:529–547. doi:[10.1016/j.mam.2012.05.007](https://doi.org/10.1016/j.mam.2012.05.007)
63. Smith DE, Clémenton B, Hediger MA (2013) Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol Aspects Med* 34:323–336. doi:[10.1016/j.mam.2012.11.003](https://doi.org/10.1016/j.mam.2012.11.003)
64. Letschert K, Faulstich H, Keller D, Keppler D (2006) Molecular characterization and inhibition of amanitin uptake into human hepatocytes. *Toxicol Sci* 91:140–149. doi:[10.1093/toxsci/kfj141](https://doi.org/10.1093/toxsci/kfj141)
65. Chen Z-S, Tiwari AK (2011) Multidrug resistance proteins (MRPs/ABCCs) in cancer chemotherapy and genetic diseases. *FEBS J* 278:3226–3245. doi:[10.1111/j.1742-4658.2011.08235.x](https://doi.org/10.1111/j.1742-4658.2011.08235.x)
66. Amin ML (2013) P-glycoprotein inhibition for optimal drug delivery. *Drug Target Insights* 7:27–34. doi:[10.4137/DTI.S12519](https://doi.org/10.4137/DTI.S12519)
67. Natarajan K, Xie Y, Baer MR, Ross DD (2012) Role of breast cancer resistance protein (BCRP/ABCG2) in cancer drug resistance. *Biochem Pharmacol* 83:1084–1103. doi:[10.1016/j.bcp.2012.01.002](https://doi.org/10.1016/j.bcp.2012.01.002)
68. Langel U (2010) Handbook of cell-penetrating peptides, 2nd edn. CRC Press, Boca Raton
69. Sagan S, Burlina F, Alves ID et al (2013) Homeoproteins and homeoprotein-derived peptides: going in and out. *Curr Pharm Des* 19:2851–2862
70. Schmidt N, Mishra A, Lai GH, Wong GCL (2010) Arginine-rich cell-penetrating peptides. *FEBS Lett* 584:1806–1813. doi:[10.1016/j.febslet.2009.11.046](https://doi.org/10.1016/j.febslet.2009.11.046)
71. Futaki S, Hirose H, Nakase I (2013) Arginine-rich peptides: methods of translocation through biological membranes. *Curr Pharm Des* 19:2863–2868
72. Tyagi M, Rusnati M, Presta M, Giacca M (2001) Internalization of HIV-1 Tat requires cell surface heparan sulfate proteoglycans. *J Biol Chem* 276:3254–3261. doi:[10.1074/jbc.M006701200](https://doi.org/10.1074/jbc.M006701200)
73. Su Y, Waring AJ, Ruchala P, Hong M (2010) Membrane-bound dynamic structure of an arginine-rich cell-penetrating peptide, the protein transduction domain of HIV TAT, from solid-state NMR. *Biochemistry (Mosc)* 49:6009–6020. doi:[10.1021/bi100642n](https://doi.org/10.1021/bi100642n)
74. Wadia JS, Stan RV, Dowdy SF (2004) Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat Med* 10:310–315. doi:[10.1038/nm996](https://doi.org/10.1038/nm996)
75. Nakase I, Tadokoro A, Kawabata N et al (2007) Interaction of arginine-rich peptides with membrane-associated proteoglycans is crucial for induction of actin organization and macropinocytosis. *Biochemistry (Mosc)* 46:492–501. doi:[10.1021/bi0612824](https://doi.org/10.1021/bi0612824)
76. Yesylevskyy S, Marrink S-J, Mark AE (2009) Alternative mechanisms for the interaction of the cell-penetrating peptides penetratin and the TAT peptide with lipid bilayers. *Biophys J* 97:40–49. doi:[10.1016/j.bpj.2009.03.059](https://doi.org/10.1016/j.bpj.2009.03.059)
77. Herce HD, Garcia AE, Litt J et al (2009) Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides. *Biophys J* 97:1917–1925. doi:[10.1016/j.bpj.2009.05.066](https://doi.org/10.1016/j.bpj.2009.05.066)
78. Mishra A, Lai GH, Schmidt NW et al (2011) Translocation of HIV TAT peptide and analogues induced by multiplexed membrane and cytoskeletal interactions. *Proc Natl Acad Sci* 108:16883–16888. doi:[10.1073/pnas.1108795108](https://doi.org/10.1073/pnas.1108795108)
79. Kawamoto S, Miyakawa T, Takasu M et al (2012) Cell-penetrating peptide induces various deformations of lipid bilayer membrane: inverted micelle, double bilayer, and transmembrane. *Int J Quantum Chem* 112:178–183. doi:[10.1002/qua.23177](https://doi.org/10.1002/qua.23177)
80. Huang K, García AE (2013) Free energy of translocating an arginine-rich cell-penetrating peptide across a lipid bilayer suggests pore formation. *Biophys J* 104:412–420. doi:[10.1016/j.bpj.2012.10.027](https://doi.org/10.1016/j.bpj.2012.10.027)
81. Jones S, Howl J (2012) Enantiomer-specific bioactivities of peptidomimetic analogues of mastoparan and mitoparan: characterization of inverso mastoparan as a highly efficient cell penetrating peptide. *Bioconjug Chem* 23:47–56. doi:[10.1021/bc2002924](https://doi.org/10.1021/bc2002924)
82. Tréhin R, Krauss U, Beck-Sickinger AG et al (2004) Cellular uptake but low permeation of human calcitonin-derived cell penetrating peptides and Tat(47–57) through well-differentiated epithelial models. *Pharm Res* 21:1248–1256. doi:[10.1023/B:PHAM.0000033013.45204.c3](https://doi.org/10.1023/B:PHAM.0000033013.45204.c3)
83. Foerg C, Merkle HP (2008) On the biomedical promise of cell penetrating peptides: limits versus prospects. *J Pharm Sci* 97:144–162. doi:[10.1002/jps.21117](https://doi.org/10.1002/jps.21117)
84. Sandvig K, van Deurs B (2005) Delivery into cells: lessons learned from plant and bacterial toxins. *Gene Ther* 12:865–872. doi:[10.1038/sj.gt.3302525](https://doi.org/10.1038/sj.gt.3302525)

85. Falnes PØ, Sandvig K (2000) Penetration of protein toxins into cells. *Curr Opin Cell Biol* 12:407–413. doi:[10.1016/S0955-0674\(00\)00109-5](https://doi.org/10.1016/S0955-0674(00)00109-5)
86. Collier RJ (2009) Membrane translocation by anthrax toxin. *Mol Aspects Med* 30:413–422. doi:[10.1016/j.mam.2009.06.003](https://doi.org/10.1016/j.mam.2009.06.003)
87. De Virgilio M, Lombardi A, Caliandro R, Fabbrini MS (2010) Ribosome-inactivating proteins: from plant defense to tumor attack. *Toxins* 2:2699–2737. doi:[10.3390/toxins2112699](https://doi.org/10.3390/toxins2112699)
88. Spooner RA, Lord JM (2012) How ricin and shiga toxin reach the cytosol of target cells: retrotranslocation from the endoplasmic reticulum. In: Mantis N (ed) Ricin shiga toxins. Springer, Berlin, pp 19–40
89. Sandvig K, Skotland T, van Deurs B, Klokk TI (2013) Retrograde transport of protein toxins through the Golgi apparatus. *Histochem Cell Biol* 140:317–326. doi:[10.1007/s00418-013-1111-z](https://doi.org/10.1007/s00418-013-1111-z)
90. Wernick NLB, Chinnappen DJ-F, Cho JA, Lencer WI (2010) Cholera toxin: an intracellular journey into the cytosol by way of the endoplasmic reticulum. *Toxins* 2:310–325. doi:[10.3390/toxins2030310](https://doi.org/10.3390/toxins2030310)
91. Cho JA, Chinnappen DJ-F, Aamar E et al (2012) Insights on the trafficking and retro-translocation of glycosphingolipid-binding bacterial toxins. *Front Cell Infect Microbiol*. doi:[10.3389/fcimb.2012.00051](https://doi.org/10.3389/fcimb.2012.00051)
92. Mercer J, Schelhaas M, Helenius A (2010) Virus entry by endocytosis. *Annu Rev Biochem* 79:803–833. doi:[10.1146/annurev-biochem-060208-104626](https://doi.org/10.1146/annurev-biochem-060208-104626)
93. Sriwilaijaroen N, Suzuki Y (2012) Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proc Jpn Acad Ser B Phys Biol Sci* 88:226–249. doi:[10.2183/pjab.88.226](https://doi.org/10.2183/pjab.88.226)
94. Tsai B (2007) Penetration of nonenveloped viruses into the cytoplasm. *Annu Rev Cell Dev Biol* 23:23–43. doi:[10.1146/annurev.cellbio.23.090506.123454](https://doi.org/10.1146/annurev.cellbio.23.090506.123454)
95. Johnson J, Banerjee M (2008) Activation, exposure and penetration of virally encoded, membrane-active polypeptides during non-enveloped virus entry. *Curr Protein Pept Sci* 9:16–27. doi:[10.2174/138920308783565732](https://doi.org/10.2174/138920308783565732)
96. Moyer CL, Nemerow GR (2011) Viral weapons of membrane destruction: variable modes of membrane penetration by non-enveloped viruses. *Curr Opin Virol* 1:44–99. doi:[10.1016/j.coviro.2011.05.002](https://doi.org/10.1016/j.coviro.2011.05.002)
97. Inoue T, Tsai B (2013) How viruses use the endoplasmic reticulum for entry, replication, and assembly. *Cold Spring Harb Perspect Biol* 5:a013250. doi:[10.1101/cshperspect.a013250](https://doi.org/10.1101/cshperspect.a013250)
98. Suomalainen M, Greber UF (2013) Uncoating of non-enveloped viruses. *Curr Opin Virol* 3:27–33. doi:[10.1016/j.coviro.2012.12.004](https://doi.org/10.1016/j.coviro.2012.12.004)
99. Veber DF, Johnson SR, Cheng H-Y et al (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45:2615–2623. doi:[10.1021/jm020017n](https://doi.org/10.1021/jm020017n)
100. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25. doi:[10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
101. Faller B, Ottaviani G, Ertl P et al (2011) Evolution of the physicochemical properties of marketed drugs: can history foretell the future? *Drug Discov Today* 16:976–984. doi:[10.1016/j.drudis.2011.07.003](https://doi.org/10.1016/j.drudis.2011.07.003)
102. Ertl P, Rohde B, Selzer P (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem* 43:3714–3717. doi:[10.1021/jm000942e](https://doi.org/10.1021/jm000942e)
103. Xiang T-X, Anderson BD (1998) Influence of chain ordering on the selectivity of dipalmitoylphosphatidylcholine bilayer membranes for permeant size and shape. *Biophys J* 75:2658–2671. doi:[10.1016/S0006-3495\(98\)77711-2](https://doi.org/10.1016/S0006-3495(98)77711-2)
104. Kuhn B, Mohr P, Stahl M (2010) Intramolecular hydrogen bonding in medicinal chemistry. *J Med Chem* 53:2601–2611. doi:[10.1021/jm100087s](https://doi.org/10.1021/jm100087s)
105. Mayer PT, Xiang T-X, Anderson BD (2000) Independence of substituent contributions to the transport of small-molecule permeants in lipid bilayer. *AAPS Pharm Sci* 2:40–52. doi:[10.1208/ps020214](https://doi.org/10.1208/ps020214)
106. Ulander J, Haymet ADJ (2003) Permeation across hydrated DPPC lipid bilayers: simulation of the titratable amphiphilic drug valproic acid. *Biophys J* 85:3475–3484. doi:[10.1016/S0006-3495\(03\)74768-7](https://doi.org/10.1016/S0006-3495(03)74768-7)
107. Xiang T-X, Anderson BD (2006) Liposomal drug transport: a molecular perspective from molecular dynamics simulations in lipid bilayers. *Adv Drug Deliv Rev* 58:1357–1378. doi:[10.1016/j.addr.2006.09.002](https://doi.org/10.1016/j.addr.2006.09.002)
108. Bennett WFD, MacCallum JL, Hinner MJ et al (2009) Molecular view of cholesterol flip-flop and chemical potential in different

- membrane environments. *J Am Chem Soc* 131:12714–12720. doi:[10.1021/ja903529f](https://doi.org/10.1021/ja903529f)
109. Maeda K, Sugiyama Y (2013) Transporter biology in drug approval: regulatory aspects. *Mol Aspects Med* 34:711–718. doi:[10.1016/j.mam.2012.10.012](https://doi.org/10.1016/j.mam.2012.10.012)
 110. Dobson PD, Patel Y, Kell DB (2009) “Metabolite-likeness” as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Discov Today* 14:31–40. doi:[10.1016/j.drudis.2008.10.011](https://doi.org/10.1016/j.drudis.2008.10.011)
 111. Dahan A, Khamis M, Agbaria R, Karaman R (2012) Targeted prodrugs in oral drug delivery: the modern molecular biopharmaceutical approach. *Expert Opin Drug Deliv* 9:1001–1013. doi:[10.1517/17425247.2012.697055](https://doi.org/10.1517/17425247.2012.697055)
 112. Majumdar S, Duvvuri S, Mitra AK (2004) Membrane transporter/receptor-targeted prodrug design: strategies for human and veterinary drug development. *Adv Drug Deliv Rev* 56:1437–1452. doi:[10.1016/j.addr.2004.02.006](https://doi.org/10.1016/j.addr.2004.02.006)
 113. Keppler A, Arrivoli C, Sironi L, Ellenberg J (2006) Fluorophores for live cell imaging of AGT fusion proteins across the visible spectrum. *Biotechniques* 41:167–170, 172, 174–175
 114. Tsien RY (1981) A non-disruptive technique for loading calcium buffers and indicators into cells. *Nature* 290:527–528
 115. Ries RS, Choi H, Blunck R et al (2004) Black lipid membranes: visualizing the structure, dynamics, and substrate dependence of membranes. *J Phys Chem B* 108:16040–16049. doi:[10.1021/jp048098h](https://doi.org/10.1021/jp048098h)
 116. Melikyan GB, Deriy BN, Ok DC, Cohen FS (1996) Voltage-dependent translocation of R18 and DiI across lipid bilayers leads to fluorescence changes. *Biophys J* 71:2680–2691. doi:[10.1016/S0006-3495\(96\)79459-6](https://doi.org/10.1016/S0006-3495(96)79459-6)
 117. Kleinfeld AM, Chu P, Storch J (1997) Flip-flop is slow and rate limiting for the movement of long chain anthroyloxy fatty acids across lipid vesicles. *Biochemistry (Mosc)* 36:5702–5711. doi:[10.1021/bi962007s](https://doi.org/10.1021/bi962007s)
 118. Homolya L, Holló Z, Germann UA et al (1993) Fluorescent cellular indicators are extruded by the multidrug resistance protein. *J Biol Chem* 268:21493–21496
 119. Chidley C, Haruki H, Pedersen MG et al (2011) A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat Chem Biol* 7:375–383. doi:[10.1038/nchembio.557](https://doi.org/10.1038/nchembio.557)
 120. Driggers EM, Hale SP, Lee J, Terrett NK (2008) The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat Rev Drug Discov* 7:608–624. doi:[10.1038/nrd2590](https://doi.org/10.1038/nrd2590)
 121. Giordanetto F, Revell JD, Knerr L et al (2013) Stapled vasoactive intestinal peptide (VIP) derivatives improve VPAC2 agonism and glucose-dependent insulin secretion. *ACS Med Chem Lett* 4:1163–1168. doi:[10.1021/ml400257h](https://doi.org/10.1021/ml400257h)
 122. Bock JE, Gavenonis J, Kritzer JA (2013) Getting in shape: controlling peptide bioactivity and bioavailability using conformational constraints. *ACS Chem Biol* 8:488–499. doi:[10.1021/cb300515u](https://doi.org/10.1021/cb300515u)
 123. Kwon Y-U, Kodadek T (2007) Quantitative comparison of the relative cell permeability of cyclic and linear peptides. *Chem Biol* 14:671–677. doi:[10.1016/j.chembiol.2007.05.006](https://doi.org/10.1016/j.chembiol.2007.05.006)
 124. White TR, Renzelman CM, Rand AC et al (2011) On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat Chem Biol* 7:810–817. doi:[10.1038/nchembio.664](https://doi.org/10.1038/nchembio.664)
 125. Malakoutikhah M, Prades R, Teixidó M, Giralt E (2010) N-methyl phenylalanine-rich peptides as highly versatile blood–brain barrier shuttles. *J Med Chem* 53:2354–2363. doi:[10.1021/jm901654x](https://doi.org/10.1021/jm901654x)
 126. Ovadia O, Greenberg S, Chatterjee J et al (2011) The effect of multiple N-methylation on intestinal permeability of cyclic hexapeptides. *Mol Pharm* 8:479–487. doi:[10.1021/mp1003306](https://doi.org/10.1021/mp1003306)
 127. Azzarito V, Long K, Murphy NS, Wilson AJ (2013) Inhibition of α -helix-mediated protein–protein interactions using designed molecules. *Nat Chem* 5:161–173. doi:[10.1038/nchem.1568](https://doi.org/10.1038/nchem.1568)
 128. Kim Y-W, Grossmann TN, Verdine GL (2011) Synthesis of all-hydrocarbon stapled α -helical peptides by ring-closing olefin metathesis. *Nat Protoc* 6:761–771. doi:[10.1038/nprot.2011.324](https://doi.org/10.1038/nprot.2011.324)
 129. Patgiri A, Menzenski MZ, Mahon AB, Arora PS (2010) Solid-phase synthesis of short α -helices stabilized by the hydrogen bond surrogate approach. *Nat Protoc* 5:1857–1865. doi:[10.1038/nprot.2010.146](https://doi.org/10.1038/nprot.2010.146)
 130. Miller SE, Kallenbach NR, Arora PS (2012) Reversible alpha-helix formation controlled by a hydrogen bond surrogate. *Tetrahedron* 68:4434–4437. doi:[10.1016/j.tet.2011.12.068](https://doi.org/10.1016/j.tet.2011.12.068)
 131. Patgiri A, Yadav KK, Arora PS, Bar-Sagi D (2011) An orthosteric inhibitor of the Ras-Sos interaction. *Nat Chem Biol* 7:585–587. doi:[10.1038/nchembio.612](https://doi.org/10.1038/nchembio.612)
 132. Okamoto T, Zobel K, Fedorova A et al (2013) Stabilizing the pro-apoptotic BimBH3

- Helix (BimSAHB) does not necessarily enhance affinity or biological activity. *ACS Chem Biol* 8:297–302. doi:[10.1021/cb3005403](https://doi.org/10.1021/cb3005403)
133. Bird GH, Gavathiotis E, LaBelle JL et al (2014) Distinct BimBH3 (BimSAHB) stapled peptides for structural and cellular studies. *ACS Chem Biol* 9:831–837. doi:[10.1021/cb4003305](https://doi.org/10.1021/cb4003305)
134. Okamoto T, Segal D, Zobel K et al (2014) Further insights into the effects of pre-organizing the BimBH3 helix. *ACS Chem Biol* 9:838–839. doi:[10.1021/cb400638p](https://doi.org/10.1021/cb400638p)
135. Verdine GL, Hilinski GJ (2012) Stapled peptides for intracellular drug targets. In: Dane Wittrup K, Verdine GL (eds) *Methods enzymol.* Academic, New York, pp 3–33
136. Bird GH, Christian Crannell W, Walensky LD (2011) Chemical synthesis of hydrocarbon-stapled peptides for protein interaction research and therapeutic targeting. *Curr Protoc Chem Biol* 3(3):99–117
137. Milletti F (2012) Cell-penetrating peptides: classes, origin, and current landscape. *Drug Discov Today* 17:850–860. doi:[10.1016/j.drudis.2012.03.002](https://doi.org/10.1016/j.drudis.2012.03.002)
138. Copolovici DM, Langel K, Eriste E, Langel U (2014) Cell-penetrating peptides: design synthesis and applications. *ACS Nano*. doi:[10.1021/nn4057269](https://doi.org/10.1021/nn4057269)
139. Appelbaum JS, LaRochelle JR, Smith BA et al (2012) Arginine topology controls escape of minimally cationic proteins from early endosomes to the cytoplasm. *Chem Biol* 19:819–830. doi:[10.1016/j.chembiol.2012.05.022](https://doi.org/10.1016/j.chembiol.2012.05.022)
140. Marschall ALJ, Frenzel A, Schirrmann T et al (2011) Targeting antibodies to the cytoplasm. *mAbs* 3:3–16. doi:[10.4161/mabs.3.1.14110](https://doi.org/10.4161/mabs.3.1.14110)
141. Gu Z, Biswas A, Zhao M, Tang Y (2011) Tailoring nanocarriers for intracellular protein delivery. *Chem Soc Rev* 40:3638–3655. doi:[10.1039/C0CS00227E](https://doi.org/10.1039/C0CS00227E)
142. Du J, Jin J, Yan M, Lu Y (2012) Synthetic nanocarriers for intracellular protein delivery. *Curr Drug Metab* 13:82–92. doi:[10.2174/138920012798356862](https://doi.org/10.2174/138920012798356862)
143. Salmaso S, Caliceti P (2013) Self assembling nanocomposites for protein delivery: supramolecular interactions of soluble polymers with protein drugs. *Int J Pharm* 440:111–123. doi:[10.1016/j.ijpharm.2011.12.029](https://doi.org/10.1016/j.ijpharm.2011.12.029)
144. Zhang Y, Yu L-C (2008) Microinjection as a tool of mechanical delivery. *Curr Opin Biotechnol* 19:506–510. doi:[10.1016/j.copbio.2008.07.005](https://doi.org/10.1016/j.copbio.2008.07.005)
145. Sharei A, Zoldan J, Adamo A et al (2013) A vector-free microfluidic platform for intracellular delivery. *Proc Natl Acad Sci* 110:2082–2087. doi:[10.1073/pnas.1218705110](https://doi.org/10.1073/pnas.1218705110)
146. Shalek AK, Robinson JT, Karp ES et al (2010) Vertical silicon nanowires as a universal platform for delivering biomolecules into living cells. *Proc Natl Acad Sci* 107:1870–1875. doi:[10.1073/pnas.0909350107](https://doi.org/10.1073/pnas.0909350107)
147. Yosef N, Shalek AK, Gaublomme JT et al (2013) Dynamic regulatory network controlling TH17 cell differentiation. *Nature* 496:461–468. doi:[10.1038/nature11981](https://doi.org/10.1038/nature11981)
148. Lo SL, Wang S (2010) Peptide-based nanocarriers for intracellular delivery of biologically active proteins. In: *Organelle-specific pharmaceutical nanotechnology*, p 323–336
149. Koren E, Torchilin VP (2012) Cell-penetrating peptides: breaking through to the other side. *Trends Mol Med* 18:385–393. doi:[10.1016/j.molmed.2012.04.012](https://doi.org/10.1016/j.molmed.2012.04.012)
150. Nakase I, Tanaka G, Futaki S (2013) Cell-penetrating peptides (CPPs) as a vector for the delivery of siRNAs into cells. *Mol Biosyst* 9:855–861. doi:[10.1039/C2MB25467K](https://doi.org/10.1039/C2MB25467K)
151. Fawell S, Seery J, Daikh Y et al (1994) Tat-mediated delivery of heterologous proteins into cells. *Proc Natl Acad Sci* 91:664–668. doi:[10.1073/pnas.91.2.664](https://doi.org/10.1073/pnas.91.2.664)
152. Nagahara H, Vocero-Akbani AM, Snyder EL et al (1998) Transduction of full-length TAT fusion proteins into mammalian cells: TAT-p27Kip1 induces cell migration. *Nat Med* 4:1449–1452. doi:[10.1038/4042](https://doi.org/10.1038/4042)
153. Morris MC, Depollier J, Mery J et al (2001) A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat Biotechnol* 19:1173–1176. doi:[10.1038/nbt1201-1173](https://doi.org/10.1038/nbt1201-1173)
154. Harford-Wright E, Lewis KM, Vink R, Ghabriel MN (2014) Evaluating the role of substance P in the growth of brain tumors. *Neuroscience* 261:85–94. doi:[10.1016/j.neuroscience.2013.12.027](https://doi.org/10.1016/j.neuroscience.2013.12.027)
155. Rizk SS, Luchniak A, Uysal S et al (2009) An engineered substance P variant for receptor-mediated delivery of synthetic antibodies into tumor cells. *Proc Natl Acad Sci* 106:11011–11015. doi:[10.1073/pnas.0904907106](https://doi.org/10.1073/pnas.0904907106)
156. Rizk SS, Misiura A, Paduch M, Kossiakoff AA (2011) Substance P derivatives as versatile tools for specific delivery of various types of biomolecular cargo. *Bioconjug Chem* 23: 42–46. doi:[10.1021/bc200496e](https://doi.org/10.1021/bc200496e)
157. Chatterjee S, Chaudhury S, McShan AC et al (2013) Structure and biophysics of type III secretion in bacteria. *Biochemistry (Mosc)* 52:2508–2517. doi:[10.1021/bi400160a](https://doi.org/10.1021/bi400160a)
158. Carleton HA, Lara-Tejero M, Liu X, Galán JE (2013) Engineering the type III secretion

- system in non-replicating bacterial minicells for antigen delivery. *Nat Commun* 4:1590. doi:[10.1038/ncomms2594](https://doi.org/10.1038/ncomms2594)
159. Doerner JF, Febvay S, Clapham DE (2012) Controlled delivery of bioactive molecules into live cells using the bacterial mechanosensitive channel MscL. *Nat Commun* 3:990. doi:[10.1038/ncomms1999](https://doi.org/10.1038/ncomms1999)
 160. Dunstone MA, Tweten RK (2012) Packing a punch: the mechanism of pore formation by cholesterol dependent cytolysins and membrane attack complex/perforin-like proteins. *Curr Opin Struct Biol* 22:342–349. doi:[10.1016/j.sbi.2012.04.008](https://doi.org/10.1016/j.sbi.2012.04.008)
 161. Provoda CJ, Stier EM, Lee K-D (2003) Tumor cell killing enabled by listeriolysin O-liposome-mediated delivery of the protein toxin gelonin. *J Biol Chem* 278:35102–35108. doi:[10.1074/jbc.M305411200](https://doi.org/10.1074/jbc.M305411200)
 162. Pirie CM, Liu DV, Wittrup KD (2013) Targeted cytolysins synergistically potentiate cytoplasmic delivery of gelonin immunotoxin. *Mol Cancer Ther* 12:1774–1782. doi:[10.1158/1535-7163.MCT-12-1023](https://doi.org/10.1158/1535-7163.MCT-12-1023)
 163. Sandvig K, van Deurs B (2002) Membrane traffic exploited by protein toxins. *Annu Rev Cell Dev Biol* 18:1–24. doi:[10.1146/annurev.cellbio.18.011502.142107](https://doi.org/10.1146/annurev.cellbio.18.011502.142107)
 164. Johannes L, Römer W (2010) Shiga toxins—from cell biology to biomedical applications. *Nat Rev Microbiol* 8:105–116. doi:[10.1038/nrmicro2279](https://doi.org/10.1038/nrmicro2279)
 165. Pastan I, Hassan R, FitzGerald DJ, Kreitman RJ (2007) Immunotoxin treatment of cancer. *Annu Rev Med* 58:221–237. doi:[10.1146/annurev.med.58.070605.115320](https://doi.org/10.1146/annurev.med.58.070605.115320)
 166. FitzGerald DJ, Wayne AS, Kreitman RJ, Pastan I (2011) Treatment of hematologic malignancies with immunotoxins and antibody-drug conjugates. *Cancer Res* 71:6300–6309. doi:[10.1158/0008-5472.CAN-11-1374](https://doi.org/10.1158/0008-5472.CAN-11-1374)
 167. Lawrence MS, Phillips KJ, Liu DR (2007) Supercharging proteins can impart unusual resilience. *J Am Chem Soc* 129:10110–10112. doi:[10.1021/ja071641y](https://doi.org/10.1021/ja071641y)
 168. Cronican JJ, Thompson DB, Beier KT et al (2010) Potent delivery of functional proteins into mammalian cells in vitro and in vivo using a supercharged protein. *ACS Chem Biol* 5:747–752. doi:[10.1021/cb1001153](https://doi.org/10.1021/cb1001153)
 169. Cronican JJ, Beier KT, Davis TN et al (2011) A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo. *Chem Biol* 18:833–838. doi:[10.1016/j.chembiol.2011.07.003](https://doi.org/10.1016/j.chembiol.2011.07.003)
 170. Weisbart RH, Noritake DT, Wong AL et al (1990) A conserved anti-DNA antibody idiotype associated with nephritis in murine and human systemic lupus erythematosus. *J Immunol* 144:2653–2658
 171. Hansen JE, Chan G, Liu Y et al (2012) Targeting cancer with a lupus autoantibody. *Sci Transl Med* 4:157ra142. doi:[10.1126/scitranslmed.3004385](https://doi.org/10.1126/scitranslmed.3004385)
 172. Lawlor MW, Armstrong D, Viola MG et al (2013) Enzyme replacement therapy rescues weakness and improves muscle pathology in mice with X-linked myotubular myopathy. *Hum Mol Genet* 22:1525–1538. doi:[10.1093/hmg/ddt003](https://doi.org/10.1093/hmg/ddt003)
 173. Kaczmarczyk SJ, Sitaraman K, Young HA et al (2011) Protein delivery using engineered virus-like particles. *Proc Natl Acad Sci* 108:16998–17003. doi:[10.1073/pnas.1101874108](https://doi.org/10.1073/pnas.1101874108)
 174. Tao P, Mahalingam M, Marasa BS et al (2013) In vitro and in vivo delivery of genes and proteins using the bacteriophage T4 DNA packaging machine. *Proc Natl Acad Sci* 110:5846–5851. doi:[10.1073/pnas.1300867110](https://doi.org/10.1073/pnas.1300867110)
 175. Mallory DL, McEwan WA, Bidgood SR et al (2010) Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc Natl Acad Sci* 107:19985–19990. doi:[10.1073/pnas.1014074107](https://doi.org/10.1073/pnas.1014074107)
 176. Torchilin V (2008) Intracellular delivery of protein and peptide therapeutics. *Drug Discov Today Technol* 5:e95–e103. doi:[10.1016/j.ddtec.2009.01.002](https://doi.org/10.1016/j.ddtec.2009.01.002)
 177. Zelphati O, Wang Y, Kitada S et al (2001) Intracellular delivery of proteins with a new lipid-mediated delivery system. *J Biol Chem* 276:35103–35110. doi:[10.1074/jbc.M104920200](https://doi.org/10.1074/jbc.M104920200)
 178. Benjaminsen RV, Mattebjerg MA, Henriksen JR et al (2013) The possible “proton sponge” effect of polyethylenimine (PEI) does not include change in lysosomal pH. *Mol Ther* 21:149–157. doi:[10.1038/mt.2012.185](https://doi.org/10.1038/mt.2012.185)
 179. Behr J-P (1997) The proton sponge: a trick to enter cells the viruses did not exploit. *Chim Int J Chem* 51:34–36
 180. Lynn DM, Langer R (2000) Degradable poly(β-amino esters): synthesis, characterization, and self-assembly with plasmid DNA. *J Am Chem Soc* 122:10761–10768. doi:[10.1021/ja0015388](https://doi.org/10.1021/ja0015388)
 181. Su X, Yang N, Wittrup KD, Irvine DJ (2013) Synergistic antitumor activity from two-stage delivery of targeted toxins and endosome-disrupting nanoparticles. *Biomacromolecules* 14:1093–1102. doi:[10.1021/bm3019906](https://doi.org/10.1021/bm3019906)
 182. Gu Z, Yan M, Hu B et al (2009) Protein nanocapsule weaved with enzymatically

- degradable polymeric network. *Nano Lett* 9:4533–4538. doi:[10.1021/nl902935b](https://doi.org/10.1021/nl902935b)
183. Yan M, Du J, Gu Z et al (2010) A novel intracellular protein delivery platform based on single-protein nanocapsules. *Nat Nanotechnol* 5:48–53. doi:[10.1038/nnano.2009.341](https://doi.org/10.1038/nnano.2009.341)
184. Biswas A, Joo K-I, Liu J et al (2011) Endoprotease-mediated intracellular protein delivery using nanocapsules. *ACS Nano* 5:1385–1394. doi:[10.1021/nn1031005](https://doi.org/10.1021/nn1031005)
185. Malmsten M (2013) Inorganic nanomaterials as delivery systems for proteins, peptides, DNA, and siRNA. *Curr Opin Colloid Interface Sci* 18:468–480. doi:[10.1016/j.cocis.2013.06.002](https://doi.org/10.1016/j.cocis.2013.06.002)
186. Loosli H-R, Kessler H, Oschkinat H et al (1985) Peptide conformations. Part 31. The conformation of cyclosporin a in the crystal and in solution. *Helv Chim Acta* 68:682–704. doi:[10.1002/hcfa.19850680319](https://doi.org/10.1002/hcfa.19850680319)
187. Bayer P, Kraft M, Ejchart A et al (1995) Structural studies of HIV-1 tat protein. *J Mol Biol* 247:529–535. doi:[10.1016/S0022-2836\(05\)80133-0](https://doi.org/10.1016/S0022-2836(05)80133-0)
188. Feld GK, Thoren KL, Kintzer AF et al (2010) Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. *Nat Struct Mol Biol* 17:1383–1390. doi:[10.1038/nsmb.1923](https://doi.org/10.1038/nsmb.1923)
189. Varghese Gupta S, Gupta D, Sun J et al (2011) Enhancing the intestinal membrane permeability of zanamivir: a carrier mediated prodrug approach. *Mol Pharm* 8:2358–2367. doi:[10.1021/mp200291x](https://doi.org/10.1021/mp200291x)

Chapter 4

Considerations and Protocols for the Synthesis of Custom Protein Labeling Probes

Ivan R. Corrêa Jr.

Abstract

Chemists and biologists have long recognized small molecule probes as powerful tools for functional genomics and proteomics studies. The possibility of specifically attaching chemical probes to individual proteins with spatial and temporal resolution has greatly improved our ability to visualize and characterize proteins in their native environment. The continued development of novel molecular probes for protein labeling is, therefore, of fundamental importance to gain new insights into biological processes in living cells and organisms. Several excellent approaches for the site-specific labeling of fusion proteins with chemical probes exist. Herein I discuss the design and generation of chemical probes for the SNAP-tag and CLIP-tag systems. The first part of this chapter is dedicated to reviewing the principles of the SNAP-tag technology, followed by a section dedicated to the development of chemical probes for unique applications, such as super-resolution imaging, protein trafficking and recycling, protein–protein interactions, and biomolecular sensing. The last part of the chapter contains experimental protocols and technical notes for the synthesis of selected SNAP-tag substrates and labeling of SNAP-tag fusion proteins *in vitro* and in living cells.

Key words Fluorescent probes, Covalent labeling, Protein modifications, Cell imaging, SNAP-tag, CLIP-tag

1 Introduction

The success of any labeling strategy for live cell imaging and proteomics applications lies in the ability to specifically confer the desired chemical or optical properties to the target protein, thereby providing means to track, manipulate, and interrogate the protein in its native environment [1–3]. In recent years, several labeling approaches have emerged that utilize specific recognition sequences to recruit chemical probes for *in situ* labeling of cellular proteins. Among the most prominent techniques are self-labeling proteins, such as SNAP-tag [4, 5], CLIP-tag [6], HaloTag [7], TMP-tag [8], and BL-tag [9], and enzymatic ligation to peptide tags, such as those mediated by phosphopantetheinyl transferases (AcpS and Sfp) [10], sortase (SrtA) [11], and lipoic acid ligase (LplA) [12].

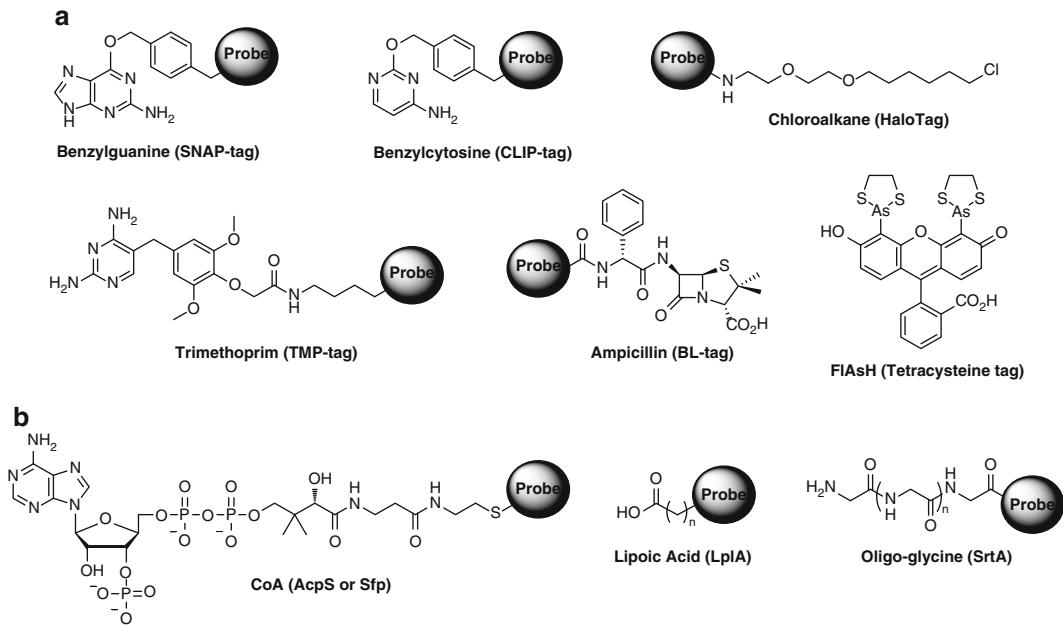


Fig. 1 Representative examples of molecular recognition motifs used in different strategies for selectively labeling proteins in living cells. **(a)** Self-labeling systems. **(b)** Enzyme-mediated labeling systems to an acceptor peptide tag

These chemoenzymatic approaches combine the convenience and specificity of genetically encoded systems with the versatility of small molecule probes. Once the fusion protein is generated, the tag can be modified with fluorescent dyes or affinity ligands, used for direct in-gel detection or Western blot assays, or immobilized on solid surfaces for purification or pull-down experiments. The small molecule probe is designed to incorporate a chemical motif that is selectively recognized by the tagged protein, thereby allowing the covalent attachment of the probe in a site-specific manner. Examples of the molecular recognition motifs for the most commonly used protein tags are illustrated in Fig. 1.

There are several excellent approaches for the site-specific labeling of fusion proteins with chemical probes, some of which are now available commercially, including the tetracysteine tag (LumioTM, Life Technologies), HaloTag[®] (Promega), TMP-tag (LigandLinkTM, Active Motif), and ACP-tag (New England Biolabs). Herein I discuss the design of small molecule probes for the SNAP-tag[®] and CLIP-tagTM systems (New England Biolabs). The first part of this chapter is dedicated to reviewing the principles of the SNAP-tag and CLIP-tag labeling systems. This is followed by a section dedicated to the design of chemical probes for unique applications, such as super-resolution imaging, protein trafficking and recycling, protein–protein interactions, and molecular biosensing.

Though the synthetic approaches referenced throughout the chapter focus on SNAP-tag substrates, the general principles of actually making the chemical probe conjugates are common across all labeling techniques. In this section, I will discuss some of these principles: For instance, the factors to be considered when selecting the appropriate chemistry to attach a probe to the protein substrate (e.g., the chemical compatibility of reactive moieties, the size and nature of the spacer separating the functional groups, and the stability of the linkage between label and substrate); the availability of reactive labels (e.g., amine- or thiol-reactive probes, azide- or alkyne-derived probes for copper-free or copper-catalyzed cycloadditions) and properties of the label (e.g., fluorescence, cell permeability, solubility); the effort required for synthetic and purification steps; and the necessary instrumentation. The last part of the chapter contains experimental protocols and technical notes for the synthesis of selected SNAP-tag substrates. The protocols include the synthesis of custom SNAP-tag substrates derived from a fluorophore active ester, an amino-functionalized quantum dot, a thiol-modified oligonucleotide, and the chemical attachment of a SNAP-tag substrate to a carboxy-functionalized resin. Additionally, a protocol for the labeling of SNAP-tag fusion proteins in mammalian cell lines is provided to illustrate the use of these substrates in living systems.

1.1 *SNAP-Tag and CLIP-Tag Labeling Systems*

The SNAP-tag is a 20-kDa engineered mutant of the human repair protein O^6 -alkylguanine-DNA alkyltransferase (hAGT) developed by Johnsson and co-workers [4, 5]. The reaction of SNAP-tag with O^6 -benzylguanine (BG) and O^6 -benzyl-4-chloropyrimidine (CP) [13] derivatives bearing a chemical or optical probe leads to the covalent and irreversible labeling of the fusion proteins. The label is transferred to the reactive cysteine residue (Cys145), yielding a stable thioether in a well-defined mechanism, with predictable stoichiometry and rapid kinetics, irrespective of the attached target protein [14]. SNAP-tag fusion proteins can be labeled with fluorescent dyes and affinity ligands [15] or immobilized on solid surfaces [16–18]. Labeling approaches based on SNAP-tag have been successful for many cellular and *in vivo* applications, including single molecule [19, 20] and super-resolution imaging [21, 22], analysis of protein function [23], identification of protein drug targets [24], determination of protein half-life in animals [25], and selective cross-linking of interacting protein partners [26, 27]. The CLIP-tag is a modified version of the SNAP-tag, which was further engineered to react with O^2 -benzylcytosine (BC) rather than with O^6 -benzylguanine derivatives [6]. The two tags have been used for orthogonal and complementary labeling of two proteins simultaneously in the same cell, for example, to visualize different generations of protein cohorts via double pulse-chase experiments [6], to sense the concentration of cell metabolites via

intramolecular FRET [28], to explore protein–protein interactions [26, 27], as well as to analyze cell surface protein GPCR heteromeric complexes [29].

A key advantage of small molecule probes over autofluorescence protein tagging is the ability to use chemistry to modulate the biophysical properties of synthetic fluorescent dyes to the needs imposed by a given biological experiment [30, 31]. Novel fluorogenic [32, 33] and photoactivatable [34–36] substrates as well as metal ion [37, 38] and cell metabolite [39, 40] fluorescent indicators have been designed to facilitate sensing and imaging experiments with high sensitivity and low background fluorescence. Owing to the covalent and quantitative nature of the labeling reaction, the SNAP-tag is well suited for the analysis and quantification of the tagged protein using common fluorescence techniques [15]. Furthermore, selective labeling of membrane or intracellular protein targets can be achieved by appropriate choice of cell permeable and impermeable substrates. A wide variety of SNAP-tag and CLIP-tag cell permeable and cell-impermeable substrates spanning the visible light spectrum as well as the near-infrared range are commercially available.

1.2 Design of Chemical Probes for Biological Applications

The SNAP-tag protein labeling system enables the specific, covalent attachment of virtually any molecule, including synthetic fluorophores, quantum dots (QDs), and gold nanoparticles to a protein of interest. The availability of chemical building blocks (Fig. 2) for synthesizing your own probe brings an added level of versatility to this labeling system. BG-NH₂ and BG-PEG-NH₂ are amine-terminated building blocks for the one-step synthesis of

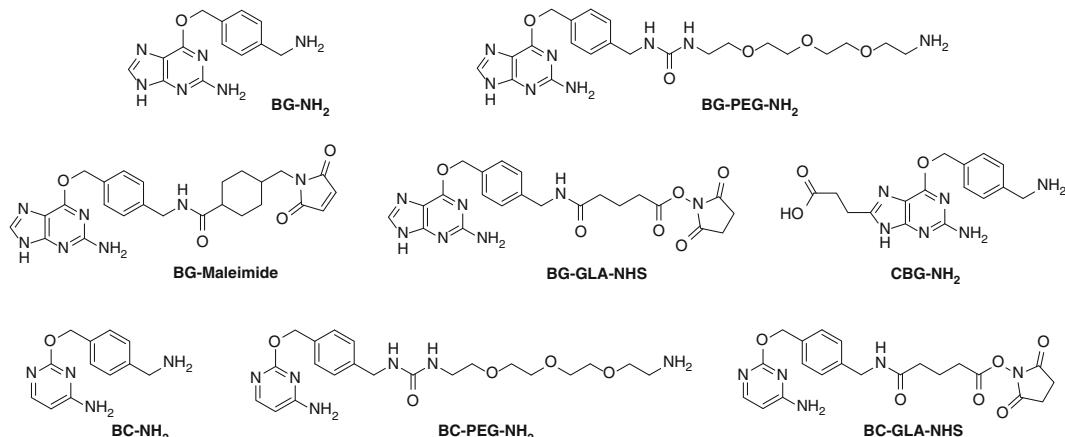


Fig. 2 Commercially available SNAP-tag and CLIP-tag reactive building blocks

SNAP-tag substrates from *N*-hydroxysuccinimide (NHS) esters or other activated carboxyl esters. BG-NH₂ is most commonly employed for the synthesis of fluorescent SNAP-tag substrates. BG-PEG-NH₂ has greater solubility than BG-NH₂ in most solvents (e.g., DMF, DMSO, and aqueous buffers), and is particularly useful for the synthesis of SNAP-tag substrates with “bulkier” substituents, affinity ligands, or solid surfaces. BG-GLA-NHS is an amine-reactive building block for the one-step synthesis of SNAP-tag substrates from amine-containing precursors including proteins, peptides, or oligonucleotides. BG-maleimide is a building block that allows the attachment of benzylguanine to thiol-containing precursors. CBG-NH₂ is a multifunctional building block used for the synthesis of dual-labeled SNAP-tag substrates with quencher/fluorophore pairs [33]. BC-NH₂, BC-PEG-NH₂, and BC-GLA-NHS are the equivalent building blocks for the synthesis of CLIP-tag substrates. Customized building blocks have also been described with azido and alkyne functionalities, for example, for the incorporation of adamantine and ferrocene to BG for supramolecular assembly of proteins on surfaces and vesicles [41, 42].

The choice of the suitable building block for your application will depend on coupling chemistries involved and the reactive functionality of your label agent [43]. In most cases, the size and polarity of the label will dictate the cell permeability of the BG conjugate. For example, neutral or positively charged fluorophores conjugates tend to be more permeant than negatively charged fluorophores conjugates. A critical aspect of detection using fluorescent chemical probes in living cells is their successful delivery through the plasma membrane into the cell. Cell impermeable substrates have been used for labeling *in vitro*, on surfaces of living cells, or after cell permeabilization/fixation. However, the use of cell-loading techniques, such as microinjection [35, 44] and transfection reagents [22], has permitted the labeling of intracellular protein targets with cell impermeable probes (e.g., near-infrared fluorescent dyes) in living cells. A particularly interesting technique, which requires no special equipment and is very simple to implement, utilizes glass beads to deliver BG-fluorophores into mammalian cells [36]. This approach relies on temporary membrane disruptions by mechanical friction caused by the contact between beads and cells, thereby creating a direct access to the cell cytoplasm. The possibility to perform intracellular labeling of fusion proteins with otherwise non-permeant probes, such as near-infrared dyes and dual-labeled fluorogenic probes, will significantly broaden the number of biological applications of the available labeling technologies. Very recently, Kai Johnsson and coworkers have introduced a novel class of cell-permeable near-infrared fluorescent probes based on a silicon-rhodamine dye which permits the

imaging of proteins in living cells and tissues without the use of any invasive cell-loading approaches [45].

The development of novel small molecule probes with unique biophysical and biochemical properties has greatly increased our ability to study proteins in living cells and whole organisms. In the following subsections, the design and application of some of these probes will be discussed. Specific examples are presented for the SNAP-tag labeling system; however, other labeling approaches may also be applicable.

1.3 Chemical Probes for Super-resolution Microscopy

The possibility to actively control the fluorescence emission at a time in a diffraction-limited region allows for the acquisition of super-resolution images for nanoscale visualization. Advancing beyond the diffraction limit requires either limiting the area of active fluorophores at a given point in time (e.g., STimulated Emission Depletion, STED) or limiting the number of fluorophores active over a larger area (e.g., PhotoActivated Localization Microscopy [PALM] and STochastic Optical Reconstruction Microscopy [STORM]) [46]. The success of any nanoscopy techniques depends strongly on the suitability of the fluorophore label—while PALM or STORM requires probes that can be cycled between fluorescent and dark states, STED fluorophores typically exhibit strong brightness, uninterrupted emission, and resistance to photobleaching. Photoswitchable fluorescent proteins have often been used for PALM-based super-resolution imaging because of their ability of being genetically targeted; however, they provide an average tenfold fewer photons before photobleaching than good small molecule emitters [46]. As site-specific labeling techniques, such as SNAP-tag, favor the spatial targeting of small molecule emitters to specific subcellular locations, they can be advantageously employed for super-resolution microscopy. A number of commercially available BG-fluorophore conjugates have been successfully used in combination with SNAP-tag to image cytoskeletal and cell membrane proteins (STED) [47], clathrin-coated pits (STORM) [22], and histone proteins (dSTORM) [48] with 20–40 nm resolution. Furthermore, a variety of chemical probes were custom designed to meet the biophysical requirements of these super-resolution techniques (Fig. 3), including silicon-rhodamines (STORM/STED)[45], photoactivatable caged rhodamines (PALM) [34], and photoswitchable substrates combining pairs of cyanine dyes (STORM) [49]. Multicolor super-resolution imaging has also been demonstrated in living cells using orthogonal labeling of SNAP-tag and CLIP-tag fusion proteins with synthetic probes [21]. While some of these customized probe conjugates can be synthesized in one chemical step from commercially available fluorophore precursors, others require a laborious multistep synthesis especially when multiple fluorophores and/or quenchers need to be assembled in the same scaffold.

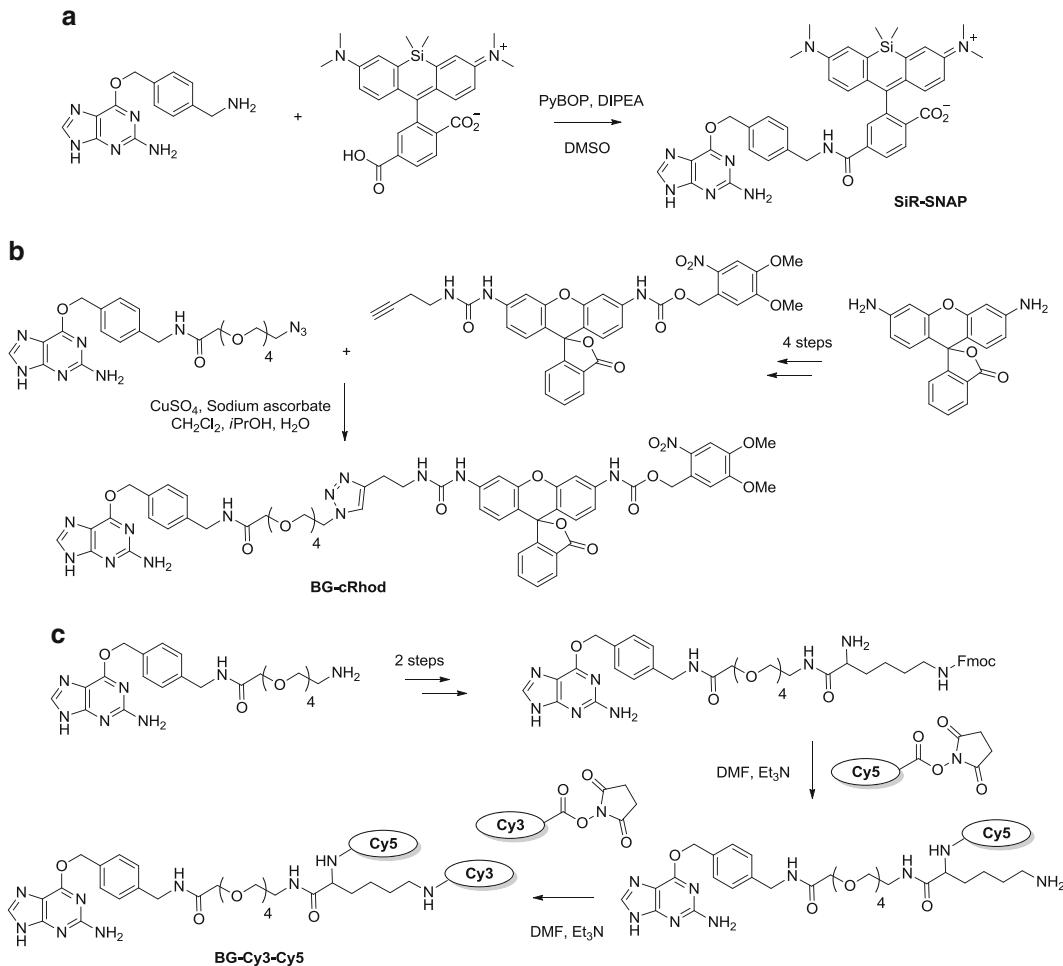


Fig. 3 Schematic of the synthesis of SNAP-tag substrates for targeted super-resolution imaging applications. (a) Synthesis of the near-infrared silicon-rhodamine SiR-SNAP via coupling reaction of BG-NH₂ to a 6-carboxy-Si-rhodamine using in situ activation mediated by PyBOP (benzotriazol-1-yl-oxytritylridinophosphonium hexafluorophosphate) and *N,N*-diisopropylethylamine [45]. (b) Synthesis of the caged rhodamine BG-cRhod via a copper-catalyzed azide-alkyne cycloaddition [34]. (c) Synthesis of the photoswitchable fluorescent probe BG-Cy3-Cy5 through a multistep coupling of the NHS ester derivatives of Cy3 and Cy5 cyanine dyes to a customized bifunctional BG building block [49]

1.4 Cleavable Probes for the Analysis of Protein Receptor Trafficking and Recycling

To help visualize membrane protein dynamics and recycling in living cells, a series of cleavable SNAP-tag substrates were developed that allow the rapid release of the fluorescent labels attached to cell surface proteins without affecting the population of labeled molecules internalized within endosomes. To achieve this goal, Donaldson and coworkers [50, 51] synthesized fluorescent probes (BG-S-S-488 and BG-S-S-594) that incorporate a cleavable disulfide bond between the BG moiety and the fluorophore label (Fig. 4). After internalization of various SNAP-tag fusion proteins

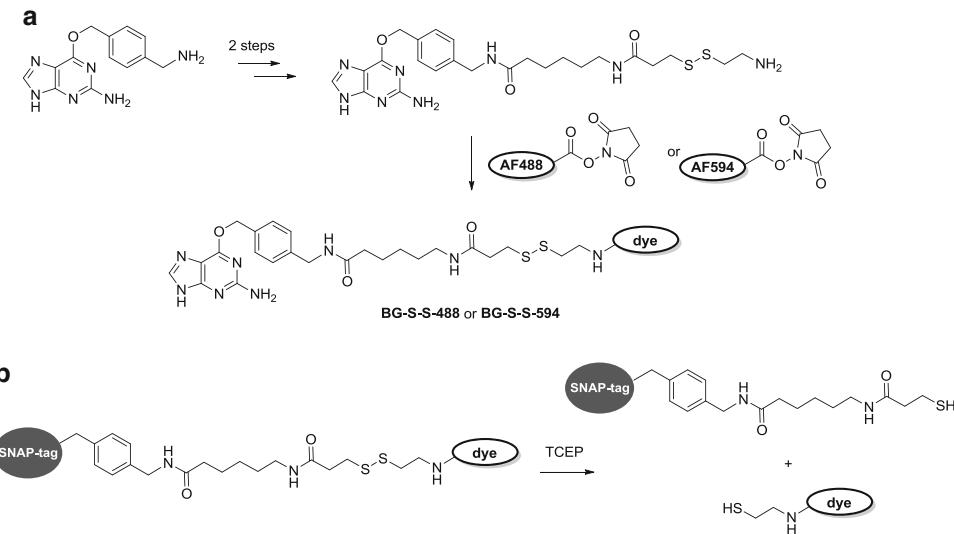


Fig. 4 Schematic of the synthesis of cleavable SNAP-tag fluorescent substrates. **(a)** The cleavable probes are synthesized using a customized BG-S-S-NH₂ building block and the corresponding fluorophore NHS esters. BG-S-S-NH₂ building block can be prepared in two steps from the reaction of BG-NH₂ with 6-(3-[2-pyridylthio]propionamido)hexanoate NHS (LC-SPDP), followed by treatment with excess of cysteamine [50]. **(b)** After labeling SNAP-tag fusion proteins with a cleavable probe, the fluorophore moiety can be readily released upon reduction of the disulfide bond (e.g., using 10–50 mM DTT, MESNA, or TCEP at pH 8.5)

labeled with cleavable probes, including the G-protein-coupled β2adrenergic receptor (β2ADR), the neurokinin-1 receptor (NK1R), and the human interleukin-2 receptor (TAC), the remaining cell surface-associated fluorescence is effectively removed by incubation with a cell-impermeable reducing agent (e.g., TCEP or MESNA) without affecting the population of internalized labeled molecules. This approach is particularly useful when studying endosomal dynamics in living cells as it permits to monitor recycling from internal compartments back to the cell surface.

1.5 Dual-Labeled Fluorogenic Probes

One of the biggest limitations of chemical-based labeling technologies is the relatively high fluorescence background caused by unreacted or nonspecifically bound substrate species. To address this issue, we and others developed SNAP-tag fluorogenic substrates (“dark” or “quenched” substrates) that generate a fluorescence response only after they have bound to the intended target fusion protein, thereby resulting in much higher target to background ratios than conventional fluorophores [32, 33]. Fluorogenic probes drastically reduce the need for extensive washing protocols and as such are particularly suitable for high-throughput imaging and real-time analysis of the dynamic processes of cellular proteins. SNAP-tag fluorogenic probes consist of benzylguanine substrates bearing an organic dye attached at the periphery of the benzylic ring and an appropriate dark quencher located on the C-8 position

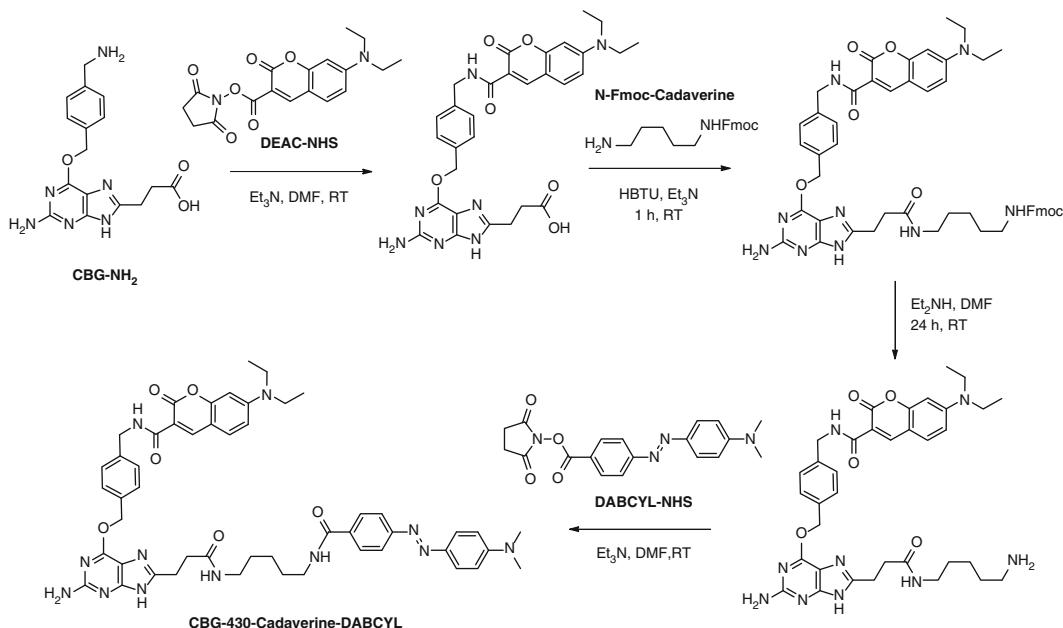


Fig. 5 Schematic of the synthesis of a SNAP-tag fluorogenic probe based on a diethylamino coumarin/dabcyl pair featuring a 1,5-pentanediamine (also known as Cadaverine) alkane linker. Fluorogenic probes are prepared in a sequential 2- or 4-step protocol (depending whether or not a spacer is installed between the guanine and quencher groups) starting from the CBG-NH₂ building block and the corresponding fluorophore/quencher NHS esters. Different alkane and PEG-based linkers can be used to optimize the balance between quenching efficiency and reaction kinetics for each probe

of the guanine ring (Fig. 5). In such probes, the fluorescence emission of the reporter dye is quenched by a fluorophore/quencher interaction via FRET and/or static quenching. Upon reaction with SNAP-tag, the quencher-linked guanine group is released free into solution leading to an increase in the relative fluorescence intensity of the dye, which remains bound to the protein tag. Several different combinations of fluorophore/quencher pairs were investigated in attempt to obtain optimal self-quenching systems across the whole visible spectrum, without significant deterioration of probe reactivity toward fusion protein. *In vitro* quenching efficiency assays have shown 76–99 % fluorescence recovery after incubation of a variety of fluorogenic probes with SNAP-tag [33].

1.6 Cross-linking Probes for Protein-Protein Interactions

Protein–protein interactions are at the core of virtually all molecular processes of any living cell. Understanding these interactions is one of the main challenges in functional proteomics and may provide the basis for new therapeutic approaches [52]. FRET strategies based on fluorescent proteins have been commonly used for studying protein–protein interactions in living cells [53]. However, fluorescent proteins suffer from a relatively low brightness and

broad absorption/emission spectra. Furthermore, the analysis of protein interactions at the surface of living cells is complicated due to fluorescence background caused by the accumulation of fluorescent proteins in intracellular compartments. To overcome these issues, researchers have developed time-resolved FRET approaches that combine the advantages of long lifetime lanthanoid cryptates with site-specific labeling technologies, and have employed these approaches to study several membrane proteins, including γ -aminobutyric (GABA) [54], muscarinic [55], ghrelin, and somatostatin receptors [56]. An alternative approach for the investigation of protein interactions in living cells and in cell extracts takes advantage of the ability to label fusion proteins with cross-linking probes. The interacting proteins are genetically fused to labeling tags, such as SNAP-tag or CLIP-tag, and specifically cross-linked using bifunctional small molecule substrates [23, 26, 27]. After cross-linking, the trapped protein complexes are isolated and analyzed by Western blot or in-gel fluorescence imaging. SNAP-tag (or CLIP-tag) cross-linking probes can be designed for the investigation of dimerization processes between the same or different protein monomers using homo- (BG-BG/BC-BC) or hetero-bifunctional (BG-BC) probes that are connected via a linker of appropriate length optionally appending a fluorophore or an affinity group of choice (Fig. 6). Tag-based approaches can be used to specifically detect weak or transient interactions that are difficult to distinguish by conventional cross-linking approaches, such as chemical or light-induced cross-linking methods, which typically suffer from poor selectivity. On the basis of a detailed kinetic analysis of the cross-linking reaction, Kai Johnsson and coworkers

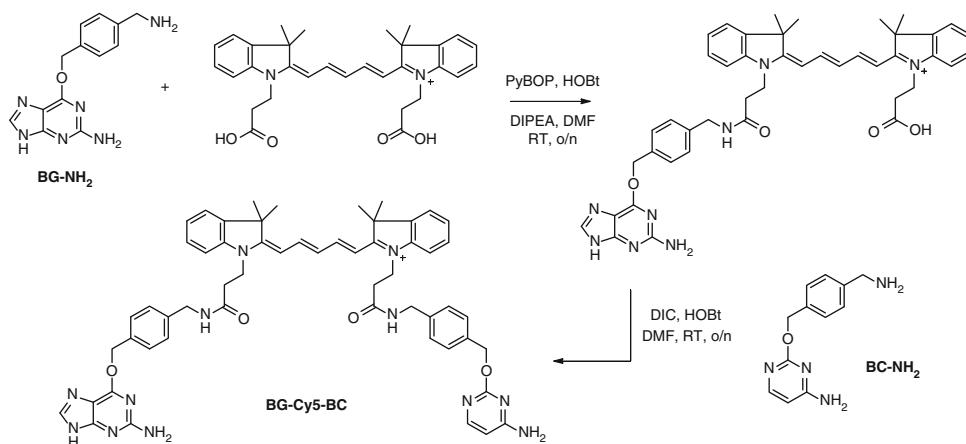


Fig. 6 Schematic of the synthesis of a SNAP-tag and CLIP-tag bifunctional cross-linker. BG-Cy5-BC was prepared in a 2-step protocol from BG-NH₂ and BC-NH₂ building blocks and a bis-carboxy-Cy5 derivative using standard peptide coupling conditions [27]. *PyBOP*: benzotriazol-1-yl-oxytritypyrrolidinophosphonium hexafluorophosphate, *HOEt*: N-Hydroxybenzotriazole, *DIPEA*: *N,N*-diisopropylethylamine, *DIC*: *N,N'*-Diisopropylcarbodiimide

showed that the cross-linking efficiency can be used as an indicator of interaction between two proteins, allowing thereby the unambiguous identification of interacting protein pairs [26].

1.7 Biosensor Probes of Ions and Cell Metabolites

In addition to the aforementioned applications, site-specific labeling techniques can be used as biochemical tools to sense dynamic cellular events in living cells. The ability to covalently incorporate biosensors into genetically encodable fusion proteins has enabled researchers to track the intracellular concentration of ions and cell metabolites with spatial and temporal resolution [57]. The use of a fusion tag allows precise targeting of molecular biosensors to a specific subcellular compartment within the native cellular environment. Two different fluorescence-based approaches have been designed for noninvasive sensing of ions and metabolites in living cells (Fig. 7). In the first approach, the presence of the analyte is detected by a change in the fluorescence emission and/or dynamic

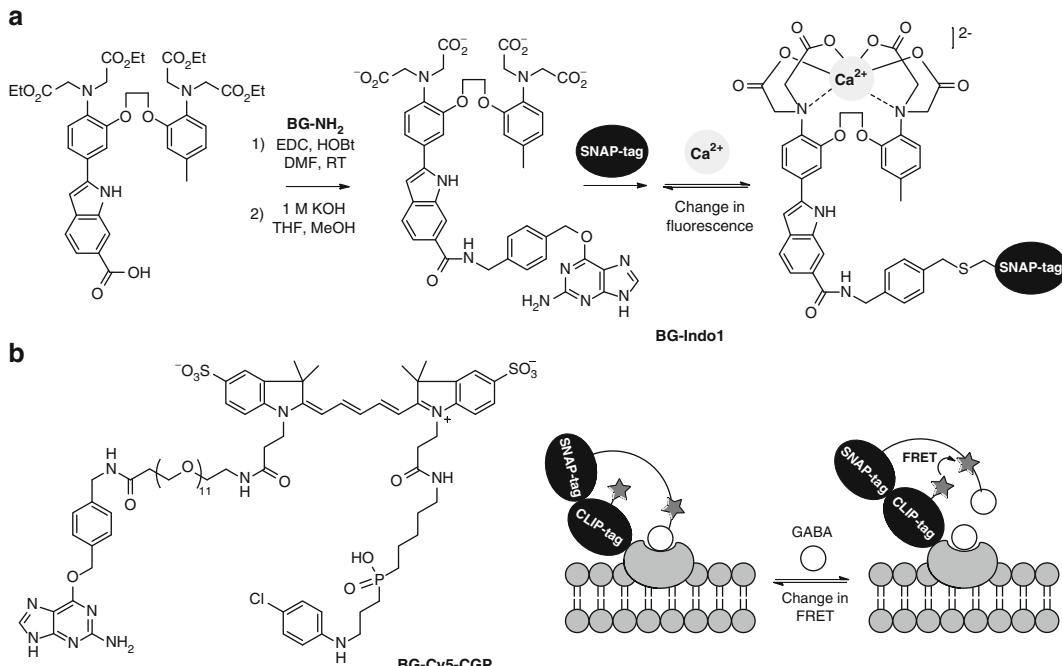


Fig. 7 Schematic of the use of SNAP-tag for the localization of biochemical sensors in living cells. (a) Synthesis of the Ca^{2+} indicator BG-Indo 1 from the coupling of BG-NH₂ with Indo-1 ethyl ester followed by alkaline hydrolysis. Both the fluorescence emission intensity and wavelength of the SNAP-Indo 1 conjugate change upon calcium binding [37]. EDC: 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide; HOBt: N-Hydroxybenzotriazole. (b) SNAP-tag substrate BG-Cy5-CGP displaying BG coupled via a short flexible linker to the Cy5 fluorophore and the GABA_B receptor antagonist CGP 51783 1 ((3-[4-chlorophenyl]methyl)amino-propyl)-(P-diethoxymethyl)-phosphinic acid). The biosensor is composed of a fusion protein between SNAP-tag, CLIP-tag, and a receptor protein for the analyte of interest (e.g., GABA_B receptor antagonist). In the absence of analyte, the intramolecular ligand keeps the fluorescent tether in a closed conformation. In the presence of sufficient concentrations of GABA_B, the system is shifted toward an open conformation which is detected through changes in FRET efficiency between the fluorescent tether and a fluorophore covalently attached to the CLIP-tag [61].

range of a synthetic fluorescent indicator covalently attached to the fusion protein. Recent examples include fluorescent sensors for intracellular concentrations of Zn²⁺ [58], Ca²⁺ [37, 38], O₂ [59], and H₂O₂ [60]. In the second approach (termed Snifit), the competitive displacement of a fluorescent intramolecular tether by the free analyte is detected by a change in the FRET of the sensor protein [39]. This approach has been used for visualizing the concentration of metabolites on the cell surface of mammalian cells, including the neurotransmitters glutamate [40] and γ-aminobutyric acid (GABA) [61].

2 Materials

2.1 Chemical Coupling

1. BG-NH₂, BG-PEG-NH₂, BG-GLA-NHS, or BG-Maleimide (New England Biolabs, Ipswich, MA).
2. Alexa Fluor® 350 Carboxylic Acid, Succinimidyl Ester (Life Technologies, Carlsbad, CA).
3. Qdot® 565 ITK™ Amino (PEG) Quantum Dots (Life Technologies, Carlsbad, CA).
4. Magnetic 3 μm micromer®-M PEG-COOH beads (Micromod, Rostock, Germany).
5. Triethylamine (Sigma-Aldrich, St. Louis, MO).
6. 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide, EDC (Sigma-Aldrich, St. Louis, MO).
7. Anhydrous N,N-dimethyl formamide, DMF (Sigma-Aldrich, St. Louis, MO).
8. Anhydrous dimethyl sulfoxide, DMSO (Sigma-Aldrich, St. Louis, MO).
9. 25 mM MES (2-(N-morpholino)ethanesulfonic acid) buffer, pH 5.0.
10. 100 mM Tris-HCl buffer, pH 7.4.
11. 50 mM borate buffer, pH 8.3.
12. 100 mM PBS (phosphate-buffered saline) buffer, pH 7.4.
13. Vivaspin with PES membrane, cutoff 100 kDa (GE Healthcare, Little Chalfont, UK).
14. NAP-5 column (GE Healthcare, Little Chalfont, UK).
15. 200 mM aqueous DTT (dithiothreitol) (Sigma-Aldrich, St. Louis, MO).

2.2 Purification of BG-Functionalized Small Molecule Probes

1. Acetonitrile, HPLC grade (EMD, Merck KGaA, Darmstadt, Germany).
2. 0.1 % trifluoroacetic acid (TFA) buffer.
3. 0.1 M triethylammonium bicarbonate (TEAB) buffer.

2.3 Labeling of SNAP-Tag Fusion Proteins

1. SNAP-tag vector and control plasmids (New England Biolabs, Ipswich, MA).
2. SNAP-tag fluorescent substrate of choice (New England Biolabs, Ipswich, MA).
3. Cultured cell line of choice (e.g., COS-7, CHO-K1, HEK293, PtK2, and NRK).
4. Culture chambers, such as 8-chamber Lab-Tek II Chambered Coverglass with #1.5 borosilicate (Nalge Nunc Int., Thermo Fisher Scientific, Waltham, MA).
5. Complete cell culture media (e.g., Gibco® Ham's F-12 K (Kaighn's) medium, Life Technologies, Carlsbad, CA).
6. 16.2 mM Hoechst 33342 (Life Technologies, Carlsbad, CA) solution.

3 Methods

3.1 Protocols for Chemical Coupling

3.1.1 Reaction Conditions for the Coupling of BG-NH₂ and Fluorophore NHS Esters

Coupling reactions typically use an excess of the BG-NH₂ (1–2 equivalents) relative to the desired NHS ester label (1 equivalent) at 5–20 mM final concentration in the presence of a 1.5-fold molar excess of triethylamine in anhydrous DMF (*see Note 1*). Mild reaction conditions in combination with high yields can also be achieved with other activation reagents for carboxylic acid labels, e.g., with dicyclohexylcarbodiimide (DCC), 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), carbonyldiimidazole (CDI), or with uronium salts such as O-(benzotriazole-1-yl)-1,1,3,3-tetramethyl uranium tetrafluoroborate (TBTU) or O-(7-azabenzotriazole-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate (HATU) [62].

1. Dissolve one 2 mg (7.4 µmol) vial of BG-NH₂ in 500 µL of anhydrous DMF (*see Note 2*).
2. Add 1.1 µL (7.9 µmol) of triethylamine.
3. Add 2.0 mg (4.9 µmol) of Alexa Fluor® 350 Carboxylic Acid, Succinimidyl Ester (*see Note 3*).
4. Mix the reaction at 30 °C overnight.
5. Remove the solvent under vacuum.
6. Dilute the resulting residue in 1 mL water/acetonitrile (9:1) and purify by reverse phase HPLC (*see Note 4*).

3.1.2 Reaction Conditions for the Coupling of BG-PEG-NH₂ and Solid Surfaces

Several strategies have been described for chemical attachment of benzylguanine (BG) derivatives to functionalized solid surfaces. Surface activation can be performed, for instance, using standard amino-coupling procedures [62]. Active ester-derivatized surfaces (e.g., NHS esters) can be directly coupled to BG amines (e.g., BG-NH₂)

or BG-PEG-NH₂) [16, 63–65]. Amino-derivatized surfaces can be conjugated to BG succinimidyl esters (e.g., BG-GLA-NHS) [66]. Streptavidin-modified surfaces can be used in combination with BG-biotin substrates (e.g., SNAP-Biotin, New England Biolabs, Ipswich, MA) [17, 67]. Other examples include the functionalization of gold surfaces with customized benzylguanine thiol reagents [18, 68], and of cyclodextrin and cucurbituril surfaces with bivalent BG adamantine or ferrocene derivatives [41, 42]. Magnetic and nonmagnetic agarose-based resins for capture and immobilization of SNAP-tag fusion proteins are commercially available (SNAP-Capture Pull-Down Resin and SNAP-Capture Magnetic Beads, New England Biolabs, Ipswich, MA). The protocol below is an example of the functionalization of micromer®-M PEG-COOH beads with BG-PEG-NH₂ [64].

1. Pre-activate a COOH solid surface of choice (for instance, 50 µL of a suspension of magnetic 3 µm micromer®-M PEG-COOH beads). The conditions for surface activation will depend on the particular solid surface.
 - (a) Wash the beads twice with MES buffer (25 mM, pH 5.0).
 - (b) Incubate the beads with a freshly prepared mixture of 25 µL EDC solution (50 mg/mL) and 25 µL NHS solution (50 mg/mL) in MES buffer (25 mM, pH 5.0) at room temperature for 30 min.
 - (c) Remove the supernatant and wash the beads twice with MES buffer (25 mM, pH 5.0).
2. Add 50 µL of a freshly prepared solution of BG-PEG-NH₂ (2 mg/mL in 10 % DMSO and 90 % PBS pH 7.4) to the activated beads (*see Note 5*).
3. Incubate at 4 °C for 16 h with slow tilt rotation (*see Note 6*).
4. Remove the supernatant and wash the beads four times with 100 µL of Tris-HCl buffer (50 mM, pH 7.4) for 5 min each (*see Notes 7 and 8*).
5. Store the beads at 4 °C before the immobilization reaction with the desired SNAP-tag fusion protein (*see Note 9*).

3.1.3 Reaction Conditions for the Coupling of BG-GLA-NHS and Quantum Dots

BG-GLA-NHS has been used to prepare SNAP-tag substrates from amine-containing precursors, including proteins [69, 70], oligonucleotides [71–73], and quantum dots [66]. Coupling solutions must be free of any other amine-containing substances such as TRIS, free amino acids, or ammonium ions. It is advisable, whenever possible, to use freshly prepared solutions of BG-GLA-NHS (typically 10–50 mM) in anhydrous DMF or DMSO (*see Note 10*). The following protocol is an example for the conjugation of BG-GLA-NHS to amino-functionalized CdSe/ZnS nanoparticles [66].

1. Dissolve one 2 mg (4.2 μ mol) vial of BG-GLA-NHS in 420 μ L of anhydrous DMF to make a 10 mM stock solution.
2. Add 0.5–1 μ M of Qdot® 565 ITK™ Amino (PEG) Quantum Dots in borate buffer (50 mM, pH 8.3). The molar ratio will depend on the loading capacity of your nanoparticle (typically one to tenfold excess BG-GLA-NHS is recommended).
3. Mix the reaction at 23 °C for 2 h.
4. Remove excess of unreacted BG-GLA-NHS by ultrafiltration (Vivaspin with PES membrane, cut-off 100 kDa) (*see Note 11*).
5. Incubate the BG-functionalized QDs with a SNAP-tag fusion protein (*see Notes 12 and 13*).

3.1.4 Reaction Conditions for the Coupling of BG-Maleimide and Oligonucleotides

Coupling reactions using BG-Maleimide are typically performed using a solvent mixture comprising DMF or DMSO and an aqueous buffer at pH 7.0–7.5 (e.g., Tris-HCl or PBS buffer). It may be advisable to carry out the reduction of any disulfide bonds by adding a 10-fold molar excess of a reducing agent such as DTT or TCEP prior to the incubation with BG-Maleimide (*see Note 14*). Several BG-oligonucleotides have been prepared by treating commercially available BG-Maleimide with thiol-modified oligonucleotides [74–76]. No detectable degradation of the resulting BG-oligonucleotide conjugates was observed after over 30 rounds of PCR [76].

1. Resuspend the 5'-thiol-modified oligonucleotide (20–40 nmol) in 0.5 mL of reduction buffer (100 mM Tris-HCl, pH 8.5, and 200 mM DTT).
2. Incubate for 1 h at room temperature.
3. Remove excess DTT by gel filtration (NAP-5 column) using PBS (pH 7.4) as elution buffer (*see Note 15*).
4. Dissolve one 2 mg (4.1 μ mol) vial of BG-Maleimide in 1.6 mL of anhydrous DMF to make a 2.5 mM stock solution (*see Note 16*).
5. Add 200 μ L of 2.5 mM BG-Maleimide stock solution to the 5'-thiol-modified oligonucleotide solution in PBS buffer (pH 7.4) (*see Note 17*).
6. Mix the reaction for 1–2 h at room temperature.
7. Remove excess of BG-Maleimide by gel filtration using deionized water as the solvent and concentrate under vacuum.
8. Purify the BG-oligonucleotide by HPLC (for instance, using a 0.1 M tetraethylammonium acetate pH 6.9/acetonitrile gradient). Samples can be analyzed by 1–3 % agarose gel electrophoresis (*see Note 18*).
9. Incubate the BG-oligonucleotide with a SNAP-tag fusion protein (*see Note 19*).

3.2 Purification of BG-Functionalized Small Molecule Probes

The purification strategy will depend on the physical characteristics (size, polarity, solubility, etc.) of the label used. Reverse-phase HPLC is method of choice for purification of small polar SNAP-tag substrates (e.g., 1–2 kDa, charged BG-fluorophore compounds). Typically, the purification is performed starting from a mixture of 5 % acetonitrile in water (equilibration time: 5–10 min) and then increasing to 95 % acetonitrile in water, using a linear gradient over 20–30 min, which depends on the polarity of your product. The flow rate conditions will depend on your HPLC column specifications: for a C18 type column, particle size 5 µm, dimensions 19 × 100 mm, the typical flow rate is 10 mL/min. Refer to the manufacturer's manual for detailed instructions on your HPLC column conditions. Detection is typically carried out at 280 nm (for benzylguanine derivatives) or 260 nm (for benzylcytosine derivatives). If you have a dual UV-Vis detector, the maximum wavelength absorption of your favorite fluorophore may also be used.

To achieve improved HPLC purification results, after the coupling reaction, the mixture containing a BG labeled compound should be first concentrated to dryness and then resuspended in a solution 9:1 water/acetonitrile (use 1 mL of this solution for a reaction performed with 2–5 mg of the building block). Make sure your product is completely dissolved in this solution. Heating at 37 °C, vortexing, and/or sonication will aid dissolution; adding 10–20 % DMSO usually helps, though an excess of DMSO will compromise the separation. The presence of additives, such as 0.1 % trifluoroacetic acid (TFA) or 0.1 M triethylammonium bicarbonate (TEAB) in the elution buffer, though not critical, in general improves the resolution of the peaks during the HPLC purification. Keep in mind, however, that some substrates are acid-sensitive and, in a number of cases, degradation of the product might be observed after purification during the sample concentration protocol. If you don't have information on the hydrolytic stability of your product, it is advisable to avoid TFA. Nevertheless, if you decide to use TFA in your purification strategy, neutralize your HPLC fractions (using aqueous NH₄OH 30 %, for instance) before sample concentration. Lyophilization is often preferred over conventional rotary evaporator concentration.

Silica gel chromatography purification can be efficiently used for BG or BC conjugates derived from nonpolar neutrally charged labels (e.g., carboxyfluorescein diacetate). In such cases, the purification can be carried out by silica gel chromatography using a mixture of dichloromethane-methanol and gradients varying from 5:1 to 20:1, depending on the polarity of your product.

3.3 Labeling of SNAP-Tag Fusion Proteins

The protocols for labeling SNAP-tag fusion proteins in living cells are based on the instructions provided by New England Biolabs, Inc. (www.neb.com). Additional information can be found in the video article available at the *Journal of Visualized Experiments*

(JoVE), which illustrates the fluorescent labeling of COS-7 expressing SNAP-tag fusion proteins [77], and in various reviews published in the literature [15].

3.3.1 Cloning and Expression of SNAP-Tag Fusion Proteins in Mammalian Cells

1. Clone the gene of interest upstream or downstream of the SNAP-tag coding sequence (*see Note 20*). Mammalian and bacterial expression plasmids encoding SNAP-tag flanked by restriction sites for cloning a gene of interest as well as well-characterized positive control plasmids are available from New England Biolabs, Inc.
2. Seed cells of choice in culture chambers and incubate in appropriate medium (e.g., complete DMEM or F-12 medium) overnight at 37 °C, 5 % CO₂. Cell density should be approximately 50–60 % confluent.
3. Transfect cells in culture with the SNAP-tag plasmid using standard transfection methods. For stable expression, begin selecting mammalian cultures (e.g., using 600–1,200 µg/mL Geneticin) 24–48 h after transfection depending on the cell line.
4. Once cell colonies are visible, incubate cells 16–24 h at 37 °C, 5 % CO₂ to allow expression of the fusion protein. The appropriate time for adequate protein expression should be determined empirically.

3.3.2 Labeling of SNAP-Tag Fusion Proteins in Mammalian Cells

1. Dissolve 50 nmol of the SNAP-tag substrate in 50 µL of DMSO to make a 1 mM stock solution (*see Note 21*). Vortex 1–5 min to ensure full dissolution of the substrate. Immediately before use, prepare labeling medium by diluting the SNAP-tag substrate stock solution 1:200 in complete medium to a final concentration of 5 µM. Mix thoroughly by pipetting up and down.
2. Carefully remove growth medium from transfected cells. Wash cells twice with complete medium and add the labeling solution (*see Note 22*). Incubate cells at 37 °C, 5 % CO₂ for 30 min.
3. While cells are incubating, prepare a nuclear counterstaining solution (e.g., a 16.2 mM solution Hoechst 33342 in 10 mL of complete F-12K) (*see Note 23*).
4. Remove the labeling medium and add 1 µL of the nuclear counterstaining solution. Incubate cells at 37 °C, 5 % CO₂ for 3 min.
5. Wash cells three times with complete medium.
6. For cell surface labels, proceed to imaging. For intracellular labels, incubate with fresh complete medium at 37 °C, 5 % CO₂ for 30 min to allow any unreacted fluorophore to diffuse out of the cells. Replace the medium and proceed to imaging (*see Note 24*).
7. Image cells using a fluorescent microscope with an appropriate filter set (*see Notes 25 and 26*).

4 Notes

1. DMF is the solvent of choice for the coupling reaction because DMF provides the most effective capacity to dissolve both BG-NH₂ or BG-PEG-NH₂ and the fluorophore labels. The use of absolute anhydrous DMF (usually supplied over molecular sieves, H₂O ≤0.01 %) is recommended as it slows down the degradation of NHS esters. At high concentrations, the solubility of the reagents might become an issue. Heating at 37 °C, vortexing, and/or sonication will aid dissolution. Anhydrous DMSO can be used in the same capacity as DMF; however, DMF is preferred for chemical synthesis applications because DMF has a lower boiling point than DMSO and, therefore, can be more easily removed under vacuum when necessary.
2. BG-NH₂ is sparingly soluble in most solvents. Heating at 37 °C, vortexing, and/or sonication may help dissolution; however, some remaining precipitate is expected and should be carried forward through the reaction. Remaining solids will continue to dissolve as the coupling reaction proceeds. Heating up to 60 °C will aid the dissolution of BG-NH₂; however, higher temperatures may be detrimental to the substrate. As an alternative, BG-PEG-NH₂ can be used for coupling reactions. The PEG-linker confers greater solubility to benzylguanine substrates.
3. A molar equivalent or excess of your desired NHS label may also be employed when needed. However, it is recommended to use an excess of BG-NH₂ in respect to the NHS label to facilitate the purification scheme. Generally speaking, it is easier to separate the excess of BG-NH₂ from the resulting BG-fluorophore conjugate than to separate an excess of a given fluorophore NHS ester (or its corresponding hydrolyzed form) from the resulting BG-fluorophore conjugate. In addition, avoiding the use of excess of the NHS labels helps minimize the costs of synthesis.
4. The purification strategy will depend on the label used. Good results have been obtained with both HPLC and silica gel chromatography. *See Sect. 3.2* for more details on the purification of BG-functionalized probes.
5. For best dissolution of the BG-PEG-NH₂ in aqueous buffers, it is recommended that you first dissolve BG-PEG-NH₂ in anhydrous DMF or DMSO (typically 10–50 mM stock solution) and then dilute this solution with the appropriate aqueous buffer. It is advisable to prepare, whenever possible, the solution immediately before starting the coupling reaction.
6. In most cases, the coupling reaction is completed within 1–6 h incubation at room temperature. In some instances, in particular

when the presence of remaining precipitate is detectable, longer incubation times may result in greater coupling efficiency.

7. Additional treatment with ethanolamine may be necessary to quench any residual activated carboxylic groups (typically 0.5–1 M solution of ethanolamine in 50 % water/isopropanol (v/v) for 1 h at room temperature). Wash the resin at least three times with deionized water for 5 min and three times with 50 % water/isopropanol (v/v) solution for 5 min.
8. The benzylguanine surface density values can be determined by UV-Vis absorption (extinction coefficient 12,900 cm⁻¹ M⁻¹ in ethanol at 285 nm) [78]. The molar absorbance of BG-NH₂ in DMSO has been reported as 7,100 cm⁻¹ M⁻¹ at 280 nm [79].
9. Once the surface is functionalized with the benzylguanine substrate, immobilization of SNAP-tag fusion proteins onto solid surfaces can be carried out by incubating a protein solution containing up to 1 mg/mL SNAP-tag fusion protein in an appropriate buffer (e.g., EDTA-free 10 mM HEPES or PBS, pH 7.5, 150 mM NaCl, 0.5 % Tween 20) containing at least 1 mM DTT for 1 h at room temperature (or overnight at 4 °C). Wash the solid surface at least three times with 1 mL of immobilization buffer for 10 min. In most cases, the labeling reaction is completed within 1 h incubation at room temperature. Longer incubation times do not necessarily result in greater labeling efficiency.
10. If the coupling reaction is to be performed in water, buffering the solution at pH 8.3 (0.1–0.2 M sodium bicarbonate or 50 mM borate buffer) has been found to be optimal for NHS ester coupling reactions. However, in some instances neutral conditions (e.g., PBS buffer pH 7) are preferred as they minimize hydrolysis of NHS esters. NHS esters readily hydrolyze in moist environments, particularly at high pHs. When using aqueous buffer conditions, it is recommended to use a two to fivefold molar excess of NHS ester due to the competition with the hydrolysis reaction. Experiments should be designed to maximize the usage of BG-GLA-NHS. Storage of unused BG-GLA-NHS solubilized in DMF or DMSO should be avoided. If unavoidable, it is advisable storing the BG-GLA-NHS solution at –20 °C for the minimum time period.
11. Optional treatment with a reagent (e.g., sulfo-NHS-acetate, Thermo Fisher Scientific, Rockford, IL) for blocking any residual primary amines may be advisable (for instance, 1 mg of sulfo-NHS-acetate dissolved in 15 µL of water is added to the BG-QD reaction mixture after 2 h and is allowed to react for another hour). Wash at least three times with the washing buffer as described.

12. Once the quantum dot is functionalized with BG-GLA-NHS, incubate the BG-QD with a SNAP-tag fusion protein solution in the desired ratio (for instance, 1:5–1:20) in an appropriate buffer (e.g., 100 mM HEPES-imidazole buffer, pH 7.4) for 1–6 h at room temperature (or overnight at 4 °C). Purification can be performed by ultrafiltration or size exclusion columns.
13. BG-GLA-NHS can be used to functionalize cell surface proteins for further conjugation with SNAP-tag by reacting BG-GLA-NHS with primary amine groups in the target proteins [69, 70]. For the functionalization of purified proteins, slowly add one- to threefold molar excess of BG-GLA-NHS to the protein solution (typically 0.05–0.2 mM) while mixing. Protein samples that have been previously dissolved in buffers containing amines should be dialyzed against aqueous NaHCO₃ or PBS. In most cases, the labeling reaction is completed within 1 h incubation at room temperature. Longer incubation times do not necessarily result in greater labeling efficiency. To increase the degree of labeling a higher ratio of NHS ester to protein can be used. The purification of protein conjugates can be performed by gel filtration or dialysis. Free BG-GLA-NHS must be removed from the labeled protein solution as it can further react with SNAP-tag. The BG-GLA-NHS conjugation degree can be determined by mass spectrometry (e.g., MALDI TOF) prior to the SNAP-tag labeling.
14. Dithiothreitol (DTT) is typically used as a reducing agent (in tenfold molar excess). Another possibility is *tris*(2-carboxyethyl) phosphine (TCEP). The excess of the reducing agent has to be removed by extraction from the solution using ethyl acetate or by dialysis or gel filtration prior to addition of the maleimide. In some instances, it may also be advisable to carry out the thiol modification in an inert atmosphere to prevent oxidation of the thiols and deoxygenate all buffers and solvents used for the thiol conjugation.
15. Even if the reducing step with DTT or TCEP is omitted, it is recommended that you desalt your oligonucleotides (e.g., using NAP-5 and NAP-10 columns) before coupling.
16. BG-Maleimide is a water-insoluble reagent and must be dissolved in an organic solvent (DMSO or DMF) before dilution to an aqueous buffer.
17. Typical coupling conditions employ a 10- to 20-fold molar excess of a freshly prepared solution of BG-Maleimide in anhydrous DMSO or DMF to the oligonucleotide solution and incubation for 1–2 h at room temperature. Longer incubation times may or may not result in greater labeling efficiency. The reaction mixture may be a suspension rather than a clear solution. However, this does not normally affect the coupling reaction.

18. The coupling efficiency may vary depending on the oligonucleotide sequence and the possibility of secondary structure formation, such as hairpins or homo-/heterodimers. The presence of a spacer (i.e., c10 modifier) between the oligonucleotide and the thiol-reacting group may be beneficial for achieving higher yields. The purity of the oligonucleotide is critical for the success of the labeling reaction. It is strongly recommended to purify the sample by precipitation prior to the coupling reaction. This coupling reaction can be carried out in organic solvents (i.e., anhydrous DMF) and using fully protected oligonucleotides. In such cases, the coupling efficiency variation due to secondary structures is greatly reduced.
19. BG-oligonucleotide can be incubated with a SNAP-tag fusion protein (e.g., in a 25:1 ratio) for 1 h at 25 °C in a typical conjugation buffer (50 mM Tris–HCl, 1 mM DTT, 100 mM NaCl, pH 7.5). Purification can be performed by size exclusion columns.
20. For cell culture, transfection, and cloning methods please refer to established protocols. The SNAP-tag vector has cleavage sites for convenient C- or N-terminal subcloning (restriction sites located upstream of SNAP-tag: NheI, EcoRV, AscI, SwaI, BsrGI, AgeI, and EcoRI; restriction sites located downstream of SNAP-tag: SbfI, BamHI, PmeI, XhoI, PacI, and NotI). The amino acid linker between the protein of interest and SNAP-tag should be limited to about ten residues to reduce nonspecific protease cleavage of long, unstructured peptides.
21. Optimal substrate concentrations and reaction times are 1–10 μM and 15–60 min, respectively, depending on the experimental conditions and expression levels of the SNAP-tag fusion protein. Although labeling concentrations ≤5 μM are recommended, no significant toxicity has been observed in proliferation or viability assays over 2 h when 20 μM substrate concentration is used.
22. Cell-impermeable substrates can be used for labeling SNAP-tag fusion proteins specifically on the surface of living cells. Unlike approaches based on autofluorescent proteins, only mature proteins localized on the cell surface—and not proteins in the secretory or recycling pathways—will be labeled. Intracellular labeling of SNAP-tag fusion proteins can be achieved using cell-permeable SNAP-tag substrates. The labeling of intracellular protein targets with cell-impermeable substrates has been achieved using microinjection [44] and other cell-loading techniques [35, 36] or after permeabilization of the cell's plasma membrane using fixation procedures.
23. The labeling of non-transfected or mock-transfected cells is usually sufficient as a negative control. A further negative control

can be generated by blocking the SNAP-tag activity in cells expressing SNAP-tag fusion protein using the nonfluorescent SNAP-tag substrate SNAP-Cell Block (bromoethylpteridine, BTP). SNAP-Cell Block can also be used in pulse-chase experiments to block the SNAP-tag reactivity between two pulse-labeling steps [80], or in quench-pulse-chase strategies to saturate all previously synthesized SNAP-tag proteins and permit the fluorescent labeling of only newly synthesized protein pools [81]. In addition to the cell-permeable SNAP-Cell Block, a cell-impermeable variant, SNAP-Surface Block, is available for selectively blocking cell surface SNAP-tag fusion proteins.

24. Labeling of SNAP-tag fusion proteins can be performed before or after fixation without significant loss of signal. Cells expressing SNAP-tag fusions have been fixed prior to labeling using standard fixation methods (e.g., formaldehyde or methanol); however, the conditions such as the fixative used and washing steps may need to be optimized to attain good results. After fixation, SNAP-tag fusion proteins can also be processed for immunofluorescence using an anti-SNAP antibody.
25. Labeled SNAP-tag has been detected in mammalian cells over 48 h, indicating stability of the signal over this period. The turnover rates of the SNAP-tag fusion protein will vary depending on the fusion partner of interest (half-life values may range from a few minutes to several hours). When either the protein turnover is fast or the fusion protein has limited stability, analyzing the cells under the microscope immediately after labeling is recommended. In some cases, the signal may be stabilized by fixing the cells prior to visualization. To visualize proteins with fast turnover rates or low thermal stability, SNAP-tag fusion proteins can be labeled for longer times at lower temperatures (for instance at 4 or 16 °C).
26. If the labeled fusion protein cannot be detected, it is most likely due to low or lack of expression. It is advisable to verify the transfection method to confirm that the cells contain the correct fusion gene of interest. If this is confirmed, check for expression of the SNAP-tag fusion protein using an antibody for the protein of interest or an anti-SNAP-tag antibody. Alternatively, a fluorescent substrate, such as SNAP-Vista Green, can be used to confirm the presence of SNAP-tag fusion in cell extracts following SDS-PAGE. If the labeling signal is weak, it may be due to insufficient exposure of the fusion protein to the fluorescent substrate. Try increasing the concentration of SNAP-tag substrate and/or the incubation time. Another possibility is that the protein may be poorly expressed or rapidly turned over. Fluorophore photobleaching may be reduced by using commercially available anti-fade reagents.

Acknowledgements

Brenda Baker and John Buswell for valuable suggestions and critical reading of the manuscript. Don Comb and Jim Ellard for their continued support of basic research at New England Biolabs.

References

1. Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21:766–776
2. Johnsson N, Johnsson K (2007) Chemical tools for biomolecular imaging. *ACS Chem Biol* 2:31–38
3. Hanek A, Corrêa IR Jr (2011) Chemical modification of proteins in living cells. In: Tschesche H (ed) *Methods in protein biochemistry*. DE GRUYTER, Berlin/Boston, pp 197–218
4. Keppler A, Gendreizig S, Gronemeyer T et al (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* 21:86–89
5. Keppler A, Pick H, Arrivoli C et al (2004) Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci U S A* 101:9955–9959
6. Gautier A, Juillerat A, Heinis C et al (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* 15: 128–136
7. Los GV, Encell LP, McDougall MG et al (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3:373–382
8. Miller LW, Cai Y, Sheetz MP et al (2005) In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat Methods* 2:255–257
9. Mizukami S (2011) Development of molecular imaging tools to investigate protein functions by chemical probe design. *Chem Pharm Bull* 59:1435–1446
10. Vivero-Pol L, George N, Krumm H et al (2005) Multicolor imaging of cell surface proteins. *J Am Chem Soc* 127:12770–12771
11. Popp MW, Ploegh HL (2011) Making and breaking peptide bonds: protein engineering using sortase. *Angew Chem Int Ed Engl* 50:5024–5032
12. Uttamapinant C, White KA, Baruah H et al (2010) A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* 107:10914–10919
13. Correa IR, Baker B, Zhang A et al (2013) Substrates for improved live-cell fluorescence labeling of SNAP-tag. *Curr Pharm Des* 19: 5414–5420
14. Mollwitz B, Brunk E, Schmitt S et al (2012) Directed evolution of the suicide protein O(6)-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* 51:986–994
15. Bodor DL, Rodriguez MG, Moreno N et al (2012) Analysis of protein turnover by quantitative SNAP-based pulse-chase imaging. *Curr Protoc Cell Biol* editorial board, Juan S Bonifacino [et al] Chapter 8, Unit8 8
16. Kindermann M, George N, Johnsson N et al (2003) Covalent and selective immobilization of fusion proteins. *J Am Chem Soc* 125: 7810–7811
17. Iversen L, Cherouati N, Berthing T et al (2008) Templated protein assembly on micro-contact-printed surface patterns. *Langmuir* 24:6375–6381
18. Engin S, Trouillet V, Franz CM et al (2010) Benzylguanine thiol self-assembled monolayers for the immobilization of SNAP-tag proteins on microcontact-printed surface structures. *Langmuir* 26:6097–6101
19. Hoskins AA, Friedman LJ, Gallagher SS et al (2011) Ordered and dynamic assembly of single spliceosomes. *Science* 331:1289–1295
20. Breitsprecher D, Jaiswal R, Bombardier JP et al (2012) Rocket launcher mechanism of collaborative actin assembly defined by single-molecule imaging. *Science* 336:1164–1168
21. Pellett PA, Sun X, Gould TJ et al (2011) Two-color STED microscopy in living cells. *Biomed Opt Express* 2:2364–2371
22. Jones SA, Shim SH, He J et al (2011) Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods* 8:499–508
23. Foraker AB, Camus SM, Evans TM et al (2012) Clathrin promotes centrosome integrity in early mitosis through stabilization of centrosomal ch-TOG. *J Cell Biol* 198:591–605
24. Chidley C, Haruki H, Pedersen MG et al (2011) A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat Chem Biol* 7:375–383

25. Bojkowska K, Santoni de Sio F, Barde I et al (2011) Measuring *in vivo* protein half-life. *Chem Biol* 18:805–815
26. Gautier A, Nakata E, Lukinavicius G et al (2009) Selective cross-linking of interacting proteins using self-labeling tags. *J Am Chem Soc* 131:17954–17962
27. Lukinavicius G, Lavogina D, Orpinell M et al (2013) Selective chemical crosslinking reveals a cep57-cep63-cep152 centrosomal complex. *Curr Biol* 23:265–270
28. Brun MA, Tan KT, Nakata E et al (2009) Semisynthetic fluorescent sensor proteins based on self-labeling protein tags. *J Am Chem Soc* 131:5873–5884
29. Doumazane E, Scholler P, Zwier JM et al (2011) A new approach to analyze cell surface protein complexes reveals specific heterodimeric metabotropic glutamate receptors. *FASEB J* 25:66–77
30. Lavis LD, Raines RT (2008) Bright ideas for chemical biology. *ACS Chem Biol* 3:142–155
31. Lukinavicius G, Johnsson K (2011) Switchable fluorophores for protein labeling in living cells. *Curr Opin Chem Biol* 15:768–774
32. Komatsu T, Johnsson K, Okuno H et al (2011) Real-time measurements of protein dynamics using fluorescence activation-coupled protein labeling method. *J Am Chem Soc* 133: 6745–6751
33. Sun X, Zhang A, Baker B et al (2011) Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. *Chembiochem* 12:2217–2226
34. Banala S, Maurel D, Manley S et al (2012) A caged, localizable rhodamine derivative for superresolution microscopy. *ACS Chem Biol* 7:289–293
35. Campos C, Kamiya M, Banala S et al (2011) Labelling cell structures and tracking cell lineage in zebrafish using SNAP-tag. *Dev Dyn* 240:820–827
36. Maurel D, Banala S, Laroche T et al (2010) Photoactivatable and photoconvertible fluorescent probes for protein labeling. *ACS Chem Biol* 5:507–516
37. Bannwarth M, Corrêa IR Jr, Sztretye M et al (2009) Indo-1 derivatives for local calcium sensing. *ACS Chem Biol* 4:179–190
38. Kamiya M, Johnsson K (2010) Localizable and highly sensitive calcium indicator based on a BODIPY fluorophore. *Anal Chem* 82:6472–6479
39. Brun MA, Griss R, Reymond L et al (2011) Semisynthesis of fluorescent metabolite sensors on cell surfaces. *J Am Chem Soc* 133: 16235–16242
40. Brun MA, Tan KT, Griss R et al (2012) A semisynthetic fluorescent sensor protein for glutamate. *J Am Chem Soc* 134:7676–7678
41. Uhlenheuer DA, Wasserberg D, Haase C et al (2012) Directed supramolecular surface assembly of SNAP-tag fusion proteins. *Chemistry* 18:6788–6794
42. Wasserberg D, Uhlenheuer DA, Neirynck P et al (2013) Immobilization of Ferrocene-modified SNAP-fusion proteins. *Int J Mol Sci* 14:4066–4080
43. Sletten EM, Bertozzi CR (2009) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* 48:6974–6998
44. Kepler A, Arrivoli C, Sironi L et al (2006) Fluorophores for live cell imaging of AGT fusion proteins across the visible spectrum. *Biotechniques* 41:167–170, 172, 174–165
45. Lukinavicius G, Umezawa K, Olivier N et al (2013) A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat Chem* 5:132–139
46. Fernandez-Suarez M, Ting AY (2008) Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Biol* 9:929–943
47. Hein B, Willig KI, Wurm CA et al (2010) Stimulated emission depletion nanoscopy of living cells using SNAP-tag fusion proteins. *Biophys J* 98:158–163
48. Klein T, Loscherger A, Proppert S et al (2011) Live-cell dSTORM with SNAP-tag fusion proteins. *Nat Methods* 8:7–9
49. Dellagiacoma C, Lukinavicius G, Bocchio N et al (2010) Targeted photoswitchable probe for nanoscopy of biological structures. *Chembiochem* 11:1361–1363
50. Cole NB, Donaldson JG (2012) Releasable SNAP-tag probes for studying endocytosis and recycling. *ACS Chem Biol* 7:464–469
51. Dutta D, Williamson CD, Cole NB et al (2012) Pitstop 2 is a potent inhibitor of clathrin-independent endocytosis. *PLoS One* 7:e45799
52. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450:1001–1009
53. Frommer WB, Davidson MW, Campbell RE (2009) Genetically encoded biosensors based on engineered fluorescent proteins. *Chem Soc Rev* 38:2833–2841
54. Maurel D, Comps-Agrar L, Brock C et al (2008) Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat Methods* 5:561–567

55. Alvarez-Curto E, Ward RJ, Pediani JD et al (2010) Ligand regulation of the quaternary organization of cell surface M3 muscarinic acetylcholine receptors analyzed by fluorescence resonance energy transfer (FRET) imaging and homogeneous time-resolved FRET. *J Biol Chem* 285:23318–23330
56. Park S, Jiang H, Zhang H et al (2012) Modification of ghrelin receptor signaling by somatostatin receptor-5 regulates insulin release. *Proc Natl Acad Sci U S A* 109:19003–19008
57. Newman RH, Fosbrink MD, Zhang J (2011) Genetically encodable fluorescent biosensors for tracking signaling dynamics in living cells. *Chem Rev* 111:3614–3666
58. Tomat E, Nolan EM, Jaworski J et al (2008) Organelle-specific zinc detection using zinpyr-labeled fusion proteins in live cells. *J Am Chem Soc* 130:15776–15777
59. Geissbuehler M, Spielmann T, Formey A et al (2010) Triplet imaging of oxygen consumption during the contraction of a single smooth muscle cell (A7r5). *Biophys J* 98:339–349
60. Srikan D, Albers AE, Nam CI et al (2010) Organelle-targetable fluorescent probes for imaging hydrogen peroxide in living cells via SNAP-Tag protein labeling. *J Am Chem Soc* 132:4455–4465
61. Masharina A, Reymond L, Maurel D et al (2012) A fluorescent sensor for GABA and synthetic GABA(B) receptor ligands. *J Am Chem Soc* 134:19026–19034
62. Montalbetti CAGN, Falque V (2005) Amide bond formation and peptide coupling. *Tetrahedron* 61:10827–10852
63. Colombo M, Mazzucchelli S, Montenegro JM et al (2012) Protein oriented ligation on nanoparticles exploiting O6-alkylguanine-DNA transferase (SNAP) genetically encoded fusion. *Small* 8:1492–1497
64. Recker T, Haamann D, Schmitt A et al (2011) Directed covalent immobilization of fluorescently labeled cytokines. *Bioconjug Chem* 22:1210–1220
65. Sielaff I, Arnold A, Godin G et al (2006) Protein function microarrays based on self-immobilizing and self-labeling fusion proteins. *Chembiochem* 7:194–202
66. Petershans A, Wedlich D, Fruk L (2011) Bioconjugation of CdSe/ZnS nanoparticles with SNAP tagged proteins. *Chem Commun* 47:10671–10673
67. Stein V, Hollfelder F (2009) An efficient method to assemble linear DNA templates for in vitro screening and selection systems. *Nucleic Acids Res* 37:e122
68. Kwok CW, Strahle U, Zhao Y et al (2011) Selective immobilization of Sonic hedgehog on benzylguanine terminated patterned self-assembled monolayers. *Biomaterials* 32: 6719–6728
69. Howland SW, Tsuji T, Gnajic S et al (2008) Inducing efficient cross-priming using antigen-coated yeast particles. *J Immunother* 31: 607–619
70. Shi G, Azoulay M, Dingli F et al (2012) SNAP-tag based proteomics approach for the study of the retrograde route. *Traffic* 13:914–925
71. Furuta K, Furuta A, Toyoshima YY et al (2013) Measuring collective transport by defined numbers of processive and nonprocessive kinesin motors. *Proc Natl Acad Sci U S A* 110:501–506
72. Sacca B, Meyer R, Erkelenz M et al (2010) Orthogonal protein decoration of DNA origami. *Angew Chem Int Ed Engl* 49: 9378–9383
73. Yang Y, Zhang CY (2012) Sensitive detection of intracellular sumoylation via SNAP tag-mediated translation and RNA polymerase-based amplification. *Anal Chem* 84: 1229–1234
74. Dohrmann PR, Manhart CM, Downey CD et al (2011) The rate of polymerase release upon filling the gap between Okazaki fragments is inadequate to support cycling during lagging strand synthesis. *J Mol Biol* 414:15–27
75. Kaltenbach M, Hollfelder F (2012) SNAP display: in vitro protein evolution in microdroplets. *Methods Mol Biol* 805:101–111
76. Kaltenbach M, Stein V, Hollfelder F (2011) SNAP dendrimers: multivalent protein display on dendrimer-like DNA for directed evolution. *Chembiochem* 12:2208–2216
77. Provost CR, Sun L (2010) Fluorescent labeling of COS-7 expressing SNAP-tag fusion proteins for live cell imaging. *J Vis Exp* pii:1876
78. Tugulu S, Arnold A, Sielaff I et al (2005) Protein-functionalized polymer brushes. *Biomacromolecules* 6:1602–1607
79. Juillerat A, Gronemeyer T, Keppler A et al (2003) Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem Biol* 10:313–317
80. Jansen LE, Black BE, Foltz DR et al (2007) Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol* 176:795–805
81. Farr GA, Hull M, Mellman I et al (2009) Membrane proteins follow multiple pathways to the basolateral cell surface in polarized epithelial cells. *J Cell Biol* 186:269–282

Chapter 5

2-Cyanobenzothiazole (CBT) Condensation for Site-Specific Labeling of Proteins at the Terminal Cysteine Residues

Lina Cui and Jianghong Rao

Abstract

Site specificity is pivotal in obtaining homogeneously labeled proteins without batch-to-batch variations. More importantly, precisely controlled modification at specific sites avoids potential pitfalls that could otherwise interfere with protein folding, structure, and function. Inspired by the chemical synthesis of d-luciferin, we have developed an efficient strategy (second-order rate constant $k_2 = 9.2 \text{ M}^{-1} \text{ s}^{-1}$) for labeling of proteins containing 1,2-aminothiol via reaction with 2-cyanobenzothiazole (CBT). In addition, the CBT condensation enjoys the convenience of protein engineering, as production of N-terminal cysteine-containing proteins has been well developed for native chemical ligation. This protocol describes the preparation of *Renilla* luciferase (rLuc) with 1,2-aminothiol at either its N- or C-terminus, and site-specific labeling of rLuc with fluorescein or ^{18}F via CBT condensation.

Key words Site-specific protein labeling, Terminal cysteine, 1,2-Aminothiol, 2-Cyanobenzothiazole (CBT), Fluorescence labeling, Radiolabeling (^{18}F)

1 Introduction

Approaches to labeling proteins with high efficiency and site specificity are highly demanded in a myriad of studies ranging from probing protein's properties and functions to biomedical imaging [1, 2]. Over the years, researchers have developed a number of strategies to engineer proteins with user-defined modification sites: a molecular tag is usually introduced to the protein via genetic encoding or chemo-enzymatic methods; the protein is then modified at the molecular tag selectively [3]. The first step is dependent on the feasibility to incorporate the reactive molecular tag, and the second step relies on the efficiency of the bioconjugation reaction between the molecular tag and the detecting molecule. Most of the current conjugation chemistries, while compatible with the first incorporation step, suffer from low reaction rates ($k_2 = 10^{-2}$ – $10^{-4} \text{ M}^{-1} \text{ s}^{-1}$) in the subsequent labeling step. This dampens the effectiveness of site-specific labeling at low concentrations such as

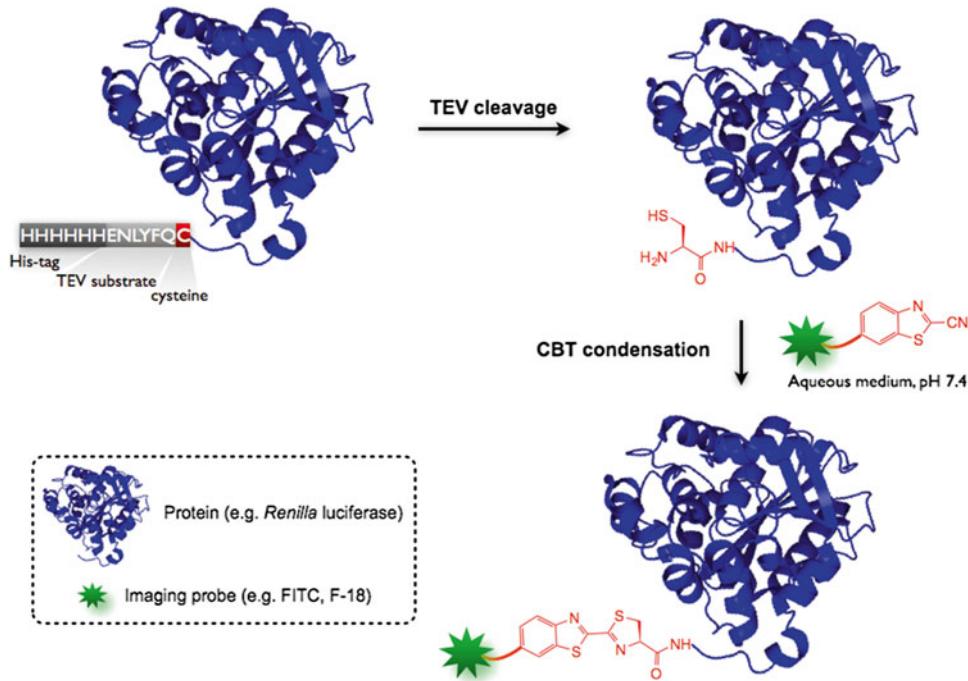


Fig. 1 Overview of the labeling method using CBT condensation

that under *in vivo* conditions or for low abundant proteins. The slow kinetics of the conjugation reactions also hampers their use when labeling proteins with imaging probes of short half-lives.

We have found the condensation reaction between D-cysteine and 2-cyanobenzothiazole (CBT), first used in the synthesis of D-luciferin [4], could proceed rapidly under physiological pH conditions in water [5]. This led us to explore the plausibility of CBT condensation as a new bioconjugation method (Fig. 1). While reacting with a cysteine-containing molecule, the second-order rate constant was determined to be $k_2 = 9.2 \text{ M}^{-1} \text{ s}^{-1}$, which is three orders of magnitude faster than the optimized Staudinger Ligation ($7.7 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$) [6]. In more comparison, the aniline-accelerated oxime ligation has a k_2 of $6.1 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$ [7, 8]; the k_2 of “click” chemistry between cyclooctyne-derivatives and azido group ranges from 10^{-3} to $1 \text{ M}^{-1} \text{ s}^{-1}$ [9]. Interestingly, CBT reacts with 1,2-, or 1,3-aminothiols to form stable condensation products; while reacting with thiol groups, only unstable or reversible products are formed. The high selectivity and fast reaction kinetics make CBT-cysteine condensation a promising bioconjugation method to label proteins at N-terminal cysteine residues. Meanwhile, protein labeling using CBT chemistry can take advantage of the existing strategies for production of proteins with N-terminal cysteine residues, which have been explored extensively in the past for protein engineering, such as that using native chemical ligation

(NCL) [10]. For example, proteins with N-terminal cysteine can be produced by cleavage of a precursor recombinant protein with a site-specific exogenous or endogenous protease to expose the cysteine residue as the new N-terminus [11–13]. Besides N-terminal cysteine residue, proteins can also be labeled at the C-terminus via introduction of 1,2-aminothiol [5]. More recently, an alternative approach, genetic encoding of 1,2-aminothiol into proteins using pyrrolylsyl-tRNA synthetase/tRNA_{CUA} pairs followed by a deprotection step, extends the utility of the CBT labeling beyond the terminal cysteine residues, making it possible to label proteins at other desirable sites [14, 15].

We have applied the CBT condensation chemistry to label synthetic cysteine-containing peptides with amino-CBT, and over 90 % of the peptide with N-terminal cysteine was labeled within 30 min in phosphate buffer at pH 7.4 at room temperature; however, no observable labeling product was detected in the reaction with peptides bearing only internal cysteine residues, suggesting the high selectivity of CBT labeling at N-terminal cysteine (1,2-aminothiol) [5]. To label proteins, we produced cysteine-containing bioluminescent protein *Renilla* luciferase (rLuc) at either the N- or the C-terminus. The N-terminal cysteine was generated by fusing a peptide substrate of tobacco etch virus (TEV) protease (ENLYFQC) to the N-terminus of *Renilla* luciferase, followed by TEV protease treatment [12]; the C-terminal cysteine was installed by expressing the rLuc fusion protein with *Mex GyrA* intein, a protein subunit that can be cleaved off via self-catalyzed rearrangement during protein-splicing, followed by intein-mediated ligation with ethylene dicysteine (two cysteines linked through their carboxylate groups by ethylenediamine) [16, 17]. Both rLuc proteins containing 1,2-aminothiol at either the N- or the C-terminus were labeled efficiently (nearly 100 %) within 1 h incubation with CBT-conjugated fluorescein or biotin (fivefold excess) at room temperature at pH 8.0, while proteins without 1,2-aminothiol were not labeled [5]. CBT condensation has also been applied to label live cell surface engineered to possess free N-terminal cysteine [5]. Currently, labeling of N-terminus of proteins with CBT derivatives has been patented by Promega (WO 2009142678A1).

The fast kinetics of the CBT labeling makes it a highly attractive tool for labeling proteins with imaging probes of short half-lives, such as positron emitting isotope fluorine-18 (¹⁸F), which has a half-life of 110 min. To demonstrate the utility of the CBT chemistry for ¹⁸F labeling, we first produced peptide and protein with N-terminal cysteine, and labeled them with ¹⁸F-conjugated CBT. Both peptide and protein were linked to ¹⁸F via CBT condensation at high radiochemical yield (80 and 12 % decay-corrected and isolated yields, respectively) and high radiochemical purity (both over 99 %), suggesting CBT chemistry can be a useful

approach for site-specific labeling of peptides or proteins with ^{18}F for medical imaging applications [18]. In addition to the fast kinetics, CBT condensation produces a single regio/stereoisomer, which is critical when precisely controlled final structure is desired [19].

Besides labeling of biomolecules, the efficient CBT condensation has found its application in: self-condensation and assembly of imaging molecules when both CBT or its analogs and a free cysteine residue are present in the molecular scaffold, in cells and in living mice [20–25]; and generation of d-luciferin using “caged” CBT and/or cysteine, which can be freed by stimuli activation, for bioluminescent detection of enzyme or reactive oxygen species (ROS) activities in mice [26, 27]. More, the CBT condensation is orthogonal to other labeling reactions, therefore simultaneous modification of proteins with different functionalities (such as a fluorescence resonance energy transfer pair) is desirable.

The following protocol describes expression of protein with cysteine (1,2-aminothiol) at its N- or C-terminus, and site-specific labeling of the protein using CBT condensation.

2 Materials

1. TOP10 *E. coli* cells.
2. LMG194 *E. coli* cells.
3. Luria–Bertani (LB) growth medium.
4. pBAD-Luc8 plasmid [28].
5. pTWIN-MBP1 (New England Biolabs).
6. 5'-Primer (5'-CC ATG GCT CAT CAT CAT CAT CAT GAA ACC TGT ATT TTC AGT GCG CTT CCA AGG TGT-3') and 3'-primer (5'-AAG CTT TTA CTG CTC GTT CTT CAG CAC-3').
7. 5'-Primer (5'-A ATT GAA TTC TGC ATC ACG GGA GAT GCT) and 3'-primer (5'-A GCT AAG CTT GGT GAG GCC AGT AGC GTG-3').
8. Lysis buffer: Tris (20 mM, pH 7.4), imidazole (20 mM), NaCl (300 mM), lysozyme (1 mg/mL), DNase I (5 $\mu\text{g}/\text{mL}$), and RNase A (10 $\mu\text{g}/\text{mL}$).
9. L-Arabinose (20 % (w/v)), filter-sterilized through a 0.2- μm syringe filter.
10. Tris buffer (20 mM, pH 7.4).
11. PBS buffer (pH 7.4).
12. Ni-NTA agarose beads (Qiagen).
13. Ni-NTA agarose beads wash buffer: Tris (20 mM, pH 7.4), imidazole (20 mM), NaCl (300 mM).

14. Ni-NTA agarose beads elution buffer: Tris (20 mM, pH 7.4), imidazole (250 mM), NaCl (300 mM).
15. NuPage denaturing 4–12 % SDS-PAGE gel (Invitrogen).
16. SDS-PAGE running buffer.
17. SDS-PAGE loading dye.
18. SDS-PAGE molecular weight markers.
19. Tobacco etch virus (TEV) protease.
20. Dialysis cassette (3,000 MWCO).
21. Centrifugal filter units (10 kDa MWCO) (Millipore, Billerica, MA).
22. Fast protein liquid chromatographic (FPLC) system.
23. Source 15Q anion exchange column.
24. NAP-10 size exclusion column (GE).
25. Sep-Pak QMA cartridge.
26. Ethylene dicysteine (*see Note 1*).
27. 2-Mercaptoethanesulfonic acid (MESA).
28. β -Mercaptoethanol.
29. 2-Cyano-6-aminobenzothiazole.
30. 2-Cyano-6-hydroxybenzothiazole.
31. Isobutyl chlorformate.
32. Boc-glycine.
33. N-Methyl morpholine.
34. Ethylene glycol di(*p*-toluenesulfonate).
35. N,N-Diisopropylethylamine (DIPEA).
36. Trifluoroacetic acid (TFA).
37. Anhydrous sodium sulfate.
38. Potassium carbonate.
39. 18-Crown-6.
40. Tetrahydrofuran (THF).
41. Ethyl acetate.
42. Dimethylformamide (DMF).
43. Dichloromethane (DCM).
44. Diethyl ether.
45. Fluorescein isothiocyanate (FITC).
46. Glutathione or Tris(2-carboxyethyl) phosphine hydrochloride (TCEP).
47. Cysteine.
48. High-performance liquid chromatographic (HPLC) system.

49. Flash chromatographic system.
50. Lyophilizer.
51. Rotary evaporator.

3 Methods

For most mature site-specific protein modification methods, one key drawback is the slow reaction rate, which requires high reaction concentrations and long labeling time. Fast kinetics is particularly important for radiolabeling of proteins, when reagents have limited half-life accompanied by their low concentrations. While efforts have been made to invent new chemical pairs for faster reactions [29], we have developed the decently fast CBT condensation reaction for protein labeling, which can take advantage of methodologies explored for the well-studied native chemical ligation. In this protocol, we will describe production of protein with N-terminal cysteine using fusion protein with subsequent protease treatment, production of cysteine-containing protein at C-terminus using intein-mediated ligation, and site-specific protein labeling using CBT-conjugated FITC and ^{18}F .

3.1 Production of *Renilla luciferase* (*rLuc*) with N-Terminal Cysteine

1. Plasmid construction. Fuse a his-tag sequence, TEV protease cleavage sequence, and a cysteine mutation at P1' position to the N-terminus of a *Renilla luciferase* mutant Luc8 (pBAD-Luc8) using 5'-primer (5'-CC ATG GCT CAT CAT CAT CAT CAT GAA ACC TGT ATT TTC AGT GCG CTT CCA AGG TGT-3') and 3'-primer (5'-AAG CTT TTA CTG CTC GTT CTT CAG CAC-3') (see Note 2). In this example, the amplified DNA fragments were inserted into pBAD vector between NcoI and HindIII.
2. Protein expression and purification. Transform TOP10 *E. coli* cells using the plasmid above, let the cells grow in LB medium (1 L) at 37 °C until the mid-log phase ($\text{OD}_{600} < 0.6$), and induce the recombinant protein expression with 0.2 % arabinose for 2 h at 37 °C. Harvest the cells by centrifugation (4,000 $\times g$, 30 min), discard the supernatant, and freeze the cell pellet at -80 °C. Thaw the cells, resuspend the cell pellet in 3 volumes of lysis buffer, and sonicate the cell suspension for lysis (<1 min). Centrifuge the cell lysate (15,000 $\times g$, 30 min), collect the supernatant, and incubate with Ni-NTA agarose resin beads (2 mL) for 2 h at 4 °C with constant mixing. Transfer the Ni-NTA agarose bead suspension to a polypropylene column, and wash the column with 20 bed volumes of wash buffer. Elute the His-tagged protein using three bed volumes of elution buffer.

3. Generation of N-terminal cysteine. Inject a solution of purified His-tagged protein (2 mg in 400 μ L, *see Note 3*) and TEV protease (200 units) in PBS buffer into a dialysis cassette (3,000 MWCO), and dialyze the mixture against PBS buffer containing β -mercaptoethanol (1 mM) for 20 h at room temperature. The TEV protease treatment leads to cleavage of the His-tag sequence and TEV protease substrate, generating N-terminal cysteine at P1' position of the desired protein. Purify the protein by addition of Ni-NTA agarose beads to the reaction solution to remove uncleaved His-tagged protein. Lyophilize the solution from the dialysis cassette to obtain rLuc with N-terminal cysteine (*see Note 4*).

3.2 Production of rLuc with C-Terminal Cysteine

1. Plasmid construction. A free cysteine residue can be introduced at the C-terminus of protein by intein-mediated ligation. First, fuse the intein GyrA to the C-terminus of rLuc by introduction of two additional restriction sites EcoR I and Hind III into pBAD-Luc8 plasmid. Amplify the GyrA mutant (N198A) gene from pTWIN-MBPI (New England Biolabs) with 5'-primer (5'-A ATT GAA TTC TGC ATC ACG GGA GAT GCT) and 3'-primer (5'-A GCT AAG CTT GGT GAG GCC AGT AGC GTG-3'). Digest the PCR product by EcoR I and Hind III, and ligate into the same enzyme-digested pBAD-Luc8 to produce pBAD-Luc8-GyrA plasmid (*see Note 5*).
2. Protein expression and purification. Transform LMG194 cell, an *E. coli* strain deficient of arabinose, with the plasmid, and grow the transformed cells in 1 L of LB media at 37 °C until OD₆₀₀ of <0.6, and induce the recombinant protein expression with 0.2 % arabinose for 4 h. Harvest the cells by centrifugation and freeze the cells at -80 °C. Thaw the cells and then lyse them for 30 min in 3 volumes of lysis buffer, followed by sonication (<1 min). Clarify the lysates by centrifugation (15,000 $\times g$, 30 min) at 4 °C, and incubate the clarified supernatant containing expressed proteins with Ni-NTA agarose beads (2 mL) with gently shaking at 4 °C for 2 h. Collect the Ni-NTA agarose beads, wash them with 20 bed volumes of wash buffer, and elute the His-tagged proteins with three bed volumes of elution buffer (*see Note 6*).
3. Introduction of cysteine at the C-terminus of the protein. Incubate the intein-rLuc fusion protein (50 μ M) with ethylene dicysteine (10 mM) and 2-mercaptoethanesulfonic acid (MESA) (20 mM) for 15 h at 4 °C in PBS buffers (pH 8.0). Purify the proteins using FPLC on Source 15Q anion exchange column (*see Note 4*).

3.3 Preparation of FITC-CBT Conjugate (Fig. 2)

1. Synthesis of CBT-Gly-NH₂ conjugate. Stir a solution containing isobutyl chlorformate (20 mg, 0.15 mmol), Boc-glycine (35 mg, 0.2 mol) and *N*-methyl morpholine (NMP) (30 mg,

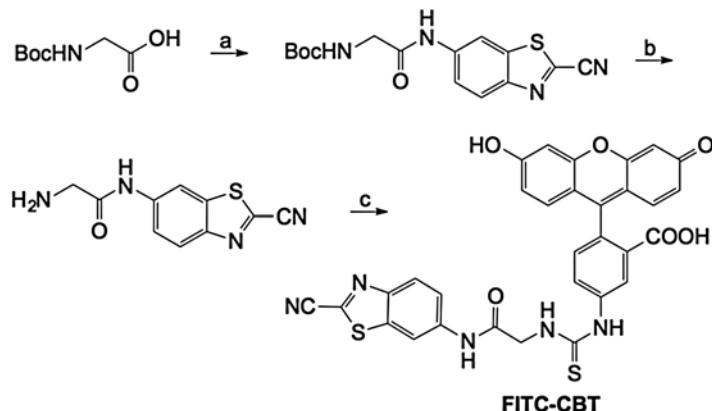


Fig. 2 Synthesis of FITC-CBT conjugate. Conditions: (a) ClCOOBu^t , NMP; then amino-CBT; (b) 20 % TFA/CH₂Cl₂; (c) FITC isomer I

0.3 mmol) in THF (4.0 mL) under N₂ at 0 °C for 20 min. Add 2-cyano-6-aminobenzothiazole (17.5 mg, 0.1 mmol) to the above mixture and stir together overnight, while allowing the mixture to warm up naturally from 0 °C to ambient temperature. Quench the reaction with saturated NaHCO₃ solution (30 mL), and extract the reaction mixture with ethyl acetate (2 × 30 mL). Combine the organic phase, dry the ethyl acetate solution with anhydrous sodium sulfate and remove the organic solvent using rotary evaporator. Purify the material using column chromatography or preparative HPLC to give pure product (typically around 80 % yield). Treat the product with 20 % TFA in CH₂Cl₂ for 1 h to give CBT-Gly-NH₂. Quench the reaction by adding cold diethyl ether to precipitate CBT-Gly-NH₂, dry the precipitates under vacuum for further use (*see Note 7*).

2. Coupling of CBT with FITC. Dissolve CBT-Gly-NH₂ (0.2 mg), FITC isomer I (0.1 mg) and DIPEA (1 μL) in DMF (0.2 mL, pH 8), and stir the mixture at ambient temperature for 2 h. Subject the reaction mixture directly to HPLC purification, lyophilize the collected fraction to give pure FITC-CBT conjugate (*see Note 8*).
1. Mix cysteine-containing protein (10 μM, obtained as described above via using other means), FITC-CBT probe (50 μM), and glutathione or TCEP (2 mM) in PBS buffer (20 μL, pH 7.4) (*see Note 9*).
2. Allow the labeling reaction to proceed for 2 h at room temperature in the dark.
3. Quench the labeling reaction with free cysteine (0.1 mM, in 20 μL PBS buffer), and pass the reaction solution through a NAP-10 size exclusion column to remove the excess FITC-CBT.
4. Verify the labeling efficiency using SDS-PAGE (*see Note 10*).

3.4 Site-Specific Labeling of Protein Using FITC-CBT (*See Note 9*)

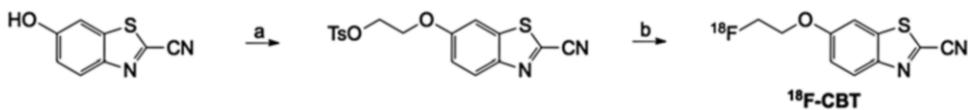


Fig. 3 Synthesis of ¹⁸F-CBT conjugate. Conditions: (a) Ethylene glycol di-tosylate, K₂CO₃, DMF, room temperature, 8 h; (b) K[¹⁸F]F, 18-crown-6, K₂CO₃, MeCN, 90 °C, 10 min

3.5 Preparation of ¹⁸F-CBT Conjugate (Fig. 3)

1. Synthesis of tosylated CBT. Dissolve 2-cyano-6-hydroxybenzothiazole (260 mg, 1.48 mmol) in DMF (1 mL), add ethylene glycol di-tosylate (1,640 mg, 4.42 mmol) and potassium carbonate (510 mg, 3.69 mmol), and stir the mixture for 24 h at room temperature. Cool the reaction mixture to 0 °C and quench the reaction with HCl (4 mL, 1 M), and extract the reaction mixture with ethyl acetate (2 × 50 mL). Combine the organic phase, dry the ethyl acetate solution with anhydrous sodium sulfate and remove the organic solvent using rotary evaporator. Purify the material using column chromatography (dichloromethane: hexanes = 4:1) or preparative HPLC to give pure product (typically around 50 % yield).
2. Prepare ¹⁸F-fluoride (1,000 mCi) by proton bombardment of 2.5 mL ¹⁸O enriched water via the ¹⁸O(p,n)¹⁸F nuclear reaction (see Note 11). Trap the ¹⁸F-fluoride onto a Sep-Pak QMA cartridge, and elute the ¹⁸F-fluoride into a dried glass containing using 18-crown-6/K₂CO₃ solution (1 mL, 15:1 MeCN/H₂O, 16.9 mg of 18-Crown-6, 4.4 mg of K₂CO₃). Dry the resulting solution azeotropically with sequential MeCN evaporation at 90 °C.
3. Synthesis of ¹⁸F-CBT. Add tosylated CBT (2 mg in 1 mL of anhydrous MeCN) to the reactor containing ¹⁸F-fluoride, and heat at 90 °C for 10 min. Cool the mixture to 30 °C, quench the reaction with HCl (0.05 M, 2.5 mL), and purify the product using HPLC. Dilute the collected ¹⁸F-CBT with water (20 mL), pass through a C18 cartridge, and elute with Et₂O (2.5 mL). Remove the Et₂O by helium stream to obtain the pure ¹⁸F-CBT (see Note 12).

3.6 Site-Specific Labeling of Protein Using ¹⁸F-CBT

1. Prepare a solution of cysteine-containing protein (rLuc, 5 nmol) in PBS buffer (150 µL, pH 7.4) containing glutathione or TCEP (2 mM).
2. Dissolve ¹⁸F-CBT conjugate in DMSO (10.7 mCi, 7.5 µL), mix with the above protein solution, and stir at 37 °C for 30 min.
3. Adjust the crude mixture to 1.0 mL with PBS buffer (pH 7.4), and pass the solution through a NAP-10 column preconditioned with PBS buffer.
4. Add 1.5 mL of PBS buffer to the column to collect the ¹⁸F-labeled protein (see Note 13).

4 Notes

1. Prepare ethylene dicysteine using standard amide bond formation condition. Stir N-Boc-S-Trityl-L-cysteine, ethylenediamine (0.5 equivalents (eq.)), HBTU (1.5 eq.), HOEt (1.5 eq.), DIPEA (3 eq.) in DMF (1 mL) for 1 h at room temperature. Quench the reaction with water, and extract the reaction mixture using ethyl acetate. Remove the ethyl acetate by rotary evaporation, and dry the product under vacuum. Treat the solid product with 20 % TFA in CH₂Cl₂ for 1 h to give CBT-Gly-NH₂. Quench the reaction by adding cold diethyl ether to precipitate ethylene dicysteine, and dry the precipitates under vacuum. The reagent can be stored stably at -20 °C for future use (good for years).
2. His-tag sequence is added here to allow easier purification of the recombinant protein. For studies where protein purification is not necessary (such as that in live cell imaging), only the insertion of protease sequence is needed.
3. Purified His-tagged protein after elution from Ni-NTA column can be used here after centrifugal replacement of Ni-NTA column elution buffer with reaction buffer using centrifugal filter units.
4. The products were analyzed by SDS-PAGE and mass spectrometry. The luciferase activity of the rLuc fusion proteins was assessed by a calibrated luminometer with coelenterazine as the substrate, and all proteins showed intact enzymatic activity.
5. A sequence of peptide (Val-Pro-Leu-Ser-Leu-Thr-Met-Gly) can also be introduced between rLuc and GyrA. Amplify the rLuc gene from pBAD-Luc8 with 5'-primer (5'-A TGC CCA TGG CTT CCA AGG TGT AC-3'), and the 3'-primer (5'-ATGC GAA TTC ACC ACC CAT TGT CAG TGA CAG AGG TAC TCC TCC CTG CTC GTT CTT CAG-3'). Digest the PCR products with Nco I and EcoR I, and ligate into the same enzyme-digested pBAD-rLuc-GyrA to give plasmids containing different peptide sequences.
6. The eluted fusion protein can be further purified using FPLC on Source 15Q anion exchange column if needed.
7. Glycine residue is added to amino-CBT because direct amidation of imaging probe with amino-CBT is challenging due to the low nucleophilicity of its amino group. With glycine spacer, imaging probes can be incorporated using routine conditions. The reagents, amino-CBT and FITC-CBT, can be stored stably at -20 °C for years.

8. Besides FITC, other activated dyes or molecules, such as *N*-hydroxysuccinimide (NHS) ester of rhodamine or biotin, can also be attached to CBT under the same condition.
9. Similar protocol has been used to label 1,2-aminothiol-containing peptides or proteins with other dyes, biotin, and peptides. Presence of glutathione or TCEP is necessary to avoid oxidative disulfide formation of the free sulfhydryl group.
10. Labeling efficiency of the CBT condensation was studied using a peptide-FITC-CBT conjugate. Upon condensation, the product protein gave a different mass (different band on gel). With two equivalents of CBT, the labeling reaction could reach 90 % completion in 1 h; with five equivalents of the CBT probe, nearly 100 % completion could be achieved within 1 h.
11. Performed using a GE PETtrace cyclotron at Lucas Center for Imaging at Stanford.
12. The isolated radiochemical yield of ¹⁸F-CBT was around 20 % (140–150 mCi, decay-corrected to end of bombardment).
13. The isolated radiochemical yield of ¹⁸F-rLuc was around 12 % (decay-corrected to end of synthesis). Overall reaction and purification steps were completed within 40 min. The specific radioactivity was 262 mCi/μmol.

Acknowledgement

This work was funded by a grant from NIGMS (R01GM086196-01), the IDEA award from Department of Defense Breast Cancer Research Program (W81XWH-09-1-0057) and the NCI ICMIC at Stanford (1P50CA114747-06).

References

1. Fernandez-Suarez M, Ting AY (2008) Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Biol* 9(12):929–943. doi:[10.1038/nrm2531](https://doi.org/10.1038/nrm2531)
2. Carter P, Merchant AM (1997) Engineering antibodies for imaging and therapy. *Curr Opin Biotechnol* 8(4):449–454. doi:[10.1016/s0958-1669\(97\)80067-5](https://doi.org/10.1016/s0958-1669(97)80067-5)
3. Jing CR, Cornish VW (2011) Chemical tags for labeling proteins inside living cells. *Acc Chem Res* 44(9):784–792. doi:[10.1021/ar200099f](https://doi.org/10.1021/ar200099f)
4. White EH, McCapra F, Field GF, McElroy WD (1961) The structure and synthesis of firefly luciferin. *J Am Chem Soc* 83(10):2402–2403. doi:[10.1021/ja01471a051](https://doi.org/10.1021/ja01471a051)
5. Ren HJ, Xiao F, Zhan K, Kim YP, Xie HX, Xia ZY, Rao J (2009) A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins. *Angew Chem Int Ed* 48(51):9658–9662. doi:[10.1002/anie.200903627](https://doi.org/10.1002/anie.200903627)
6. Soellner MB, Nilsson BL, Raines RT (2006) Reaction mechanism and kinetics of the traceless staudinger ligation. *J Am Chem Soc* 128(27):8820–8828. doi:[10.1021/ja060484k](https://doi.org/10.1021/ja060484k)
7. Dirksen A, Hackeng TM, Dawson PE (2006) Nucleophilic catalysis of oxime ligation. *Angew Chem Int Ed* 45(45):7581–7584. doi:[10.1002/anie.200602877](https://doi.org/10.1002/anie.200602877)
8. Dirksen A, Dirksen S, Hackeng TM, Dawson PE (2006) Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. *J Am Chem Soc* 128(27):8820–8828. doi:[10.1021/ja060484k](https://doi.org/10.1021/ja060484k)

- Soc 128(49):15602–15603. doi:[10.1021/ja067189k](https://doi.org/10.1021/ja067189k)
9. Sletten EM, Bertozzi CR (2011) From mechanism to mouse: a tale of two bioorthogonal reactions. *Acc Chem Res* 44(9):666–676. doi:[10.1021/ar200148z](https://doi.org/10.1021/ar200148z)
 10. Dawson PE, Muir TW, Clarklewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. *Science* 266(5186):776–779
 11. Erlanson DA, Chytil M, Verdine GL (1996) The leucine zipper domain controls the orientation of AP-1 in the NFAT center dot AP-1 center dot DNA complex. *Chem Biol* 3(12):981–991. doi:[10.1016/s1074-5521\(96\)90165-9](https://doi.org/10.1016/s1074-5521(96)90165-9)
 12. Tolbert TJ, Wong C-H (2002) New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew Chem Int Ed* 41(12): 2171–2174. doi:[10.1002/1521-3773\(20020617\)41:12<2171::aid-anie2171>3.0.co;2-q](https://doi.org/10.1002/1521-3773(20020617)41:12<2171::aid-anie2171>3.0.co;2-q)
 13. Gentle IE, De Souza DP, Baca M (2004) Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug Chem* 15(3):658–663. doi:[10.1021/bc049965o](https://doi.org/10.1021/bc049965o)
 14. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J Am Chem Soc* 133(30):11418–11421. doi:[10.1021/ja203111c](https://doi.org/10.1021/ja203111c)
 15. Li X, Fekner T, Ottesen JJ, Chan MK (2009) A pyrrolysine analogue for site-specific protein ubiquitination. *Angew Chem* 121(48):9348–9351. doi:[10.1002/ange.200904472](https://doi.org/10.1002/ange.200904472)
 16. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat Methods* 3(6):429–438. doi:[10.1038/nmeth886](https://doi.org/10.1038/nmeth886)
 17. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* 95(12):6705–6710. doi:[10.1073/pnas.95.12.6705](https://doi.org/10.1073/pnas.95.12.6705)
 18. Jeon J, Shen B, Xiong L, Miao Z, Lee KH, Rao J, Chin FT (2012) Efficient method for site-specific ¹⁸F-labeling of biomolecules using the rapid condensation reaction between 2-cyanobenzothiazole and cysteine. *Bioconjug Chem* 23(9):1902–1908. doi:[10.1021/bc300273m](https://doi.org/10.1021/bc300273m)
 19. Cheng Y, Peng H, Chen W, Ni N, Ke B, Dai C, Wang B (2013) Rapid and specific post-synthesis modification of DNA through a biocompatible condensation of 1,2-aminothiols with 2-cyanobenzothiazole. *Chemistry* 19(12):4036–4042. doi:[10.1002/chem.201201677](https://doi.org/10.1002/chem.201201677)
 20. Liang GL, Ren HJ, Rao JH (2010) A biocompatible condensation reaction for controlled assembly of nanostructures in living cells. *Nat Chem* 2(1):54–60. doi:[10.1038/ncchem.480](https://doi.org/10.1038/ncchem.480)
 21. Liang GL, Ronald J, Chen YX, Ye DJ, Pandit P, Ma ML, Rutt B, Rao JH (2011) Controlled self-assembling of gadolinium nanoparticles as smart molecular magnetic resonance imaging contrast agents. *Angew Chem Int Ed* 50(28):6283–6286. doi:[10.1002/anie.201007018](https://doi.org/10.1002/anie.201007018)
 22. Ye DJ, Liang GL, Ma ML, Rao JH (2011) Controlling intracellular macrocyclization for the imaging of protease activity. *Angew Chem Int Ed* 50(10):2275–2279. doi:[10.1002/anie.201006140](https://doi.org/10.1002/anie.201006140)
 23. Shen B, Jeon J, Palner M, Ye D, Shuhendler A, Chin FT, Rao J (2013) Positron emission tomography imaging of drug-induced tumor apoptosis with a caspase-triggered nanoaggregation probe. *Angew Chem Int Ed* 52(40):10511–10514. doi:[10.1002/anie.201303422](https://doi.org/10.1002/anie.201303422)
 24. Ye D, Shuhendler AJ, Cui L, Tong L, Tee SS, Tikhomirov G, Felsher DW, Rao J (2014) Bioorthogonal cyclization-mediated in situ self-assembly of small-molecule probes for imaging caspase activity in living mice. *Nat Chem* 6 (6):519–526. doi:[10.1038/ncchem.1920](https://doi.org/10.1038/ncchem.1920)
 25. Ye D, Shuhendler AJ, Pandit P, Brewer KD, Tee SS, Cui L, Tikhomirov G, Rutt B, Rao J (2014) Caspase-responsive smart gadolinium-based contrast agent for magnetic resonance imaging of drug-induced apoptosis. *Chem Sci* 5(10):3845–3852. doi:[10.1039/c4sc01392a](https://doi.org/10.1039/c4sc01392a)
 26. Van de Bittner GC, Bertozzi CR, Chang CJ (2013) Strategy for dual-analyte luciferin imaging: in vivo bioluminescence detection of hydrogen peroxide and caspase activity in a murine model of acute inflammation. *J Am Chem Soc* 135(5):1783–1795. doi:[10.1021/ja309078t](https://doi.org/10.1021/ja309078t)
 27. Godinat A, Park HM, Miller SC, Cheng K, Hanahan D, Sanman LE, Bogyo M, Yu A, Nikitin GF, Stahl A, Dubikovskaya EA (2013) A biocompatible in vivo ligation reaction and its application for noninvasive bioluminescent imaging of protease activity in living mice. *ACS Chem Biol.* doi:[10.1021/cb3007314](https://doi.org/10.1021/cb3007314)
 28. Loening AM, Fenn TD, Wu AM, Gambhir SS (2006) Consensus guided mutagenesis of *Renilla* luciferase yields enhanced stability and light output. *Protein Eng Des Sel* 19(9):391–400. doi:[10.1093/protein/gzl023](https://doi.org/10.1093/protein/gzl023)
 29. Schmidt MJ, Summerer D (2012) A need for speed: genetic encoding of rapid cycloaddition chemistries for protein labelling in living cells. *Chembiochem* 13(11):1553–1557. doi:[10.1002/cbic.201200321](https://doi.org/10.1002/cbic.201200321)

Chapter 6

Fluorescent Labeling for Patch-Clamp Fluorometry (PCF) Measurements of Real-Time Protein Motion in Ion Channels

Thomas K. Berger and Ehud Y. Isacoff

Abstract

Understanding the function of ion channels is a major goal of molecular neurophysiology. While standard electrophysiological methods are invaluable tools to investigate the gating of ion channels, the structural rearrangements that mediate the way a channel senses physiological signals and opens and closes its gates cannot be measured electrically in a direct way. Here, we describe a method, based on site-specific labeling of a channel of interest with an environmentally sensitive fluorophore, which makes it possible to monitor conformational changes of ion channels in biological membranes in real time.

Key words Patch-clamp fluorometry (PCF), Voltage-clamp fluorometry (VCF), Ion channels, Structural dynamics, Cysteine-reactive probes, Environmentally sensitive fluorophores

1 Introduction

The ability to express ion channels in heterologous systems like mammalian cell lines or *Xenopus laevis* frog oocytes greatly facilitates the study of ion channels and transmembrane receptors. High expression of the channel or receptor of interest, combined with relatively little expression of endogenous channels, allows good signal-to-noise ratios in electrophysiological recordings. The standard way of probing a channel's functional characteristics is to electrically record currents induced by specific physiological stimuli, e.g., changes in membrane potential imposed by a voltage clamp or agonists. In this setting, the current reflects the channel's transitions between closed, open and, inactivated or desensitized states, i.e., the gating steps. Channels, however, can exist in several distinct closed and open states, reflecting two kinds of partial activation, such as full activation of only a subset of their multiple subunits (e.g., incomplete occupancy of ligand binding sites) or incomplete activation of an individual subunit (e.g., voltage

activation to an intermediate state) or inactivation or desensitization of a channel that is only partially activated. Transitions to and between many of these intermediate states can be electrophysiologically silent because they do not change membrane conductance and may move too little protein charge through the membrane electric field to be detected. Yet these transitions involve conformational changes in the membrane protein. The technique of Voltage-Clamp Fluorometry (VCF), established in 1996 [1], and the patch-clamp variant Patch-Clamp Fluorometry (PCF) [2], makes it possible to visualize these conformational changes using an environmentally sensitive fluorophore that is covalently attached in a site-directed manner to a moveable part of the channel. Solvent polarity, interaction with nearby amino-acid side chains on the protein and with membrane lipids and position in the membrane electric field can strongly influence the quantum yield, inter-system conversion, and spectra of a fluorophore [3, 4]. A variant on the method is to label a channel at more than one site with a pair of fluorophores, where one acts as a donor and the second as an acceptor so that changes in their proximity or relative angle can be detected as changes in fluorescence resonance energy transfer (FRET) [5, 6]. Thus, conformational changes in the protein that reposition the fluorophore from one nano-environment to another can strongly influence its fluorescence. The fluorescence signature is virtually unique to each amino acid attachment site, even though the same transition can be detected at many sites [7, 8]. For example, a fluorophore at a group of attachment sites may report on a particular functional transition, but the direction and magnitude of the fluorescence change will differ between the sites, and some of them may report on other rearrangements, but not necessarily the same ones. This diversity is particularly startling when one sees a great diversity of fluorescence changes even when comparing the same fluorophore at neighboring positions in the 3D crystal structure [9]. The change in fluorescence intensity for fluorophores can be as large as 2,000 % [10].

Measurements can be made from many channels at once [1] or one channel at a time [10, 11]. DNA or RNA encoding the channel protein is delivered into a non-excitable cell that has few channels of its own so that the channel can be studied in an environment where it is present at much higher densities than native channels. As with earlier electrophysiological analysis, the most common cell to use is the *Xenopus* oocyte. Small organic fluorophores, such as rhodamine, fluorescein, bodipy, alexa, or cyanine dyes, can be attached to specific amino acids of the channel after it has been made by the cell and shipped to its plasma membrane. The fluorophore is targeted to a specific site in the channel protein by mutating the channel to remove all of its externally exposed native cysteines and introducing a single cysteine residue to serve as an anchoring site. The fluorophore selectively forms a permanent

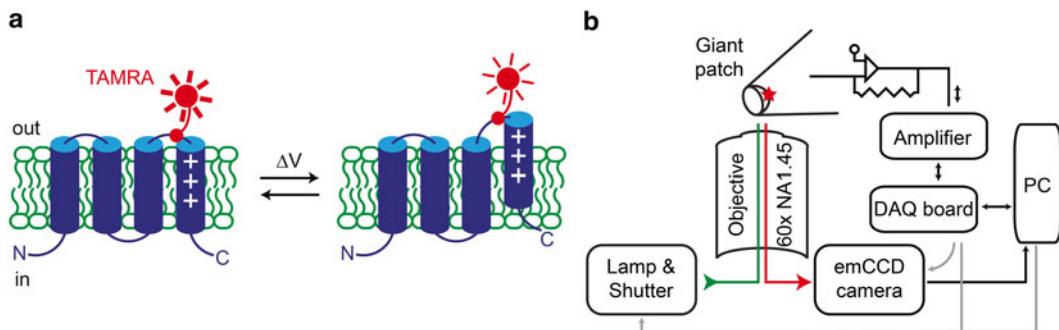


Fig. 1 (a) Cartoon of fluorescence changes induced by protein motion. An environmentally sensitive fluorophore (red) is attached via a thiol-reactive linker to an accessible cysteine at the extracellular part of the ion channel (blue). Motions of the ion channel in response to a stimulus (e.g., change of membrane potential) can position the fluorophore in a different environment, thereby changing the fluorescence intensity. (b) Cartoon of the experimental setup. A giant membrane patch is excised from a labeled oocyte and observed with a 60 \times objective with a high numerical aperture (NA 1.45). The fluorophores in the patch cone (symbolized with a red star) are excited with a xenon lamp and their fluorescence output is recorded with a CCD camera. The patch-clamp amplifier is connected to a PC via a data acquisition board (DAQ board). The whole system can be controlled with a single PC. Gray arrows depict control signals (TTL pulses), black arrows the flow of data

covalent bond with the introduced cysteine through cysteine-specific thiol chemistry. Once the channels are fluorescently labeled the cell is voltage clamped on a microscope, the fluorophore is excited through an objective and its fluorescence emission is captured through the same objective and detected by a photomultiplier, photodiode, or CCD camera (Fig. 1).

The key to VCF is that channel fluorescence (and thus structural state) is monitored at the same time as voltage jumps or agonist application is used to activate the channel and current measurements are obtained to detect the functional state of the channel, thus making it possible to relate structural transitions to functional transitions.

VCF and PCF are time-resolved fluorescence assays. Since the excited state lifetimes of most fluorophores are in the nanosecond range [12] and the fastest ion-channel motions occur on the microsecond scale, VCF and PCF report on the protein structure at a rate much faster than the structural rearrangements take place, thus providing a real-time readout of the protein motion. Thus, when the fluorophore is attached to the so-called pore helix in a voltage-gated K⁺ channel, the fluorescence changes over hundreds of milliseconds, in parallel with the slow closure of the channel inactivation gate, thereby identifying this piece of the pore as a gating element involved in channel inactivation [7, 13, 14]. In contrast, when the fluorophore is attached to the charged S4 segment of a voltage-gated K⁺ channel, the change in fluorescence intensity occurs in milliseconds, and has the voltage dependence and kinetics of the gating current which precedes the opening of the pore [1, 14, 15].

But VCF and PCF provide more than just kinetic information about which part of the protein moves when. They also reveal steady-state information about the fraction of channels that have undergone that transition at each dose of ligand or voltage step. Thus, one can compare the fluorescence-voltage ($F-V$) relation for a fluorophore at a particular attachment site to the gating charge-voltage ($Q-V$) relation or conductance-voltage ($G-V$) relation or inactivation-voltage relation or dose-response relation and determine the energetics of the conformational rearrangement and confirm the match to a particular transition deduced from the kinetic match.

Because small fluorophores can be attached at many individual spots it is possible to reconstruct global gating rearrangements of the protein from motions sensed at multiple individual sites [8, 9]. One particularly powerful application is that a fluorophore can be attached selectively to one subunit to study the influence of other, unlabeled, subunits on its conformational changes and thereby identify cooperative transitions [16–18].

VCF and PCF have revealed fundamental aspects of channel gating, which could not have been learned from other methods. These include: (a) the identity of S4 as the voltage sensor of voltage-gated ion channels, along with evidence that S4 moves in a series of independent steps in the four subunits of the channel, with a final cooperative step corresponding to the opening of the gate [1, 15, 17, 19], (b) that gating speed can be regulated by changing the rate of S4 motion [20], (c) that to open the HCN channel only two out of the four S4 voltage sensors must move [21], (d) that the four nonidentical subunits of Na^+ channels activate in a specific sequence [22], (e) that voltage-gated proton channels, which are bundled into pairs, gate cooperatively [18, 23], and (f) that the voltage-sensing phosphatase is a monomer in which a single voltage-sensing domain undergoes a complex series of voltage-sensing rearrangements [24, 25]. Environment-sensing fluorescence has also shown that GABA_A receptor channels and β -adrenoceptor G-protein coupled receptors are activated at different rates by agonists that have distinct actions and place the receptor in distinct conformations [26, 27] and has provided important insights about the function of transporters [28, 29].

VCF was first established using the two-electrode voltage clamp technique (TEVC) [1, 15]. The PCF variant came later, first in excised inside-out patches using changes in accessibility to a soluble quencher to change the environment of the fluorophore [2], later in whole cell recordings relying on the environmental sensitivity without an added quencher [30]. Here, we describe the technique of how to label the extracellular site of ion channels expressed in *Xenopus laevis* oocytes and to monitor fluorescence changes in excised patches, exemplified with the voltage-gated proton channel Hv1.

2 Materials

Unless noted otherwise, all solutions are prepared and stored at room temperature and should be sterile filtered with 0.2 mm filters to reduce contaminations. Water should be prepared by purifying deionized water.

2.1 Solutions

1. ND96 solution: Oocyte are incubated in ND96 solution which contains 96 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂, 5 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), and 2.5 mM sodium pyruvate, as well as Gentamicin at pH 7.6. Add about 80 ml to a 100 ml graduated cylinder or glass beaker. Weigh and sequentially add under stirring 0.56 g NaCl, 0.015 g KCl, 0.04 g CaCl₂ 6H₂O, 0.02 g MgCl₂ 6H₂O, 0.12 g HEPES, 0.028 g sodium pyruvate, and 0.1 ml Gentamicin solution (50 mg/ml solution). Titrate with NaOH to pH 7.6. Make up to 100 ml with water.
2. Stock solution of the fluorophore: The stock solution of the fluorophore, e.g., 2-((5(6)-Tetramethyl-rhodamine)carboxyl-amino)ethyl Methanethiosulfonate (MTS-TAMRA, Toronto Research Chemicals, Toronto, ON), is prepared unfiltered under low illumination with *N,N*-Dimethylformamide (DMF, Sigma-Aldrich) to reach a concentration of 25 mM fluorophore. MTS-TAMRA is available as 1 mg in a glass vial. Add 70.4 µl DMF to the vial and shake well to obtain a 25 mM stock solution. Aliquots of 10–20 µl are stored in lightproof tubes (e.g., tin foil wrapped) inside a closed container containing desiccants at –20 °C.
3. Labeling solution: The labeling solution is prepared with buffer which contains 92 mM KCl, 0.75 mM CaCl₂, 1 mM MgCl₂, 10 mM HEPES at pH 7.5. Add about 80 ml to a 100 ml graduated cylinder or glass beaker. Weigh and sequentially add under stirring 0.69 g KCl, 0.016 g CaCl₂ 6H₂O, 0.02 g MgCl₂ 6H₂O, 0.24 g HEPES. Titrate with KOH to pH 7.5. Make up to 100 ml with water.
4. Recording solution: The solution for electrophysiological recordings is chosen according to the specific ion channel that is investigated. For the voltage-gated proton channel Hv1, we use a solution with a high pH buffer concentration and low concentration of metallic cations. It contains (in mM) 100 HEPES, 30 methanesulfonic acid (MS), 5 tetraethylammonium chloride (TEACl), 5 Ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetraacetic acid (EGTA), adjusted to pH 7.5 with TEA hydroxide (TEAOH, ~35 mM). Add about 80 ml to a 100 ml graduated cylinder or glass beaker. Weigh and sequentially add under stirring 2.38 g HEPES, 0.195 ml MS, 0.083 g TEACl. Titrate with

TEAOH to about pH 7, which requires a few ml TEAOH. Add 0.19 g EGTA and stir for 10 min. Finish titration with TEAOH to pH 7.5.

2.2 RNA Injection Setup

1. Stereoscope (e.g., Olympus SZ51).
2. Nanoliter injector (e.g., Drummond Nanoject II, Drummond Scientific, Broomall, PA) mounted onto a manual micromanipulator (e.g., WPI, Sarasota, FL). The micromanipulator can be mounted onto a magnetic stand that is positioned on a heavy magnetic metal plate next to the stereoscope.
3. Injection capillaries (e.g., 3-000-203-G/X, Drummond Scientific).
4. Light mineral oil (e.g., Sigma M8410).
5. Pasteur pipettes with fire-polished tips of around 2 mm in diameter are used to handle oocytes.
6. A pair of fine forceps (e.g., Dumont #5, FST, Foster City, CA), which is used for devitellinization of oocytes just prior to recording.
7. Forceps to break injection capillaries.
8. Plastic petri dishes.
9. Parafilm M.
10. Small glass vials.
11. Standard lab shaker.
12. Incubator for oocytes, capable of cooling to 12 °C.

2.3 Electro- physiological Setup

1. Air table (e.g., TMC 63-500, Peabody, MA).
2. Inverted microscope Olympus IX71 (Olympus, Tokyo, Japan), equipped with objective LUCPLFLN20X for sealing and excising the patch, and PLAPON 60XOTIRFM NA 1.45 objective for fluorometry measurements, filter cube matching the excitation and emission spectra of the fluorophore.
3. Standard stable xenon illumination, e.g., Polychrome 5 (Till Photonics, Martinsried, Germany), attenuated with a neutral density filter (ND 1) controlled by a shutter system (e.g., Uniblitz, Vincent Associates, Rochester, VA) (*see Note 1*).
4. Recording chamber with matching cover glass, thickness #1 or #0 (Ted Pella, Redding, CA). Perfusion system with a gravity-driven inflow, outflow via vacuum pump.
5. Immersion oil for fluorescence microscopy (e.g., Olympus IMMOIL-F30CC).
6. Patch-clamp amplifier (e.g., Axopatch 200B, Molecular Devices, Union City, CA) with low-noise headstage mounted on motorized micromanipulator (e.g., Sutter MP-285, Sutter, Novato, CA).

7. Data acquisition board (e.g., Digidata 1440, Molecular Devices).
8. Acquisition software (e.g., pClamp 10, Molecular Devices).
9. Fast and sensitive CCD camera for acquisition of fluorescence changes (e.g., Andor Luca S, Andor Technology, Belfast, UK) run with software that can be triggered externally (e.g., Andor Solis) (*see Note 2*).
10. Pipette puller (e.g., P-97 by Sutter).
11. Microforge (e.g., Narishige MF-830, Narishige, Tokyo, Japan).
12. Borosilicate glass capillaries (e.g., G150TF-4, Warner Instruments, Hamden, CT).
13. Soda lime capillaries (e.g., Plain blue tip glass capillaries, Globe Scientific, Paramus, NJ).

3 Methods

All steps are carried out at room temperature unless specified otherwise.

3.1 Site-Directed Mutagenesis

The cDNA encoding the ion channel of interest is inserted into a vector like pGEMHE or pSD64TF, which is suited for the translation of RNA in oocytes. The ion channel needs to contain a single cysteine at the extracellular, accessible site of interest in order to be labeled with a thiol-linked fluorophore. Extracellular, accessible native cysteines should be removed in order to prevent background labeling at multiple sites. Standard protocols for site-directed mutagenesis using PCR can be used, e.g., QuickChange (Stratagene, La Jolla, CA).

3.2 RNA Preparation

DNA is linearized with a restriction endonuclease downstream of the 3' end of the polyA tail and transcribed using a RNA transcription kits (e.g., mMessage mMACHINE, Ambion, Austin, TX) (*see Note 3*).

3.3 RNA Injection into Oocytes

1. *Xenopus laevis* oocytes are prepared from frogs the day before RNA injection or can be bought ready-to-inject (e.g., Ecocyte Bioscience, Castrop-Rauxel, Germany). They are stored at 16 °C in ND96 in petri dishes or six-well plates.
2. Make injection pipettes from the injection capillaries with the puller. Break the tip with forceps so that the opening is sharp and has a diameter of around 20–50 µm (*see Note 4*).
3. Fill an injection pipette with mineral oil using a steel needle mounted onto a standard syringe. Make sure to fill the pipette without air bubbles.
4. Mount the oil-filled injection pipette on the injector with the piston retracted 1 cm from the maximal outward position.

5. Place a piece of some square cm of parafilm with the inner side on top below the stereoscope, and put some drops of ethanol between the parafilm and the stereoscope base to keep the parafilm flat.
6. Place 1–2 μl of RNA at the correct concentration (depending on the construct, e.g., 0.5 $\mu\text{g}/\mu\text{l}$) on parafilm.
7. Backfill the RNA drop with the injector.
8. Set the injector to 50 nl injection volume and on “fast”.
9. Scratch small squares (approximately 1 \times 1 mm) into a petri dish with forceps.
10. Fill the scratched dish with ND96, place oocytes along a line in that dish.
11. Penetrate the oocytes with the injection pipette at the vegetale pol (white side) and inject the RNA.
12. (Optional) Blocking native cysteines with 2 mM nonfluorescent glycine-maleimide: Add about 1.5 mg glycine maleimide to 5 ml ND96 and shake well. Fill 1–2 ml of ND96 containing glycine maleimide in a small glass vial. Incubate oocytes for 15 min in this solution. Wash the oocytes in three separate glass vials with ND96 by putting them from one vial to the next one while transferring as little solution as possible (*see Note 5*).
13. Incubate injected oocytes for 1–4 days at 18 °C. The optimal incubation time for the investigated ion channel has to be determined empirically (*see Note 6*).

3.4 Oocyte Labeling

1. Select oocytes that are in good, healthy condition (*see Note 7*).
2. Pipette a small volume (0.5 ml) of labeling solution in a small glass vial on ice.
3. Add 0.5–1 μl fluorophore stock solution under dim light and shake well. This corresponds to 25–50 μM fluorophore concentration.
4. Carefully add 5–20 oocytes without transferring too much solution.
5. Gently shake on ice for ~30 min (*see Note 8*). During this incubation time, Subheading 3.5 (see below) can be started.
6. Wash the oocytes in three separate glass vials with ND96 by putting them from one vial to the next one while transferring as little solution as possible.
7. Store labeled oocytes until recording at 12 °C.

3.5 Production of Patch Pipettes

1. Pull patch pipettes with a tip diameter of 10–25 μm . In case of using the Sutter puller, use the Borosilicate glass and a program for standard patch pipettes (~1 μm tip diameter) and decrease the heat value for the last line by 15 % or more while

setting the velocity parameter to a high value to make sure the glass breaks (*see Note 9*).

2. Place a large bead of soda lime glass onto the platinum wire of the microforge. This can be done by using a pipette with a long and thin tip, melt it with the microforge so that the glass retracts and forms a bead of increasing size which is eventually placed onto the platinum wire (*see Note 10*).
3. Fire polish the patch pipettes above the soda lime glass bead until you see at the pipette tip some slight inward-bending of the rim. Make sure that the pipette tip is located exactly above the glass bead. Store 5–10 of these patch pipettes in a dust free container, e.g., a large petri dish with strips of modeling clay to hold the pipettes.
1. Devitellinize an oocyte. Transfer an oocyte from the glass vial into a petri dish filled with ND96. Under the stereoscope, lift the oocyte with one fine forceps at its very top. It might take several trials to grab the oocyte exclusively at its vitelline membrane, without clamping parts of the cell membrane at the same time. Once it is lifted, apply the second fine forceps between the first forceps and the oocyte itself, grab the vitelline membrane as well and pull the two forceps apart, thereby stripping off the vitelline membrane from the oocyte.
2. Take the devitellinized oocyte with a Pasteur pipette and put it into a petri dish filled with the recording solution with minimal solution transfer.
3. Transfer the oocyte into the recording chamber (which is filled with the recording solution) of the patch-clamp setup. Position it towards the patch pipette in the desired orientation (*see Note 11*).
4. Mount a patch pipette filled with pipette solution on the pipette holder of the head stage, apply positive pressure onto the pipette (~0.1 mbar), immerse into the bath, and position the tip close to the equator of the oocyte under visual control with the 20 \times objective.
5. Use the voltage-clamp configuration, adjust the pipette offset to 0 mV, and apply continuous short voltage steps (2 mV, 15 ms) to monitor the electrical resistance over the pipette.
6. Approach the oocyte until a significant increase of electrical resistance occurs.
7. Release positive pressure while moving a final short step forward with the patch pipette.
8. Gently apply negative pressure and negative voltages until a gigaseal forms (*see Note 12*).

3.6 Patch-Clamp Fluorometry

9. Slowly remove patch pipette from oocyte until the patched membrane is completely disconnected from the oocyte (*see Note 13*).
10. Make sure that the patch contains your expressed channel by applying voltage and observe the current.
11. Center pipette into the middle of the field of view.
12. Change to the 60× objective (*see Note 14*) and focus few micrometer above the cover slip.
13. Move down the patch pipette slowly into focus. The patch pipette should be positioned a few micrometer above the cover slip.
14. Switch to the xenon light source and video image the patch pipette with long exposure time (0.3–0.5 s).
15. Focus onto the brightest and biggest plane of the patch cone. Acquire a single frame, then switch off the xenon light and set the shutter on external trigger mode.
16. Define the region of interest, the binning, frame rate, and file name in the camera software, set the camera to await external trigger (*see Note 15*).
17. Start the acquisition with the patch-clamp software (pClamp), which is set up to trigger the imaging software.

3.7 Data Analysis

1. Load the acquired data to a program for image analysis (e.g., Matlab, The Mathworks, Natick, MA).
2. Define more precisely the region of interest, i.e., the region around the most intense signal of the patch cone (Fig. 2b). This can be best done when using a temporal average of all recorded frames.
3. Take the spatial average of the region of interest for each frame and plot the data as a function of time and stimulus (Fig. 2c, d).

4 Notes

1. For a compromise between an acceptable signal-to-noise ratio and little fluorophore photobleaching one should use relatively low light intensities (around 100–400 $\mu\text{W}/\text{mm}^2$).
2. It is also possible to use a photomultiplier tube (PMT) for the acquisition of fluorescence data. The advantage is the higher bandwidth of up to a few kHz, i.e., faster signals can be recorded as compared to current CCDs (up to 0.5 kHz). However, it can difficult to exclude background fluorescence that can come from debris at the outside of the patch pipette (*see Fig. 2a*).
3. RNA is susceptible to degradation by the ubiquitous RNases. Make sure to always wear gloves when handling RNA, use

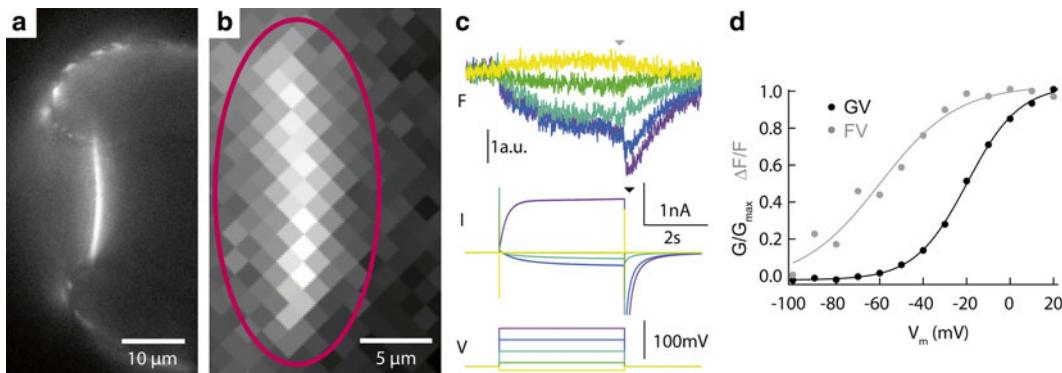


Fig. 2 (a) Excised inside-out patch of a membrane expressing ciHv1-L245C labeled with MTS-TAMRA and illuminated with 550 nm wavelength and acquired with long exposure time (0.3 s, no binning). The membrane is strongly labeled and clearly visible. Some spots of fluorescence at the outer glass rim are also visible. This background should be excluded from analysis. (b) Patch cone acquired at 8×8 binning, pink ellipse highlights the approximate region of interest chosen offline for subsequent analysis (spatial averaging). (c) Fluorescent signal (spatial average, top traces) in response to different voltage steps (bottom traces) in matching color code. Current responses are in the middle. For clarity, only five different stimuli are shown. (d) Conductance-Voltage (GV) and steady-state Fluorescence-Voltage (FV) relations. Panels a and b are pictures from the same patch, and data from panels c and d are from another patch

nuclease-free pipette tips and clean the bench and gloves with a decontaminating solution (e.g., RNase Away, Molecular BioProducts, San Diego, CA).

4. Injection pipettes with a small tip make it easier to penetrate and are gentler to the oocyte. However, when highly concentrated RNA solution is used, the actual injection volume might be inaccurate and slightly larger tips should be used.
5. Blocking native cysteines can increase the signal-to-noise ratio in case the expression level of the heterologously expressed ion channel is low or the change of fluorescence is small.
6. Alternatively, oocytes can be incubated at 12 °C for several days, which allows translation of the protein, and the night before recording oocytes can be stored at 20 °C or room temperature, which allows protein trafficking to the plasma membrane.
7. Oocytes should have uniform white and dark poles. The dark pole can be black or brown.
8. The labeling is done on ice to increase specificity of the labeling process and prevent endocytosis. Optimal incubation time depends on the reactivity of the probe and the accessibility of the free sulphhydryl group of the cysteine. Some oocytes (~10 %), depending on the incubation time, do accumulate the fluorophore intracellularly, possibly via endocytosis. This is easily seen by coloration of the oocytes. These oocytes should be discarded.

9. It might take other modifications of the puller program to get satisfying patch pipettes. Several adjustments might be necessary to get appropriate and reproducible pipettes of the right size which can be laborious. For more information, refer to the puller's manual. See also http://www.sutter.com/contact/faqs/pipette_cookbook.pdf for further advice.
10. The polishing with the bead is absolutely crucial for good and relatively fast sealing. A bead is good for 20–50 patch pipettes and should be redone thereafter. Multiple beads can be placed next to each other in order to save platinum wire.
11. *Xenopus* oocytes tend to have higher expression levels on the animal (dark) pole compared to the vegetal pole, so the side with the best expression level can be chosen easily.
12. The sealing process can take a long time and takes a lot of practice. However, if no gigaseal is obtained within 5 min, the patch pipette should be replaced and a new trial started. Ideally, the gigaseal should be above 10 GΩ which seems to be necessary for giant patch stability.
13. In some cases, especially when the sealing process required long duration or high levels of negative pressure, the patched membrane invaginates into the patch pipette and intracellular components accumulate between the very tip of the pipette and the patch cone and get co-excised. This can decrease the quality of electrical and fluorescence signal.
14. Apply the immersion oil onto the 60× objective just prior to positioning it into the optical axis of the microscope. It can be tricky to reach the objective and place an appropriate amount of oil on the objective, which is right under the recording chamber. It can be helpful to put a longer and thinner adapter tip on the standard oil flask for a more controlled oil dispensing.
15. Fast CCD cameras can go up to 200 frames/s (with sufficient net exposure times per frame) when a small region of interest and a binning of 4×4 pixels are chosen, a binning of 8×8 might allow for 500 frames/s. Note that most cameras have differences in vertical and horizontal read-out speeds, i.e., the orientation of a rectangular region of interest can matter.

Acknowledgement

This work was supported by NIH grant R01 NS35549 (E.Y.I.), and a postdoctoral fellowship of the Swiss National Science Foundation (SNSF, PA00P3_134163) (T.K.B.).

References

1. Mannuzzu LM, Moronne MM, Isacoff EY (1996) Direct physical measure of conformational rearrangement underlying potassium channel gating. *Science* 271:213–216
2. Zheng J, Zagotta WN (2000) Gating rearrangements in cyclic nucleotide-gated channels revealed by patch-clamp fluorometry. *Neuron* 28:369–374
3. Asamoah OK, Wuskell JP, Loew LM, Bezanilla F (2003) A fluorometric approach to local electric field measurements in a voltage-gated ion channel. *Neuron* 37:85–97
4. Cohen BE, Pralle A, Yao X, Swaminath G, Gandhi CS, Jan YN et al (2005) A fluorescent probe designed for studying protein conformational change. *Proc Natl Acad Sci U S A* 102:965–970
5. Glauner KS, Mannuzzu LM, Gandhi CS, Isacoff EY (1999) Spectroscopic mapping of voltage sensor movement in the Shaker potassium channel. *Nature* 402:813–817
6. Cha A, Snyder GE, Selvin PR, Bezanilla F (1999) Atomic scale movement of the voltage-sensing region in a potassium channel measured via spectroscopy. *Nature* 402:809–813
7. Loots E, Isacoff EY (2000) Molecular coupling of S4 to a K(+) channel's slow inactivation gate. *J Gen Physiol* 116:623–636
8. Gandhi CS, Clark E, Loots E, Pralle A, Isacoff EY (2000) The orientation and molecular movement of a k(+) channel voltage-sensing domain. *Neuron* 40:515–525
9. Pathak MM, Yarov-Yarovoy V, Agarwal G, Roux B, Barth P, Kohout S et al (2007) Closing in on the resting state of the Shaker K(+) channel. *Neuron* 56:124–140
10. Sonnleitner A, Mannuzzu LM, Terakawa S, Isacoff EY (2002) Structural rearrangements in single ion channels detected optically in living cells. *Proc Natl Acad Sci U S A* 99: 12759–12764
11. Blunck R, McGuire H, Hyde HC, Bezanilla F (2008) Fluorescence detection of the movement of single KcsA subunits reveals cooperativity. *Proc Natl Acad Sci U S A* 105:20263–20268
12. Lakowicz J (2006) Principles of fluorescence spectroscopy. Springer, New York
13. Loots E, Isacoff EY (1998) Protein rearrangements underlying slow inactivation of the Shaker K+ channel. *J Gen Physiol* 112:377–389
14. Gandhi CS, Loots E, Isacoff EY (2000) Reconstructing voltage sensor-pore interaction from a fluorescence scan of a voltage-gated K+ channel. *Neuron* 27:585–595
15. Cha A, Bezanilla F (1997) Characterizing voltage-dependent conformational changes in the Shaker K+ channel with fluorescence. *Neuron* 19:1127–1140
16. Mannuzzu LM, Isacoff EY (2000) Independence and cooperativity in rearrangements of a potassium channel voltage sensor revealed by single subunit fluorescence. *J Gen Physiol* 115:257–268
17. Pathak M, Kurtz L, Tombola F, Isacoff E (2005) The cooperative voltage sensor motion that gates a potassium channel. *J Gen Physiol* 125:57–69
18. Tombola F, Ulbrich MH, Kohout SC, Isacoff EY (2010) The opening of the two pores of the Hv1 voltage-gated proton channel is tuned by cooperativity. *Nat Struct Mol Biol* 17:44–50
19. Baker OS, Larsson HP, Mannuzzu LM, Isacoff EY (1998) Three transmembrane conformations and sequence-dependent displacement of the S4 domain in Shaker K+ channel gating. *Neuron* 20:1283–1294
20. Schönher R, Mannuzzu LM, Isacoff EY, Heinemann SH (2002) Conformational switch between slow and fast gating modes: allosteric regulation of voltage sensor mobility in the EAG K+ channel. *Neuron* 35:935–949
21. Bruening-Wright A, Elinder F, Larsson HP (2007) Kinetic relationship between the voltage sensor and the activation gate in spHCN channels. *J Gen Physiol* 130:71–81
22. Chanda B, Bezanilla F (2002) Tracking voltage-dependent conformational changes in skeletal muscle sodium channel during activation. *J Gen Physiol* 120:629–645
23. Gonzalez C, Koch HP, Drum BM, Larsson HP (2010) Strong cooperativity between subunits in voltage-gated proton channels. *Nat Struct Mol Biol* 17:51–56
24. Kohout SC, Ulbrich MH, Bell SC, Isacoff EY (2008) Subunit organization and functional transitions in Ci-VSP. *Nat Struct Mol Biol* 15:106–108
25. Villalba-Galea CA, Sandtner W, Starace DM, Bezanilla F (2008) S4-based voltage sensors have three major conformations. *Proc Natl Acad Sci U S A* 105:17600–17607
26. Chang Y, Weiss DS (2002) Site-specific fluorescence reveals distinct structural changes with GABA receptor activation and antagonism. *Nat Neurosci* 5:1163–1168
27. Yao X, Parnot C, Deupi X, Ratnala VRP, Swaminath G, Farrens D et al (2006) Coupling ligand structure to specific conformational switches in the beta2-adrenoceptor. *Nat Chem Biol* 2:417–422
28. Larsson HP, Tzingounis AV, Koch HP, Kavanaugh MP (2004) Fluorometric measure-

- ments of conformational changes in glutamate transporters. *Proc Natl Acad Sci U S A* 101:3951–3956
29. Dempski RE, Friedrich T, Bamberg E (2009) Voltage clamp fluorometry: combining fluorescence and electrophysiological methods to examine the structure–function of the Na(+) / K(+)-ATPase. *Biochim Biophys Acta* 1787:714–720
30. Blunck R, Starace DM, Correa AM, Bezanilla F (2004) Detecting rearrangements of shaker and NaChBac in real-time with fluorescence spectroscopy in patch-clamped mammalian cells. *Biophys J* 86:3966–3980

Chapter 7

Fluorescent Labeling of SNAP-Tagged Proteins in Cells

Gražvydas Lukinavičius, Luc Reymond, and Kai Johnsson

Abstract

One of the most prominent self-labeling tags is SNAP-tag. It is an in vitro evolution product of the human DNA repair protein O^6 -alkylguanine-DNA alkyltransferase (hAGT) that reacts specifically with benzylguanine (BG) and benzylchloropyrimidine (CP) derivatives, leading to covalent labeling of SNAP-tag with a synthetic probe (Gronemeyer et al., Protein Eng Des Sel 19:309–316, 2006; Curr Opin Biotechnol 16:453–458, 2005; Keppler et al., Nat Biotechnol 21:86–89, 2003; Proc Natl Acad Sci U S A 101:9955–9959, 2004). SNAP-tag is well suited for the analysis and quantification of fused target protein using fluorescence microscopy techniques. It provides a simple, robust, and versatile approach to the imaging of fusion proteins under a wide range of experimental conditions.

Key words Snap-tag, Synthetic fluorophores, Living and fixed cells, Covalent labeling, Self-labeling tags, Fluorescence microscopy, Episomal protein expression

1 Introduction

A powerful approach for studying protein function inside living cells is their specific labeling with synthetic fluorescent reporter groups [1, 2]. One of the most frequently used methodologies is based on so-called self-labeling tags. A popular example of such a tag is SNAP-tag, an in vitro evolution product of the human DNA repair protein O^6 -alkylguanine-DNA alkyltransferase (hAGT) that reacts specifically and rapidly with benzylguanine (BG) and benzylchloropyrimidine (CP) derivatives (Fig. 1a), leading to covalent labeling of SNAP-tag with a synthetic probe [3–9]. Proteins of interest (POI) fused to SNAP-tag react only once with a single substrate molecule generating fluorescently tagged fusion proteins (Fig. 1b). The nontoxic nature of BG and CP derivatives together with the molecular specificity of SNAP-tag makes it suitable for a broad range of applications including *in vivo* imaging (Fig. 2). Examples of recent applications include single molecule [10, 11] and super-resolution imaging [12, 13], analysis of protein function [14], targeted protein inactivation [15], protein–protein interactions

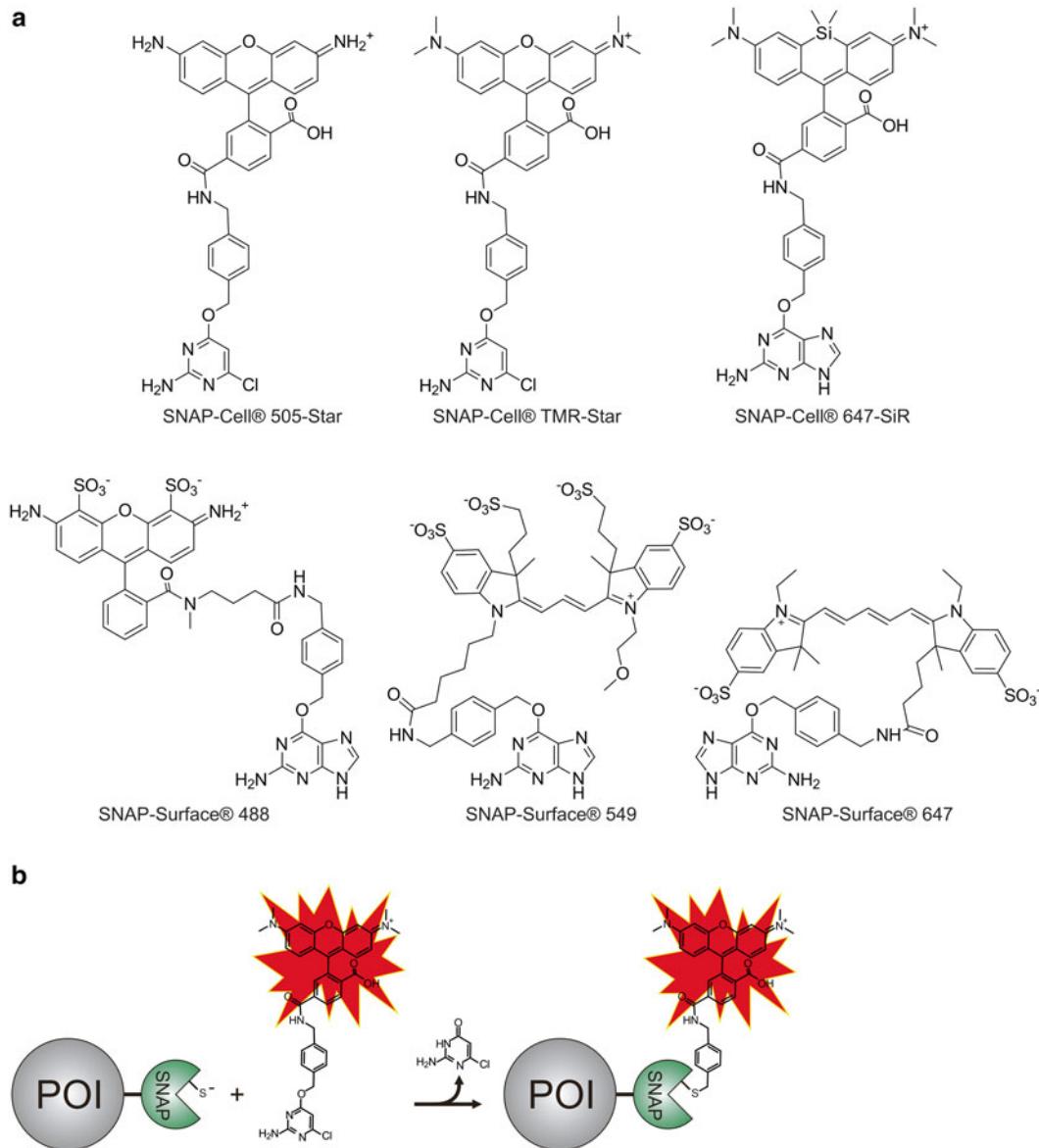


Fig. 1 SNAP-tag protein labeling technology. **(a)** Chemical structures of fluorescent substrates used for SNAP-tagged protein labeling: *top row*—examples of cell permeable substrates, *bottom row*—examples of cell nonpermeable substrates. **(b)** Scheme showing principle of protein of interest (POI) labeling via SNAP-tag

[16, 17], protein–drug interactions [18, 19], and the determination of protein half-life in animals [20]. Additionally, similar hAGT-based tag, named CLIP-tag, was developed recently [21]. It reacts specifically with O^2 -benzylcytosine derivatives. SNAP-tag and CLIP-tag possess orthogonal substrate specificities, SNAP and CLIP fusion proteins can be labeled simultaneously and specifically

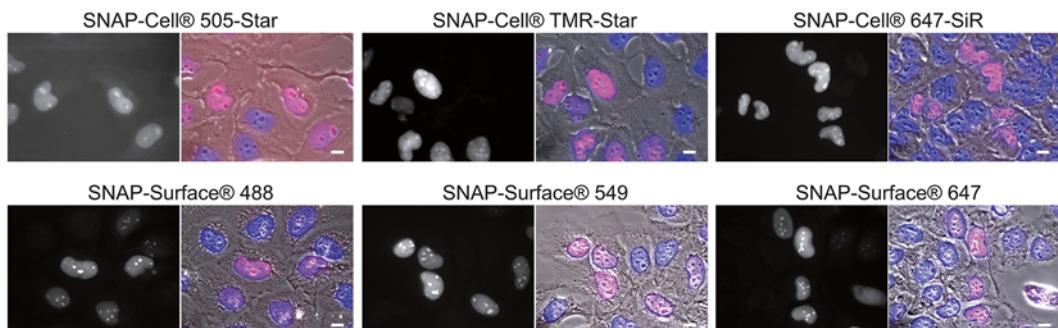


Fig. 2 Staining performance of SNAP-tag substrates. Live U2OS (*top row*) or fixed cells (*bottom row*) expressing nuclear localized SNAP-tag construct were stained with the substrates and Hoechst 33342. In the overlay image Hoechst 33342 is presented in *blue*, SNAP-tag substrate *red*, and transmission phase contrast image in *gray*. Panel to the *left* of overlay image represents SNAP-tag substrate image in *gray*. Scale bar 10 μm (color figure online)

with different molecular probes in living cells. Examples of such application include simultaneous labeling of two different fusion proteins [21], selective cross-linking (S-CROSS) of interacting proteins [16, 17, 19] and simultaneous measurement of protein SUMOylation at the single-molecule level [22].

2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a resistance of 18 M Ω cm at 25 °C) and analytical-grade reagents. All water solutions are filtered through 0.22 μm pore size membrane syringe filter directly after preparation.

1. U2OS cells from ATCC (HTB-96™). It is recommended to prepare multiple frozen stocks after obtaining cells. Details on cell line cultivation and cryopreservation are available on ATCC web site.
2. Growth medium: high-glucose DMEM without glutamine, pyruvate, and phenol red. Before usage, supplement with 10 % fetal bovine serum, 2 mM GlutaMAX™-I and 1 mM sodium pyruvate (*see Note 1*).
3. Trypsin/EDTA solution. Aliquots and store at -20 °C.
4. Opti-MEM® I reduced serum medium. Store at +4 °C.
5. Lipofectamine 2000. Store at +4 °C. Freezing might change performance.
6. 1 mg/ml Puromycin stock solution (1,000×). Aliquots and store at -20 °C.
7. 0.1 mg/ml Doxycycline stock solution (1,000×). Aliquots and store at -20 °C.

8. 1 mg/ml Hoechst 33342 stock solution (1,000×). Aliquots and store at -20 °C.
9. PBS buffer without Ca²⁺, Mg²⁺, or phenol red. Store in the dark at room temperature.
10. Hanks balanced salt solution (HBSS) without phenol red. Store in the dark at room temperature.
11. Lyophilized powder of albumin from bovine serum (BSA). Store at +4 °C.
12. EGTA, molecular biology grade. Store at room temperature.
13. PIPES, molecular biology grade. Store at room temperature.
14. BRB80 extraction buffer: 80 mM K-PIPES, pH 6.8, 1 mM MgCl₂; 1 mM EGTA, 0.2 % IGEPAL-630. Prepare freshly before experiment from the stock solutions of 0.5 M K-PIPES, pH 6.8, 1 M MgCl₂, 0.5 M EGTA, and IGEPAL-630 (*see Notes 2 and 3*).
15. Methanol, analytical reagent grade. Store at -20 °C (*see Note 4*).
16. 1 % w/v BSA solution in PBS. Store at +4 °C.
17. PBS-T wash buffer: PBS buffer supplemented with 0.05 % TX-100.
18. Staining buffer: PBS buffer supplemented with 1 % w/v BSA. DMSO solution of the substrate is added just before labeling procedure at concentration indicated in protocol part.
19. SNAP-tag substrates: SNAP-Cell® 505-Star (New England Biolabs), SNAP-Cell® TMR-Star (New England Biolabs), SNAP-Surface® 488 (New England Biolabs), SNAP-Surface® 549 (New England Biolabs), and SNAP-Surface® 647 (New England Biolabs). SNAP-Cell® 647-SiR (New England Biolabs) substrate synthesis is described in [23]. SNAP-tag substrates are dissolved in dry DMSO at final concentration of 1 mM and stored at -20 °C.
20. Mounting media: 90 ml of glycerol mixed with 10 ml of 10× PBS and dissolved 2–4 g of propyl gallate (*see Note 5*).

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Generation of Cell Lines Expressing SNAP-Tagged Protein

1. U2OS cells are used for generation of cell lines expressing SNAP-tagged proteins of interest. Cells are cultured in growth medium in a humidified 5 % CO₂ incubator at 37 °C. 10 ml of medium is used for 25 cm² dish (*see Note 6*).
2. Prepare U2OS cells for transfection by splitting the cultured cells 24 h before transfection. Remove growth medium

from confluent monolayer of cells in 25 cm² dish. Wash cells with 5–10 ml of PBS buffer and add 1 ml of trypsin/EDTA solution. Incubate for 5 min at 37 °C and suspend detached cells in 11 ml of growth medium. Prepare six-well plate containing 2 ml of fresh growth medium in each well. Add 1 ml of cell suspension to each well and incubate overnight (*see Note 7*).

3. Next day dissolve 3–4 µl of Lipofectamine 2000 and 2–4 µg of DNA (pEBTet plasmid) in 100 µl Opti-MEM I separately. Incubate for 5 min at room temperature. Prepare Lipofectamine 2000 and DNA complex by mixing both components and incubating for 15 min at room temperature (*see Notes 8 and 9*).
4. Prepare U2OS cells for transfection by replacing growth medium with 1 ml of Opti-MEM I medium. Add prepared Lipofectamine 2000 and DNA complex solution and incubate for 6 h in a humidified 5 % CO₂ incubator at 37 °C.
5. After 6 h replace Opti-MEM I medium with growth medium and incubate for additional 24–48 h in a humidified 5 % CO₂ incubator at 37 °C (*see Note 10*).
6. Episomal plasmid pEBTet contains gene which renders transfected U2OS cells resistant to puromycin. Select for these cells by replacing growth medium with growth medium containing 1 µg/ml of puromycin. Selective medium has to be replaced each 2 days for duration of 4–6 days (*see Note 11*).
7. Wash cells with 2 ml of PBS buffer and add 0.4 ml of trypsin/EDTA solution. Incubate for 5 min at 37 °C, suspend detached cells, and transfer suspension into 25 cm² dish with 10 ml of growth medium and 1 µg/ml puromycin. Selected cells can be passaged every 4–5 days (1:10 dilution) for about 2 months without major loss of transgene expression level (*see Note 12*).

3.2 Labeling of SNAP-Tagged Proteins in Living Cells

1. Prepare U2OS cells for microscopy by splitting the cultured cells 24–48 h before experiment. Remove growth medium from confluent monolayer of cells in 25 cm² dish. Wash cells with 5–10 ml of PBS buffer and add 1 ml of trypsin/EDTA solution. Incubate for 5 min at 37 °C and suspend detached cells in 11 ml of growth medium (*see Notes 13 and 14*).
2. Prepare glass bottom six-well plate containing 2 ml of fresh growth medium supplemented with 1 µg/ml puromycin and 0.1 µg/ml doxycycline in each well. Add 0.5 ml of cell suspension to each well of plate. Incubate in a humidified 5 % CO₂ incubator at 37 °C for 24–48 h (*see Note 15*).
3. Stain cells with cell permeable substrates by replacing growth medium with growth medium containing 0.3–5 µM substrate. 1 µg/ml Hoechst 33342 can be included in the growth

medium together with substrate (*see Note 16*). Incubate cells for 1 h in a humidified 5 % CO₂ incubator at 37 °C (*see Note 17*). Afterwards wash cells two times with 1 ml of HBSS followed by 3–5 min incubation at room temperature. Replace HBSS after last wash with growth medium and incubate for additional 1 h in a humidified 5 % CO₂ incubator at 37 °C. Samples are ready for living cell imaging after this step.

3.3 Labeling of SNAP-Tagged Proteins in Fixed Cells

1. Prepare cells for fixation, remove the growth medium, and add precooled to -20 °C methanol and incubate for 3–10 min at -20 °C in freezer (*see Notes 18 and 19*). Take six-well plate from freezer and wash two times cells with 2 ml PBS buffer.
2. Incubate for 60 min in 2 ml of 1 % BSA in PBS solution (*see Note 20*). Remove BSA solution and stain DNA by incubating with 1 ml of 1 µg/ml Hoechst 33342 PBS solution for 1 min at room temperature. Wash excess of dye three times with 2 ml PBS-T wash buffer.
3. SNAP-tagged proteins can also be labeled after methanol fixation (*see Note 21*). Replace PBS with 1 ml of staining buffer containing cell not permeable SNAP-tag substrate (0.5–2 µM). Incubate for 1 h at room temperature (*see Note 22*). Wash excess of dye 2–3 times (incubating 3–5 min each time) with 2 ml PBS-T wash buffer.
4. This step is optional and can be performed if additional antibody-based staining is needed (*see Note 23*). Remove PBS-T wash buffer and put 0.5 ml of primary antibodies diluted in PBS with 1 % BSA. Incubate samples overnight at 4 °C. Wash excess of antibody three times 3–5 min with 1 ml PBS-T wash buffer and add dilutions of secondary antibodies in PBS with 1 % BSA. Incubate for 4–6 h at room temperature. Wash excess of antibody 3–5 min times with 1 ml PBS-T wash buffer. Samples are ready for imaging after last wash (*see Note 24*).

3.4 Labeling of SNAP-Tagged Proteins for Stimulated Emission Depletion (STED) Microscopy

Stimulated emission depletion (STED) microscopy [24] becomes more and more popular among the biologists who want to investigate cellular processes beyond the diffraction limit. Recent developments of fluorophores [17, 23, 25], parallelization of acquisition [26], time-resolved detection [27], and multicolor imaging [28, 29] pave the way to myriads of biological applications. Sample preparation for the STED microscopy is the same as described above (*see Subheadings 3.2 and 3.3*). However, only rhodamine-class fluorophores are recommended to be used. They display good enough photostability and brightness to be used for STED imaging (*see Note 25*). Image comparison of SNAP-tagged centrosomal proteins obtained with microscope operating in confocal or STED mode is provided in Fig. 3.

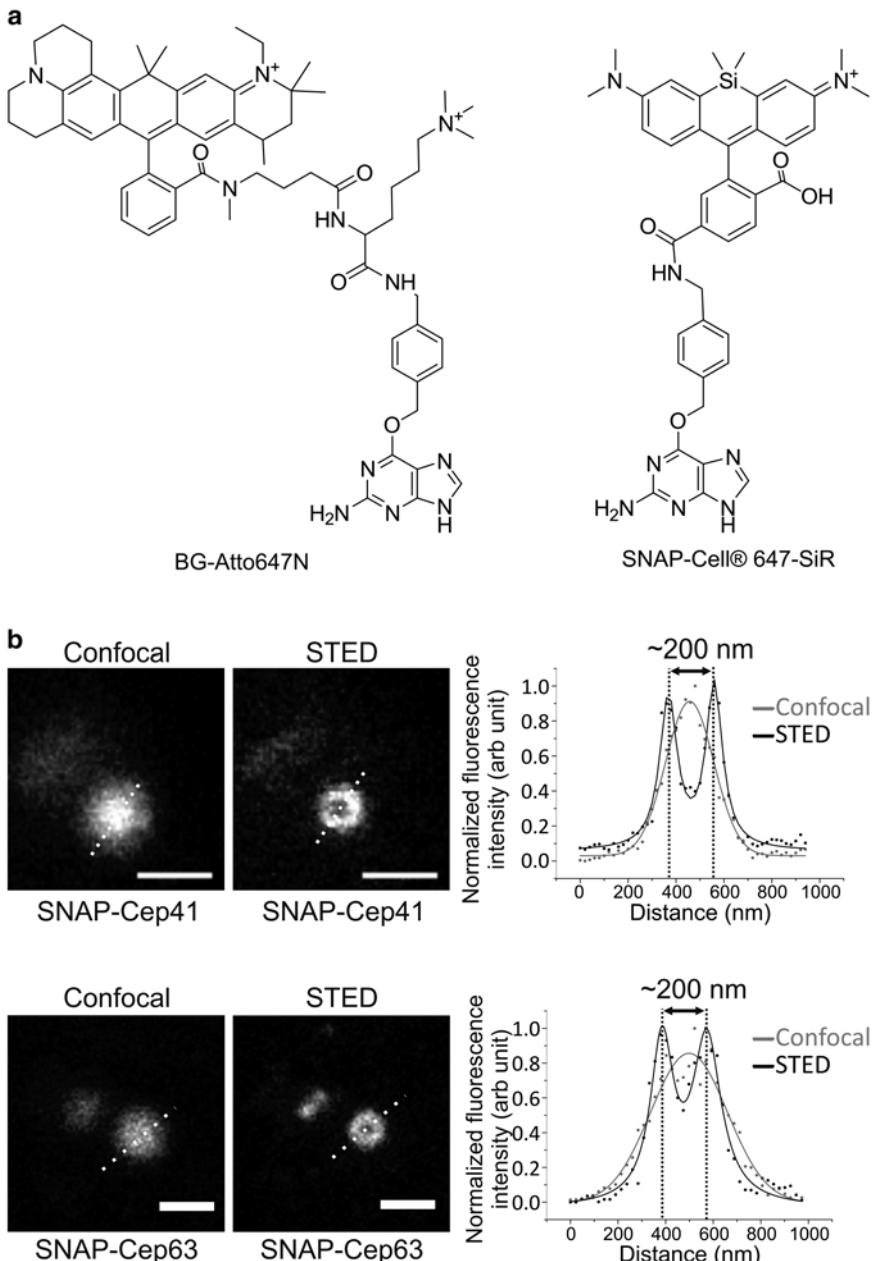


Fig. 3 STED imaging of SNAP-tagged proteins. **(a)** Chemical structures of STED compatible fluorescent substrates used for SNAP-tagged protein labeling: BG-Atto647N—example of cell nonpermeable substrate, SNAP-Cell® 647-SiR—example of cell permeable substrate. **(b)** Comparison of confocal and STED images of SNAP-tagged centrosomal proteins. SNAP-Cep41 expressing cells stained with SNAP-Cell® 647-SiR and imaged without fixing. SNAP-Cep63 expressing cells stained with BG-Atto647N after fixation. *Left panel* shows obtained images and profile line, *right panel* shows intensity profile of the line. Note, that improvement in resolutions leads to detection of doughnut shaped structure

4 Notes

1. Phenol red is interfering with imaging of living cells and cells should be cultivated in the medium without it. Alternatively, cells can be propagated in growth medium containing phenol red, but this medium has to be replaced by fresh growth medium just before imaging experiment.
2. 0.5 M K-PIPES stock: 15.1 g of piperazine-*N,N'*-bis(2-ethanesulfonic acid) in 90 ml of deionized water. Adjust pH to 6.8 with 10 M potassium hydroxide. Long standing K-PIPES solution tends to develop yellow-brown color, but it is not interfering with quality of obtained results.
3. 0.5 M EGTA solution: 19.0 g of ethylene glycol-bis(2-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid in 90 ml of deionized water. Adjust pH to 7.0 using 10 M sodium hydroxide. EGTA displays low solubility in water as free acid and dissolves completely when pH is close to 7.0.
4. Cold methanol can be supplemented with 5 mM EGTA pH 7.0 solution in order to improve cytoskeleton structure preservation during fixation.
5. Dissolving propyl gallate in 90 % glycerol solution might take long time. It is recommended to sonicate solution or place tube in +50 °C water bath and periodically mix it.
6. It is recommended to use U2OS cell line for generation of SNAP-tagged protein expressing cells since there is considerable variation of expression level in-between difference cell lines [30]. For example, HEK 293 displays higher expression levels and HeLa displays lower expression levels compared to U2OS.
7. Added volume of trypsinated cells is adjusted so that ~20 % confluence is obtained if all cells are adhered to the surface of new dish.
8. Transfection efficiency is dependent on many factors. It is recommended to determine the best transfection conditions before performing this experiment.
9. Generation of episomal expression vector pEBTet encoding SNAP-tagged protein of interest (POI). Expression vector is generated by LR recombination via attL and attR sites which is a part of Gateway® Cloning Technology available commercially from Life Technologies (Fig. 4). It contains Epstein–Barr virus origin of replication (oriP), which is capable of continuous episomal propagation in the mammalian cell lines in the presence of plasmid-encoded EBNA-1 protein. Cells without episome could be easily eliminated by selection with puromycin. Protein expression is driven from inducible CMV+TetO2

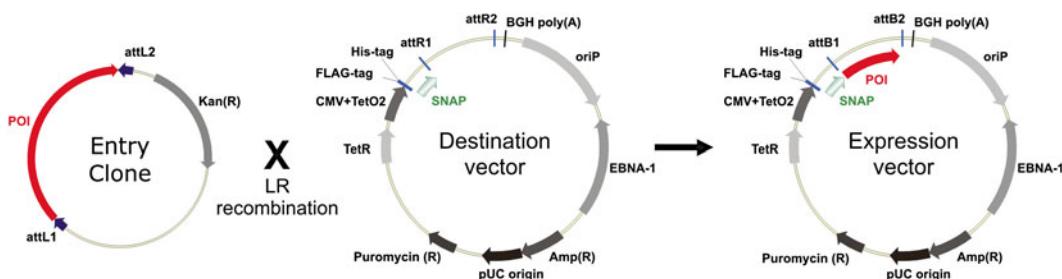


Fig. 4 Construction of the fused proteins expressing plasmids using Gateway™ cloning system

promoter [31]. Additionally, destination/expression vector can be propagated in *E. coli* cells since it contains pUC origin of replication and antibiotic resistance genes. Construction of destination vector is described in [17]. Entry clone plasmids are commercially available from Life technologies or GeneCopoeia.

10. Prolonged exposure to Opti-MEM I and Lipofectamine 2000 complexes result in considerable death of the cells. U2OS cell line is very well tolerating up to 6 h exposure, but it cannot be incubated overnight.
11. Puromycin is an aminonucleoside antibiotic, which causes premature chain termination during translation. Resistance gene encodes puromycin *N*-acetyltransferase which inactivates cytotoxic puromycin by acetylating it [32].
12. Expression of transgene is not stable due to silencing by various mechanisms. It is recommended not to split cells more than 10–15 times during the experiment.
13. For obtaining good quality microscopy images it is necessary that cell culture is not reaching complete confluence at the day of imaging. It is recommended to start with ~10–20 % of confluence 48 h before the imaging experiment.
14. Do not consume all the cells when setting up of an imaging experiment. The remaining cell suspension can be used for cell line propagation. This allows replication of experimental data.
15. Episomal pEBTet expression vector contains modified CMV promoter which is triggered “ON” in the presence of tetracycline or doxycycline in the medium. Keep tetracycline or doxycycline concentrations as low as possible since they display certain toxicity [33].
16. Hoechst 33342 counterstaining is recommended to include in most of the imaging experiments since it produces strong fluorescence signal, which is convenient for finding the cells and focusing.

17. It is recommended to use low concentrations (0.3–0.5 μM) of fluorescent substrates. High concentrations will result in higher background staining which is difficult to eliminate even with extensive washing. Fluorogenic SNAP-Cell® 647-SiR substrate is an exception and can be applied at concentrations up to 5 μM without need of extensive washing. Properties of cell permeable substrates are listed in Table 1.
18. SNAP-tag is highly stable and retains its activity even after fixation with cold methanol or paraformaldehyde. Fixation by cold methanol procedure is described here in more details, but paraformaldehyde or other fixatives can be used depending on which cell structure has to be preserved [35].
19. For the visualization of cell skeleton it is recommended to include preextraction step with BRB80 extraction buffer before applying cold methanol [12, 17]. It is done by replacing growth medium with 2 ml of BRB extraction buffer for 0.5 min at room temperature. Cold methanol is applied directly after this step.
20. Blocking with 1 % BSA in PBS reduces background staining in all the following steps. Hoechst 33342 staining can be combined together with incubation with 1 % of BSA in PBS.
21. SNAP-tag can be labeled with cell permeable substrates before fixation and covalent labeling survives fixation procedure very well. In general, labeling before fixation gives stronger specific signal compared to postfixation labeling [17].
22. It is recommended to use low concentrations (0.5–2 μM) of fluorescent substrates. All cell impermeable substrates are highly charged molecules with low off-target binding, but prolonged exposure to high concentrations of them will result in higher background staining which elimination requires extensive washing. Properties of cell impermeable substrates are listed in Table 2.

Table 1
Properties of common cell permeable SNAP-tag substrates

Substrate name	Excitation (nm)	Emission (nm)	QY ^a	$\epsilon \cdot 10^4$ ($\text{cm}^{-1} \text{M}^{-1}$)	Rate constant $\cdot 10^3$ ($\text{M}^{-1} \text{s}^{-1}$)	Ref.
SNAP-Cell® 505-Star	504	532	0.92	7.4	12.4	[34]
SNAP-Cell® TMR-Star	554	580	0.68	9.1	15.5	[34]
SNAP-Cell® 647-SiR	650	670	0.39	10.0	20.0	[23]

^aOf the unconjugated dye

Table 2
Properties of common cell impermeable SNAP-tag substrates

Substrate name	Excitation (nm)	Emission (nm)	QY ^a	$\epsilon \cdot 10^4$ (cm ⁻¹ M ⁻¹)	Rate constant 10 ³ (M ⁻¹ s ⁻¹)	Ref.
SNAP-Surface® 488	506	526	0.80	9.0	12.1	[36]
SNAP-Surface® 549	560	575	n.d.	15	11.1	[36]
SNAP-Surface® 647	660	673	0.25	25	n.d.	[8]
BG-Atto647N	644	669	0.65	15	2.9	[17]

^aOf the unconjugated dye

- 23. SNAP-tag labeling is not interfering with antibody staining. Both techniques can be combined to obtain multicolor images [12, 17].
- 24. Fixed samples do not require special mounting media for wide field microscopy. Suggested fluorophores are stable enough in simple PBS solution. Samples prepared for confocal or STED microscopy should be mounted in 90 % glycerol in PBS containing 2–4 % w/v of propyl gallate.
- 25. SNAP-Surface® 549 and SNAP-Surface® 647 substrates are derivatives of cyanines and bleach extremely fast under STED imaging conditions.

References

- Giepmans BN, Adams SR, Ellisman MH, Tsien RY (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312:217–224
- van de Linde S, Heilemann M, Sauer M (2012) Live-cell super-resolution imaging with synthetic fluorophores. *Annu Rev Phys Chem* 63:519–540
- Gronemeyer T, Chidley C, Juillerat A, Heinis C, Johnsson K (2006) Directed evolution of O⁶-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng Des Sel* 19:309–316
- Gronemeyer T, Godin G, Johnsson K (2005) Adding value to fusion proteins through covalent labelling. *Curr Opin Biotechnol* 16:453–458
- Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat Biotechnol* 21:86–89
- Keppler A, Pick H, Arrivoli C, Vogel H, Johnsson K (2004) Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci U S A* 101:9955–9959
- Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21:766–776
- Keppler A, Arrivoli C, Sironi L, Ellenberg J (2006) Fluorophores for live cell imaging of AGT fusion proteins across the visible spectrum. *Biotechniques* 41:167–170, 172, 174–175
- Reymond L, Lukinavicius G, Umezawa K, Maurel D, Brun MA, Masharina A, Bojkowska K, Mollwitz B, Schena A, Griss R, Johnsson K (2011) Visualizing biochemical activities in living cells through chemistry. *Chimia (Aarau)* 65:868–871
- Breitsprecher D, Jaiswal R, Bombardier JP, Gould CJ, Gelles J, Goode BL (2012) Rocket launcher mechanism of collaborative actin assembly defined by single-molecule imaging. *Science* 336:1164–1168
- Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ (2011) Ordered and dynamic assembly of single spliceosomes. *Science* 331:1289–1295
- Dellagiacoma C, Lukinavicius G, Bocchio N, Banala S, Geissbuhler S, Marki I, Johnsson K,

- Lasser T (2010) Targeted photoswitchable probe for nanoscopy of biological structures. *Chembiochem* 11:1361–1363
13. Jones SA, Shim SH, He J, Zhuang X (2011) Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods* 8:499–508
 14. Foraker AB, Camus SM, Evans TM, Majeed SR, Chen CY, Taner SB, Correa IR Jr, Doxsey SJ, Brodsky FM (2012) Clathrin promotes centrosome integrity in early mitosis through stabilization of centrosomal ch-TOG. *J Cell Biol* 198:591–605
 15. Keppler A, Ellenberg J (2009) Chromophore-assisted laser inactivation of alpha- and gamma-tubulin SNAP-tag fusion proteins inside living cells. *ACS Chem Biol* 4:127–138
 16. Gautier A, Nakata E, Lukinavicius G, Tan KT, Johnsson K (2009) Selective cross-linking of interacting proteins using self-labeling tags. *J Am Chem Soc* 131:17954–17962
 17. Lukinavicius G, Lavogina D, Orpinell M, Umezawa K, Reymond L, Garin N, Gonczy P, Johnsson K (2013) Selective chemical cross-linking reveals a Cep57-Cep63-Cep152 centrosomal complex. *Curr Biol* 23:265–270
 18. Chidley C, Haruki H, Pedersen MG, Muller E, Johnsson K (2011) A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat Chem Biol* 7:375–383
 19. Haruki H, Gonzalez MR, Johnsson K (2012) Exploiting ligand–protein conjugates to monitor ligand–receptor interactions. *PLoS One* 7:e37598
 20. Bojkowska K, Santoni de Sio F, Barde I, Offner S, Verp S, Heinis C, Johnsson K, Trono D (2011) Measuring *in vivo* protein half-life. *Chem Biol* 18:805–815
 21. Gautier A, Juillerat A, Heinis C, Correa IR Jr, Kindermann M, Beaufils F, Johnsson K (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* 15:128–136
 22. Yang Y, Zhang CY (2013) Simultaneous measurement of SUMOylation using SNAP/CLIP-tag-mediated translation at the single-molecule level. *Angew Chem Int Ed Engl* 52:691–694
 23. Lukinavicius G, Umezawa K, Olivier N, Honigmann A, Yang G, Plass T, Mueller V, Reymond L, Correa IR Jr, Luo ZG, Schultz C, Lemke EA, Heppenstall P, Eggeling C, Manley S, Johnsson K (2013) A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat Chem* 5:132–139
 24. Hell SW, Wichmann J (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt Lett* 19:780–782
 25. Kolmakov K, Wurm CA, Hennig R, Rapp E, Jakobs S, Belov VN, Hell SW (2012) Red-emitting rhodamines with hydroxylated, sulfonated, and phosphorylated dye residues and their use in fluorescence nanoscopy. *Chemistry* 18:12986–12998
 26. Chmyrov A, Keller J, Grotjohann T, Ratz M, d'Este E, Jakobs S, Eggeling C, Hell SW (2013) Nanoscopy with more than 100,000 ‘doughnuts’. *Nat Methods* 10:737–740
 27. Vicidomini G, Schonle A, Ta H, Han KY, Moneron G, Eggeling C, Hell SW (2013) STED nanoscopy with time-gated detection: theoretical and experimental aspects. *PLoS One* 8:e54421
 28. Gottfert F, Wurm CA, Mueller V, Berning S, Cordes VC, Honigmann A, Hell SW (2013) Coaligned dual-channel STED nanoscopy and molecular diffusion analysis at 20 nm resolution. *Biophys J* 105:L01–L03
 29. Pellett PA, Sun X, Gould TJ, Rothman JE, Xu MQ, Correa IR Jr, Bewersdorf J (2011) Two-color STED microscopy in living cells. *Biomed Opt Express* 2:2364–2371
 30. Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren BZ, Lahn BT (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* 5:e10611
 31. Bach M, Grigat S, Pawlik B, Fork C, Utermohlen O, Pal S, Banczyk D, Lazar A, Schomig E, Grundemann D (2007) Fast set-up of doxycycline-inducible protein expression in human cell lines with a single plasmid based on Epstein–Barr virus replication and the simple tetracycline repressor. *FEBS J* 274:783–790
 32. Vara J, Perez-Gonzalez JA, Jimenez A (1985) Biosynthesis of puromycin by *Streptomyces alboniger*: characterization of puromycin N-acetyltransferase. *Biochemistry* 24:8074–8081
 33. Sekeroglu ZA, Afan F, Sekeroglu V (2012) Genotoxic and cytotoxic effects of doxycycline in cultured human peripheral blood lymphocytes. *Drug Chem Toxicol* 35:334–340
 34. Correa IR, Baker B, Zhang A, Sun L, Provost CR, Lukinavicius G, Reymond L, Johnsson K, Xu MQ (2013) Substrates for improved live-cell fluorescence labeling of SNAP-tag. *Curr Pharm Des* 19:5414–5420
 35. Luther PW, Bloch RJ (1989) Formaldehyde-amine fixatives for immunocytochemistry of cultured *Xenopus* myocytes. *J Histochem Cytochem* 37:75–82
 36. Sun X, Zhang A, Baker B, Sun L, Howard A, Buswell J, Maurel D, Masharina A, Johnsson K, Noren CJ, Xu MQ, Correa IR Jr (2011) Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. *Chembiochem* 12:2217–2226

Chapter 8

HaloTag Technology for Specific and Covalent Labeling of Fusion Proteins

Hélène A. Benink and Marjeta Urh

Abstract

Appending proteins of interest to fluorescent protein tags such as GFP has revolutionized how proteins are studied in the cellular environment. Over the last few decades many varieties of fluorescent proteins have been generated, each bringing new capability to research. However, taking full advantage of standard fluorescent proteins with advanced and differential features requires significant effort on the part of the researcher. This approach necessitates that many genetic fusions be generated and confirmed to function properly in cells with the same protein of interest. To lessen this burden, a newer category of protein fusion tags termed “self-labeling protein tags” has been developed. This approach utilizes a single protein tag, the function of which can be altered by attaching various chemical moieties (fluorescent labels, affinity handles, etc.). In this way a single genetically encoded protein fusion can easily be given functional diversity and adaptability as supplied by synthetic chemistry. Here we present protein labeling methods using HaloTag technology; comprised of HaloTag protein and the collection of small molecules designed to bind it specifically and provide it with varied functionalities. For imaging purposes these small molecules, termed HaloTag ligands, contain distinct fluorophores. Due to covalent and rapid binding between HaloTag protein and its ligands, labeling is permanent and efficient. Many of these ligands have been optimized for permeability across cellular membranes allowing for live cell labeling and imaging analysis. Nonpermeable ligands have also been developed for specific labeling of surface proteins. Overall, HaloTag is a versatile technology that empowers the end user to label a protein of interest with the choice of different fluorophores while alleviating the need for generation of multiple genetic fusions.

Key words Protein labeling, Self-labeling proteins, HaloTag technology, HaloTag Ligands, Live cell imaging, Fluorescent dyes

1 Introduction

Modifying proteins of interest with protein fusion tags in order to equip proteins with additional functionalities has been an invaluable technique for analysis of proteins. Fluorescent protein tags such as GFP have had a profound impact on how protein function is studied in living cells [1–3]. More recently efforts have been made towards development of self-labeling tags where the fusion tag enables the labeling of a protein with synthetic probes which

bind to the protein tag in a specific manner [4–10]. Typically, these tag-specific synthetic compounds will consist of two groups; the binding group designed for specific binding to the tag, and the functional group, which can be any of a number of entities, such as a fluorescent dye or an affinity handle. This approach allows more flexibility to the researcher as attachment of a wide variety of different, but specific, compounds to the protein fusion is possible and thus enables labeling the fusion protein with different fluorophores or other functionalities.

It is important to note that these self-labeling protein tags are not fluorescent on their own; they become fluorescent only after binding of the compound appended with the fluorophore. As a result, the same fusion protein can be labeled with different fluorophores (colors) assuming that the protein populations subjected to labeling can be separated by their subcellular location or time of labeling. This feature allows for development of unique applications that make use of differential spatial and temporal labeling in cells. Such labeling techniques are typically very challenging or unfeasible when utilizing standard fluorescent labeling technologies [11]. Furthermore, in addition to protein labeling with fluorescence, self-labeling tags also enable attachment to surfaces modified with the tag specific molecule and development of many other applications allowing multifaceted analysis of proteins [12–15].

Here we describe a method for labeling of fusion proteins that relies upon the use of a self-labeling protein fusion tag, termed HaloTag [9, 10]. HaloTag can be genetically fused, either N- or C-terminally, to any protein of interest (Fig. 1a). HaloTag has been engineered to form an irreversible, covalent bond with its ligands, which include a category of ligands carrying a variety of fluorophores, termed the HaloTag Fluorescent Ligands (Fig. 1b) [9–12, 16, 17].

Ligands carrying eight distinct fluorophores are available enabling labeling of fusion proteins in different fluorescent colors. These ligands have been shown not to have any deleterious affects on cellular health, in cultured cells, including primary and stem cells. This was shown by exposing cells to labeling with a variety of HaloTag ligands for extended periods of time without any effect on cell viability [11]. The majority of the ligands were developed and optimized for rapid permeation across cellular membranes for intracellular labeling of proteins in live cells. In addition to these permeable ligands, two nonpermeable fluorescent HaloTag ligands were developed to allow specific labeling of membrane proteins. In this case it is important that the protein of interest is fused with HaloTag in the appropriate orientation to position HaloTag on the cellular surface. Both the rapid intracellular and surface labeling is accomplished using the protocol described in Subheading 3.1 (Fig. 2). Labeling can be done on cells that have been transfected with a recombinant vector containing a protein of interest fused to HaloTag resulting in either transient or stable expression.

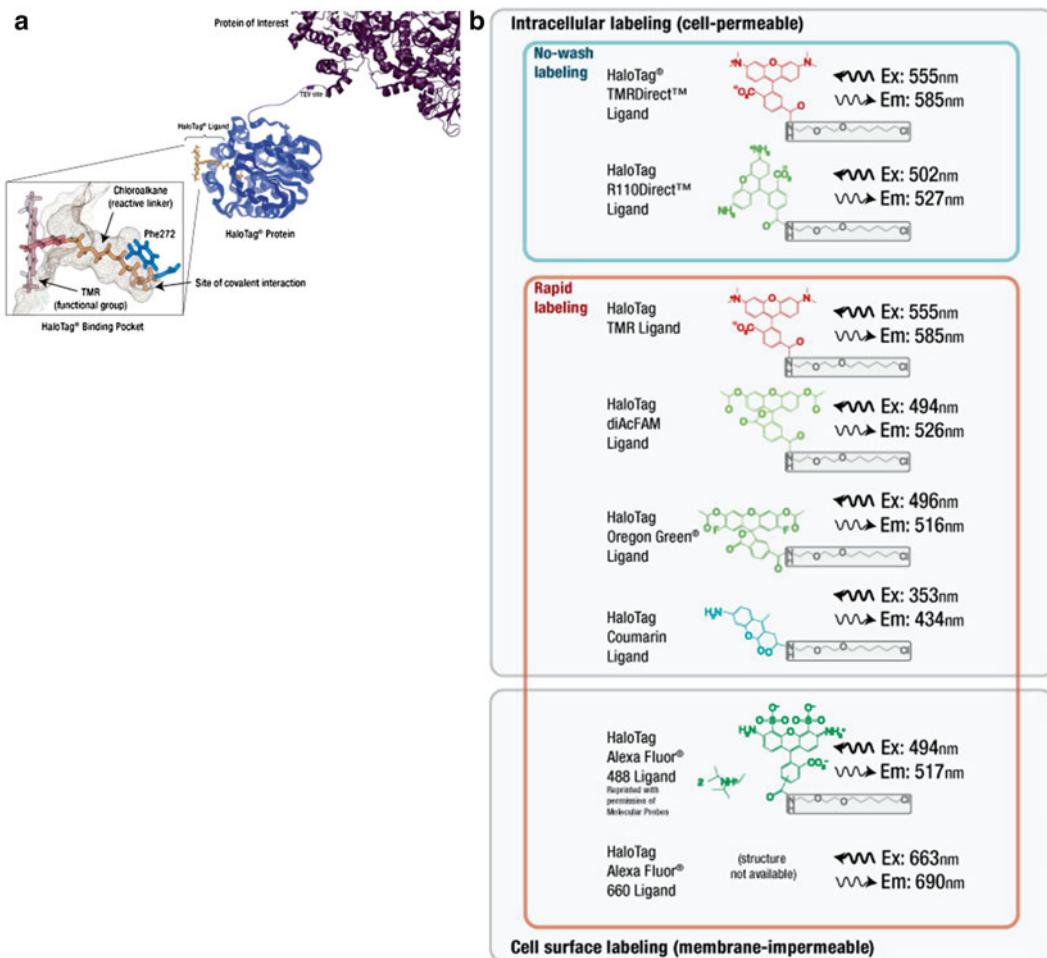


Fig. 1 HaloTag fusion protein model and available fluorescent ligands. **(a)** Model for HaloTag protein fusion displays the linker to scale with TEV cleavage site and the binding pocket for the covalent attachment of HaloTag ligands. **(b)** Available fluorescent HaloTag ligands give the choice of “color,” cell permeability, and type of labeling protocol. Ligands within the *top gray box* are cell-permeable while those included in the *lower gray box* are impermeable. Ligands within the *blue box* were developed to be used in the No-Wash protocol (Subheading 3.2) while those in the *red box* were developed for the Rapid labeling protocol (Subheading 3.1) (color figure online)

Given the highly specific, irreversible, and rapid binding of HaloTag to its ligands, proteins are efficiently and permanently labeled, during the short exposure of cells to the HaloTag ligand added to the cell medium. Due to the extremely low nonspecific binding of HaloTag ligands to proteins other than HaloTag, unbound ligands can be easily and efficiently washed out of the cells. Notably, following this protocol, even proteins expressed at very low concentrations can be efficiently labeled without concern of dissociation of HaloTag ligand from the protein. In addition, for convenience and accommodation of different workflows two additional ligands, the TMRDirect and R110Direct HaloTag ligands have been

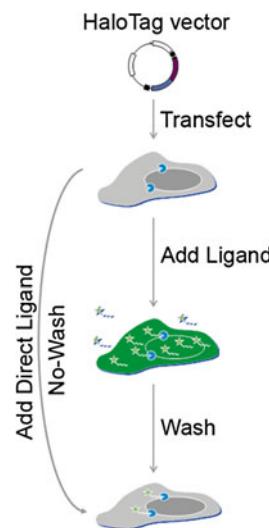


Fig. 2 Schematic showing Rapid and No-Wash options for HaloTag labeling in live cells

developed to use in the No-Wash labeling protocol described in Subheading 3.2. Both the time of labeling and concentration of these ligands have been optimized such that the need for a washing step is eliminated (Fig. 2).

Once a HaloTag fusion protein is labeled with one of the fluorescent ligands, it stays permanently labeled due to covalent binding of the ligand to HaloTag. Thus the same protein molecule cannot be labeled with a different HaloTag ligand. However, newly synthesized HaloTag fusion proteins that are expressed after the initial labeling, and removal of unbound ligand, are available for labeling with any of the HaloTag ligands. Thus the newly synthesized protein can be labeled with a ligand containing a different fluorophore. This enables differential labeling of two protein populations and study of protein synthesis and degradation. This approach is also known as pulse-chase labeling. Similarly, by combining nonpermeable and permeable HaloTag ligands in different colors (e.g. nonpermeable HaloTag Alexa Fluor® 488 ligand and permeable red HaloTag®TMR) we can differentially label surface and intracellular protein pools. In this case, cells are first treated with nonpermeable Alexa488 to label surface exposed proteins in green, followed by addition of red TMR HaloTag ligand which will specifically label the remaining unlabeled intracellular proteins. This approach allows for analysis of protein trafficking [11].

In addition to labeling live cells, HaloTag protein labeling can also be applied to cellular lysates or proteins expressed using *in vitro* expression systems (Subheading 3.3). Further, quantitative downstream analysis for all of these different labeling strategies can take the form of SDS-PAGE followed by a fluorescent gel scan. This is possible due to the covalent nature of the bond between

HaloTag and its ligand as well as the inherent stability of the fluorescent dyes. For the same reason cells carrying labeled HaloTag-protein fusions can be further analyzed using immunocytochemistry (ICC) for other cellular proteins. The simplicity and flexibility of the technology makes this system amenable for development of other imaging applications and recently, *in vivo* animal imaging using HaloTag has been reported [18, 19].

2 Materials

2.1 Protein Labeling in Live Cells

1. Complete culture medium appropriate for your cells warmed to 37 °C.
2. 1× PBS buffer: 137 mM NaCl, 2.68 mM KCl, 1.47 mM KH₂PO₄, 8.1 mM Na₂HPO₄ in filtered water, pH 7.5.
3. Appropriate HaloTag fluorescent ligand as designated in each of the protocols below.

2.2 Fixation and Other Sample Preparation

1. Cell fixation buffer: 4 % paraformaldehyde, 0.2 M sucrose, 1× PBS (pH 7.5). Prepare fresh for each use.
2. 4× SDS-sample buffer: 0.24 M Tris-HCl, 2 % SDS, 50.4 % glycerol, 0.4 M DTT, 3 mM bromophenol blue in filtered water. Titrate to pH 6.8 using HCl.
3. 1× SDS-sample buffer: A 1:4 dilution of 4× SDS Sample Buffer in filtered water.

2.3 Equipment for Imaging and Vector-Based Materials

1. Confocal microscope or wide-field fluorescent microscope equipped with appropriate filter sets. Available HaloTag® Ligands span the visible spectrum.
2. If preparing the HaloTag fusion protein clone by PCR amplification, the following mammalian HaloTag fusion vectors are recommended; for an N-terminal HaloTag fusion: pFN21A HaloTag CMV Flexi Vector (Promega) and for a C-terminal HaloTag construct pFC14A HaloTag CMV Flexi Vector (Promega).
3. HaloTag fusion protein vectors are also commercially available from Kazusa DNA Research Institute and Promega.

3 Methods

3.1 Rapid Labeling of HaloTag-Fusion Proteins in Live Cells (15–60 min)

This protocol is intended for labeling live cells with HaloTag® TMR, diAcFAM, Oregon Green®, Coumarin, Alexa Fluor® 488 or Alexa Fluor® 660 Ligands.

1. Prepare a 1:200 dilution of HaloTag® Ligand in warm culture medium just prior to addition to cells. This is a 5× working stock solution (*see Note 1*).

2. Label cells by replacing one-fifth of the existing volume of medium with the 5× HaloTag® Ligand working stock solution, and mix gently. This results in the recommended final 1:1,000 labeling concentrations of 5 µM TMR; 3.5 µM Alexa Fluor® 660; 1 µM diAcFAM, Oregon Green® or Alexa Fluor® 488; and 10 µM Coumarin Ligand.
 3. Incubate for 15 min in a 37 °C + CO₂ cell culture incubator.
 4. For cell-permeant ligands, gently replace the ligand-containing medium with an equal (or greater) volume of warm fresh medium. Repeat this two times for a total of three complete rinses, and proceed to **step 5** (for pulse-chase *see Note 2* and for example *see Fig. 3*).
- For cell-impermeant ligands (Alexa Fluor®-containing ligands) replace the ligand-containing medium with an equal (or greater) volume of warm fresh medium twice, and proceed to **step 7**. Because they are cell-impermeant, these ligands do not require washing out of unbound ligand.
5. Incubate cells in complete culture medium at 37 °C + CO₂ in a cell culture incubator for 30 min to wash out unbound ligand (*see Note 3*).

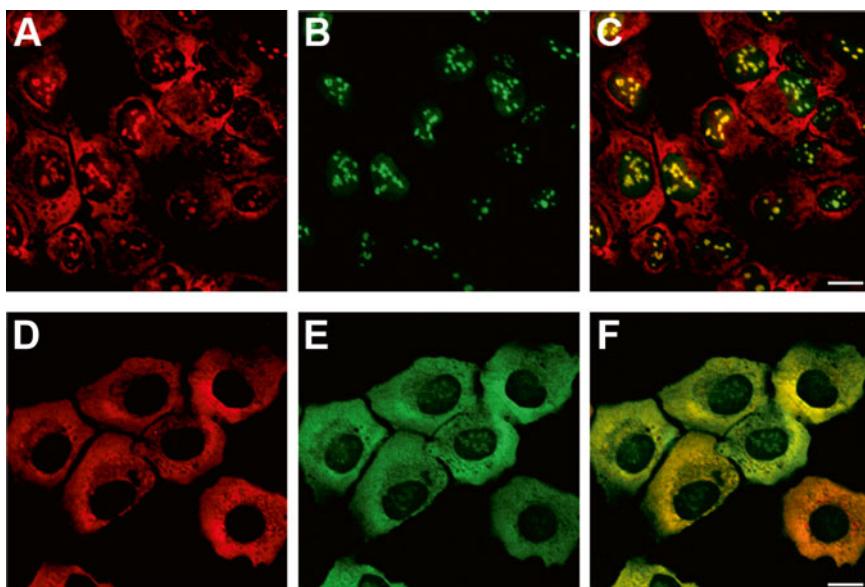


Fig. 3 Pulse-chase labeling (Subheading 3.1) reveals ribosomal subunit translocation over time. U2OS cells stably expressing RPS9-HaloTag were serum starved for 18 h and then labeled using the TMR ligand (panels **a** and **d**). After recovery in complete media cells were chase labeled with Oregon Green ligand for 3 h (panel **b**) or 24 h (panel **e**) prior to imaging. Panel **c** is an overlay of panels **a** and **b**. Panel **f** is an overlay of panels **d** and **e**. All labeling steps were done using the Rapid labeling protocol. Images were acquired on an Olympus FV500 confocal equipped with an environmental chamber. Scale bars 20 µm

6. Replace the medium with an equal volume of fresh warm culture medium. Use of medium lacking phenol red may improve imaging.
7. Transfer to a microscope, and capture images (Fig. 3; for cell fixation/ICC *see Note 4*; for SDS-PAGE analysis *see Note 5*).

3.2 No-Wash Labeling of Proteins in Live Cells (Overnight)

This protocol is intended for labeling of live cells with the cell-permeant HaloTag® TMRDirect™ or R110Direct™ Ligand.

1. Prepare a 1:200 dilution of HaloTag® TMRDirect™ or R110Direct™ Ligand in warm culture medium just prior to addition to cells. This is a 5× working stock solution (*see Note 1*).
2. Replace one-fifth of the existing volume of medium with the 5× HaloTag® ligand working stock solution, and mix gently. This process results in the recommended final 1:1,000 labeling concentration of 100 nM HaloTag® TMRDirect™ or R110Direct™ Ligand (*see Note 6*).
3. Incubate overnight in a 37 °C + CO₂ cell culture incubator.
4. Gently replace the ligand-containing medium with an equal (or greater) volume of warm fresh medium, or fix cells. Due to the low ligand concentration unbound ligand washout time is not required.
5. Transfer to an imaging device, and capture images (Fig. 4; for cell fixation/ICC *see Note 4*; for SDS-PAGE analysis *see Note 5*).

3.3 Labeling Proteins in Mammalian Cell Lysates or Proteins Expressed in Cell-Free Systems

Fluorescent labeling of HaloTag® fusion protein with the HaloTag® TMRDirect™ Ligand provides a rapid and convenient method to monitor protein expression and purification efficiency. HaloTag® TMRDirect™ Ligand is used in the example below because it is provided with the HaloTag® Mammalian Protein Detection and Purification System. If desired, another HaloTag Ligand can easily be substituted for this ligand as long as it is used at the same final concentration. The detection and quantification can be performed

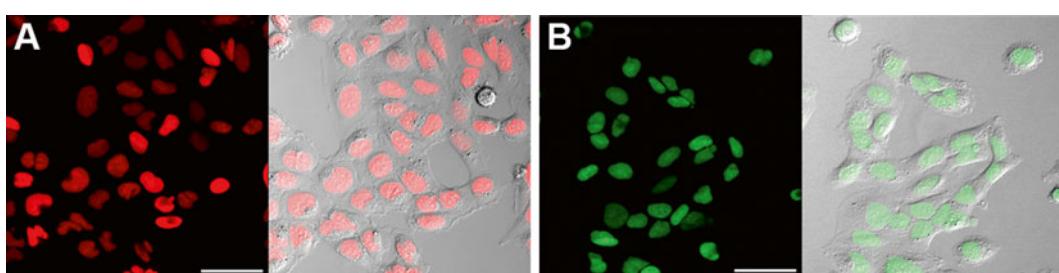


Fig. 4 No-Wash labeling (Subheading 3.2) results in specific labeling of a nuclear targeted HaloTag protein. U2OS cells stably expressing HaloTag-NLS₃ were labeled with either TMRDirect ligand (panel **a**) or R110Direct ligand (panel **b**) using the No-Wash labeling protocol. Images were acquired on an Olympus FV500 confocal equipped with an environmental chamber. Scale bars 50 μm

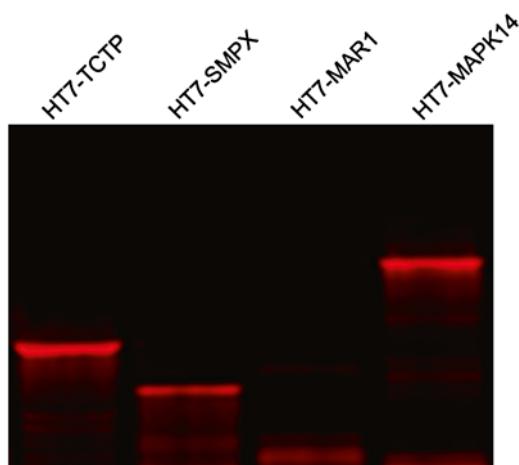


Fig. 5 Mammalian lysates expressing designated HaloTag fusion proteins were labeled with HaloTag® TMRDirect™ ligand following the protocol described in Subheading 3.3. The labeled lysate was boiled and 10 µl was loaded onto an SDS-polyacrylamide gel and proteins were resolved in the gel electrophoresis. Proteins were visualized with fluorescent detection scanner, the Typhoon®, GE Healthcare Bio-sciences (excitation 532 nm; emission 580 nm)

on fluorescent scanners using appropriate filter sets for each particular HaloTag ligand.

1. Dilute the HaloTag® TMRDirect™ Ligand stock solution (100 µM) twofold in DMSO to make a 50 µM working solution that can be stored, protected from light, at -20 °C; alternatively, the stock solution can be prepared in 1× PBS but cannot be stored.
2. Combine 10 µl of lysate containing the HaloTag® fusion protein or the equivalent amount of unbound fraction with 19 µl of any buffer of choice (e.g. PBS or TBS) and 1 µl of 50 µM HaloTag® TMRDirect™ Ligand.
3. Incubate at room temperature for 15 min protected from light.
4. Add 10 µl of 4× SDS sample buffer, and heat at 70 °C for 3 min.
5. Load 10 µl onto an SDS-polyacrylamide gel and run (*see Note 7*).
6. Scan the gel on a fluorescent scanner and quantitate band intensities (Fig. 5).

4 Notes

1. Dilution of the ligand to a 5× labeling stock prior to addition to cells allows even mixing with less perturbation to cells.
2. Termed pulse-chase labeling; in cases where protein trafficking or turnover is of interest, a second ligand of distinct fluorescence can be used to label cells sequentially (chase) to the first

one (pulse) with either a biological stimulation or time in between the two labeling steps (for example *see* Fig. 3).

3. Wash out of unbound ligand is an active cellular process and therefore dependent on both cell type and cell health. Cell background can be alleviated by making sure cells are in their optimal complete media (not stressed), are of low passage number (high passages tend to be less healthy) and, if necessary, allowing longer wash incubation time.
4. The bond between HaloTag® protein and the ligands is stable and HaloTag® Ligands contain stable dyes that continue to be bright after fixation/ICC protocols. For cell fixation, replace media with warm fixation buffer (37 °C) and leave for 10 min at room temperature. ICC protocols can then be followed as needed.
5. For direct quantification of fusion protein cells can be directly harvested after labeling (no washing is needed as unbound ligand will run off bottom of gel), run on a gel and then scanned using a fluorescent scanner. This is done by replacing media with 1× sample buffer, boiling samples at 95 °C for 5 min and running SDS-PAGE.
6. HaloTag® Direct™ Ligands were developed for the No-Wash labeling protocol in order to accommodate workflows for which less distinct labeling steps are preferred. As such the ligands can be added at the same time as transfection reagents or at the time of plating in the case of cells already expressing a HaloTag fusion, as long as they are allowed an overnight incubation to label.
7. When running a gel, the dye front might contain fluorescent material that can complicate detection. It might be necessary to run the gel until the dye front migrates off of the gel or cut the dye front off of the bottom of the gel before scanning it.

References

1. Zhang J, Campbell RE, Ting AY et al (2002) Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Biol* 3:906–918
2. Lippincott-Schwartz J, Patterson GH (2003) Development and use of fluorescent protein markers in living cells. *Science* 300:87–91
3. Miyawaki A, Sawano A, Kogure T (2003) Lighting up cells: labeling proteins with fluorophores. *Nat Cell Biol* 5(Suppl):S1–S7
4. Gautier A, Juillerat A, Heinis C et al (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* 15: 128–136
5. Chapman S, Oparka KJ, Roberts AG (2005) New tools for *in vivo* fluorescence tagging. *Curr Opin Plant Biol* 8:565–573
6. Gronemeyer T, Godin G, Johnsson K (2005) Adding value to fusion proteins through covalent labeling. *Curr Opin Biotechnol* 16: 453–458
7. Griffin BA, Adams SR, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281: 269–272
8. Kepler A, Gendreizig S, Gronemeyer T et al (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat Biotechnol* 21:86–89
9. Los GV, Encell LP, McDougall MG et al (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3(6):373–382

10. Encell LP, Ohana RF, Zimmerman K et al (2012) Development of a dehalogenase-based protein fusion tag capable of rapid, selective and covalent attachment to customizable ligands. *Curr Chem Genomics* 6(Suppl 1-M7): 55–71
11. Svendsen S, Zimprich C, McDougall MG et al (2008) Spatial separation and bidirectional trafficking of proteins using a multi-functional reporter. *BMC Cell Biol* 9:17–31
12. Urh M, Rosenberg M (2012) HaloTag, a platform technology for protein analysis. *Curr Chem Genomics* 6:72–78
13. Nath N, Hurst R, Hook B et al (2008) Improving protein array performance: focus on washing and storage conditions. *J Proteome Res* 10:4475–4482
14. Hurst R, Hook B, Slater MR et al (2009) Protein–protein interaction studies on protein arrays: effect of detection strategies on signal-to-background ratios. *Anal Biochem* 392: 45–53
15. Daniels DL, Méndez J, Mosley AL et al (2012) Examining the complexity of human RNA polymerase complexes using HaloTag technology coupled to label free quantitative proteomics. *J Proteome Res* 11:564–575
16. Benink HA, McDougall MG, Klaubert DH et al (2009) Direct pH measurements by using subcellular targeting of 5(and 6-) carboxyseminaphthorhodafluor in mammalian cells. *Biotechniques* 47:769–774
17. Schroder J, Benink H, Dyba M et al (2009) *In vivo* labeling method using a genetic construct for nanoscale. *Biophys J* 96:L1–L3
18. Hong H, Benink HA, Zhang Y et al (2011) HaloTag: a novel reporter gene for positron emission tomography. *Am J Transl Res* 3: 392–403
19. Kosaka N, Ogawa M, Choyke PL et al (2009) *In vivo* stable tumor-specific painting in various colors using dehalogenase-based protein-tag fluorescent ligands. *Bioconjug Chem* 20: 1367–1374

Chapter 9

Ligation of Synthetic Peptides to Proteins Using Semisynthetic Protein *trans*-Splicing

Julian C.J. Matern, Anne-Lena Bachmann, Ilka V. Thiel, Gerrit Volkmann, Alexandra Wasmuth, Jens Binschik, and Henning D. Mootz

Abstract

Protein *trans*-splicing using split inteins is a powerful and convenient reaction to chemically modify recombinantly expressed proteins under mild conditions. In particular, semisynthetic protein *trans*-splicing with one intein fragment short enough to be accessible by solid-phase peptide synthesis can be used to transfer a short peptide segment with the desired synthetic moiety to the protein of interest. In this chapter, we provide detailed protocols for two such split intein systems. The M86 mutant of the *Ssp* DnaB intein and the MX1 mutant of the AccL-TerL intein are two highly engineered split inteins with very short N-terminal intein fragments of only 11 and 25 amino acids, respectively, and allow the efficient N-terminal labeling of proteins.

Key words Protein semisynthesis, Protein labeling, Intein, Protein splicing, Peptide ligation, Synthetic label, Peptide synthesis, Protein expression, Fluorophore

1 Introduction

1.1 Intein-Mediated Ligation vs. Chemical Ligation

The incorporation of synthetic moieties into recombinant proteins is of great importance for probing protein structure and function and for equipping proteins with new properties. Applications range from basic research to therapeutic proteins and synthetic biology. Currently available and widely used techniques include classical bioconjugation, primarily on cysteine and lysine side chains, unnatural amino acids mutagenesis using suppressor tRNAs, chemical modification of fused tag peptides or proteins and ligation approaches to link two peptides or proteins with a peptide bond.

The latter ligation approaches offer the possibility to incorporate a synthetic peptide including desired chemical modifications into the backbone of a protein. They can be brought about either by a chemical reaction or by chemoenzymatic processes. Chemical reactions, for example Native Chemical Ligation (NCL) [1], Expressed Protein Ligation (EPL) [2], traceless Staudinger

Ligation [3, 4], and KAHA Ligation [5], are based on the chemoselective reaction between two functional groups, of which at least one is abiotic, and that have to be introduced by synthetic means. In contrast, the selectivity of chemoenzymatic reactions relies on the biocatalyst and can allow to completely circumvent the requirement for unnatural functional groups. Particularly useful for protein semisynthesis are the sortase-mediated ligation [6] and the protein *trans*-splicing reaction catalyzed by split inteins [7–9], which is the focus of this chapter.

A split intein mediates the ligation of the flanking sequences (exteins; Ex^N and Ex^C), as shown in Fig. 1, with concomitant excision of the N-terminal and C-terminal intein fragments (Int^N and Int^C). Generally, protein *trans*-splicing allows one to link two polypeptide sequences from separate molecules. It can serve to link a polypeptide that has been prepared to contain a synthetic moiety or a larger stretch of synthetic sequence to a protein that has not undergone any chemical manipulations.

In this chapter we describe split intein systems useful for protein semisynthesis by linking a synthetic peptide with a recombinantly expressed protein. The key feature of these inteins is their short N-terminal fragment that can be easily accessed by solid-phase peptide synthesis. Thus, unusual building blocks can be incorporated into the flanking extein sequence (Ex^N) in a straightforward way and will be transferred to the protein of interest (POI) during protein *trans*-splicing. In the accompanying chapter we describe the use of fully recombinant split intein fragments in which specific residues in the flanking tag sequences (referred to as Chem-tag) are selectively derivatized by simple bioconjugation reactions. This modified Chem-tag is then transferred to the POI.

Advantages of protein *trans*-splicing over chemical ligation procedures include the high affinity between intein fragments. K_D values in the range between low nanomolar [10, 11] and single-digit micromolar [12, 13] were reported and the reactions can be carried out at these concentrations. In contrast, NCL usually requires high micromolar to millimolar concentrations. Furthermore, protein *trans*-splicing is carried out under very mild

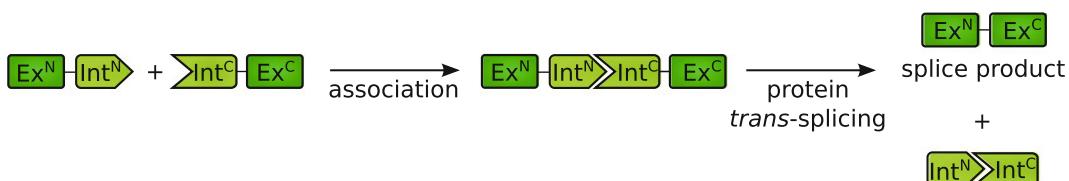


Fig. 1 Scheme of protein *trans*-splicing *in trans*. Protein *trans*-splicing is a posttranslational reaction in which the intein links the two flanking sequences, termed N- and C-terminal exteins (Ex^N and Ex^C), with a peptide bond and undergoes concomitant excision. The active intein is formed after association and folding of the intein fragments (Int^N and Int^C)

and native conditions. No potentially harmful additives like copper ions or thiol catalysts are required and the reaction is best performed at neutral pH. The recognition between two complementary intein fragments is highly specific and therefore the reaction can also be carried out in complex mixtures [14] and even inside living cells [15].

The conserved intein domain is in total about 130–150 amino acids in length. Currently available split inteins with one fragment short enough for convenient synthesis are either naturally split or have been artificially created from the more abundant contiguous *cis*-inteins [16]. Split inteins with either a short Int^N fragment or a short Int^C fragment are known and thus both N- and C-terminal labeling is feasible [15, 17–19, 20–23]. The synthetic peptide containing the short intein fragment is reacted with the respective complementary intein fragment fused to the protein of interest (POI). This fusion protein is obtained by recombinant protein expression.

In the following, we will focus on two split intein systems useful for N-terminal labeling as illustrated in Fig. 2 that have been developed in our laboratory [13, 19, 24, 23]. In both cases, the POI thus represents the Ex^C sequence and is expressed as an Int^C-POI fusion protein in recombinant form and purified. The complementary intein fragment is synthesized as an Ex^N-Int^N peptide, in which Ex^N represents the desired chemical modification including zero, one or more than one amino acids that may be kept as flanking residues of the intein [25]. Upon mixing both components, spontaneous protein *trans*-splicing will occur and the reaction product Ex^N-POI can be purified from the reaction mixture.

1.2 The Artificially Split Ssp DnaB Intein and Its M86 Mutant

The *Ssp* DnaB intein was identified as a contiguous intein in the DNA helicase B of *Synechocystis* sp. PCC6803. Wu et al. removed the endonuclease domain between amino acids 107 and 381 and

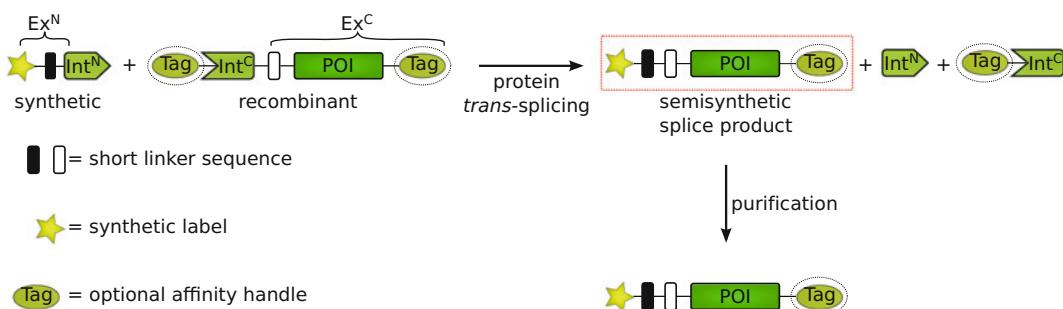


Fig. 2 Concept of N-terminal protein labeling by semisynthetic protein *trans*-splicing. A synthetic peptide including the Int^N sequence and the desired synthetic label is reacted with a recombinantly produced fusion protein harboring the protein of interest (POI) and the complementary Int^C fragment. Affinity tags may be included to facilitate protein purification. Short linker sequences with a few amino acids representing natively flanking residues of the intein can be helpful to achieve high yields in the protein *trans*-splicing reaction

created a mini-intein of 154 aa [26]. This intein was later split following position 11 [27] and shown to be active in *in vitro trans*-splicing reactions using a purified Int^C-POI fusion and a short synthetic Ex^N-Int^N peptide [23, 24].

A further improvement as ligation tool was achieved by directed protein evolution. The resulting M86 mutant of the Int^C fragment displayed a significantly increased rate in the protein *trans*-splicing reaction, higher splicing yields of up to 90 %, a lower K_D between the intein fragments (0.1 μM) and a significantly reduced dependence on the amino acid context directly flanking the intein [13]. The fragments of this intein are hereafter referred to as M86^N and M86^C.

1.3 The Naturally Split AceL-TerL Intein and Its MX1 Mutant

The AceL-TerL intein (Ace Lake Terminase Large subunit) originates from a phage and was identified in metagenomic databases as the first atypically split, naturally occurring intein [19]. All other naturally split inteins reported before were split at a position where often a homing endonuclease is inserted in *cis*-inteins, and which gives rise to an Int^N of about 100 aa and an Int^C of about 35 aa [7, 15, 28]. In contrast, the Int^N and Int^C fragments of the AceL-TerL intein consist of 25 aa and 103 aa. The Int^N thus represents the shortest naturally occurring intein fragment reported so far.

The AceL-TerL intein was shown to be highly active in protein *trans*-splicing with a synthetic peptide containing the Int^N. The finding that no C-terminal cleavage side reactions could be observed at the optimal reaction temperature of 8 °C may reflect the natural character of the intein fragments. Furthermore, mutants selected from directed protein evolution experiments exhibited significantly increased rates both at 37 and at 8 °C with the MX1 mutant being best characterized for the labeling of a diverse set of proteins [19]. The fusion of a maltose-binding protein (MBP) was found to be very well tolerated. MBP-Int^C-POI fusion proteins exhibited improved solubility and showed higher activity of the intein [19]. The fragments of the mutant AceL-TerL intein are hereafter referred to as MX1^N and MX1^C.

2 Material

2.1 Plasmid Construction

1. Plasmid pIT21 (encoding M86^C-Trx-His₆) or pGV160 (encoding MBP-MX1^C-Ubc9-His₆).
2. Restriction enzymes *Bam*HI, *Sal*I, *Hind*III, *Kpn*I-FD, plus corresponding buffers and primers.
3. T4-DNA-ligase.

2.2 Solid-Phase Peptide Synthesis

1. The following building blocks were used: Fmoc-Ala-OH·H₂O, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH,

Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH, 5,6-Carboxyfluoresceine.

2. Fmoc-L-Ala-Wang-Tentagel resin (subs.: 0.25 mmol/g).
3. Reagent K: 81.5 % (v/v) TFA, 5 % (v/v) thioanisole, 5 % (w/v) phenol, 5 % (v/v) ddH₂O, 2.5 % (v/v) EDT, 1 % TIS.
1. Competent *E. coli* BL21 Gold (DE3) cells.
2. LB medium; 5 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract, pH 7.5.
3. Ampicillin stock solution (50 mg/L), sterile.
4. Isopropyl-β-thiogalactoside stock solution (400 mM), sterile.
5. Ni²⁺-NTA affinity chromatography
 - (a) Ni²⁺-NTA agarose.
 - (b) Ni²⁺-NTA buffer A; 50 mM Tris-HCl, 300 mM NaCl, pH 8.0.
 - (c) 5 M imidazole stock solution, pH 8.0.
6. Amylose affinity chromatography
 - (a) Amylose resin.
 - (b) Amylose column buffer; 20 mM Tris-HCl, 200 mM NaCl, pH 7.4.
 - (c) Amylose elution buffer; 20 mM Tris-HCl, 200 mM NaCl, 10 mM maltose, pH 7.4.
7. Dialysis tube (MWCO = 14 kDa).

2.4 Protein trans-Splicing Reaction

1. Splice buffer: 50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, pH 7.0.
2. 4× SDS loading buffer: 500 mM Tris-HCl, 8 % (w/v) SDS, 40 % (v/v) glycerol, 20 % (v/v) β-mercaptoethanol, 4 g/L bromophenol blue, pH 6.8.
3. Freshly prepared DTT stock solution (50 mM). Dissolve 7.7 mg in 1 mL ddH₂O.

3 Method

The sample reaction described here using the M86 intein will result in a protein of interest (POI) N-terminally labeled with a carboxy-fluorescein moiety (=Fl). The exact sequence of the semisynthetic protein will be Fl-KKESGSIE-POI-His₆. The sequence KKESGSIE corresponds to additional amino acids flanking the intein that remain in the splice product (*see Note 1*). Using the MX1 intein with the constructs described here will give rise to the semisynthetic protein Fl-KKEFECEFL-POI.

3.1 Synthesis of the Synthetic Int^N Peptide

Prepare the Ex^N-Int^N peptide using solid-phase peptide synthesis and purify it. For the M86 mutant of the *Ssp* DnaB intein prepare the peptide Fl-KKESG-CISGDSLISLA-OH (Fl=5,6-carboxyfluorescein; Int^N sequence is underlined), and for the MX1 mutant of the AceL-TerL intein use Fl-KKEFE-CVYGDTMVETEDGKIKIEDLYKRLA-OH. The amino acids ESG and EFE, respectively, represent three native flanking amino acids upstream of the two inteins. The two lysines were used to improve the solubility of the peptides during purification and splicing.

No peptides with special functional groups like a thioester are required that can be more demanding in the synthesis. Therefore the peptides could be synthesized by local peptide synthesis facilities or commercial suppliers. In the following, we described our standard protocol using the Fmoc strategy on Fmoc-L-Ala-Wang-TG resin (subs.: 0.25 mmol/g; Iris Biotech) in a 0.1 mmol scale.

1. Synthesize the peptide using a Fmoc-L-Ala-Wang-TG resin. Use the amino acids in a fivefold excess and activate with 5 equiv. HBTU + Oxima-Pure/DMF, and 10 equiv. DIPEA/NMP. Perform the Fmoc deprotection using 20 % piperidine in DMF.
2. Cleavage from the resin is carried out using Reagent K. To this end, treat the resin with 10 mL Reagent K for 1 h and precipitate the cleaved peptide by addition of four volumes of -20 °C cold diethyl ether. Incubate the mixture at -20 °C for 1 h. Afterwards, collect the pellet by centrifugation and wash it using 20 mL -20 °C diethyl ether. Collect the pellet by centrifugation afterwards. Repeat this washing procedure two times. Add 10 mL ddH₂O to the pellet and flash-freeze the pellet using liquid nitrogen. Afterwards, dry the peptide using a lyophilizer.
3. Dissolve the peptide in ddH₂O with 5 % acetonitrile and 0.05 % TFA and purify by preparative HPLC on a PLRP column using a standard ddH₂O/acetonitrile (+0.05 % TFA) gradient.
4. Subsequently, dry the purified peptide using a lyophilizer and store at -20 °C. Prior to use, dissolve the peptide in splice buffer to create a 100 µM stock solution.

3.2 Preparation of the Recombinant Int^C Fusion Protein

Use expression vectors pIT21 and pGV160 (*see* Fig. 3) to express the Int^C fusion proteins M86^C-POI-His₆ and MBP-MX1^C-POI-His₆, respectively. These vectors encode for the proteins of interest (POI) thioredoxin (Trx) and the SUMO-specific E2 conjugating enzyme Ubc9, respectively, and serve as template to insert other protein encoding DNA fragments. Note that the MBP fusion may also be included for the M86 mutant and may also be omitted for the MX1 mutant [19], depending on optimization procedures for the individual POI. More remarks on alternative designs of the Int^C fusions can be found in Notes 1–3.

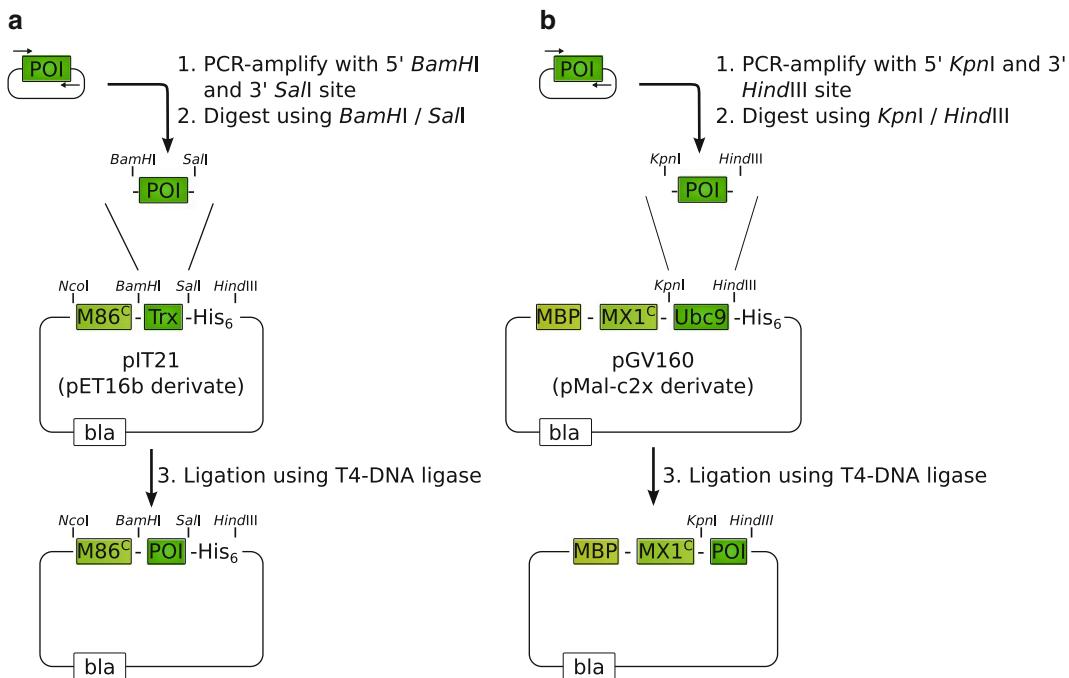


Fig. 3 Cloning strategy for expression vectors. The plasmids pIT21 and pGV160 serve as templates to express Int^c fusions with the protein of interest (POI), based on (a) the M86^c intein fragment and (b) the MX1^c intein fragment, respectively. The cloning strategy shown in (b) will result in the loss of the sequence encoding the hexahistidine tag. Note that different affinity tag strategies will be realized for the two presented options

1. Use pIT21 to generate an expression plasmid for the M86^c-POI fusion protein (see Fig. 3a). Three restriction sites are available in the plasmid. By employing *Bam*HI and *Sal*I sites to insert the DNA fragment encoding the gene of interest, the linker sequence (amino acids SIE) between the M86^c fragment and the POI will be kept and the resulting protein will feature a C-terminal His₆-tag for purification. The linker represents native flanking residues of the intein. Keeping them may be beneficial for high splicing yields. By using the *Bam*HI and *Hind*III sites the C-terminal His₆-tag will be removed and the inserted gene fragment should contain a stop codon to terminate translation.
2. Plasmid pGV160 serves as a template to construct fusions with the MX1^c fragment (see Fig. 3b). A DNA fragment encoding the POI can be inserted using the *Kpn*I and *Hind*III restriction sites. The MBP in the resulting fusion protein MBP-MX1^c-POI will improve expression levels and solubility and can serve as a purification tag for affinity chromatography on amylose resins. The C-terminal His₆-tag still encoded in pGV160 will be omitted in the cloning procedure. An additional C-terminal purification tag may be introduced with the insert fragment. A stop codon to terminate translation must be introduced with the insert fragment.

3. Transform competent *E. coli* BL21 (DE3) cells with the generated plasmid. Spread the transformed bacteria to a LB-agar plate containing 100 mg/L ampicillin. Incubate the plate overnight at 37 °C, until visible colonies have grown.
4. Transfer a single colony to 6 mL liquid LB-medium containing 100 mg/L ampicillin and incubate at 37 °C overnight in a shaker or roller incubator.
5. Use the overnight culture to inoculate the main culture (600 mL LB-medium with the respective antibiotic as described before, add 2 % glucose for constructs based on pGV160) in a ratio of 1:100 (v/v). Incubate the culture in a rotary shaker at 37 °C under aerobic conditions until it has reached an optical density ($\lambda=600$ nm) of 0.6.
6. Before inducing expression, remove a sample of 1 mL for analytical purposes (pre-induction). Pellet the cells of this sample by centrifugation (1 min at $20,000 \times g$), discard the supernatant, resuspend the cells in 60 μ L 1× SDS sample buffer and store it at -20 °C.
7. Reduce the temperature to 28 °C and let the culture cool down for 10 min while still shaking. Then, induce the protein production by adding IPTG to final-concentration of 0.4 mM (add 0.6 mL from a 400 mM stock solution). Incubate the culture for 4 h at 28 °C to allow high levels of protein to be produced. These parameters may be optimized for different proteins.
8. Before harvesting the cells, take another sample of 1 mL (post-induction) and treat as described before.
9. Harvest the cells by centrifugation (30 min at $3,000 \times g$ and 4 °C) in a precooled centrifuge rotor. Discard the supernatant. The pellet can be stored at -20 °C for a prolonged period of time, either before or after resuspension in the buffer for cell lysis (see first step in the next protocol).

3.3 Purification of the Recombinant Proteins

3.3.1 Purification of Fusion Proteins Based on Expression Vector pLT21

The protocols to purify the fusion proteins by affinity chromatography are standard protocols. They differ for the two intein systems because of the different tags used in these examples. Purification can also be achieved by other chromatography techniques or be improved by combining different methods.

In case of the M86^C-constructs, purification is done by IMAC, as described in the following for a gravity column protocol. It can also be done using a chromatography system and/or gradients of the applied buffers:

1. Resuspend the cell pellet in ice cold 10 mL Ni-NTA buffer A with 20 mM imidazole (40 μ L from a 5 M stock solution) and rupture the cells using your preferred method.

2. Separate the soluble fraction from insoluble material and cell debris by centrifugation (20 min at $25,000 \times g$ and 4 °C).
3. Apply the supernatant to a pre-calibrated Ni²⁺-NTA column (bed volume 2.5 mL). Take a sample of 6 µL of the collected flowthrough for later analysis and mix it with 2 µL of 4× SDS sample buffer.
4. Wash the column with the bound protein using 25 mL Ni²⁺-NTA buffer A with 20 mM imidazole (10 column volumes (CV)).
5. Subsequently, the protein can be eluted from the column by addition of Ni²⁺-NTA buffer A with 250 mM imidazole (10 mL Ni-NTA buffer A + 500 µL of a 5 M imidazole stock solution; 4 CV). Collect eight fractions with a volume of 1 mL each and place on ice. Take an aliquot of 6 µL from each fraction for later analysis and mix it with 2 µL of 4× SDS sample buffer.

3.3.2 Purification of Proteins Based on the Expression Vector pGV160

Use amylose column material to purify the MBP-tagged MX1^C fusion protein, as described in the following.

1. Resuspend the cell pellet in ice-cold 10 mL ACB and rupture the cells using your preferred method.
2. Remove insoluble material and cell debris by centrifugation (20 min at $25,000 \times g$ and 4 °C).
3. In the meantime, prepare the amylose-column (bed-volume of 2 mL) by equilibration with 10 mL ACB.
4. After centrifugation, apply the supernatant to the column and collect the flowthrough. Take a sample of 6 µL for later analysis and mix it with 2 µL of 4× SDS sample buffer.
5. Wash the column with 20 mL ACB (10 CV). Collect the wash fractions.
6. Subsequently retrieve the purified protein from the column by addition of 8 mL ACB with 10 mM maltose. Collect eight fractions of 1 mL each and place on ice. Take of each a sample of 6 µL and add 2 µL of 4× SDS sample buffer.

3.4 Analysis of the Samples of the Protein Purification

1. Boil all samples at 95 °C for 10 min and analyze them by SDS-PAGE. Apply 2.5 µL of pre- and post-induction, pellet, flowthrough, and wash-fraction samples. Load 6 µL of the eluate samples. Analyze the gel by Coomassie staining.
2. Pool all elution fractions containing the purified Int^C-POI fusion protein.
3. For further experiments, exchange the buffer to splice buffer with 10 % glycerol (v/v) by dialysis. Perform at least three dialysis steps (use the 100-fold volume of buffer). Reducing agents such as DTT or TCEP (at a concentration of 1–2 mM) may be added in the first dialysis step. It is sufficient to include the glycerol in the last dialysis step. Glycerol protects the samples from precipitation when frozen.

- Determine the protein concentration and aliquot the protein solution in suitable volumes. Flash-freeze the aliquots in liquid nitrogen and store at -80 °C. Thaw on ice for later use.

3.5 Protein trans-Splicing Reactions

The described split intein systems typically allow a close to quantitative protein *trans*-splicing reaction (80–95 %) within 2–4 h (*see also Note 4*). We perform the reaction at a concentration of 10–50 µM (typically 15 µM) of the Int^C-POI construct and an excess of the Ex^N-Int^N peptide (20–30 µM higher than Int^C-POI; typically 45 µM). These concentrations fit to protein concentrations usually obtained from affinity chromatography purifications and allow for convenient analysis of such reactions by loading 5–10 µL onto an SDS-PAGE gel with subsequent Coomassie staining. Note that the reactions can also be performed at significantly lower concentrations and that the efficiency will ultimately depend on the affinity between the intein fragments. To prevent formation of disulfides, add 2 mM DTT or 1 mM TCEP to the sample (*see Note 4*). Typical concentrations of stock solutions and reaction mixtures are given in Table 1.

- Create a 100 µM stock solution of the Ex^N-Int^N peptide in splice buffer.
- Premix all components except the Int^C-POI construct according to Table 1 and pre-incubate the mixture for 10 min at 25 or 8 °C.
- Add the recombinant protein as the last component to start the reaction. Immediately after mixing the components, remove a sample of 6 µL, mix it with 2 µL of 4× SDS sample buffer. This is the sample for the time point at 0 h. Incubate the remaining mixture for 4 h at 25 °C (for the M86 intein) or at 8 °C (for the MX1 intein). Take another sample of 6 µL and treat it as described above.
- Place the reaction cups on ice. Note that this will not formally stop the reaction and splicing as well as cleavage side-reactions could still proceed.

Table 1
Pipetting scheme for protein *trans*-splicing reaction

Component	Stock-concentration	Volume (µL)	Final concentration
Int ^C -POI-His ₆	30 µM	50	15 µM
Peptide: Ex ^N -Int ^N	100 µM	45	45 µM
DTT	50 mM	4	2 mM
Splice buffer		1	
Total		100	

5. Analyze the two samples by SDS-PAGE with Coomassie staining to estimate the degree of modification. When analyzing the gel under UV illumination prior to Coomassie staining, the transfer of a fluorophore moiety can be observed.

The reactions of the sample proteins IT21 and GV160 with their respective counterparts are depicted in Fig. 4. Comments on purification procedures for the splice product can be found in Notes 2, 3 and 5.

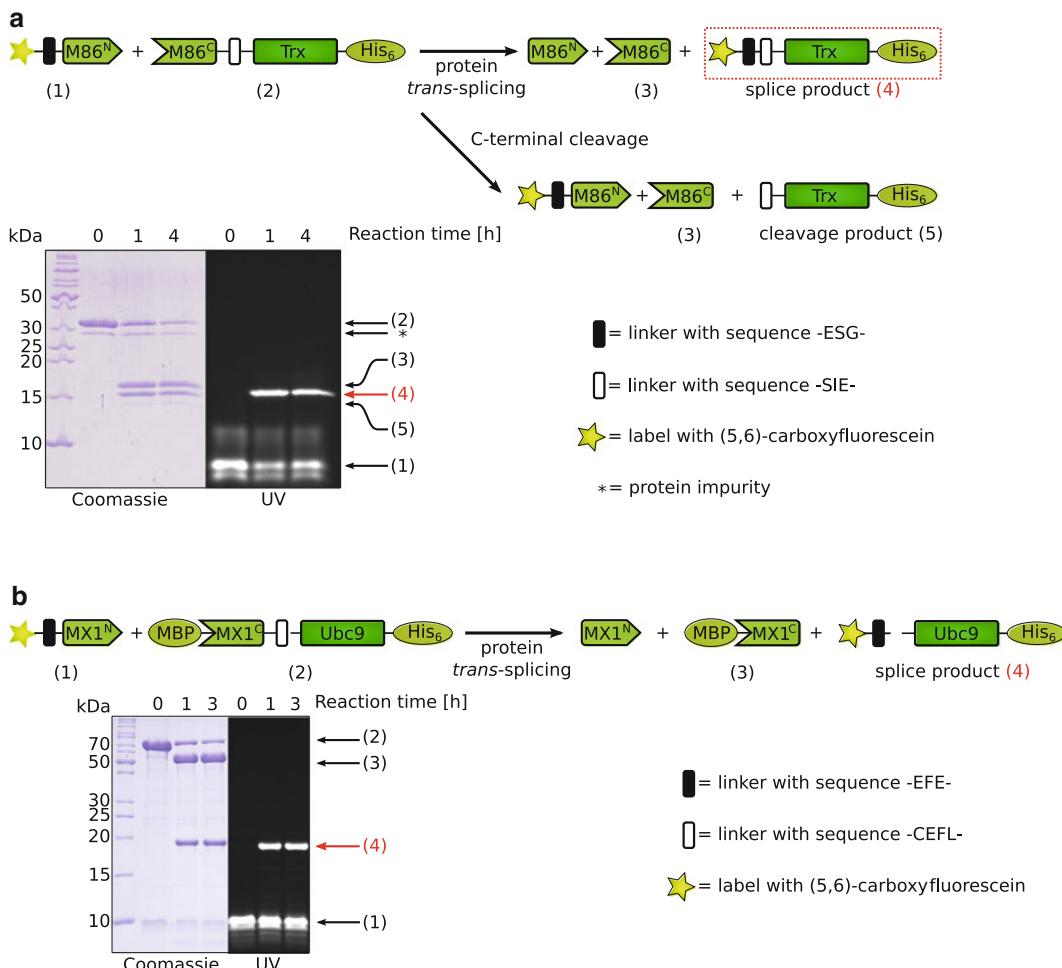


Fig. 4 Protein trans-splicing reactions of selected sample proteins. (a) Reaction of M86^C-SIE-Trx-His₆ with the peptide Fl-KKESG-M86^N. Protein (15 μM) and peptide (45 μM) were incubated for 4 h at 25 °C in splice buffer with 1 mM TCEP. Small amounts of the C-terminal cleavage product can be seen. (b) Reaction of fusion protein MBP-MX1^C-CEFL-Ubc9-His₆ with the peptide Fl-KKEFE-MX1^N. Protein (15 μM) and peptide (45 μM) were incubated for 3 h at 8 °C in splice buffer with 2 mM DTT. No formation of the C-terminal cleavage product could be observed. Schemes illustrate the reactions. Shown are SDS-PAGE gels to monitor the reaction progress analyzed under UV illumination and Coomassie staining, as indicated

4 Notes

1. The described protocols provide access to a semisynthetic, labeled protein. However, due to the peptides and fusion proteins used in the depicted protocols, a short extra sequence corresponding to the natively flanking amino acids of the two intein fragments will remain in the labeled protein. For the two examples described in this chapter, this sequence is X-ESG-SIE-POI for the M86 intein and X-EFE-CEFL-POI for the MX1 intein (X=Fl-KK-). A dependency of the activity is plausible, because the flanking sequences can be regarded as covalently fused substrates of the intein that have undergone coevolution. In general, the structure–function relationship between these sequences and intein activity is not well understood and requires testing for the individual cases. However, only the single residue at the +1 position of the intein is strictly required for the splice reaction. This residue is a Ser in case of the M86 intein and a Cys in case of the MX1 intein. Other residues at the flanking positions may either be tolerated or may lead to a significant reduction in splice activity of the intein [25, 29, 30]. Splicing was reported with unnatural buildings blocks such as sarcosine (*N*-methyl glycine) or β-alanine immediately flanking the upstream splice junction at the -1 position [12]. The Ex^N residues may even be completely omitted in selected cases. We could show for the M86 intein that a fluorophore based on the coumarin structure could be coupled directly to the Int^N sequence and be efficiently transferred in the protein *trans*-splicing reaction, thus allowed for virtually traceless labeling [25]. As a rule of thumb, 2–5 native amino acids on each side should be kept or preferentially conservative amino acids substitutions should be undertaken in these regions. For the very recently reported MX1 intein this dependency has not yet been investigated [19]. The M86 intein was actually selected to exhibit high sequence tolerance and was indeed shown to be capable of splicing in several different sequence contexts [13].
2. The strategy for the purification of the desired, labeled POI (the splice product) from the reaction mixture will depend on its biophysical properties and on the affinity tags used. For example, for the purification of Fl-Trx-His₆, we used anion exchange and Ni-NTA affinity chromatography [23]. In case of the MX1 mutant of the AceL-TerL intein and the TycB1 protein as the POI we used a gel filtration column to separate off the small synthetic peptide, followed by an amylose affinity column to remove unspliced precursor and the MBP-MX1^C by-product [19].

3. To prepare an untagged labeled protein, but still purify the Int^C-POI fusion construct by affinity chromatography, the Int^C fragment may be tagged N-terminally. This was realized for the MBP-MX1^C-POI construct described in this chapter. Likewise an MBP-tag [19, 31], StreptagII [32], or His₆-tag [14] may be fused to the N terminus of the Int^C fragment for this purpose.
4. Typical by-products that can be formed during the protein *trans*-splicing reaction result from partial N- or C-terminal cleavage. Also (see Note 6) for potential by-products arising from difficult regions in the synthesis of the MX1^N sequence. These reactions occur when the intein complex is formed but a peptide bond at one intein splice junction is cleaved before the two exteins are linked. C-terminal cleavage is caused by premature succinimide formation through cyclization of the Asn residue found at the last position of the intein. For the AceL-TerL intein and its MX1 mutant we could not observe this reaction at 8 °C but to a considerable degree at 37 °C [19]. The artificially split *Ssp* DnaB intein and its M86 mutant is more prone to this side-reaction [13], but it can also suppressed at lower temperature (at the cost that also the rate of the splicing reaction will be decreased) [19]. N-terminal cleavage occurs by nucleophilic attack of the initially formed thioester or the branched intermediate. Nucleophiles can be water or thiols like DTT (note that this kind of thiolytic with thiols forming more stable thioesters like MESNA is deliberately used to prepare the thioester building blocks for Expressed Protein Ligation [2]). Thus, although we routinely use 2 mM DTT as a reducing agents in our assays, better results in terms of low N-terminal cleavage levels may be achieved by replacing it with 1–2 mM TCEP. Mind to adjust the pH of your TCEP stock solution, as the solution may be strongly acidic. Generally, reducing agents can be removed completely from the splice reaction as long as the catalytic cysteines of the intein are not oxidized. Only over time undesired disulfide bridges may be observed with the Int^N or Int^C fragment due to air oxidation. Finally, cleavage at the N-terminal splice junction may also induce C-terminal cleavage, as documented for the *Ssp* DnaB intein [14].
5. At least for the M86^C fragment of the *Ssp* DnaB intein we know that renaturation from denaturing conditions is possible and thus the fusion protein may be treated with 8 M urea, for example for purification under denaturing conditions, followed by urea removal by dialysis.
6. The sequence of the MX1^N fragment contains the motif Asp-Gly, a classical sequence for aspartimide formation, which can in fact be observed after the synthesis. The problem may be alleviated by applying appropriate measures, for example using a backbone-protected building block.

Acknowledgements

We thank Xiang-Qin Liu (Dalhousie University, Canada) for collaboration on initial work on the M86 mutant and Shmuel Pietrovski (Weizmann Institute, Israel) for collaboration on the AceL-TerL intein. We acknowledge funding by DFG (grants MO1073/3-2; SPP1623, MO1073/5-1, and Cells-in-Motion excellence cluster, EXC1003) and Fonds der Chemischen Industrie.

References

- Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* 266(5186):776–779
- Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* 95(12):6705–6710
- Nilsson BL, Kiesling LL, Raines RT (2000) Staudinger ligation: a peptide from a thioester and azide. *Org Lett* 2(13):1939–1941
- Saxon E, Armstrong JI, Bertozzi CR (2000) A “traceless” Staudinger ligation for the chemoselective synthesis of amide bonds. *Org Lett* 2(14):2141–2143
- Bode JW, Fox RM, Baucom KD (2006) Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and alpha-ketoacids. *Angew Chem Int Ed* 45(8):1248–1252. doi:[10.1002/anie.200503991](https://doi.org/10.1002/anie.200503991)
- Mao HY, Hart SA, Schink A, Pollok BA (2004) Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* 126(9):2670–2671. doi:[10.1021/ja039915e](https://doi.org/10.1021/ja039915e)
- Noren CJ, Wang JM, Perler FB (2000) Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed* 39(3):450–466
- Volkmann G, Mootz HD (2013) Recent progress in intein research: from mechanism to directed evolution and applications. *Cell Mol Life Sci* 70(7):1185–1206. doi:[10.1007/s0018-012-1120-4](https://doi.org/10.1007/s0018-012-1120-4)
- Shah NH, Muir TW (2014) Inteins: nature’s gift to protein chemists. *Chem Sci* 5:446–461. doi:[10.1039/C3SC52951G](https://doi.org/10.1039/C3SC52951G)
- Shi J, Muir TW (2005) Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J Am Chem Soc* 127(17):6198–6206
- Shah NH, Eryilmaz E, Cowburn D, Muir TW (2013) Extein residues play an intimate role in the rate-limiting step of protein trans-splicing. *J Am Chem Soc* 135(15):5839–5847
- Schwarzer D, Ludwig C, Thiel IV, Mootz HD (2012) Probing intein-catalyzed thioester formation by unnatural amino acid substitutions in the active site. *Biochemistry* 51(1):233–242
- Appleby-Tagoe JH, Thiel IV, Wang Y, Mootz HD, Liu XQ (2011) Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. *J Biol Chem* 286(39):34440–34447
- Bischik J, Zettler J, Mootz HD (2011) Photocontrol of protein activity mediated by the cleavage reaction of a split intein. *Angew Chem Int Ed Engl* 50(14):3249–3252
- Giriat I, Muir TW (2003) Protein semi-synthesis in living cells. *J Am Chem Soc* 125(24):7180–7181. doi:[10.1021/ja0347361](https://doi.org/10.1021/ja0347361)
- Mootz HD (2009) Split inteins as versatile tools for protein semisynthesis. *Chembiochem* 10(16):2579–2589. doi:[10.1002/cbic.200900370](https://doi.org/10.1002/cbic.200900370)
- Volkmann G, Liu XQ (2009) Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein. *PLoS One* 4(12):e8381. doi:[10.1371/journal.pone.0008381](https://doi.org/10.1371/journal.pone.0008381)
- Appleby JH, Zhou K, Volkmann G, Liu X-Q (2009) Novel split intein for trans-splicing synthetic peptide onto C terminus of protein. *J Biol Chem* 284(10):6194–6199
- Thiel IV, Volkmann G, Pietrovski S, Mootz HD (2014) An atypical naturally split intein engineered for highly efficient protein labeling. *Angew Chem Int Ed Engl*. doi:[10.1002/anie.201307969](https://doi.org/10.1002/anie.201307969)
- Aranko AS, Zuger S, Buchinger E, Iwai H (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS One* 4(4):e5185. doi:[10.1371/journal.pone.0005185](https://doi.org/10.1371/journal.pone.0005185)
- Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K, Iwai H (2009) Solution structure of DnaE intein from *Nostoc punctiforme*:

- structural basis for the design of a new split intein suitable for site-specific chemical modification. *FEBS Lett* 583(9):1451–1456. doi:[10.1016/j.febslet.2009.03.058](https://doi.org/10.1016/j.febslet.2009.03.058)
22. Borra R, Dong D, Elnagar AY, Woldemariam GA, Camarero JA (2012) In-cell fluorescence activation and labeling of proteins mediated by FRET-quenched split inteins. *J Am Chem Soc* 134(14):6344–6353. doi:[10.1021/ja300209u](https://doi.org/10.1021/ja300209u)
23. Ludwig C, Pfeiff M, Linne U, Mootz HD (2006) Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. *Angew Chem Int Ed Engl* 45(31):5218–5221
24. Ludwig C, Schwarzer D, Mootz HD (2008) Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction. *J Biol Chem* 283(37):25264–25272
25. Wasmuth A, Ludwig C, Mootz HD (2013) Structure–activity studies on the upstream splice junction of a semisynthetic intein. *Bioorg Med Chem* 21(12):3495–3503
26. Wu H, Xu M-Q, Liu X-Q (1998) Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim et Biophys Acta* 1387(1–2):422–432
27. Sun W, Yang J, Liu X-Q (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem* 279(34):35281–35286
28. Brenzel S, Kurpiers T, Mootz HD (2006) Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split *Ssp* DnaB intein and comparison to the split *Sce* VMA intein. *Biochemistry* 45(6):1571–1578
29. Cherian M, Pedamallu CS, Tori K, Perler F (2013) Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. *J Biol Chem* 288(9):6202–6211
30. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M (2009) Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci U S A* 106(27):11005–11010. doi:[10.1073/pnas.0904366106](https://doi.org/10.1073/pnas.0904366106)
31. Kurpiers T, Mootz HD (2007) Regioselective cysteine bioconjugation by appending a labeled cysteine tag to a protein by using protein splicing in trans. *Angew Chem Int Ed Engl* 46(27):5234–5237
32. Zettler J, Schutz V, Mootz HD (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* 583(5):909–914. doi:[10.1016/j.febslet.2009.02.003](https://doi.org/10.1016/j.febslet.2009.02.003)

Chapter 10

Chemical-Tag Labeling of Proteins Using Fully Recombinant Split Inteins

Anne-Lena Bachmann, Julian C.J. Matern, Vivien Schütz,
and Henning D. Mootz

Abstract

Chemical-tag labeling of proteins involving split inteins is an approach for the selective chemical modification of proteins without the requirement of any chemical synthesis to be performed. In a two-step protocol, a very short tag fused to a split intein auxiliary protein is first labeled in a bioconjugation reaction with a synthetic moiety either at its N-terminus (amine-tag) or at the side chain of an unnatural amino acid (click-tag). The labeled protein is then mixed with the protein of interest fused to the complementary intein fragment. In the resulting spontaneous protein *trans*-splicing reaction the split intein fragments remove themselves and ligate the tag to the protein of interest in a virtually traceless fashion. The reaction can be performed either using a purified protein of interest or to label a protein in the context of a living cell. All protein components are recombinantly expressed and all chemical reagents are commercially available.

Key words Bioconjugation, Protein labeling, Intein, Protein splicing, Click chemistry, Synthetic label, Protein expression, Fluorophore

1 Introduction

Appending synthetic moieties to proteins has ever since been of interest in both basic research on protein biochemistry and in applied fields, such as diagnostics, therapy, and bionanotechnology. Also known as bioconjugation, classical approaches for the modification of native functional groups in proteins, with chemistries for cysteine and lysine side chains still are the most widely used, face the problem of chemo- and regioselectivity [1]. When more than one amine or thiol group is present in the protein, regioselectivity is generally difficult or impossible to control. The resulting mixtures are usually highly undesirable, for example when producing antibody-drug conjugates for therapeutic applications. For the same reasons, classical bioconjugation approaches are mostly not useful in more complex sample mixtures, e.g. the chemical labeling of proteins in or on living cells for fluorescence microscopy studies, with few exceptions [2].

In recent years, many new methodologies have been developed to address these limitations using new bioconjugation reactions [3–5] and also combinations with molecular biology. In particular, in the light of the success story of genetic fusions with fluorescent proteins like GFP, various fusion tags have been developed that allow for chemical modification, for example the cysteine-rich FLASH-tag [6], the SNAP/CLIP-tags [7], Halo-Tag [8], DHFR-tag [9], and BL-tag [10], or tags for sortase [11, 12] biotin-ligase [13], lipoic acid ligase [14], and 4'-phosphopantetheine-transferase [15]. In these cases, the selectivity problem of classical bioconjugation approaches is circumvented by the genetic fusion of the tag to the protein of interest and its unique self-modification, recognition by cognate transferases or ligases, or its chemical composition. These tags provide advantages over fluorescent proteins that mainly rely on the synthetic nature of the probes that can be attached to them, which are not limited to fluorophores but can also include, for example, biotin or PEG polymers. Synthetic fluorophores can offer beneficial properties compared to GFP, like higher photostability and smaller size. Furthermore, in some cases the tags are significantly shorter than GFP.

Other methodologies involve the total or semisynthesis of proteins by solid-phase peptide synthesis and ligation techniques like Native Chemical Ligation (NCL) [16–18]. They allow not only the labeling of a protein with a synthetic probe but also both subtle and massive modifications within the peptide chain of virtually any kind, for example to introduce posttranslational modifications with native structure.

The tRNA suppression approach for the incorporation of unnatural amino acids during protein biosynthesis has found wide acceptance in recent years [19, 20]. This methodology is per definition of potentially very general utility as the position of the unnatural amino acid in the protein can be chosen freely. Many of the available unnatural amino acids allow for subsequent bioconjugation through bioorthogonal reactions, e.g., copper-catalyzed alkyne-azide cycloaddition (CuAAC).

In this chapter, we describe an approach termed chemical tag labeling using protein *trans*-splicing by split inteins [21]. The selectivity problem is solved by the split intein tag, which removes itself during the reaction, and therefore allows for virtually a traceless protein modification. Chemical tag labeling as described here offers a unique combination of advantages, including a very short tag of only a few amino acids with variable sequence, fully recombinant nature of the proteins involved, use of simple and commercially available chemical reagents that are also used in classical bioconjugations, mild and native reaction conditions with very low split intein reagent concentrations (nano to micromolar), and modularity in the two-step protocol of the chemical modification and the protein *trans*-splicing reaction that is beneficial for adaptation to various proteins of interest (POI).

Protein *trans*-splicing (PTS) is a self-processive reaction between two precursor polypeptides or proteins, each harboring one of a pair of split intein fragments [22–24]. After association and folding of the split intein fragments into the active intein domain, the two flanking sequences, termed N- and C-terminal exteins, are linked with a concomitant removal of the intein fragments (see Fig. 1). The only scar remaining at the spliced junction in the splice product is strictly a cysteine, serine, or threonine residue (dependent on the intein) important for the splicing mechanism.

We had previously reported a simple protein labeling concept that was based on the combination of cysteine bioconjugation with protein *trans*-splicing and that was also described in a previous volume of this book series [25–28]. One of the extein sequences was reduced to a short tag, termed Cys-tag, of only a few amino acids but including a single cysteine. This Int^{C} -Cys-tag construct was recombinantly expressed, purified and the cysteine chemoselectively alkylated using standard reagents to introduce fluorophores, biotin, etc. The modified Cys-tag auxiliary protein was then used in a PTS reaction with a $\text{POI}-\text{Int}^{\text{N}}$ construct (POI =protein of interest). Additional cysteine residues in the POI remain unlabeled because they do not come in contact with the thiol-specific reagent. Thus the regioselectivity problem was solved.

A prerequisite for this method is a split intein fragment free of cysteine residues to avoid reaction during the tag labeling step. In particular, the intein fragment must not have a catalytic cysteine at one of the splice junctions. Inteins that fulfill this criteria are rare and the inteins used for the Cys-tag concept, i.e. the *Ssp* DnaB intein [27], the *Mxe* GyrA intein [26], and the *Psp*-GBD Pol intein [28], were all artificially split from originally contiguous, *cis*-splicing inteins. More recently, split inteins exhibiting significantly better properties with regard to solubility, splicing yields and kinetics have been reported. Most of these are naturally split inteins [29–32].

The currently best characterized naturally split intein with excellent splicing rates and yields is the *Npu* DnaE intein from *Nostoc punctiforme* PCC73102, as shown in Fig. 1 [33, 34]. The Int^{N} and Int^{C} fragments of the *Npu* DnaE intein exhibit high

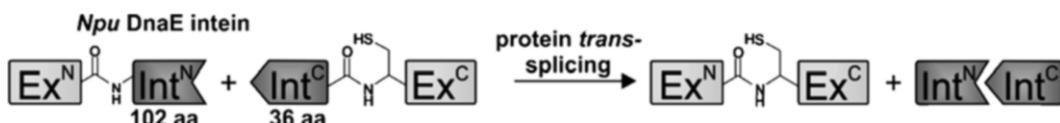


Fig. 1 Scheme of protein *trans*-splicing by the *Npu* DnaE intein. Protein *trans*-splicing is a posttranslational, self-processing reaction in which the two polypeptides, termed N- and C-terminal exteins (Ex^{N} and Ex^{C}) are linked with a peptide bond while the two split intein fragments, Int^{N} and Int^{C} , are excised out of the precursor proteins. The active intein domain is reconstituted after association of the Int^{N} and Int^{C} pieces. All functional information required for the splicing reaction resides in the intein sequence and the downstream flanking amino acid (+1 residue), which is a cysteine, serine, or threonine. The *Npu* intein from *Nostoc punctiforme* PCC73102 is a well-studied, efficiently and rapidly splicing intein and employs a +1 cysteine

affinity with a K_D in the low nanomolar range [35] and the intein has been reported in many in vitro and cellular applications [36–41]. However, this intein contains catalytic cysteine residues at both splice junctions and their mutation to serine significantly impairs or even abrogates splicing activity [34], thus precluding the Cys-tag approach. In order to allow combination with the *Npu* DnaE intein, the chemical tag concept was extended to amine-tag and click-tag approaches with compatible labeling chemistries [21], which are described in this chapter. Figure 2 shows the general procedure for N-terminal labeling of a POI. In particular, we describe here the protocols involving the chemical modification of the tag on the Int^N fragment of the *Npu* DnaE intein using

1. Acylation of the N-terminal amine group, termed amine-tag approach (Fig. 3), or

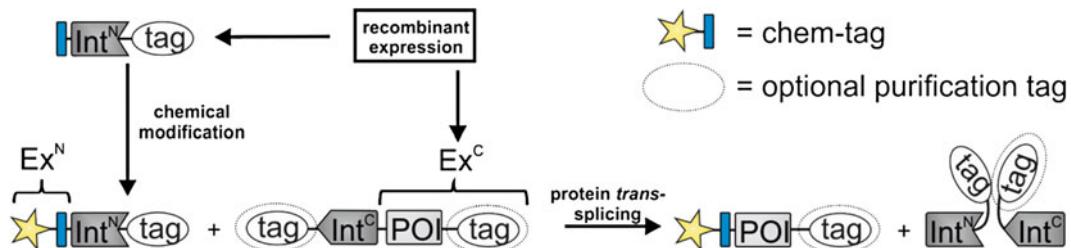


Fig. 2 Concept of chemical-tag labeling of proteins using the fully recombinant *Npu* DnaE intein. The Int^N fragment is expressed in fusion with a short peptide sequence, termed chemical tag or chem-tag. The latter is then chemically modified with a selective bioconjugation reaction. In a second step, the purified and labeled chem-tag- Int^N construct is added to an Int^C -POI fusion protein (POI = protein of interest). Spontaneous protein *trans-splicing* gives rise to the splice product, the desired labeled POI. Affinity tags associated at the inteins may further facilitate the subsequent purification of the labeled POI

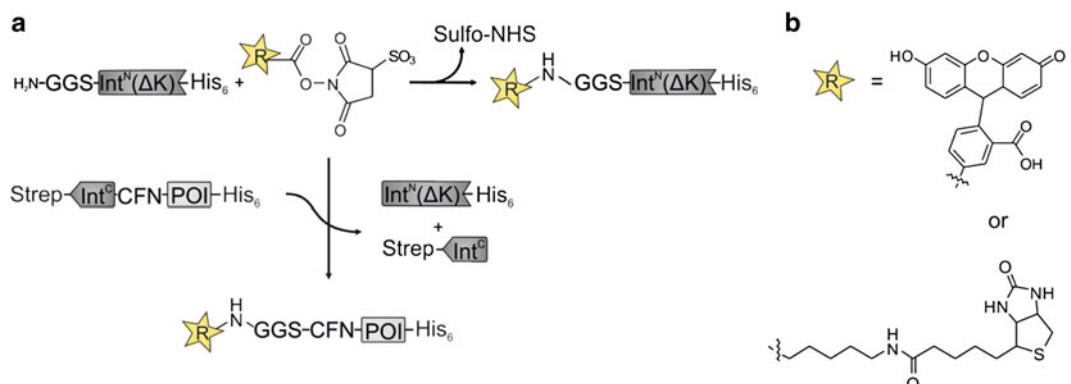


Fig. 3 Reaction scheme for amine-tag approach. (a) All internal lysines were removed in the Int^N fragment. The N-terminal amine group of the GGS- Int^N - His_6 construct is modified with amine-reactive reagents. This approach allows the use of simple amine-reactive bioconjugation reagents without the problem of labeling lysine residues in the POI. (b) Chemical structures of the labeling reagents described here

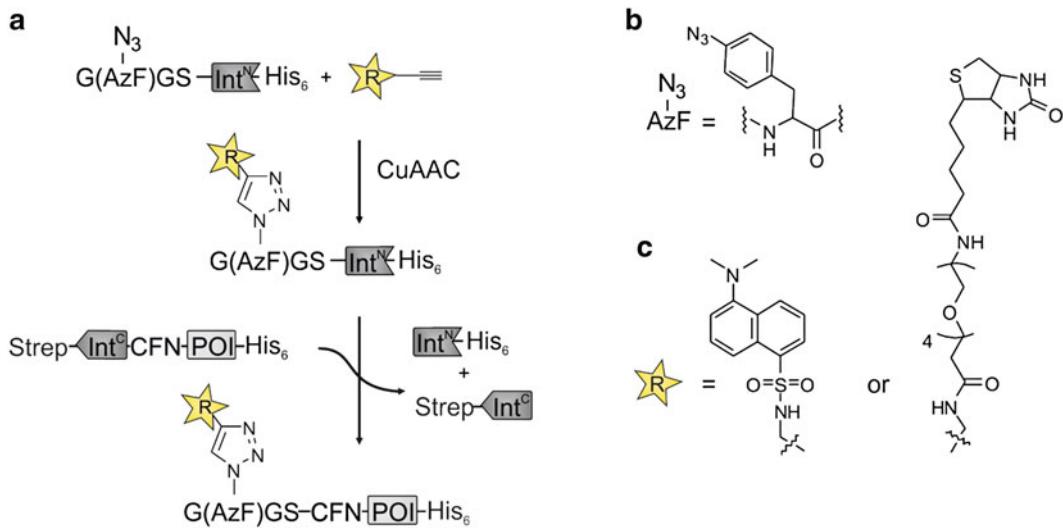


Fig. 4 Reaction scheme for click-tag approach. (a) The unnatural amino acid *p*-azidophenylalanine (AzF) is incorporated into the short click-tag fused to the Int^{N} fragment. Tag labeling is performed by the copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC). One potential advantage of this approach over direct incorporation of the unnatural amino acid is that the POI is not exposed to the CuAAC reaction conditions involving harmful and toxic copper ions. (b) Chemical structure of the incorporated unnatural amino acid AzF. (c) Chemical structures of the labeling reagents described here

2. CuAAC-mediated modification of the unnatural amino acid *p*-azidophenylalanine (AzF) within the tag, termed click-tag approach (Fig. 4).

In the accompanying chapter we describe another split-intein-related protocol which uses one fully synthetic intein fragment that contains the synthetic label to be transferred to the POI [42]. For this alternative approach termed semisynthetic protein *trans-splicing* one intein fragment needs to be as short as possible to be conveniently amenable to solid-phase peptide synthesis [43]. Most split inteins fulfilling this criterion (max 25 aa) are artificially split, except for one recent example [31], and are therefore outperformed by the naturally split *Npu* DnaE intein, in particular with regard to intein fragment affinity and fusion protein solubility.

2 Material

2.1 Plasmid Construction

1. Use plasmid pVS26 (encoding MGGS- $\text{Int}^{\text{N}}(\Delta K)$ - His_6) for the amine-tag approach [21]; plasmid pVS35 (encoding MG(Uaa) GS- Int^{N} - His_6) for the click-tag approach [21]. Plasmid pVS41 (Strep- Int^{C} -eGFP- His_6 ; see Fig. 5a) serves in both approaches as a template to construct the expression plasmid for the purified protein of interest (POI) and pVS37 (Ig κ -HA-Trx- Int^{C} -myc-TM-mCherry; see Fig. 5b) as a template for labeling of a protein on the cell surface [21] (see Note 1).

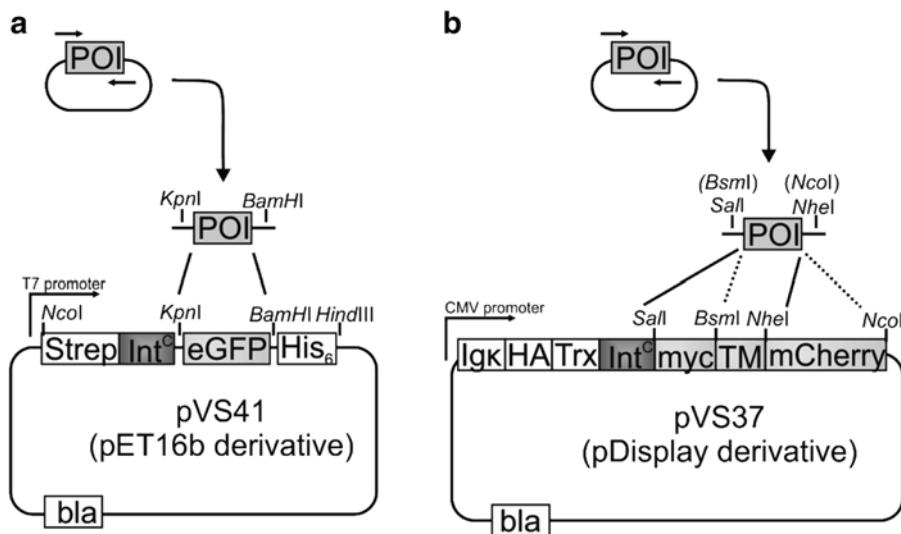


Fig. 5 Cloning strategy for expression vectors. The plasmid pVS41 is based on pET16b (Novagen) and serves as template to express Int^c fusions with the protein of interest (POI) in *E. coli* to perform splice reactions in vitro. Replace the insert encoding eGFP using the *Kpn*I site and the *Bam*HII or *Hind*III sites. Plasmid pVS37 is a shuttle vector for mammalian cells based on pDisplay (Invitrogen). The insert encoding the POI can be inserted with either keeping the transmembrane region TM and the mCherry fluorescent protein or with deleting either of them. (Strep = StreptagII; Igκ = signal sequence from antibody light chain; HA = hemagglutinin tag; Trx = *E. coli* thioredoxin; myc = myc tag; TM = transmembrane region from the PDGF receptor)

2. Restriction enzymes *Kpn*I, *Bam*HII, *Hind*III, *Sal*I, *Bsm*I, *Nhe*I, or *Nco*I, corresponding buffer and primers.
3. T4-DNA-Ligase.

2.2 Protein Purification

1. Competent *E. coli* BL21 (DE3) cells.
2. LB medium: 5 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract, pH 7.5.
3. Minimal media: 9.79 g/L Na₂HPO₄, 13.6 g/L KH₂PO₄, 0.50 g/L NaCl, 5–10 % Glycerin, 0.1 mM CaCl₂, 100 mM MgSO₄, 30 mg/L Thiamin (filter sterilize), 0.1 % NH₄Cl (filter sterilize), 0.2 % Glucose (w/v, filter sterilize), 22 nM Fe(III)Cl₃ (filter sterilize).
4. Ampicillin stock solution (50 mg/L) and chloramphenicol stock solution (34 mg/L), filter sterilize.
5. Isopropyl-β-thiogalactoside stock solution (400 mM, filter sterilize), Ni²⁺-NTA affinity chromatography.
 - (a) Ni²⁺-NTA agarose.
 - (b) Ni²⁺-NTA buffer A; 50 mM Tris-HCl, 300 mM NaCl, pH 8.0.
 - (c) 5 M imidazole stock solution, pH 8.0.

6. Ni²⁺-NTA affinity chromatography under denaturing conditions.
 - (a) All buffers described in the previous point additionally containing 8 M urea.
7. Dialysis tube (cut-off 14 kDa).

2.3 Chemical Modification of Amine-Tag

1. Borate buffer: 50 mM sodium borate, pH 8.5.
2. Phosphate buffered saline (PBS): 136 mM NaCl, 2.7 mM KCl, 8.2 mM Na₂HPO₄, 1.5 mM KH₂PO₄ pH 7.4.
3. 1 mg/100 µL stock solution *N*-hydroxysuccinimide(NHS)-fluorescein in PBS buffer (pH 8.0).
4. 50 mM NH₄Cl.

2.4 Chemical Modification of Click-Tag

1. Plasmid pEVOL or pSUP encoding the orthogonal AzF-specific tRNA-Synthetase/tRNA_{CUA} pair [44, 45].
2. *p*-Azidophenylalanine (AzF).
3. Alkyne reagents, e.g. dansyl-alkyne (5 mM stock solution in DMF).
4. 50 mM stock solution of tris(2-carboxyethyl)phosphine (TCEP) in H₂O. Adjust to pH 7.0.
5. 1.7 mM stock solution of Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) in DMSO/tBuOH (1:4).
6. 50 mM CuSO₄ in H₂O.
7. 1 M EDTA stock solution in H₂O.

2.5 Transient Transfection of N2a Cells

1. Neuro2a mice cells.
2. Serum-free DMEM (Dulbecco's Modified Eagle's Media) media.
3. Gene Juice transfecting agent.

2.6 Protein trans-Splicing Reaction

1. Splice buffer: 50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, pH 7.0.
2. 4× SDS loading buffer: 500 mM Tris-HCl, 8 % (w/v) SDS, 40 % (v/v) glycerol, 20 % (v/v) β-mercaptoethanol, 4 g/L bromophenol blue, pH 6.8.
3. Freshly prepared DTT stock solution (50 mM). Dissolve 7.7 mg in 1 mL splice buffer.

3 Methods

3.1 Generating the Int^c-POI Fusion Construct

1. See Fig. 5 for the cloning strategies. Amplify the gene encoding your protein of interest (POI) by polymerase chain reaction (PCR) with a 5'- and 3'-terminal extension for the restriction enzymes *Kpn*I and *Bam*HI to generate the

construct Strep-Int^C-CFN-POI-His₆. The short amino acid sequence CFN corresponding to the three natively flanking extein amino acids of the intein will be kept. Alternatively, use the restriction enzymes *Kpn*I and *Hind*III to remove the His₆-tag encoding region. In this case, a stop codon to terminate translation has to be introduced with the insert fragment (stop codon overlaps with *Hind*III site, TAAGCTT) (see Note 2).

To generate a construct suitable for cell surface splicing amplify the gene of interest with the restriction sites *Sal*I and *Nhe*I to generate construct Igκ-HA-Trx-Int^C-CFN-POI-mCherry or *Bsm*I and *Nhe*I to keep the myc tag. Using *Nco*I instead of *Nhe*I will lead to removal of the fragment encoding mCherry.

2. Purify the PCR product by agarose gel electrophoresis and DNA extraction kit of your choice. Digest the PCR product and the respective plasmid with the restriction enzymes.
3. Purify the digested PCR product and vector backbone.
4. Ligate both fragments with T4 DNA ligase to generate a complete plasmid encoding for your Int^C-POI fusion construct. Isolate desired plasmids following transformation of *E. coli* cells.

1. Transform competent *E. coli* BL21 (DE3) cells with pVS26 for amine-tag and pVS35 for click-tag equipped Int^N fragments, respectively. Transform with pVS41 or the generated plasmid derivative encoding your protein of interest for the Int^C-POI fusions. Spread the transformed bacteria onto a LB agar plate containing 100 mg/L ampicillin. Incubate the plate overnight at 37 °C, until visible colonies have grown. In case of pVS35 cotransform with plasmid encoding the orthogonal AzF-specific tRNA-synthetase/tRNA_{CUA} pair which contains a chloramphenicol resistance gene in *E. coli* BL21 (DE3) cells [44, 45] and prepare minimal media for expression.
2. Transfer a single colony to 6–8 mL liquid LB medium containing 100 mg/L ampicillin or additionally 34 mg/L chloramphenicol (in case of construct pVS35) and incubate at 37 °C overnight in a shaker or roller incubator.
3. Use the overnight culture to inoculate the main culture (600 mL LB medium or 600 mL minimal media with the respective antibiotics as described before) in a ratio of 1:100 (v/v) or 1:50 (v/v) for minimal media. Incubate the culture in a rotary shaker at 37 °C under aerobic conditions until it has reached an optical density at 600 nm of 0.6.
4. For incorporation of AzF in minimal media spin down cells (10 min, 3300×*g*) and resuspend them carefully in 100 mL minimal media. Add AzF (1 mM final concentration) and incubate the culture for 5–15 min before induction.

3.2 Expression and Purification of Recombinant Proteins

5. Take 1 mL sample as pre-induction sample, pellet these cells by centrifugation (1 min at $20,000 \times g$) and discard the supernatant. Resuspend cells in 60 μ L 1 \times SDS sample buffer and store at -20 °C.
6. Cool down each culture for 10 min to 28 °C while still shaking and induce the protein production by adding IPTG to final concentration of 0.4 mM (add 0.6 mL from a 400 mM stock solution for 600 mL expression). Incubate the culture for 4–5 h at 28 °C.

These parameters could vary for different proteins.

7. Take 1 mL post-induction sample before harvesting the cells, and treat as described before.
8. Harvest the cells by centrifugation (30 min at $3,000 \times g$ and 4 °C) in a precooled centrifuge rotor. Discard the supernatant. The pellet can be stored at -20 °C for a prolonged period of time, either before or after resuspension in the buffer for cell lysis.
9. Resuspend the cell pellets in 10 mL Ni²⁺-NTA buffer A with 8 M urea and 20 mM imidazole (add 40 μ L from a 5 M stock solution). Purification of pVS42 or derivatives could be also purified under native conditions at 4 °C. A second purification step of this construct by using the StrepII-tag is possible.
10. Rupture the cells using your preferred method and separate the soluble fraction from insoluble material and cell debris by centrifugation (20 min at $25,000 \times g$ and 4 °C).
11. Described is a protocol for manual, gravity-flow chromatography. We use disposable columns with the diameter of 1.5 cm and fill these with Ni²⁺-NTA agarose. Alternatively, an FPLC chromatography system with a buffer gradient can be used. Load the supernatant to an equilibrated Ni²⁺-NTA column (bed volume 2.5 mL). Take a sample of 6 μ L of the collected flowthrough for later analysis and mix it with 2 μ L of 4 \times SDS sample buffer.
12. Wash the column two times with 25 mL (ten column volumes; CV) Ni²⁺-NTA buffer A with 20 mM imidazole. Wash the column with 15 mL Ni²⁺-NTA buffer A containing 40 mM imidazole. Elute the protein by addition of Ni²⁺-NTA buffer A with 250 mM imidazole from column. Collect ten fractions of 1 mL each and place on ice. Take 9 μ L from each fraction and mix them with 3 μ L of 4 \times SDS sample buffer. For proteins from pVS26 and pVS35 include 8 M urea in all Ni²⁺-NTA buffers to purify under denaturating conditions.
13. Analyze all samples by SDS-PAGE to check purity and solubility. Load 2.5 μ L of pre- and post-induction, pellet, flowthrough, wash fraction samples, and 4 μ L of the elution samples onto the gel. Stain the gel with coomassie brilliant blue and pool all

elution fractions containing purified protein. Dialyze the pooled fractions of the amine-tag-Int^N-His₆ and click-tag-Int^N-His₆ constructs against the buffer preferred for the chemical conjugation step (*see* Subheadings 3.4 and 3.5). Dialyze the Strep-Int^C-POI-His₆ protein three times against a 100-fold excess volume of splice buffer. Add 2 mM DTT during the first dialysis step. Exchange buffer twice and add 10 % glycerol to the last dialysis step (to protect the purified protein samples upon freezing for storage). If protein purity is not sufficient, the Strep-Int^C-POI-His₆ protein can be purified in a second step by streptactin affinity chromatography or other suitable chromatography techniques according to the characteristics of your POI. Dialyze proteins from pVS26 and pVS35 at room temperature to remove urea (*see Note 3*).

14. Determine the protein concentration of the dialyzed samples using the calculated molecular extinction coefficient at 280 nm and aliquot protein solution in suitable volumes. Flash-freeze them in liquid nitrogen and store the proteins at -80 °C.

3.3 Transient Transfection of N2a Cells

1. Seed 10⁵ cells on a 35 mm cell culture dish.
2. Transfect cells with a solution containing 4 µL Gene Juice and 1 µg DNA in a serum-free DMEM (total volume 100 µL).
3. Cultivate cells for another 24 h.

3.4 Amine-Tag Modification

1. Dialyze purified protein (GGS-Int^N-His₆) against ddH₂O, flash-freeze it in liquid nitrogen and lyophilize.
2. Dissolve 1 mg/mL protein in borate buffer (pH 8.5).
3. Add a 24-fold molar excess (1.8 mM) of NHS-fluorescein in DMF or a fivefold excess of sulfo-NHS-LC-Biotin in PBS buffer.
4. Incubate the reaction for 2–4 h at 4 °C.
5. Quench by adding 50 mM NH₄Cl for 30 min and purify the protein by Ni²⁺-NTA chromatography.
6. Perform a buffer exchange to splice buffer for in vitro splicing and PBS buffer for cell surface splicing afterwards (*see Note 4*).

3.5 Click-Tag Modification

1. Mix purified protein (G(AzF)GS-Int^N-His₆) at 20–100 µM in PBS buffer with the alkyne reagent (two to fivefold excess from a 10 mM stock solution in DMF). Test the reaction in a reaction volume of 50 µL (*see Note 5*).
2. Add the following compounds from their stock solutions with the mentioned final concentrations: 1 mM TCEP, a mixture of TBTA (100 µM) in DMSO/tBuOH (1:4) and 1 mM CuSO₄ in H₂O.
3. Incubate the reaction for 30 min at room temperature. Note that longer reaction times may cause precipitation of the

protein. Quench by adding EDTA to a final concentration of 10 mM.

4. Dialyze to separate off from copper ions and excess small molecule reagents. Remove EDTA in the last dialysis step to prepare for Ni²⁺-NTA chromatography.
5. Purify the protein by Ni²⁺-NTA and dialyze against PBS buffer or splicing buffer.
1. Set up the protein of interest (Strep-Int^C-POI-His₆; at 10–20 μM final concentration) in splicing buffer with 2 mM DTT. Test the splice reaction in a 50–100 μL volume. Pre-incubate at 25 °C for 10 min.
2. Add a pre-incubated (25 °C, 10 min) modified amine-tag or click-tag protein to the same final concentration (10–20 μM) in splice buffer with 2 mM DTT (see Note 6). In order to possibly increase reaction yields, the amine-tag or click-tag proteins can also be used at three to fivefold molar excess.
3. Remove an aliquot from each protein component before mixing in the same ratio and mix both aliquots directly in 4× SDS buffer. This is the prereaction sample for later analysis.
4. Start splicing reaction by mixing the prepared proteins and incubate the reaction at 25 °C.
5. Take samples at defined time points, mix with 4× SDS buffer, and boil at 95 °C for 10 min. Check progress of the reaction by SDS-PAGE followed by Coomassie-staining or western blot analysis (see Notes 7 and 8).

3.6.2 Modification of Cell Surface Presented Proteins

1. Wash N2a cells once with serum-free DMEM 24 h after transfection.
2. Add 5 μM (final concentration) of the amine-tag or click-tag Int^N fusion protein in serum-free DMEM. Either include DTT to a final concentration of 1 mM or use Int^N proteins that were freshly reduced with DTT followed removal of the reducing agent, for example by dialysis.
3. Incubate for 1 h at 37 °C to allow for protein *trans*-splicing. Note that shorter times down to 10 min have been reported for this intein and may be sufficient [37].
4. Wash the cells two times with serum-free DMEM and three times with PBS buffer. When using click-tag protein G(AzF-biotin)GS-Int^N-His₆ incubate with streptavidin-Alexa488 conjugate for 1 h to analyze spliced proteins by fluorescence microscopy.
5. Fix the cells with 1 mL 0.5 % paraformaldehyde in PBS at 37 °C for 15 min and wash five times with PBS buffer. Finally add 2 mL of PBS buffer. Monitor the Alexa488 signal by fluorescence microscopy with an argon-laser (488 nm) (see Note 9).

4 Notes

1. Although inteins generally can splice out of heterologous sequences, they exhibit some preference for the native sequence context. Therefore, as a rule of thumb, three to five amino acids are typically kept on either side of the intein. These residues are part of the extein sequences and will appear in the splice product. It might therefore be of interest to minimize or alter this sequence stretch to obtain a splice product with the desired sequence. Only the +1 residue flanking the intein C-terminally is strictly required for efficient protein splicing. This residue is a Cys in case of the *Npu* DnaE intein. Several studies have shown the high tolerance of the *Npu* DnaE intein for other flanking amino acids. The accepted variability in the C-terminal residues is more restricted than in the N-terminal residues [33, 46, 47]. Therefore we kept the sequence CFN at the C-terminal splice junction. The N-terminal tags MGGS and MG(AzF)GS of the amine-tag and click-tag approaches, respectively, represent minimal, nonnative flanking residues. At least for the MGGS tag we could show by mass spectrometry that the starting methionine is removed after expression in *E. coli* to give a GGS tag. Other tag sequences will very likely be possible, if required, but remain to be tested.
2. After the splice reaction a reaction mixture is obtained. Next to the desired splice product with the transferred labeled tag, the liberated intein fragments and unreacted precursor proteins will remain. The intein fragments will form a high-affinity, noncovalent complex. In order to purify the desired splice product subsequent chromatographic purification steps are required and the optimal procedure will depend on the nature of the protein of interest (POI). The internal affinity tags fused to the C-terminus of the Int^N and the N-terminus of the Int^C fragments may also be useful to remove precursor proteins of intein fragment by-products.
3. For the click-tag approach no amine-containing buffers such as Tris–HCl should be used because these inactivate the copper-catalyst. We have good experience with phosphate and HEPES buffer.
4. Chemical modification of the chemical tags may not always be complete. AzF is not fully metabolically stable and may be reduced to *p*-aminophenylalanine. Acylation of the N-terminal amine group in the amine-tag approach may not always be driven to completion. The Int^N fragment of the *Npu* DnaE intein is intrinsically disordered to a large part [53] and this may lead to partial sequestration of the N-terminus and therefore reduced accessibility.
5. The chemical labeling strategies allow a high flexibility with regard to the labeling reagents. There is a broad range of commercial com-

pounds available, e.g. NHS-esters, isothiocyanate reagents like FITC, and sulfonyl chlorides for amine modification [1], as well as terminal alkynes for CuAAC [48, 49], various forms of cyclic alkynes used in strain-promoted alkyne-azide cycloaddition (SPAAC) [50] and phosphines for the Staudinger ligation [51]; all of which could be employed for labeling of the click-tag. We have so far only explored the CuAAC reaction because of its robustness and strict bioorthogonality with regard to cysteine residues (SPAAC reagents may label cysteines) [52]. Since the copper catalyst is only used to label the tagged intein fragment in the first step and not in contact with the POI during the protein *trans*-splicing reaction of the second step, its potential impact on protein integrity or on cellular toxicity is circumvented.

6. DTT is used as a reducing agent, however, it can also act as a nucleophile and cleave the thioester formed by the intein during the splicing reaction. If the protein context should slow down the reaction considerably, the cleavage reaction may become significant. This problem can be circumvented by used TCEP as reducing agent.
7. The use of the amine-tag is limited to N-terminal protein labeling. However, for the click-tag a similar approach would be conceivable for C-terminal labeling. Furthermore, by using different established unnatural amino acids incorporated with the suppression technology, also other bioorthogonal reactions for their modification would be possible.
8. The substitution of the four lysines by arginines in the *Npu* DnaE Int^N for the amine-tag strategy reduced the rates of protein *trans*-splicing by three to fourfold and the obtainable yields by 5–10 % compared to the wildtype intein [21]. Nevertheless, the mutated intein is still among the best known split intein systems.
9. To label proteins on the surface of living cells we have only tested the described example with the transmembrane region of the PDGF receptor so far. Due to the modularity of the approach we would expect that other transmembrane proteins will also be accessible, however, this remains to be tested. The thioredoxin protein fused N-terminally to the Int^C fragment was found to be beneficial to correct protein localization, probably because a stably fold protein domain would help to shuttle the intrinsically disordered Int^C fragment [53] through the secretory pathway.

Acknowledgements

We thank Peter G. Schultz (The Scripps Research Institute) for providing plasmids for AzF incorporation. Financial support was kindly provided by the DFG (SPP1623).

References

1. Hermanson GT (2010) Bioconjugate techniques, 2nd edn. Elsevier Science, Burlington
2. Banghart M, Borges K, Isacoff E, Trauner D, Kramer RH (2004) Light-activated ion channels for remote control of neuronal firing. *Nat Neurosci* 7(12):1381–1386
3. Stephanopoulos N, Francis MB (2011) Choosing an effective protein bioconjugation strategy. *Nat Chem Biol* 7(12):876–884
4. Lin YA, Chalker JM, Davis BG (2009) Olefin metathesis for site-selective protein modification. *Chembiochem* 10(6):959–969
5. Chalker JM, Bernardes GJ, Lin YA, Davis BG (2009) Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem Asian J* 4(5):630–640
6. Griffin BA, Adams SR, Jones J, Tsien RY (2000) Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Methods Enzymol* 327:565–578
7. Gautier A, Juillerat A, Heinis C, Correa IR Jr, Kindermann M, Beaufils F, Johnsson K (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* 15(2):128–136
8. Los GV, Wood K (2007) The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol Biol* 356:195–208
9. Chen Z, Jing C, Gallagher SS, Sheetz MP, Cornish VW (2012) Second-generation covalent TMP-tag for live cell imaging. *J Am Chem Soc* 134(33):13692–13699
10. Mizukami S, Watanabe S, Hori Y, Kikuchi K (2009) Covalent protein labeling based on noncatalytic beta-lactamase and a designed FRET substrate. *J Am Chem Soc* 131(14):5016–5017
11. Mao H, Hart SA, Schink A, Pollok BA (2004) Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* 126(9):2670–2671
12. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3(11):707–708
13. Chen I, Howarth M, Lin W, Ting AY (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* 2(2):99–104
14. Uttamapinant C, Sanchez MI, Liu DS, Yao JZ, Ting AY (2013) Site-specific protein labeling using PRIME and chelation-assisted click chemistry. *Nat Protoc* 8(8):1620–1634
15. Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* 102(44):15815–15820
16. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* 266(5186):776–779
17. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* 72:249–289
18. Hackenberger CP, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. *Angew Chem Int Ed Engl* 47(52):10030–10074
19. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413–444
20. Davis L, Chin JW (2012) Designer proteins: applications of genetic code expansion in cell biology. *Nat Rev Mol Cell Biol* 13(3):168–182
21. Schütz V, Mootz HD (2014) Click-tag and amine-tag: new chemical tag approaches for efficient protein labeling in vitro and on live cells using the naturally split Npu DnaE intein. *Angew Chem Int Ed Engl* 53:4113–4117
22. Volkmann G, Mootz HD (2013) Recent progress in intein research: from mechanism to directed evolution and applications. *Cell Mol Life Sci* 70(7):1185–1206
23. Noren CJ, Wang J, Perler FB (2000) Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed Engl* 39(3):450–466
24. Shah NH, Muir TW (2014) Inteins: nature's gift to protein chemists. *Chem Sci* 5:446–461
25. Dhar T, Kurpiers T, Mootz HD (2011) Extending the scope of site-specific cysteine bioconjugation by appending a prelabelled cysteine tag to proteins using protein trans-splicing. *Methods Mol Biol* 751:131–142
26. Kurpiers T, Mootz HD (2008) Site-specific chemical modification of proteins with a prelabelled cysteine tag using the artificially split Mxe GyrA intein. *Chembiochem* 9(14):2317–2325
27. Kurpiers T, Mootz HD (2007) Regioselective cysteine bioconjugation by appending a labeled cysteine tag to a protein by using protein splicing in trans. *Angew Chem Int Ed Engl* 46(27):5234–5237
28. Brenzel S, Cebi M, Reiss P, Koert U, Mootz HD (2009) Expanding the scope of protein trans-splicing to fragment ligation of an integral membrane protein: towards modulation of porin-based ion channels by chemical modification. *Chembiochem* 10(6):983–986

29. Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res* 37(8):2560–2573
30. Carvajal-Vallejos P, Pallisse R, Mootz HD, Schmidt SR (2012) Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem* 287(34):28686–28696
31. Thiel IV, Volkmann G, Pietrokovski S, Mootz HD (2014) An atypical naturally split intein engineered for highly efficient protein labeling. *Angew Chem Int Ed Engl* 53(5):1306–1310
32. Shah NH, Dann GP, Vila-Perello M, Liu Z, Muir TW (2012) Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. *J Am Chem Soc* 134(28):11338–11341
33. Iwai H, Zuger S, Jin J, Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett* 580(7):1853–1858
34. Zettler J, Schutz V, Mootz HD (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* 583(5):909–914
35. Shah NH, Vila-Perello M, Muir TW (2011) Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein. *Angew Chem Int Ed Engl* 50(29):6511–6515
36. Mohlmann S, Bringmann P, Greven S, Harrenga A (2011) Site-specific modification of ED-B-targeting antibody using intein-fusion technology. *BMC Biotechnol* 11:76
37. Dhar T, Mootz HD (2011) Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the Npu DnaE intein. *Chem Commun (Camb)* 47(11):3063–3065
38. Vila-Perello M, Liu Z, Shah NH, Willis JA, Idoyaga J, Muir TW (2013) Streamlined expressed protein ligation using split inteins. *J Am Chem Soc* 135(1):286–292
39. Borra R, Dong D, Elnagar AY, Woldemariam GA, Camarero JA (2012) In-cell fluorescence activation and labeling of proteins mediated by FRET-quenched split inteins. *J Am Chem Soc* 134(14):6344–6353
40. Subramanyam P, Chang DD, Fang K, Xie W, Marks AR, Colecraft HM (2013) Manipulating L-type calcium channels in cardiomyocytes using split-intein protein transsplicing. *Proc Natl Acad Sci U S A* 110(38):15461–15466
41. Ramirez M, Valdes N, Guan D, Chen Z (2013) Engineering split intein DnaE from *Nostoc punctiforme* for rapid protein purification. *Protein Eng Des Sel* 26(3):215–223
42. Matern CJ, Bachmann A-L, Thiel IV, Volkmann G, Wasmuth A, Binschik J, Mootz HD (2014) Ligation of synthetic peptides to proteins using semisynthetic protein trans-splicing. In: Gautier A, Hinner M (eds) *Site-specific protein labeling*. Methods Mol Biol 1266:129–143
43. Mootz HD (2009) Split inteins as versatile tools for protein semisynthesis. *Chembiochem* 10(16):2579–2589
44. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc* 124(31): 9026–9027
45. Young TS, Ahmad I, Yin JA, Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J Mol Biol* 395(2):361–374
46. Cherian M, Pedamallu CS, Tori K, Perler F (2013) Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. *J Biol Chem* 288(9):6202–6211
47. Shah NH, Eryilmaz E, Cowburn D, Muir TW (2013) Extein residues play an intimate role in the rate-limiting step of protein trans-splicing. *J Am Chem Soc* 135(15):5839–5847
48. Tornoe CW, Christensen C, Meldal M (2002) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* 67(9):3057–3064
49. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew Chem Int Ed Engl* 41(14):2596–2599
50. de Almeida G, Sletten EM, Nakamura H, Palaniappan KK, Bertozzi CR (2012) Thiacycloalkynes for copper-free click chemistry. *Angew Chem Int Ed Engl* 51(10): 2443–2447
51. Agard NJ, Baskin JM, Prescher JA, Lo A, Bertozzi CR (2006) A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol* 1(10):644–648
52. Yang H, Srivastava P, Zhang C, Lewis JC (2014) A general method for artificial metalloenzyme formation through strain-promoted azide-alkyne cycloaddition. *Chembiochem* 15(2):223–227
53. Shah NH, Eryilmaz E, Cowburn D, Muir TW (2013) Naturally split inteins assemble through a “capture and collapse” mechanism. *J Am Chem Soc* 135(49):18673–18681

Chapter 11

Phage Selection Assisted by Sfp Phosphopantetheinyl Transferase-Catalyzed Site-Specific Protein Labeling

Bo Zhao, Keya Zhang, Karan Bhuripanyo, Yiyang Wang, Han Zhou, Mengnan Zhang, and Jun Yin

Abstract

Phosphopantetheinyl transferases (PPTase) Sfp and AcpS catalyze a highly efficient reaction that conjugates chemical probes of diverse structures to proteins. PPTases have been widely used for site-specific protein labeling and live cell imaging of the target proteins. Here we describe the use of PPTase-catalyzed protein labeling in protein engineering by facilitating high-throughput phage selection.

Key words Phosphopantetheinyl transferase-catalyzed site-specific protein labeling, Phage display, Protein engineering, Protein immobilization

1 Introduction

Protein labeling with small molecules expands the diversity of the functional groups anchored on the peptide chain. The labelled proteins can function as reporters to register their locations in the cell, to reveal their partnerships with other cellular proteins, and to record their life cycle including expression, posttranslational modification, and degradation in various cellular processes. A good protein labeling method prefers site-specific attachment of the small molecule labels to the target proteins so that the position and stoichiometry of the label can be precisely defined. A good labeling method should also be versatile so that the labels of diverse chemical structures and functionalities can be attached to the target proteins based on their tasks in studying cell biology. A good labeling method should also be fast and of high efficiency so that the labeled proteins can be tracked in real time and in the live cell.

The protein labeling reaction catalyzed by Sfp phosphopantetheinyl transferase can fulfill all these criterions [1]. The native activity of Sfp is to transfer the phosphopantetheinyl (Ppant) group of coenzyme A (CoA) to a specific serine residue of the peptidyl

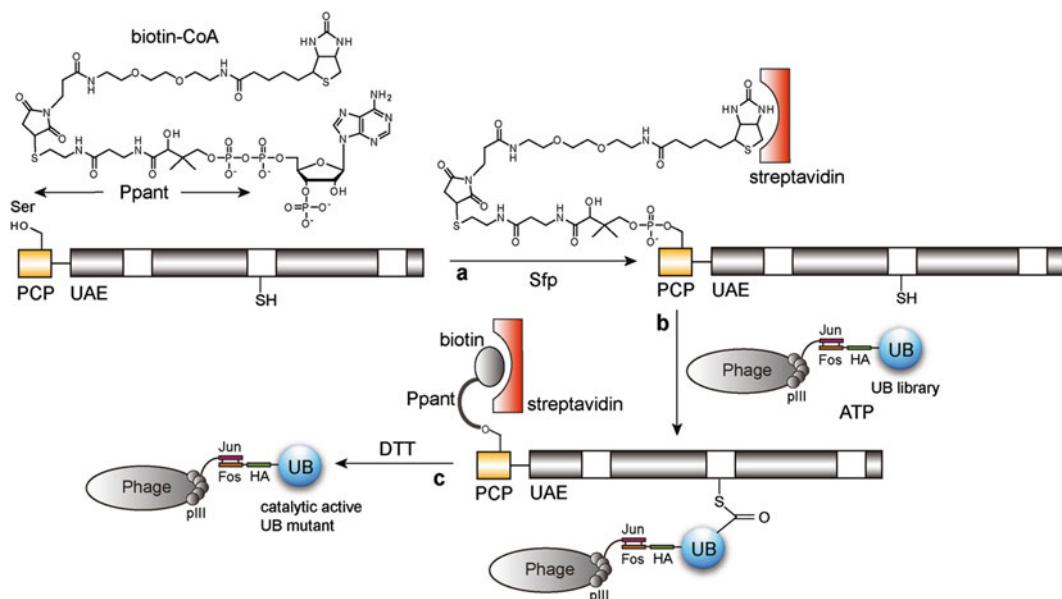


Fig. 1 Sfp-catalyzed labeling of PCP-UAE fusion protein with the biotin-CoA conjugate and phage selection with the UAE enzyme immobilized on the streptavidin plate. (a) Biotin-Ppant group is transferred from biotin-CoA by Sfp to a Ser residue in the PCP domain fused to the N-terminus of UAE. Biotin-labeled UAE is then immobilized on the streptavidin plate. (b) Phage library displaying UB variants reacts with UAE on the plate to form UB~UAE conjugates. (c) After washing the plate, phage displaying catalytic active UB variants is eluted by DTT that cleaves the thioester bond

carrier protein (PCP) domains embedded in the nonribosomal peptide synthetase (NRPS). The Ppant modification of PCP activates NRPS for natural product biosynthesis [2]. Sfp was found to be very promiscuous with the chemical functionalities attached to the terminal thiol of CoA; besides the Ppant arm itself, the enzyme can recognize small molecule—CoA conjugates as substrates and attach small molecule labels to PCP through the Ppant linker (Fig. 1) [3, 4]. Sfp-catalyzed protein labeling is also fast, nearly of quantitative yield, and can be performed on the surface of live cells. All these features make Sfp a very useful tool for site-specific protein labeling.

We previously demonstrated that Sfp can be used to conjugate diverse chemical labels to the target proteins [4, 5], to image cell surface proteins by fluorescent resonance transfer (FRET) [6], and to profile natural product biosynthetic clusters in a single bacteria or in the metagenome [7]. We have also developed small peptide tags named ybbR and S6 that are 11-residue long for protein labeling by Sfp [8, 9]. We later identified an A1 peptide tag that can be specifically labeled with AcPpant, a phosphopantetheinyl transferase (PPTase) from *E. coli* [8]. The A1-AcpS pair and the S6-Sfp pair are orthogonal to each other so that distinctive cell surface receptors can be labeled with different fluorophores to track their movements

in the same cell [8]. Other researchers have used Sfp and AcpS to image polarized secretion of proteins on the yeast cell wall [10], to immobilize proteins on the hydrogel or glass slides or on nanoparticles [11–13], to attach fluorescent labels to neurotoxins, chemokines, and Hedgehog receptors to reveal their trafficking in the cell [14–16]. We have compiled a detailed protocol for Sfp-catalyzed protein labeling for cell imaging studies and it has been reported elsewhere [1]. We recently developed efficient methods for protein engineering by phage display and yeast cell surface display using biotin-labeled proteins generated by Sfp [17–19]. In this chapter, we present the methods of using Sfp to label proteins to facilitate phage selection.

Our research demonstrated that Sfp-catalyzed protein labeling is especially suitable for conjugating affinity probes to proteins to facilitate protein engineering by phage and yeast cell selection. This is because Sfp-catalyzed protein labeling is very fast and is almost of quantitative yield. Target proteins can be freshly labeled with biotin or other chemical probes and used directly for phage or yeast selection without an additional purification step. This significantly increases the selection efficiency and the likelihood of success of the protein engineering experiment. We previously showed that we can use Sfp to label Nrf2 with biotin to select for Keap1 variants that have high affinity for cancer-related Nrf2 mutants by yeast cell sorting [19]. Here we provide an example to use Sfp-labeled protein for phage selection. We use Sfp to site-specifically label ubiquitin (UB) activating enzyme (UAE) with biotin-Ppant to select for phages displaying UB variants that are reactive with UAE (Fig. 1) [17, 18].

UAE is the first enzyme in the UB transfer cascade for protein ubiquitination [20, 21]. It activates the C-terminal carboxylate of UB so that UB can be covalently conjugated to the amino group of the Lys side chain on the cellular proteins. In the activation reaction, UAE binds to both UB and ATP to catalyze the condensation reaction between the C-terminal carboxylate of UB and ATP to form UB-AMP conjugate. A catalytic Cys residue of UAE then reacts with UB-AMP to form UB~UAE thioester in which the C-terminal carboxylate of UB is covalently bound to the catalytic Cys of UAE (Fig. 1). UAE bound UB can then be further transferred to E2 and E3 in the enzymatic cascade on its way to be attached to substrate proteins. In this protocol, we use phage display to profile the interaction between UB C-terminal sequences with UAE. We construct a fusion protein with the PCP domain fused to the N-terminus of UAE. We then use Sfp and biotin-CoA conjugate to label the PCP-UAE fusion with the biotin-Ppant group and immobilize biotin-labeled UAE on the streptavidin plate (Fig. 1). We add the phage library to PCP-UAE coated plate in the presence of ATP to initiate the activation reaction. UB variants that can be recognized by UAE for the activation reaction would

form UB~UAE thioester conjugate. This covalent interaction retains the catalytically active phage particles on the plate and they can then be eluted by dithiothreitol (Fig. 1). Following such a procedure, we identified UB variants with alternative C-terminal sequences that are reactive with UAE [17].

Based on this protocol, readers can develop their own methods to label the target proteins with various chemical probes by Sfp for phage selection.

2 Materials

The materials needed for the synthesis and purification of biotin-CoA conjugate, expression of Sfp, labeling target proteins with Sfp, and phage selection are covered below. All aqueous solutions were prepared with deionized water from a commercial water purifier with a conductivity of $18\text{ M}\Omega$ or higher.

2.1 Materials for the Synthesis of Biotin-CoA Conjugate

1. Biotin maleimide powder. Store at $4\text{ }^{\circ}\text{C}$ until use.
2. Coenzyme A powder in the form of trilithium salt. Store at $-20\text{ }^{\circ}\text{C}$ until use.
3. Dimethylsulfoxide (DMSO), analytical grade.
4. Conjugation reaction buffer (50 mM MES acetate, pH 6.0). Dissolve 9.76 g MES acetate in 1 L water and adjust pH to 6.0.

2.2 Materials for the Purification of Biotin-CoA Conjugate by HPLC

1. Solution A, 0.1 % trifluoroacetic acid (TFA) in water.
2. Solution B, 0.1 % TFA in acetonitrile (ACN). HPLC-grade ACN is used.
3. High performance liquid chromatography (HPLC) machine equipped with a preparation scale C18 reverse phase column.

2.3 Materials for the Expression and Purification of the Sfp Enzyme

1. The plasmid for Sfp expression. The Sfp gene was cloned into the pET29 expression plasmid with a C-terminal $6\times\text{His}$ tag. The plasmid also has a kanamycin resistant marker.
2. Luria-Bertani (LB) media. Dissolve 25 g LB powder in 1 L water. Autoclave the media with a liquid cycle for 20 min.
3. LB-Agar plate supplemented with kanamycin. Mix 25 g LB and 15 g agar in 1 L water. Autoclave the media with a liquid cycle for 20 min. After the solution is cooled down to around $50\text{ }^{\circ}\text{C}$, add kanamycin to a final concentration of 50 $\mu\text{g}/\text{mL}$. Mix melted LB-agar well and pour approximately 40 mL media to each plate. Close the lid and leave the plates at room temperature to cool off. Inverse the plates and store the plates at $4\text{ }^{\circ}\text{C}$ after the agar solidifies.
4. Kanamycin solution (50 mg/mL). Dissolve 50 mg of kanamycin in 1 mL water. Filter sterilize the solution and store the solution in a $-20\text{ }^{\circ}\text{C}$ freezer.

5. Isopropyl β -D-1-thiogalactopyranoside (IPTG) (0.1 M). Dissolve 1.2 g IPTG in deionized water to a final volume of 50 mL. Filter sterilize and store in aliquots at -20 °C.
6. Lysis buffer for Ni-NTA affinity chromatography (50 mM NaH₂PO₄, 300 mM NaCl, and 5 mM imidazole). Dissolve 6.9 g NaH₂PO₄, 17.6 g NaCl, and 0.34 g imidazole in 1 L water. Adjust pH to 8.0.
7. Wash buffer for Ni-NTA affinity chromatography (50 mM NaH₂PO₄, 300 mM NaCl, and 20 mM imidazole). Dissolve 6.9 g NaH₂PO₄, 17.6 g NaCl, and 1.36 g imidazole in 1 L water. Adjust pH to 8.0.
8. Elution buffer for Ni-NTA affinity chromatography (50 mM NaH₂PO₄, 300 mM NaCl, and 200 mM imidazole). Dissolve 6.9 g NaH₂PO₄, 17.6 g NaCl, and 13.6 g imidazole in 1 L water. Adjust pH to 8.0.
9. Dialysis buffer for Sfp (50 mM Tris-HCl, 10 mM MgCl₂, 5 mM dithiothreitol (DTT), 5 % (v/v) glycerol). Dissolve 6.1 g Tris-OH, 2.1 g MgCl₂, 0.77 g DTT, and 50 mL glycerol in 1 L water. Adjust pH to 8.0.
10. Sfp labeling reaction buffer (50 mM HEPES, 10 mM MgCl₂, pH 7.5). Dissolve 11.9 g HEPES and 0.95 g MgCl₂ in 1 L water. Adjust pH to 7.5.

2.4 Materials for Phage Preparation and Selection

1. 2YT medium. Dissolve 31 g of 2YT powder in 1 L water. Autoclave the media with a liquid cycle for 20 min.
2. Ampicillin (100 mg/mL). Dissolve 1 g ampicillin powder in 10 mL water. Filter sterilize the solution and store the aliquots in -20 °C freezer.
3. PEG solution. Dissolve 100 g PEG 8000 and 75 g NaCl in 500 mL H₂O. Filter sterilize the solution and store it at room temperature.
4. TBS buffer (10 mM Tris-HCl, 150 mM NaCl, pH 7.5). Dissolve 8.8 g NaCl in 1 L water with the addition of 10 mL 1 M Tris-HCl, pH 7.5.
5. TBS-T buffer (10 mM Tris-HCl, 150 mM NaCl, pH 7.5, 0.05 % (v/v) Tween 20, and 0.05 % (v/v) Triton X-100). Dissolve 0.5 mL Tween 20 and 0.5 mL Triton X-100 in 1 L TBS.
6. UAE reaction buffer (1.5 % BSA, 1 mM ATP, and 50 mM MgCl₂ in TBS). Dissolve 0.15 g bovine serum albumin (BSA), 5.5 mg ATP, and 47.6 mg MgCl₂ in 10 mL TBS.
7. Phage elution buffer (20 mM DTT in TBS). Dissolve 0.31 g DTT in TBS. Filter sterilize the solution and store the solution at 4 °C.

3 Methods

3.1 Synthesizing Biotin-CoA Conjugate

1. Dissolve biotin maleimide (10 mg, 0.019 mmol) in 300 μ L DMSO.
2. Dissolve Coenzyme A trilithium salt (18.2 mg, 0.023 mmol) in 2 mL conjugation reaction buffer.
3. Mix the biotin maleimide and coenzyme A solution and stir the solution at room temperature for overnight.
4. Purify the biotin-CoA product by preparative HPLC on a reversed-phase C18 column. Wash the column with a gradient 0–60 % ACN (solution B) in 0.1 % TFA/water (solution A) over 35 min.
5. Lyophilize the fraction with the biotin-CoA product.
6. Use matrix-assisted laser desorption/ionization (MALDI) operating in the positive ion mode to confirm the identity of biotin-CoA.
7. Dissolve biotin-CoA in water to a concentration of 10 mM. Aliquot and store the solution at –20 °C before use.

3.2 Expressing and Purifying the Sfp Enzyme

1. Transform the pET29-Sfp expression plasmid into *E. coli* BL21(DE3) pLysS chemically competent cells. Plate out the transformation mixture on an LB Agra plate supplemented with 50 μ g/mL kanamycin. Incubate the plates at 37 °C for overnight.
2. Pick a single colony and inoculate 10 mL LB culture with 50 μ g/mL kanamycin. Incubate the culture in a 37 °C shaker for overnight.
3. Use the overnight culture to inoculate 1 L LB with kanamycin. Shake the culture at 37 °C for 4–6 h until the OD₆₀₀ of the culture reaches 0.5.
4. Reduce the temperature of the shaker to 20 °C. Shake for an additional 30 min. Add IPTG to the culture to a final concentration of 1 mM. Incubate the culture overnight.
5. On the next day, harvest the cells by centrifugation (4,000 $\times g$, 15 min). Pour out the supernatant and resuspend the cell pellets in 6 mL lysis buffer with 1 unit/mL DNase I.
6. Lysis the cell by passing the cell suspension twice through a French press machine at a pressure of 16,000 psi.
7. Collect the lysate and precipitate the cell debris by high speed centrifugation (95,000 $\times g$, 30 min).
8. Save the clarified supernatant from the centrifugation bottle. Incubate the supernatant with 1 mL Ni-NTA resin for 3–4 h at 4 °C.

9. Load the supernatant with Ni-NTA resin onto a gravity column. Drain the cell lysate through the column. Wash the resin twice, each time with 10 mL lysis buffer. Wash the resin once with 10 mL wash buffer.
10. Elute the Sfp protein from the column with 6 mL elution buffer. Adjust the flow to 1 drop/s during elution.
11. Add the elution fractions to a dialysis tube with a molecular weight cutoff (MWCO) of 10 kD and dialyze the eluent against the dialysis buffer for Sfp. Allow 4 h or more to equilibrate. Put the dialysis tube into fresh solution and dialyze again for 4 h or more.
12. Concentrate the Sfp solution to more than 5 mg/mL by a Centriprep concentrator. Store the aliquots at -80 °C.

3.3 Labeling PCP-UAE Protein with Biotin-CoA Catalyzed by Sfp and Immobilizing Biotin-Labeled Protein on the Streptavidin Plate

1. Set up 100 µL labeling reaction with 5 µM PCP-UAE, 5 µM biotin-CoA, and 0.3 µM Sfp in the Sfp reaction buffer.
2. Incubate the reaction at room temperature for 1 h.
3. Add 1/3 volume of 3 % BSA in TBS to the labeling reaction so that the final concentration of BSA is 1 % in the reaction mixture.
4. Add 100 µL labeling reaction mixture with 1 % BSA to each well of the streptavidin plate.
5. Equilibrate the plate for 1 h at room temperature to allow the binding of biotin-labeled protein to the plate.
6. Wash the plate with TBS for three times, each time with 250 µL TBS for each well in the plate.
7. Add 150 µL TBS to each well of the plate. Cover the plate with parafilm. Store the plate at 4 °C before use (*see Notes 1 and 2*).

3.4 Preparing the Phage Library for Selection

The phage selection experiment takes 2 days. On day 1, the DNA of the phagemid library is transformed into *E. coli* cells and the production of the UB variants is induced overnight for their display on the phage surface. On the second day, phage library is harvested, and PCP-UAE coated streptavidin plate is used for the selection of catalytically active UB variants.

1. Transform the phagemid library of UB into SS320 super competent cells infected with M13KO7 helper phage. Add the transformed cells to the SOC media and grow for 1 h at 37 °C with shaking.
2. Use the cell culture to inoculate 100 mL 2YT supplemented with ampicillin (100 µg/mL) and kanamycin (50 µg/mL). Shake the cell culture overnight at 37 °C.
3. On the next day, spin down the cells with centrifugation at 2,795 × g. Pour the supernatant into a 500 mL centrifuge

bottle. Add $\frac{1}{4}$ volume of PEG solution to the bottle. Mix the PEG thoroughly with the supernatant. Incubate the bottle on ice for 1 h.

4. Centrifuge the bottle at $9,055 \times g$ for 30 min to precipitate the phage particles. Discard the supernatant. Drain the residue PEG solution from the bottle.
5. Resuspend the phage pellet in 2 mL TBS.
6. Add the phage solution to Eppendorf tubes and centrifuge at $11,336 \times g$ to remove the cell debris.
7. Save the supernatant from the tube as the phage solution for the selection reaction. Store the phage solution at 4°C .

3.5 Phage Selection with Biotin-Labeled PCP-UAE Immobilized on the Streptavidin Plate

1. Empty solution in the wells of the streptavidin plate immobilized with biotin-labeled PCP-UAE.
2. Dilute phage solution in the UAE reaction buffer to about 1×10^{11} phage/mL.
3. Add 100 μL phage solution to each well of the streptavidin plate. Incubate the UB displayed phage with immobilized UAE for 1 h to allow the formation of UB~UAE conjugate.
4. Remove the supernatant in the well by pipetting and wash the wells with TBS-T for 30 times, each time with 250 μL TBS-T for each well.
5. Wash the wells with TBS for an additional 30 times.
6. Add 100 μL phage elution buffer containing DTT to each well and incubate at room temperature for 10 min.
7. Add the eluent to 10 mL of log phase *E. coli* XL1-Blue cells and shake at 37°C for 1 h to allow phage infection of the cells.
8. Plate out the cell culture on LB agar plate supplemented with 2 % (w/v) glucose and 100 $\mu\text{g}/\text{mL}$ ampicillin. Incubate the plates overnight at 37°C .
9. On the next day, scrape the colonies on the plate with a sterilized spatula.
10. Extract plasmid DNA from the cells with a QIAprep Plasmid Miniprep kit. The phagemid DNA is then used for the next round of phage amplification and selection.
11. In parallel to the selection reaction, also set up control reactions excluding key components such as ATP or PCP-UAE immobilization on the streptavidin plate. After each round of selection, titer phage particles eluted from the selection and the control reactions. Successful phage selection should show stepwise enrichment of the phage particles from the eluent of the selection reaction comparing to the controls. During iterative rounds of selection, the number of the input phage particles, the concentration of E1 enzymes, and the reaction time can also be decreased in each round to increase the stringency of phage selection.

4 Notes

1. Because Sfp-catalyzed protein labeling is highly specific, the biotin-labeled protein in the reaction mixture can be directly used for binding to the streptavidin plate despite the presence of other components of the labeling reaction such as the Sfp enzyme, excess biotin-CoA, and AMP. To increase the selection efficiency and avoid nonspecific binding of biotin-CoA with proteins on the phage surface, the labeling reaction mixture can be purified by desalting with a Centriprep concentrator or by passing through a desalting column.
2. We typically use freshly labeled biotin-UAE conjugate for the selection of UB variants by phage display. In this way the enzymatic activity of UAE would be the highest for the selection reaction. On the other hand the biotin-Ppant conjugate attached to the PCP tag is very stable and the labeled protein can be stored for years in a -20 °C freezer without losing the label on the protein.

Acknowledgements

This work was supported by a National Science Foundation CAREER award (1057092) and a National Institute of Health grant 1R01GM104498 to J.Y.

References

1. Yin J, Lin AJ, Golan DE, Walsh CT (2006) Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat Protoc* 1: 280–285
2. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT (1996) A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem Biol* 3:923–936
3. La Clair JJ, Foley TL, Schegg TR, Regan CM, Burkart MD (2004) Manipulation of carrier proteins in antibiotic biosynthesis. *Chem Biol* 11:195–201
4. Yin J, Liu F, Li X, Walsh CT (2004) Labeling proteins with small molecules by site-specific posttranslational modification. *J Am Chem Soc* 126:7754–7755
5. Yin J, Liu F, Schinke M, Daly C, Walsh CT (2004) Phagemid encoded small molecules for high throughput screening of chemical libraries. *J Am Chem Soc* 126:13570–13571
6. Yin J, Lin AJ, Buckett PD, Wessling-Resnick M, Golan DE, Walsh CT (2005) Single-cell FRET imaging of transferrin receptor trafficking dynamics by Sfp-catalyzed, site-specific protein labeling. *Chem Biol* 12:999–1006
7. Yin J, Straight PD, Hrvatin S, Dorrestein PC, Bumpus SB, Jao C, Kelleher NL, Kolter R, Walsh CT (2007) Genome-wide high-throughput mining of natural-product biosynthetic gene clusters by phage display. *Chem Biol* 14:303–312
8. Zhou Z, Cironi P, Lin AJ, Xu Y, Hrvatin S, Golan DE, Silver PA, Walsh CT, Yin J (2007) Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem Biol* 2:337–346
9. Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* 102:15815–15820
10. George N, Pick H, Vogel H, Johnsson N, Johnsson K (2004) Specific labeling of cell

- surface proteins with chemically diverse compounds. *J Am Chem Soc* 126:8896–8897
11. Waichman S, Bhagawati M, Podoplelova Y, Reichel A, Brunk A, Paterok D, Pichler J (2010) Functional immobilization and patterning of proteins by an enzymatic transfer reaction. *Anal Chem* 82:1478–1485
 12. Wong LS, Okrasa K, Micklefield J (2010) Site-selective immobilisation of functional enzymes on to polystyrene nanoparticles. *Org Biomol Chem* 8:782–787
 13. Wong LS, Thirlway J, Micklefield J (2008) Direct site-selective covalent protein immobilization catalyzed by a phosphopantetheinyl transferase. *J Am Chem Soc* 130: 12456–12464
 14. Band PA, Blais S, Neubert TA, Cardozo TJ, Ichtchenko K (2010) Recombinant derivatives of botulinum neurotoxin A engineered for trafficking studies and neuronal delivery. *Protein Expr Purif* 71:62–73
 15. Kawamura T, Stephens B, Qin L, Yin X, Dores MR, Smith TH, Grimsey N, Abagyan R, Trejo J, Kufareva I, Fuster MM, Salanga CL, Handel TM (2014) A general method for site specific fluorescent labeling of recombinant chemo-kines. *PLoS One* 9:e81454
 16. Wang Y, Zhou Z, Walsh CT, McMahon AP (2009) Selective translocation of intracellular Smoothened to the primary cilium in response to Hedgehog pathway modulation. *Proc Natl Acad Sci U S A* 106:2623–2628
 17. Zhao B, Bhuripanyo K, Schneider J, Zhang K, Schindelin H, Boone D, Yin J (2012) Specificity of the E1-E2-E3 enzymatic cascade for ubiquitin C-terminal sequences identified by phage display. *ACS Chem Biol* 7:2027–2035
 18. Zhao B, Bhuripanyo K, Zhang K, Kiyokawa H, Schindelin H, Yin J (2012) Orthogonal ubiquitin transfer through engineered E1-E2 cascades for protein ubiquitination. *Chem Biol* 19:1265–1277
 19. Zhang K, Li H, Bhuripanyo K, Zhao B, Chen TF, Zheng N, Yin J (2013) Engineering new protein-protein interactions on the beta-propeller fold by yeast cell surface display. *Chembiochem* 14:426–430
 20. Lee I, Schindelin H (2008) Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* 134:268–278
 21. Schulman BA, Harper JW (2009) Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* 10:319–331

Chapter 12

Site-Specific Biotinylation of Purified Proteins Using BirA

Michael Fairhead and Mark Howarth

Abstract

The binding between biotin and streptavidin or avidin is one of the strongest known non-covalent biological interactions. The (strept)avidin-biotin interaction has been widely used for decades in biological research and biotechnology. Therefore labeling of purified proteins by biotin is a powerful way to achieve protein capture, immobilization, and functionalization, as well as multimerizing or bridging molecules. Chemical biotinylation often generates heterogeneous products, which may have impaired function. Enzymatic biotinylation with *E. coli* biotin ligase (BirA) is highly specific in covalently attaching biotin to the 15 amino acid AviTag peptide, giving a homogeneous product with high yield. AviTag can conveniently be added genetically at the N-terminus, C-terminus, or in exposed loops of a target protein. We describe here procedures for AviTag insertion by inverse PCR, purification of BirA fused to glutathione-S-transferase (GST-BirA) from *E. coli*, BirA biotinylation of purified protein, and gel-shift analysis by SDS-PAGE to quantify the extent of biotinylation.

Key words Neutravidin, Streptavidin-biotin, Femtomolar, Nanotechnology, Bionanotechnology

1 Introduction

Biotin is a cofactor for carboxylase enzymes, present in all living organisms [1]. Streptavidin binds to biotin with a K_d of 4×10^{-14} M [2]. Streptavidin-biotin binding is rapid, specific, and can still occur under conditions where most other proteins have denatured, such as high temperatures or 6 M guanidinium hydrochloride or 1 % sodium dodecyl sulfate (SDS) [3]. A breakthrough for the use of biotin for protein modification was harnessing the cell's natural machinery for biotin conjugation, using the *E. coli* enzyme BirA to achieve precise biotin modification [4]. The natural substrate of BirA is the Biotin Carboxyl Carrier Protein (BCCP), requiring fusion of at least 75 residues to the target protein [4]. However, phage display selection enabled the development of the AviTag (also known as the Acceptor Peptide, AP), which is superior to BCCP as a BirA substrate but only 15 amino acids in length [5] (see Fig. 1), so extending the range of protein sites amenable to site-specific enzymatic biotinylation.

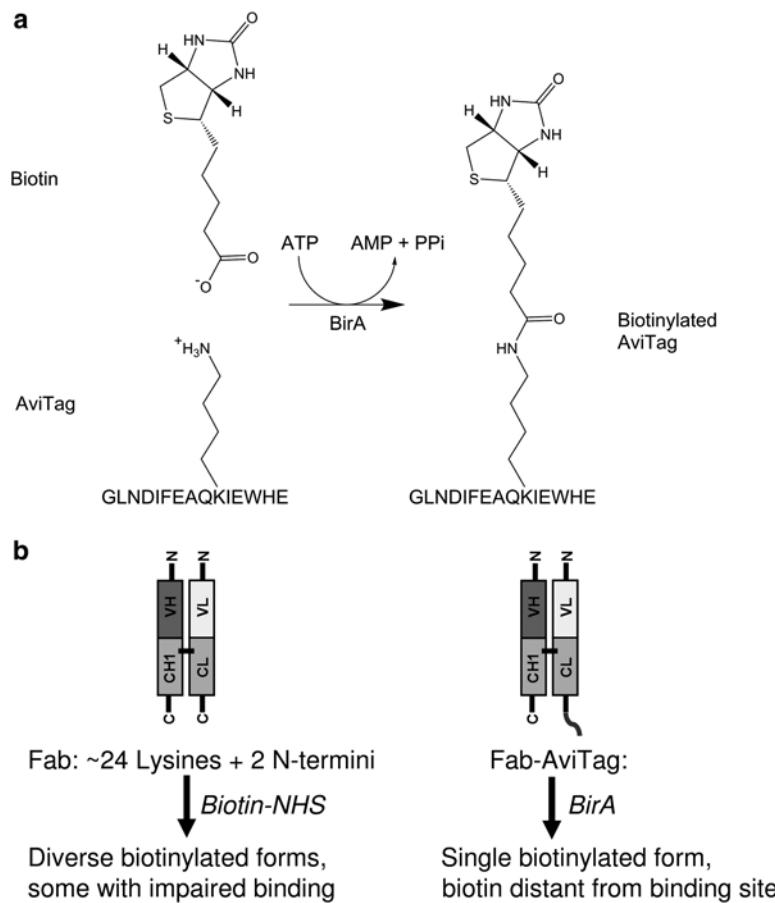


Fig. 1 Principle of BirA use. (a) Biotin ligase (BirA) reaction, covalently linking free biotin to the lysine of AviTag. (b) Advantage of labeling with BirA compared to labeling with amine-reactive biotin *N*-hydroxysuccinimide (NHS) esters, illustrated with regard to a Fab antibody fragment

More recent work has established that BirA can biotinylate such substrate peptides specifically in the cytosol [6], secretory pathway, and at the cell surface in mammalian and invertebrate systems [7–10]. A detailed protocol for labeling with BirA at the mammalian cell surface for fluorescent imaging has recently been published [11].

Biotinylation of purified proteins has been applied in a wide range of areas of biochemistry and cell biology (*see Fig. 2*):

- Tetramerization—enhancing the avidity of ligand binding. For example, MHC class I tetramerized by streptavidin enabled stable binding to the T cell receptor and so allowed monitoring of the immune response and isolation of anti-pathogen or anti-cancer T cells [12].

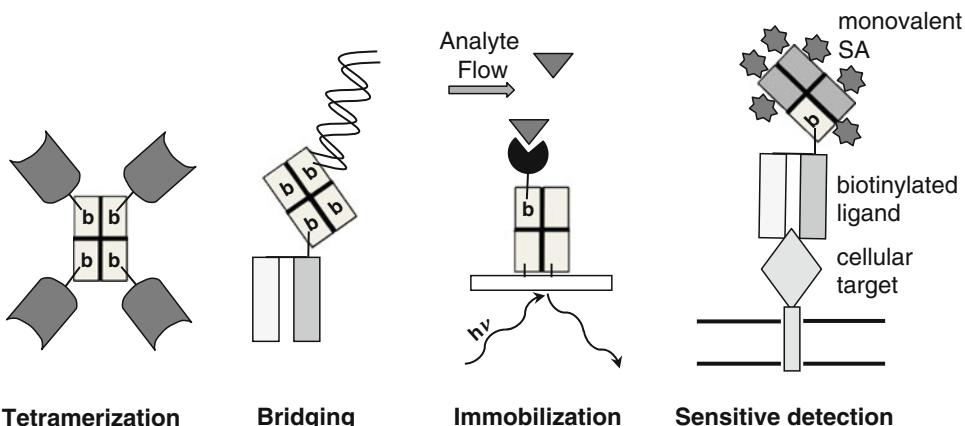


Fig. 2 Common applications of BirA biotinylation of purified proteins

- Bridging—for nanoassembly, streptavidin is often used as a bridge between one biotinylated protein and another biotinylated molecule, such as DNA, sugars, lipids, or small-molecule drugs [13].
- Immobilization—giving precise attachment that is stable over time, to a wide range of pH values, and to force. BirA-biotinylated proteins are commonly used for capture on chromatography columns, chips (e.g., for surface plasmon resonance or next-generation sequencing) [14], atomic force microscope tips [15], or nanoparticles (e.g., quantum dots or magnetic particles) [16].
- Sensitive detection—an in vitro biotinylated protein can be added to cells and subsequently recognized with high affinity by streptavidin conjugates [17]. Use of monovalent streptavidin facilitates efficient measurement of the absolute number of biotin binding sites on cells [16].

An important advance in BirA labeling is its use for electron microscopy [18]. Biotin ligase from *E. coli* or other species can also ligate to a peptide tag biotin analogs, including desthiobiotin for reversible streptavidin binding [19], or analogs containing functional groups for bio-orthogonal reaction: keto [20], azido, and alkyne groups [21]. However, only small changes to the structure of biotin could be tolerated by biotin ligase and so the related ligase LplA has proved more amenable to direct incorporation of fluorophores [22].

Engineering of streptavidin is important in extending the usefulness of BirA-labeling; in particular variants with controlled valency (e.g., monovalent streptavidin, mSA), enabling precise control over assembly of biotin conjugates [11, 23]. In addition, we generated a streptavidin variant with tenfold lower off-rate for biotin and enhanced thermal stability (traptavidin) [24, 25].

New applications of BirA have been for labeling specific protein populations—by targeting BirA to a specific chromatin-associated protein, particular AviTag-linked nucleosome populations were biotinylated [26]. By targeting BirA to one synaptic membrane, AviTag-proteins on the opposite synaptic membrane were biotinylated, allowing imaging of specific protein–protein interactions at synapses [27]. Through expressing a BirA-substrate peptide on a nuclear envelope protein and BirA in specific tissues of *Arabidopsis thaliana*, *Caenorhabditis elegans*, or *Drosophila melanogaster*, nuclei from specific cell types could be isolated by streptavidin pull-down [28, 29]. Also the use of enzymes to achieve *promiscuous* biotinylation (a BirA mutant or a peroxidase) has enabled labeling of untagged proteins in particular cellular regions or compartments [30–32].

The convenience and high yield of BirA labeling must be considered against certain limitations:

- A peptide tag must be introduced into the target protein. For site-specific biotinylation while only changing a single residue, suppressor tRNA bearing a biotinylated amino acid can be used (although some protein locations were not well tolerated) [33]. However, biotinylation via artificial amino acid incorporation brings disadvantages of more complex expression and of uncertainty in percentage incorporation—the initial assessment of biotinylation yield in *Xenopus* oocytes was done indirectly via electrophysiology and radioactive streptavidin binding [33]. *p*-Aminophenylalanine-linked biotin conjugates on tRNA showed improved protein incorporation in cell-free translation [34] (reagents are available from RiNA GmbH or Cosmo Bio Co. Ltd.). Biotinylation can also be achieved directly at the N-terminus, such as with subtiligase [35], or at the C-terminus using inteins [36].
- The binding partner of biotin, streptavidin or avidin, does not interact covalently and is not a good fusion partner. Covalent linkage to peptide tags can now be achieved using split inteins [37, 38], sortase [39], and SpyCatcher [40], although they have not yet demonstrated the high sensitivity of detection shown by streptavidin or avidin. A monomeric streptavidin has been developed that is suitable as a fusion tag [41]. A key future development will be to improve monomeric streptavidin’s binding affinity to that of the original tetrameric streptavidin.

2 Materials

2.1 Equipment

1. Incubators and shakers appropriate for growing bacterial cultures.
2. Centrifuges: floor-standing centrifuge capable of spinning at $5,000 \times g$ on 1 L bacterial culture and benchtop centrifuge capable of spinning 1.5 mL tubes at $20,000 \times g$.

3. Sonicator or other cell-disrupting apparatus (e.g., French press).
4. UV–Vis spectrophotometer or Nanodrop for protein quantification.
5. Electrophoresis apparatus for running SDS-PAGE.
6. PCR machine.

2.2 Proteins, DNA, and Other Reagents

1. Streptavidin (commercially available from several sources including Thermo Scientific, Sigma, and Roche) (*see Note 1*).
2. pGEX-GST-BirA plasmid [42] (a kind gift from Chris O’Callaghan, University of Oxford); alternative expression vectors containing Maltose Binding Protein-BirA or His₆-BirA are available through Addgene (www.addgene.org).
3. Complete Protease Inhibitor Cocktail tablets (Roche) for inhibiting *E. coli* proteases during purification of GST-BirA.
4. 100 mM PMSF solution: 17.4 mg of phenylmethylsulfonyl fluoride (PMSF) in 1 mL of isopropanol. Store at -20 °C. (CAUTION: PMSF is toxic. PMSF should be added to the aqueous buffer just prior to use, as it has a short half-life in aqueous solutions.)
5. Glutathione-HiCap resin for purification of GST-tagged proteins (Qiagen).
6. Target protein with AviTag peptide sequence (*see Note 2*).
7. KOD hot start polymerase (Merck Millipore).
8. T4 DNA ligase (NEB).
9. T4 polynucleotide kinase (NEB).
10. DpnI (NEB).
11. 50 mM d-Biotin solution: 12.2 mg of d-biotin in 1 mL of anhydrous DMSO. Store at -20 °C.
12. 100 mM ATP solution: 55.1 mg of adenosine 5'-triphosphate disodium salt hydrate in 1 mL MilliQ water. Store in aliquots at -80 °C (*see Note 3*).
13. 1 M magnesium chloride solution: 203 mg of magnesium chloride hexahydrate in 1 mL MilliQ water. Store at room temperature.
14. 100 mM DTT solution: 15.4 mg of dithiothreitol in 1 mL MilliQ water. Aliquot and store at -20 °C. Make freshly each week.
15. 420 mM IPTG solution: 1 g isopropyl-β-D-thiogalactopyranoside in 10 mL MilliQ water. Syringe-filter and store at -20 °C.
16. 100 mg/mL Ampicillin solution: 1 g of ampicillin sodium salt in 10 mL MilliQ water. Syringe-filter and store at -20 °C.
17. 20 % glucose solution: 200 g/L of d-glucose in MilliQ water. Autoclave and store at room temperature.

2.3 Buffers, Media, and Cells

1. Phosphate buffered saline (PBS): 1.44 g/L di-sodium hydrogen phosphate, 0.24 g/L potassium di-hydrogen phosphate, 0.2 g/L KCl, 8 g/L NaCl, pH 7.4.
2. PBS-L buffer: PBS, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL lysozyme, 1 % Triton X-100 (make fresh each day).
3. PBS-EW buffer: PBS, 1 mM DTT and 1 mM EDTA.
4. Elution buffer: 50 mM Tris-HCl pH 8.0, 0.4 M NaCl, 50 mM reduced glutathione, 1 mM DTT. Make this buffer fresh on each occasion.
5. Luria Bertani broth (LB): 10 g/L bacto-tryptone, 5 g/L yeast extract, 5 g/L NaCl. Autoclave and store at room temperature.
6. *E. coli* strain suitable for protein expression, e.g., BL21 [DE3] RIPL (Agilent).
7. 2× SDS-PAGE buffer (nonreducing): 4 % SDS, 20 % glycerol, 0.12 M Tris-HCl pH 6.8. Store aliquots at -20 °C.

3 Methods

The methods described below utilize a glutathione-*S*-transferase-BirA fusion protein, but are adaptable to His₆-tagged or Maltose Binding Protein (MBP) fusion constructs. All three constructs express well but GST-BirA can be efficiently removed from the biotinylated substrate after reaction, by passing through glutathione-agarose.

3.1 GST-BirA Production

1. Transform an appropriate *E. coli* expression strain (e.g., BL21) with the pGEX-GST-BirA plasmid.
2. Grow a 10 mL overnight culture from a single colony in LB plus 10 µL of 100 mg/mL ampicillin and 200 µL of 20 % glucose.
3. Use 8 mL of the overnight culture to inoculate 800 mL LB plus 0.8 mL 100 mg/mL ampicillin and 30 mL 20 % glucose in a 2 L baffled flask.
4. Grow at 37 °C with 200 rpm shaking to an OD₆₀₀ of 0.5.
5. Induce protein expression by addition of 0.8 mL of 420 mM IPTG solution.
6. Continue growth at 25 °C with 200 rpm shaking overnight.
7. Harvest cells by centrifugation for 10 min at 5,000×*g* at 4 °C.
8. Resuspend cells in 15 mL of PBS and freeze at -80 °C.
9. Thaw cells on ice and add 0.17 mL of 10 mg/mL lysozyme, one Complete Protease Inhibitor Cocktail tablet, 0.17 mL of 100 mM PMSF, 1.7 mL of 10 % Triton X-100, 0.17 mL of 100 mM EDTA, and 0.17 mL of 100 mM DTT.

10. Incubate 30 min on ice and freeze again at -80 °C to help cell lysis.
11. Thaw cells and add 15 mL cold PBS-L buffer. Hereafter, keep the sample at 4 °C at all stages.
12. Sonicate to reduce viscosity (e.g., 3–5× 30 s bursts on ice). (Caution: wear appropriate ear protection.)
13. Centrifuge lysed cells at 20,000× g for 30 min.
14. Collect the supernatant and add 1 mL of glutathione-HiCap resin to the supernatant, mixing end-over-end for 30 min at 4 °C.
15. Centrifuge resin for 2 min at 1,000× g and discard supernatant.
16. Wash resin with 30 mL PBS-EW. Centrifuge resin for 2 min at 1,000× g and repeat the wash.
17. Elute GST-BirA with 2 mL Elution buffer and incubate for 30 min at 4 °C.
18. Centrifuge resin for 2 min at 1,000× g and collect supernatant.
19. Check purity by SDS-PAGE (14 % polyacrylamide) (*see Fig. 3*) and concentration via OD₂₈₀ (GST-BirA has an ϵ_{280} of 90,550 M⁻¹ cm⁻¹).

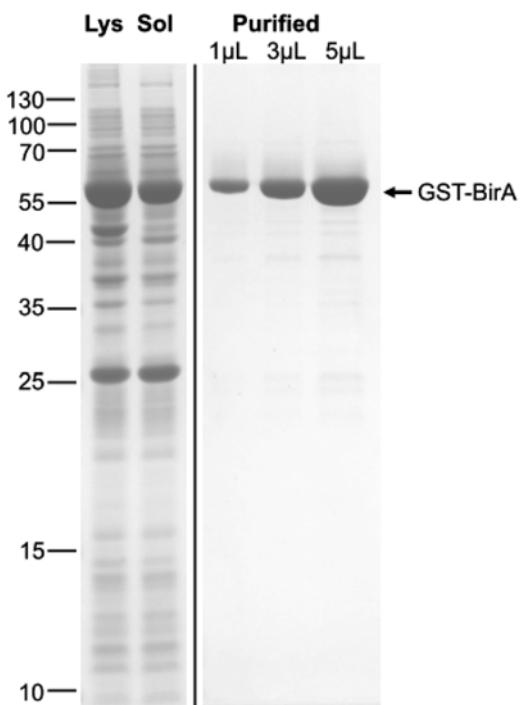


Fig. 3 Expression and purification of GST-BirA. 14 % SDS-PAGE with Coomassie staining of samples of the lysate (Lys) and soluble fraction (Sol) of *E. coli* expressing GST-BirA and varying amounts of the protein preparation purified with glutathione-resin

- Concentrate by ultrafiltration to ~50 µM and store in single-use aliquots at -80 °C. Concentrations of GST-BirA much greater than 50 µM may crash out. Final yield should be 10–20 mg/L of expression culture. After thawing, aliquots stored at 4 °C should be used within 1 week.

3.2 Generation of AviTag Protein Constructs

A variety of standard molecular biology methods can be used to add the AviTag (*see Note 2*) to an appropriate site in a target protein (*see Note 4*). For certain experiments it may also be valuable to clone a negative control peptide that is not biotinylated by BirA (*see Note 5*). We suggest using a modified inverse PCR mutagenesis [43] (*see Fig. 4*) or Site-directed Ligase-Independent Mutagenesis (SLIM) reaction [44], which enables the insertion of the substrate peptide without requiring any restriction sites nearby. Below is an example inverse PCR mutagenesis protocol.

- Forward and reverse primers for peptide insertion should be designed to each have 18–25 bp matching the parental sequence and have a calculated annealing temperature (to the parent sequence) of at least 55 °C (*see Fig. 4*).
- Assemble the following reaction mixture in a PCR tube: 29.5 µL MilliQ water, 1.5 µL DMSO, 5 µL KOD polymerase buffer, 5 µL 25 mM MgSO₄, 1 µL 15 µM forward primer, 1 µL 15 µM reverse primer, 1 µL 100 ng/µL template plasmid DNA, 5 µL 2 mM dNTP mix, and finally 1 µL KOD hot start polymerase.

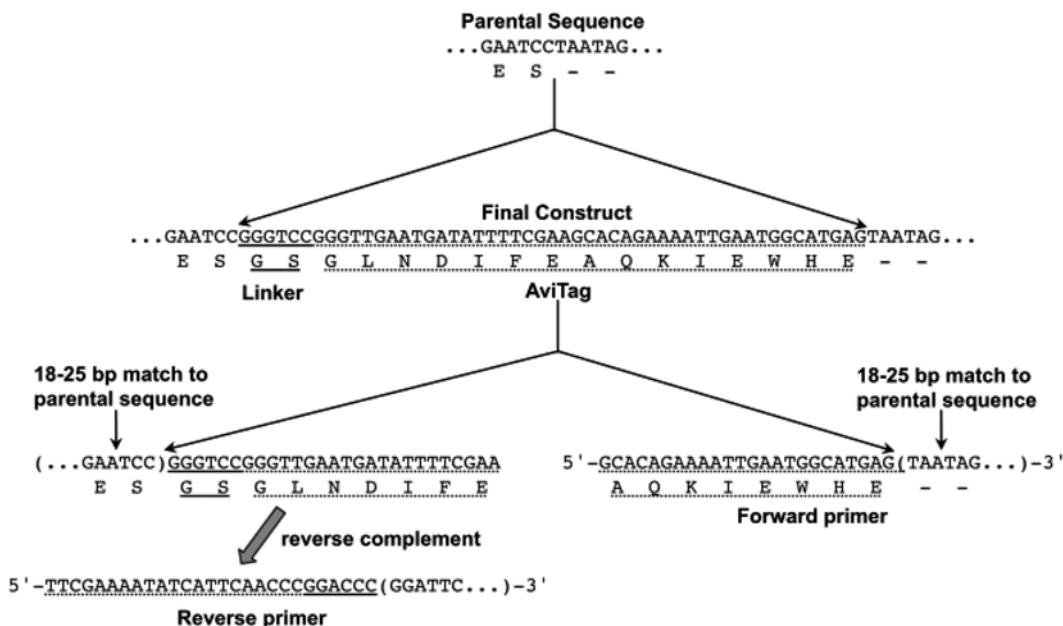


Fig. 4 Design of primers for AviTag insertion using the inverse PCR mutagenesis method

3. After transferring the tube to a PCR machine, perform an initial denaturing step of 3 min at 95 °C, followed by 12 cycles of 95 °C for 30 s, 55 °C for 30 s, and 68 °C for 30 s/kb of target plasmid DNA.
4. Add 1 µL of 20 U/µL DpnI enzyme to the PCR mix and incubate at 37 °C for 1 h.
5. Run an aliquot of the reaction on a 0.7 % agarose gel to confirm the success and fidelity of the PCR (a clean band should be observed corresponding to the size of the linearized target plasmid DNA).
6. To 2 µL of the PCR product, add 14 µL MilliQ water, followed by 2 µL of 10× T4 DNA ligase buffer, 1 µL T4 polynucleotide kinase, and 1 µL of T4 DNA ligase.
7. Incubate the sample for 1 h at room temperature and transform an appropriate strain of competent *E. coli* (e.g., DH5α, XL1-Blue, JM109) with 5 µL of the ligation reaction. Cells with competency of at least 10⁷ cfu/µg should be sufficient.
8. After validating the construct by sequencing, the AviTag-fused protein can be overexpressed in the appropriate cell system (commonly *E. coli*, baculovirus, or HEK 293T cells).

3.3 Biotinylation of AviTag-Fused Proteins Using BirA

1. To 100 µM AviTag-fused protein in 952 µL of PBS, add 5 µL 1 M magnesium chloride, 20 µL 100 mM ATP, 20 µL 50 µM GST-BirA, and 3 µL 50 mM d-Biotin (*see Note 6*).
2. Incubate sample for 1 h at 30 °C with gentle mixing on a rocking platform.
3. Add the same amount of fresh biotin and GST-BirA and incubate for a further hour.
4. GST-BirA may be removed by incubation of the sample with 0.1 mL of a 50 % slurry of glutathione-HiCap resin in PBS for 30 min at room temperature, followed by centrifugation and collection of the supernatant [45].
5. Dialyze the sample into PBS or other suitable buffer, for storage and to remove the excess biotin.
6. The biotinylation of the target protein is generally irreversible in vitro; apparent loss of biotinylation is most likely to reflect proteolysis separating the biotinylation site from the rest of the target protein.

3.4 Testing the Extent of Protein Biotinylation by a Streptavidin Gel-Shift

The efficiency of the biotinylation reaction has been examined by Western blotting [6] or other enzymatic or ligand-displacement assays [46], but these approaches are time-consuming and only indirectly allow quantitation. A rapid and easily quantified alternative is to saturate the target protein with streptavidin and study the gel-shift in SDS-PAGE (*see Fig. 5*). Provided the gel does not get

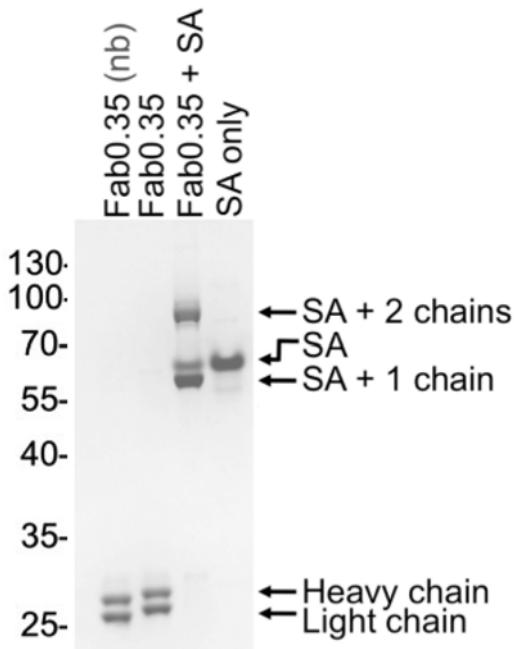


Fig. 5 Testing the extent of biotinylation by SDS-PAGE gel-shift. Coomassie-stained SDS-PAGE of an antibody fragment (Fab0.35) with an AviTag on the C-terminus of both the heavy and light chains. The lanes represent non-biotinylated Fab (nb), biotinylated Fab, biotinylated Fab with streptavidin (SA), and streptavidin alone. Streptavidin has four binding sites and so may associate with one or two chains of the biotinylated target, but this does not affect the calculation of the depletion of the original target protein band

excessively warm during the run, streptavidin will retain its native tetramer structure and remain bound to biotin conjugates under normal SDS-PAGE conditions [16]. A streptavidin monomer (i.e., one biotin binding site) has a calculated ϵ_{280} of $41,940\text{ M}^{-1}\text{ cm}^{-1}$.

1. Prepare a PCR tube containing 5 μL of 10 μM biotinylated target protein and add 10 μL of 2 \times SDS-PAGE buffer.
2. Heat samples at 95 °C for 5 min in a PCR block with a heated lid.
3. Allow the sample to cool to room temperature and briefly centrifuge.
4. After this boiling and cooling, add 5 μL of PBS containing a small molar excess (two- to fivefold) of streptavidin to the samples and incubate at room temperature for 5 min (it is advisable to run a control lane of streptavidin without the target protein).
5. Run samples on an appropriate SDS-PAGE gel (the streptavidin tetramer, running at 50–60 kDa, is clearly visible on 10, 12, 14, 16 % gels) (*see Note 7*).

6. Stain the gel with InstantBlue or Coomassie blue and visualize. If desired, quantify the degree of biotinylation by densitometry, measuring the change in intensity of the relevant protein band with and without addition of streptavidin (*see Note 8*). In the lane containing biotinylated protein and streptavidin, the presence of a band corresponding to free streptavidin verifies that streptavidin was indeed provided in excess and so all biotinylated protein will have been bound. Streptavidin may sometimes increase in mobility upon binding to biotin conjugates, according to the size and charge of the biotin conjugate (*see Fig. 5*).

4 Notes

1. Instead of streptavidin, other high-affinity biotin-binding proteins may be used to bind to enzymatically biotinylated proteins. Avidin is not recommended because its positive charge promotes nonspecific binding to cells and DNA, but neutravidin should be satisfactory for many applications [47].
2. Several peptide sequences have been described for BirA-mediated biotinylation. These are based on those first described by Schatz and coworkers [5, 48], who found a 13 amino acid peptide to be the minimal substrate peptide for BirA (LX§IFEAQKIEWR, where X = any and § = any but not L, V, I, W, F, or Y). This sequence was further optimized to improve the rate of biotinylation, resulting in AviTag (GLNDIFEAQKIEWHE). AviTag works at either the N or C terminus of the target protein [46]. A close 15 residue relative, termed BioTag (ALNDIFEAQKIEWHA), is also used in some papers [10, 48]. BLRP (biotin ligase recognition peptide) contains a core of AviTag and is 23 residues: (MAGGLNDIFEAQKIEWHEDTGGS) [5, 49]. Another popular target is the 15 residue “BirA Substrate Peptide” (BSP), LHHILDAQKMWVNHR [42, 48]. A further consideration is whether some flexibility should be added between the AviTag and the target protein. We would suggest including a flexible two residue GS linker between the AviTag and the target protein or any other surrounding peptide tag or domain. In the unlikely event that constructs with N-terminal or C-terminal AviTag do not enable biotinylation or yield low amounts of protein, first try increasing the spacer to 6 residues and then it may be worth trying BSP [42]. Vectors are also available containing N- or C-terminal AviTag sequences from Avidity or from Genecopoeia (for bacterial, mammalian, or cell-free expression; some plasmids have BirA downstream for coexpression).

3. Prepare single-use aliquots: freeze–thawing damages ATP stocks. Also, ATP will be hydrolyzed at pH greater than 8.5.
4. Selected examples of successful biotinylation following BirA-substrate peptide insertion in protein loops: the *E. coli* flagellar hook [50], Cystic Fibrosis Transmembrane Regulator (CFTR) [51], and Dicer [52]. For Dicer, Lau et al. use streptavidin to highlight features for Cryo-electron Microscopy and describe several functional and some nonfunctional peptide insertion sites, advising insertion in short loops disordered in the crystal structure or less highly conserved [52].
5. The Lys to Ala mutant of AviTag (GLNDIFEAQAIIEWHE) serves as an effective negative control sequence that will not be biotinylated [8, 20]. Note that AviTag-fusions expressed in *E. coli* may have some biotinylation from the cell's own BirA, but this reaction may often not reach completion, even in strains with BirA overexpressed (AVB101, Avidity) [53]. Also, adding BirA to an AviTag-fusion in the absence of ATP or biotin may still allow some biotinylation to take place, because of biotin-AMP pre-bound to the purified protein [8, 20].
6. Other buffers may be used for biotinylation. Schatz et al. prefer 50 mM bicine pH 8.3, maintaining low [NaCl] [46], but in our hands biotinylation in PBS is still quantitative. It is preferable to have the AviTag-fusion at concentration >40 μM when incubating with BirA; otherwise biotinylation is less efficient [46]. The biotinylation reaction may be run on a smaller scale; the only issue is that losses from dialysis become more significant when working with a low total amount of protein.
7. If your target protein happens to have exactly the same mobility as streptavidin, use a different percentage gel. The target protein is unfolded and will run according to its molecular weight, but streptavidin remains folded and runs at a different height on different percentage gels.
8. With incomplete biotinylation, it is possible to purify the biotinylated fraction using monomeric avidin (a chemically modified version of avidin with reversible biotin binding) [45], but we would suggest that it is preferable to modify the biotinylation reaction until the reaction does go to completion.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC). We thank Jayati Jain (Howarth laboratory) for providing Fig. 5.

References

- Chapman-Smith A, Cronan JE Jr (1999) In vivo enzymatic protein biotinylation. *Biomol Eng* 16:119–125
- Green NM (1990) Avidin and streptavidin. *Methods Enzymol* 184:51–67
- Sano T, Vajda S, Cantor CR (1998) Genetic engineering of streptavidin, a versatile affinity tag. *J Chromatogr B Biomed Sci Appl* 715:85–91
- Cronan JE Jr (1990) Biotinylation of proteins in vivo. A post-translational modification to label, purify, and study proteins. *J Biol Chem* 265:10327–10333
- Beckett D, Kovaleva E, Schatz PJ (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* 8:921–929
- de Boer E et al (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* 100:7480–7485
- Parrott MB, Barry MA (2001) Metabolic biotinylation of secreted and cell surface proteins from mammalian cells. *Biochem Biophys Res Commun* 281:993–1000
- Howarth M, Takao K, Hayashi Y, Ting AY (2005) Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* 102:7583–7588
- Yang J, Jaramillo A, Shi R, Kwok WW, Mohanakumar T (2004) In vivo biotinylation of the major histocompatibility complex (MHC) class II/peptide complex by coexpression of BirA enzyme for the generation of MHC class II/tetramers. *Hum Immunol* 65:692–699
- Ooi SL, Henikoff JG, Henikoff S (2010) A native chromatin purification system for epigenomic profiling in *Caenorhabditis elegans*. *Nucleic Acids Res* 38:e26
- Howarth M, Ting AY (2008) Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat Protoc* 3:534–545
- Sims S, Willberg C, Kleinerman P (2010) MHC-peptide tetramers for the analysis of antigen-specific T cells. *Expert Rev Vaccines* 9:765–774
- Valadon P et al (2010) Designed auto-assembly of nanostreptabodies for rapid tissue-specific targeting in vivo. *J Biol Chem* 285:713–722
- Williams JG et al (2008) An artificial processivity clamp made with streptavidin facilitates oriented attachment of polymerase-DNA complexes to surfaces. *Nucleic Acids Res* 36:e121
- Rakshit S, Zhang Y, Manibog K, Shafraz O, Sivasankar S (2012) Ideal, catch, and slip bonds in cadherin adhesion. *Proc Natl Acad Sci U S A* 109:18815–18820
- Jain J, Veggiani G, Howarth M (2013) Cholesterol loading and ultrastable protein interactions determine the level of tumor marker required for optimal isolation of cancer cells. *Cancer Res* 73:2310–2321
- Sung K, Maloney MT, Yang J, Wu C (2011) A novel method for producing mono-biotinylated, biologically active neurotrophic factors: an essential reagent for single molecule study of axonal transport. *J Neurosci Methods* 200:121–128
- Viens A et al (2008) Use of protein biotinylation in vivo for immunoelectron microscopic localization of a specific protein isoform. *J Histochem Cytochem* 56:911–919
- Wu SC, Wong SL (2004) Development of an enzymatic method for site-specific incorporation of desthiobiotin to recombinant proteins in vitro. *Anal Biochem* 331:340–348
- Chen I, Howarth M, Lin W, Ting AY (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* 2:99–104
- Slavoff SA, Chen I, Choi YA, Ting AY (2008) Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *J Am Chem Soc* 130:1160–1162
- Uttamapinant C et al (2010) A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* 107:10914–10919
- Howarth M et al (2006) A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* 3:267–273
- Chivers CE et al (2010) A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat Methods* 7:391–393
- Chivers CE, Koner AL, Lowe ED, Howarth M (2011) How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochem J* 435:55–63
- Lau PN, Cheung P (2013) Elucidating combinatorial histone modifications and crosstalks by coupling histone-modifying enzyme with biotin ligase activity. *Nucleic Acids Res* 41:e49
- Liu DS, Loh KH, Lam SS, White KA, Ting AY (2013) Imaging trans-cellular neurexin-neuroligin

- interactions by enzymatic probe ligation. *PLoS One* 8:e52823
28. Deal RB, Henikoff S (2011) The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat Protoc* 6:56–68
 29. Steiner FA, Talbert PB, Kasinathan S, Deal RB, Henikoff S (2012) Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res* 22:766–777
 30. Cronan JE (2005) Targeted and proximity-dependent promiscuous protein biotinylation by a mutant *Escherichia coli* biotin protein ligase. *J Nutr Biochem* 16:416–418
 31. Roux KJ, Kim DI, Raida M, Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196:801–810
 32. Martell JD et al (2012) Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol* 30:1143
 33. Gallivan JP, Lester HA, Dougherty DA (1997) Site-specific incorporation of biotinylated amino acids to identify surface-exposed residues in integral membrane proteins. *Chem Biol* 4:739–749
 34. Watanabe T, Muranaka N, Iijima I, Hohsaka T (2007) Position-specific incorporation of biotinylated non-natural amino acids into a protein in a cell-free translation system. *Biochem Biophys Res Commun* 361:794–799
 35. Yoshihara HA, Mahrus S, Wells JA (2008) Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg Med Chem Lett* 18:6000–6003
 36. Lesaicherre ML, Lue RYP, Chen GYJ, Zhu Q, Yao SQ (2002) Intein-mediated biotinylation of proteins and its application in a protein microarray. *J Am Chem Soc* 124:8768–8769
 37. Carvajal-Vallejos P, Pallisse R, Mootz HD, Schmidt SR (2012) Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem* 287:28686–28696
 38. Shah NH, Dann GP, Vila-Perello M, Liu Z, Muir TW (2012) Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. *J Am Chem Soc* 134:11338–11341
 39. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3:707–708
 40. Zakeri B et al (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* 109:E690–E697
 41. Lim KH, Huang H, Pralle A, Park S (2013) Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnol Bioeng* 110:57–67
 42. O'Callaghan CA et al (1999) BirA enzyme: production and application in the study of membrane receptor-ligand interactions by site-specific biotinylation. *Anal Biochem* 266:9–15
 43. Gama L, Breitwieser GE (2002) Generation of epitope-tagged proteins by inverse polymerase chain reaction mutagenesis. *Methods Mol Biol* 182:77–83
 44. Chiu J, March PE, Lee R, Tillett D (2004) Site-directed, ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res* 32:e174
 45. Saviranta P, Haavisto T, Rappu P, Karp M, Lovgren T (1998) In vitro enzymatic biotinylation of recombinant fab fragments through a peptide acceptor tail. *Bioconjug Chem* 9:725–735
 46. Cull MG, Schatz PJ (2000) Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol* 326:430–440
 47. Marttila AT et al (2000) Recombinant NeutraLite avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. *FEBS Lett* 467:31–36
 48. Schatz PJ (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Biotechnology (N Y)* 11:1138–1143
 49. Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 456:125–129
 50. Brown MT et al (2012) Flagellar hook flexibility is essential for bundle formation in swimming *Escherichia coli* cells. *J Bacteriol* 194:3495–3501
 51. Bates IR et al (2006) Membrane lateral diffusion and capture of CFTR within transient confinement zones. *Biophys J* 91:1046–1058
 52. Lau PW, Potter CS, Carragher B, MacRae IJ (2012) DOLORS: versatile strategy for internal labeling and domain localization in electron microscopy. *Structure* 20:1995–2002
 53. Li Y, Sousa R (2012) Expression and purification of *E. coli* BirA biotin ligase for in vitro biotinylation. *Protein Expr Purif* 82:162–167

Chapter 13

Site-Specific Labeling of Proteins via Sortase: Protocols for the Molecular Biologist

Maximilian Wei-Lin Popp

Abstract

Creation of site-specifically labeled protein bioconjugates is an important tool for the molecular biologist and cell biologist. Chemical labeling methods, while versatile with respect to the types of moieties that can be attached, suffer from lack of specificity, often targeting multiple positions within a protein. Here we describe protocols for the chemoenzymatic labeling of proteins at the C-terminus using the bacterial transpeptidase, sortase A. We detail a protocol for the purification of an improved pentamutant variant of the *Staphylococcus aureus* enzyme (SrtA 5°) that exhibits vastly improved kinetics relative to the wild-type enzyme. Importantly, a protocol for the construction of peptide probes compatible with sortase labeling using techniques that can be adapted to any cellular/molecular biology lab with no existing infrastructure for synthetic chemistry is described. Finally, we provide an example of how to optimize the labeling reaction using the improved SrtA 5° variant.

Key words Sortase, Transpeptidation, Site-specific labeling, Sortagging, Chemoenzymatic labeling, VHH

1 Introduction

Sortases are a class of bacterial transpeptidases that are exploited by gram-positive bacteria to anchor a diverse array of virulence factors to their peptidoglycan cell wall [1]. The reaction that ensues upon recognition of an appropriate peptide sequence involves breaking of an amide bond, followed by reformation of a new peptide bond, and this property has been leveraged for the site-specific labeling of proteins. The *Staphylococcus aureus* sortase A enzyme recognizes an LPXTG motif in substrate proteins and cleaves the peptide bond between the threonine and glycine residues using a key catalytic cysteine, generating a thioacyl intermediate in a mechanism reminiscent of cysteine proteases [2]. However, in stark contrast to cysteine proteases, resolution of the acyl-enzyme intermediate by water proceeds poorly. Instead, it is nucleophilic attack by the α -amine of an incoming oligoglycine-based nucleophile, present on a lipid-linked cell wall precursor, that resolves the acyl-enzyme

intermediate. This transpeptidation reaction is fully portable and can be adapted for routine laboratory use using recombinantly produced sortase A, a chemically synthesized oligoglycine-based nucleophile carrying a payload of choice, and a protein of interest bearing the requisite LPXTG cleavage site (Fig. 1a). Owing to the partial ionization of the catalytic cysteine residue at physiological pH [3, 4], the wild-type sortase A enzyme exhibits sluggish kinetics, a problem that can be circumvented through addition of near-stoichiometric quantities in labeling reactions. A useful addition to the sortase labeling toolbox, however, is the generation of an evolved sortase pentamutant (SrtA 5°) variant [5] with increased transpeptidation kinetics.

The limited requirements for substrate design have been detailed elsewhere [6] and for C-terminal protein labeling involve only the addition of the LPXTG (usually LPETG) to the protein of interest in a flexible and solvent-exposed position. The glycine in the LPXTG motif should be placed in peptide linkage to another residue, and positioning a hexahistidine affinity purification tag downstream of the LPXTG site (Fig. 2) is a convenient way to both satisfy this criterion and provide a means to purify the transpeptidation product from unreacted input material and similarly hexahistidine-tagged sortase enzyme [7]. The small size and non-perturbing composition of the sortase recognition motif make the technique a useful complement to other methods detailed in this volume. Reaction conditions are mild and modification of the protein to be labeled does not demand installation of protein-sized domains. Synthesis of compatible nucleophiles bearing nearly any non-genetically templated moiety does not demand any specialized equipment or skills (*see* Subheadings 3.2 and 3.3). Labeling, however, takes place either at the C-terminus (as detailed here), the N-terminus [8, 9], or in an unstructured loop whose integrity is dispensable for function of the target protein [10–12].

Sortase has now been used to label a wide range of structurally and functionally diverse substrates including purified proteins [7, 9–11, 13–20], viruses [12, 21–23], cell surface proteins [8, 12, 15, 19, 24, 25], and peptides [26] using derivatized glycine-based probes bearing a similarly diverse array of functionalities [27]. Applications range from the creation of biological probes to biotechnological uses, to use in synthetic chemistry schemes. The sortase reaction has even been successfully used in the yeast and mammalian cytosol as well as the endoplasmic reticulum [28] to circularize proteins and create protein–protein fusions. The main limitation to conducting intracellular labeling reactions with non-genetically encoded probes is the delivery of sufficient quantities of nucleophile across the plasma membrane, a problem that is likely solvable through the creative use of masking groups and/or cell-penetrating peptides. The main strength of the sortase method is the ease of implementation. Once conditions are identified for

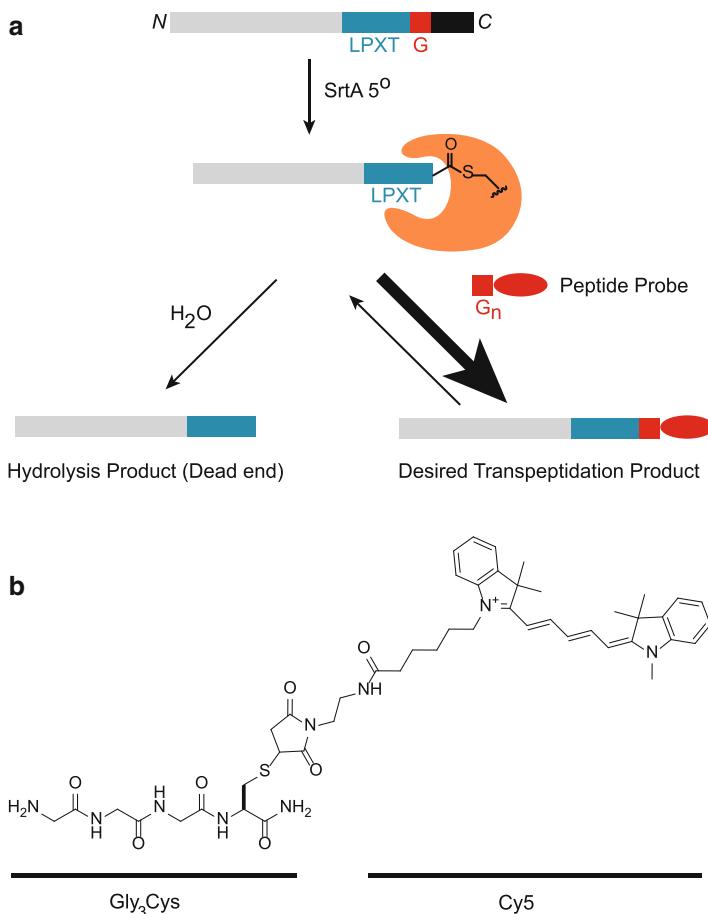


Fig. 1 Sortase labeling schematic and probe construction. **(a)** Sortase labeling proceeds via enzymatic recognition of an LPXTG site within a target protein (*top*). SrtA cleaves the amide bond between the threonine and glycine in this recognition sequence, and forms an acyl enzyme intermediate (*middle*) that is either inefficiently resolved by water (*bottom left*) to yield the hydrolysis product or in the presence of an oligoglycine-based nucleophile probe is resolved to form the desired transpeptidation product (*bottom right*). See text for details. **(b)** Example of a sortase-compatible probe consisting of a Gly₃Cys scaffold linked to a Cy5 moiety by a thioether bond. **(c)** Setup of a standard vortexer, available in most molecular biology labs, for use in peptide synthesis with disposable fritted polypropylene syringes

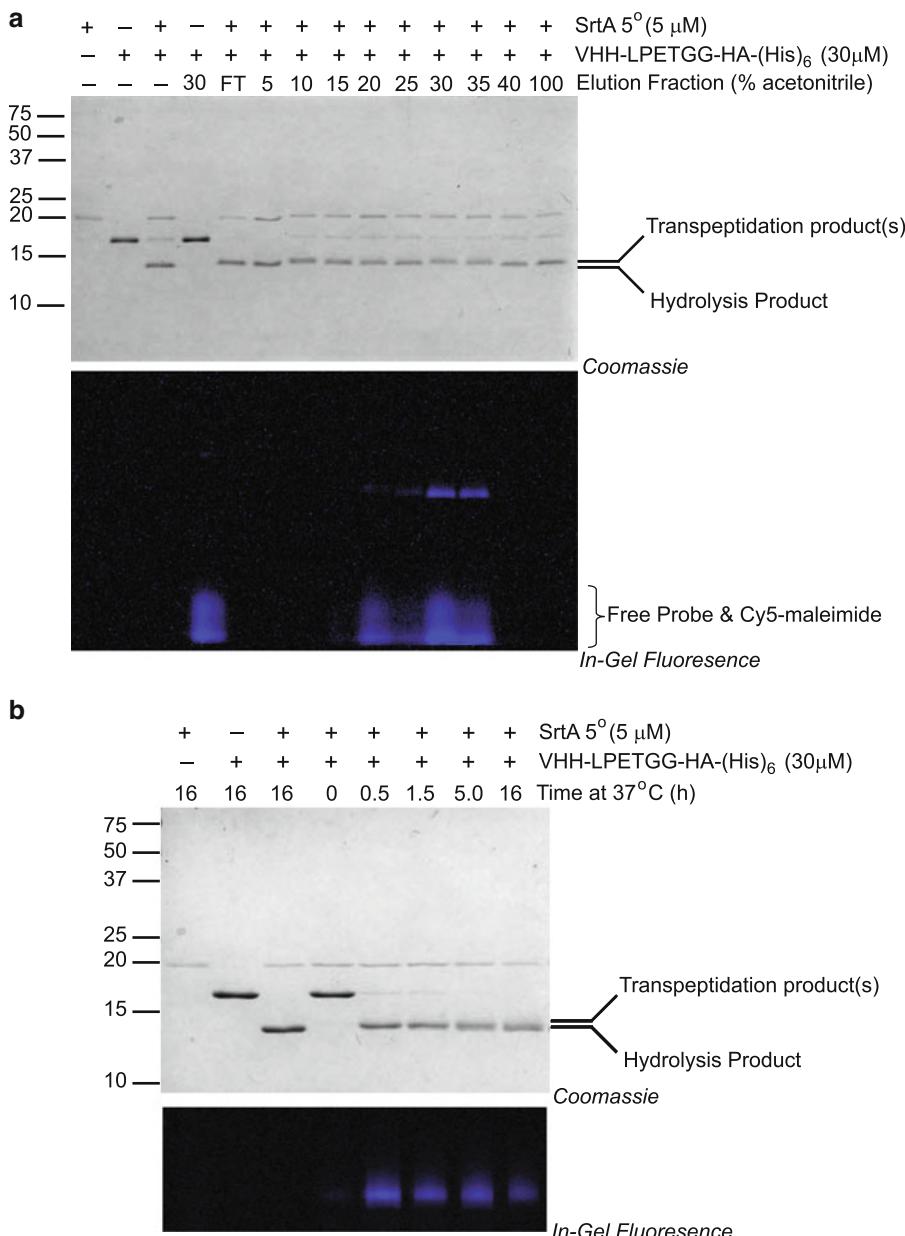


Fig. 2 Examples of simple probe purification and a substrate labeling time course. **(a)** The G₃C(Cy5) probe was purified as described in Subheading 3.3 and each elution fraction was subjected to sortase labeling using the test protein VHH-LPETGG-HA-(His)₆ as a substrate [18], as described in Subheading 3.4. Reactions were allowed to proceed for 1 h at 37 °C before quenching with sample buffer. Products were resolved by 17 % Tris-Tricine SDS-PAGE and visualized by both in-gel fluorescent imaging using a Typhoon 9410 imager and Coomassie staining. Note that the 30 and 35 % fractions contain active probe and these are pooled for further use. **(b)** Labeling of VHH-LPETGG-HA-(His)₆ was performed as detailed in Subheading 3.4 using 500 µM G₃C(Cy5) probe. Products were resolved by 17 % Tris-Tricine SDS-PAGE and visualized by both in-gel fluorescent imaging using a Typhoon 9410 imager and Coomassie staining. Note that the hydrolysis product accumulates at long time points (16 h) with a concomitant decrease in transpeptidation product quantity (in-gel fluorescence, last lane) and increased mobility in SDS-PAGE. The reaction is essentially complete between 30 and 90 min and this should be used as the optimized reaction time that yields the desired transpeptidation product

labeling a protein of interest, nearly all glycine-based nucleophiles behave similarly in sortase labeling reactions—a property that has been leveraged for high-throughput generation of bioconjugates [17]. Still virtually unexploited in protein engineering are the pilus-building sortases [29] that catalyze isopeptide bond formation. These enzymes and their ability to potentially label internal positions in proteins could be a useful addition to the sortase toolbox. The hallmark of a generally applicable and useful method is the migration of the technology from specialist labs to end users who seek to address diverse questions, and this has been the case for the sortase labeling technique [21, 22, 30–32]. Because the most common end users of this technology will be cellular/molecular biology labs instead of specialist labs equipped for chemical synthesis, we present a protocol for the creation of sortase-compatible probes and their use, using only inexpensive and commonly available equipment and techniques, with the expectation that this will enable the use of sortase labeling by a broader audience.

2 Materials

2.1 Protein Expression and Purification

1. BL21(DE3)pLysS competent bacteria.
2. 2YT medium: 16 g/L Bacto-tryptone, 10 g/L yeast extract, 5 g/L NaCl.
3. 1,000× Kanamycin: Dissolve at 30 mg/mL in water.
4. 1 M Isopropyl β -D-thiogalactopyranoside (IPTG): Dissolve in water.
5. Phosphate-buffered saline (PBS), pH 7.4: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄.
6. 10 mg/mL DNaseI: Dissolve in water.
7. Lysis buffer: 50 mM Tris·Cl pH 7.5, 150 mM NaCl, 20 mM imidazole.
8. Elution buffer: 50 mM Tris·Cl pH 7.5, 150 mM NaCl, 500 mM imidazole.
9. Purification buffer: 20 mM Tris·Cl pH 7.5, 150 mM NaCl.
10. Branson Sonifier or French Press Cell.
11. Ni-NTA agarose slurry.
12. Centrifugal Protein Concentrators 3 kDa MWCO.
13. Bradford Reagent.

2.2 Probe Synthesis

1. Disposable fritted polypropylene peptide synthesis syringe (Torviq).
2. Vortexer outfitted with a multi-tube platform (*see* Fig. 1c).
3. Speed-vac.

4. Fmoc deprotection cocktail: 20 % Piperidine (Sigma) in *N*-Methyl-2-pyrrolidone (NMP; Acros Organics).
5. Fmoc-Cys(Trt)-OH coupling cocktail: Dissolve five equivalents of Fmoc-Cys(Trt)-OH (Novabiochem/EMD Millipore) relative to resin loading and 4.9 equivalents of Benzotriazol-1-yl-oxytritypyrrolidinophosphonium hexafluorophosphate (PyBOP; Novabiochem/EMD Millipore) relative to resin loading in minimal NMP. For some amino acids, this may necessitate agitation by vortexing. Add ten equivalents of Diisopropylethylamine (DIPEA; Sigma) relative to resin loading to the solution and mix. Prepare immediately before use.
6. Fmoc-Gly-OH coupling cocktail: Dissolve five equivalents of Fmoc-Gly-OH (Novabiochem/EMD Millipore) relative to resin loading and 4.9 equivalents of Benzotriazol-1-yl-oxytritypyrrolidinophosphonium hexafluorophosphate (PyBOP; Novabiochem/EMD Millipore) relative to resin loading in minimal NMP. For some amino acids, this may necessitate agitation by vortexing. Add ten equivalents of Diisopropylethylamine (DIPEA; Sigma) relative to resin loading to the solution and mix. Prepare immediately before use.
7. Cleavage cocktail: 94 % Trifluoroacetic acid (TFA; Alfa Aesar), 3 % Triisopropylsilane (TIPS; Sigma), 3 % 2-Mercaptoethanol (β -Mercaptoethanol; Sigma).
8. Cy5 maleimide coupling solution: 1 mg of Cy5 maleimide (Lumiprobe) dissolved in 250 μ L of PBS.
9. Ninhydrin/Kaiser Test Kit (AnaSpec).

2.3 Probe Purification

1. Milli-Q-grade water containing 0.1 % trifluoroacetic acid ($H_2O/0.1\% TFA$).
2. Bond Elut JR C18 Cartridge, 500 mg Bed (Agilent).

2.4 Labeling

1. Purified SrtA 5°.
2. Purified LPXTG containing target protein.
3. Purified oligoglycine-based probe, dissolved in water or DMSO.
4. 10× Labeling buffer: 500 mM Tris-HCl pH 7.5, 1.5 M NaCl, 100 mM CaCl₂.

3 Methods

3.1 Protein Expression and Purification

The SrtA 5° variant has been cloned into pET29a (kanamycin resistant) and carries a C-terminal hexahistidine tag for purification [5]. The test protein described here (a camelid VHH domain) also carries a hexahistidine tag for purification, located downstream of the requisite LPETG cleavage site [18]. Both SrtA 5° and the

VHH substrate can be produced and purified using this simplified protocol; however production and purification protocols may have to be tailored to the specific protein to be labeled (*see Notes 1–4*). The protocol outlined here is a standard Ni-NTA purification scheme [33], using Tris-based buffers instead of phosphate-based buffers (*see Notes 4* and *7*).

1. Transform plasmid into BL21(DE3)pLysS cells and plate on selective media.
2. Pick a single colony and grow to saturation (overnight) in 30 mL of 2YT media containing kanamycin.
3. Inoculate 1 L of 2YT media containing antibiotics with 10 mL of the overnight culture and shake at 30 °C until an OD₆₀₀ of 0.6 is reached.
4. Add IPTG to a final concentration of 400 μM, lower temperature to 25 °C, and continue to grow cultures overnight.
5. Harvest bacteria by centrifugation at 6,000×*g* for 20 min. Decant supernatant and wash pellet with 30 mL PBS. Centrifuge again at 6,000×*g* for 20 min. Decant PBS and transfer pellet to a 50 mL conical tube.
6. Resuspend pellet in 25 mL of ice-cold lysis buffer with 20 μg/mL DNaseI.
7. Lyse bacteria using a Branson sonifier (output=6, duty cycle=50) with five rounds of sonication for 1 min each. Alternatively pass the bacteria through a prechilled French press cell twice at 1,250 psi.
8. Clarify the lysate by centrifugation at 12,000×*g* for 30 min. Save supernatant. Resuspend pellet in an additional 25 mL of ice-cold lysis buffer with 20 μg/mL DNaseI and repeat **step 7**. Pool supernatants.
9. Meanwhile, prepare a Ni-NTA gravity chromatography column using a disposable plastic column or a reusable glass column. Pack 4 mL of 50 % Ni-NTA slurry (2 mL final bed volume) and equilibrate with at least ten volumes of ice-cold lysis buffer.
10. Add the combined 50 mL of bacterial lysate to the column, being careful not to disturb the resin bed.
11. Wash the column with at least 100 column volumes of ice-cold lysis buffer.
12. Elute with two column volumes of ice-cold elution buffer and collect the eluate.
13. SrtA 5° may be further purified by size-exclusion chromatography using purification buffer (for example, over a Hi-Load Superdex 75 16/60 column using an appropriate FPLC system) to remove imidazole and any contaminating proteins. However, the Ni-NTA eluate is often of sufficient purity to

- label proteins after simple dialysis against purification buffer to remove imidazole.
14. Concentrate the purified protein using centrifugal concentrators. Protein concentration is determined by the Bradford method (*see Notes 1–5*).

3.2 Probe Synthesis

Although trivial for a synthetic chemist, generation of the peptide probes is the most challenging step for the molecular biologist. Solid-phase peptide synthesis is a standard technique [34] and as such, commercial vendors exist that will synthesize and purify peptides. Many will also incorporate dyes and other moieties; whether they are able to provide the moiety of interest, however, depends on the company. The sortase-compatible probes are easily constructed by hand and are a cost-effective alternative to commercial synthesis. The following protocol for the synthesis of a glycine-based Cy5 probe (Fig. 1b) consists of two parts. First, a GGGC scaffold peptide is synthesized on solid phase, followed by coupling of the liberated peptide to Cy5 maleimide, while in solution. With the scaffold in hand, the user can generate probes with virtually any non-genetically encoded moiety of interest.

1. Outfit a standard vortexer with a multi-tube foam platform as shown (Fig. 1c) in an appropriate fume hood. Attach rubber bands so that the peptide synthesis syringe can be affixed to the vortexer.
2. Accurately weigh out Rink amide resin, 100–200 mesh, and note the loading. Resins with a loading of 0.6–0.7 mmol/g are suitable for the short peptide synthesized in this protocol. Synthesis on a 150–300 μmol scale provides ample peptide scaffold for many probes and this protocol may be scaled up or down as appropriate. Add the pre-weighed resin to a disposable fritted polypropylene peptide synthesis syringe.
3. Attach an 18 G needle to the syringe and add 10 mL of dichloromethane (methylene chloride, DCM) to the beads. After removing the needle and replacing it with a plastic stopper, shake on the vortexer for 20 min to solvate the resin. Eject the DCM and add 10 mL of *N*-methyl-2-pyrrolidone (NMP), shaking for 20 min. Remove the NMP wash.
4. To remove the first Fmoc protecting group from the resin, freshly prepare Fmoc deprotection cocktail. Incubate the beads in 10 mL of Fmoc deprotection cocktail for 15 min with agitation. Repeat this step.
5. Wash the beads three times in NMP (10 mL) for 10 min each, followed by three washes with DCM (10 mL) and three more washes with NMP (10 mL). The deprotected resin now has an available amine where the first amino acid—Fmoc-Cys(Trt)-OH—will be attached.

6. Freshly prepare Fmoc-Cys(Trt)-OH coupling cocktail. Immediately add the solution to the resin. Agitate for at least 3 h at room temperature. Peptide couplings can proceed overnight to provide a convenient stopping point.
7. Dispose of the coupling solution. Wash the beads three times in NMP (10 mL) for 10 min each, followed by three washes with DCM (10 mL). Eject all DCM.
8. Check coupling efficiency by ninhydrin (Kaiser) test. In a microfuge tube, add 10 µL of reagent 1, 10 µL of reagent 2, and 10 µL of reagent 3. Carefully remove the plunger from the syringe and stick a pipette tip into the resin. Only a few beads are needed and what sticks to the pipette tip will be more than sufficient. Pipette up and down into the ninhydrin test mix to transfer the beads to the microfuge tube. Close the lid securely with a tube lock and boil for 3 min in a 100 °C heating block. If the first amino acid has been completely coupled to the resin, no available amine will remain and a colorless solution (negative test) is present. If the solution is purple (positive test), coupling was incomplete. In either case, wash the resin three times with NMP (10 mL). If the test was positive, repeat the coupling step (**step 6**). Proceed to the next step if the test is negative.
9. Repeat **steps 4** through **8** to remove the Fmoc attached to the cysteine residue and couple the next amino acid—Fmoc-Gly-OH—to the cysteine α -amine using the Fmoc-Gly-OH coupling cocktail.
10. Repeat **steps 4** through **8** to remove the Fmoc attached to the first glycine residue and couple the next amino acid—Fmoc-Gly-OH—to the first glycine α -amine using Fmoc-Gly-OH coupling cocktail.
11. Repeat **steps 4** through **8** to remove the Fmoc attached to the second glycine residue and couple the final amino acid—Fmoc-Gly-OH—to the second glycine α -amine using Fmoc-Gly-OH coupling cocktail.
12. Remove the terminal Fmoc group by repeating **steps 4** and **5**.
13. The peptide must be liberated from the resin and in the process, the trityl group protecting the cysteine sulfhydryl will be removed. Freshly prepare cleavage cocktail. Wash the resin three times with DCM. Suspend the resin in 3 mL of cleavage cocktail and agitate for 1 h. Safeguard the eluate in a glass container and repeat the cleavage procedure two more times, pooling all of the eluates (9 mL total).
14. During the 3-h cleavage reaction, chill three 50 mL Falcon tubes filled with diethyl ether in a bucket with dry ice inside of the fume hood.

15. Add 3 mL of pooled eluate, dropwise, to each tube of diethyl ether to precipitate the peptide. Incubate on dry ice for a further 30 min and then centrifuge at $4,000 \times g$ for 15 min in a swinging bucket centrifuge. Discard the supernatant and repeat two more times.
16. Dissolve and combine the three peptide pellets in 1 mL of methanol. Add this dropwise to 50 mL of chilled diethyl ether and centrifuge at $4,000 \times g$ for 15 min. Repeat this methanol precipitation two more times to remove reducing agent from the crude peptide. Remove the final ether wash as completely as possible.
17. Preweigh a microfuge tube. Dissolve the peptide pellet in a minimal amount of water and transfer to the microfuge tube. Remove the lid from a second microfuge tube, place it on the tube containing the peptide, and poke a small hole in the top with a narrow-bore (27 G) needle. Use a speed-vac to concentrate the peptide solution to dryness. The second cap protects from losses that occur if residual diethyl ether violently evaporates under the reduced pressure of the speed-vac. Weigh the tube with peptide and subtract the tube weight to arrive at an estimation of peptide quantity and yield.
18. The crude sulphydryl-containing peptide can now be coupled to the maleimide-containing Cy5 dye (*see Note 6*). Add 3 M equivalents of peptide to 1 M equivalent of freshly prepared Cy5-maleimide coupling solution. Incubate overnight at room temperature with agitation in a foil-wrapped tube.
19. Add two equivalents of dithiothreitol (DTT) to the reaction to quench any unreacted Cy5 maleimide and incubate for a further 2 h with agitation.

3.3 Probe Purification

The dye-containing peptide conjugate must be resolved from the excess free peptide, since free peptide will effectively compete to resolve the acyl-enzyme intermediate formed by sortase during labeling (Fig. 1). Owing to the excellent robustness and specificity of the sortase enzyme, this is the only impurity that must be removed. If available, the best option is to use a reversed-phase high-performance liquid chromatography system (RP-HPLC) equipped with a C18 column for purification. However, this is often beyond the scope of most molecular biology labs and so the following protocol exploits the fact that the GGGC scaffold peptide binds weakly if at all to inexpensive disposable syringe-driven cartridges filled with C18 bonded silica sorbents. In contrast, the dye conjugate is more retentive and can be separated from the scaffold peptide.

1. Attach a Bond Elut JR C18 Cartridge to a 10 mL syringe and slowly and gently pass 10 mL of acetonitrile over the resin bed.

For all steps, the solvent should be ejected gently with minimal pressure and should exit the cartridge as drops rather than as a stream.

2. Repeat with 10 mL of H₂O/0.1 % TFA to equilibrate the resin.
3. Dilute 75 µL of the crude G₃C(Cy5) peptide containing reaction into 900 µL of H₂O/0.1 % TFA. Gently pass this through the resin and collect the flow-through (FT).
4. Wash the resin with 1 mL of H₂O/0.1 % TFA.
5. Begin a step gradient elution of the G₃C(Cy5) peptide from the resin. Wash with 1 mL each of 5, 10, 15, 20, 25, 30, 35, 40, and 100 % acetonitrile in H₂O/0.1 % TFA. Collect each eluate fraction for testing.
6. Identity of the fraction containing the dye-peptide conjugate is best done by mass spectrometry (MALDI or LC-MS). However, this is beyond the resources of many labs. As an alternative, each fraction can simply be tested in a sortase reaction (Fig. 2a). Concentrate each fraction in preweighed tubes down to dryness using a speed-vac. Redissolve each fraction in water such that the apparent concentration is 5 mM (based on the predicted molecular weight of the probe). Test each fraction in a sortase reaction as outlined in Subheading 3.4 (*see Notes 7 and 8*). An essential control is to omit sortase, mixing only the peptide fraction with the target protein to ensure that no active Cy5 maleimide has been carried over into the purified fractions (Fig. 2a, lane 4). Once active fractions are identified, these are pooled for further use.

3.4 Labeling Optimization

All labeling reactions are subject to optimization. All substrates tested to date have been successfully labeled, but the kinetics of labeling can vary by substrate. Although labeling with SrtA 5° requires far less enzyme than the wild-type counterpart, an additional consideration is that the improved kinetics may favor hydrolysis upon long incubation times (Fig. 2b, Lane 8). This is easily circumvented by simply terminating the reaction upon complete conversion to the transpeptidation product, but before hydrolysis product accumulates. Thus a time course of the reaction is essential for optimization and is outlined below.

1. Mix labeling buffer (to a final concentration of 1×), 30 µM substrate, 5 µM SrtA 5°, and 500 µM probe together (*see Note 7*). Incubate at 37 °C (*see Note 8*). Withdraw aliquots at various time points, quenching with SDS-PAGE sample buffer. Analyze samples by SDS-PAGE (Fig. 2b) or by an appropriate method tailored to the protein of interest (i.e., mass spectrometry, spectroscopy, fluorimetry, immunoblot).

4 Notes

1. *Staphylococcus aureus* sortase A 5° is produced in excellent, albeit slightly lower yield (15–20 mg/L of culture) than its wild-type counterpart and is available upon request [5] (David Liu's Lab, Harvard University Chemistry Department. E-mail: drliu@fas.harvard.edu).
2. *Staphylococcus aureus* sortase A and its variants are extremely soluble. They are routinely concentrated to stocks of >2 mM and have been concentrated up to 4 mM without solubility issues.
3. The VHH substrate described here and elsewhere [18] is likewise produced in excellent yield (~40 mg/L of culture) and can be concentrated to >2 mM without solubility issues. It is available upon request (Hidde Ploegh's lab, MIT Biology Department/Whitehead Institute. E-mail: ploegh@wi.mit.edu) and represents an easily produced positive control protein for test labeling reactions.
4. Sortase A is purified in the rigorous absence of protease inhibitors because these may react with the catalytic cysteine residue. Target proteins may be purified in the presence of protease inhibitors if desired, provided that they are removed by chromatography/dialysis before the labeling reaction is performed.
5. Protein concentration of sortase A and its variants should be determined by Bradford assay. Estimation of concentration by predicted molar extinction coefficient and absorbance at 280 nm is highly inaccurate for sortase A.
6. After liberation of the cysteine-containing peptide from the resin and subsequent precipitation, the peptide should be used immediately to avoid oxidation and formation of disulfide bonds. If this is a concern, the peptide can be shaken with immobilized Tris(2-carboxyethyl)phosphine (TCEP) beads (Pierce/Thermo Scientific) at room temperature for 10–15 min before use.
7. The reaction buffer used during sortase labeling contains calcium chloride and should not be added to proteins in phosphate-containing buffers. This leads to precipitation of calcium phosphate and often precipitation of the protein. All purified components should be diluted in Tris- or HEPES-containing buffers.
8. The concentration of each reaction component can be adjusted to speed or slow the reaction appropriately. For instance, decreasing the amount of SrtA 5° leads to slower transpeptidation product accumulation and also slower hydrolysis product accumulation. Increasing the amount of nucleophile may help

to avoid hydrolysis product accumulation. Decreasing the reaction temperature to room temperature and even below can also be used for target proteins with solubility issues at 37 °C, but this usually necessitates longer incubation times to achieve quantitative conversion to the transpeptidation product.

Acknowledgements

MWP is supported by an HHMI fellowship from the Damon Runyon Cancer Research Foundation, DRG-2119-12. I gratefully thank Hidde Ploegh, David Liu, and Lynne Maquat for advice, reagents, and support.

References

1. Marraffini LA, Dedent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* 70(1):192–221
2. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* 96(22):12424–12429
3. Connolly KM, Smith BT, Pilpa R, Ilangovan U, Jung ME, Clubb RT (2003) Sortase from *Staphylococcus aureus* does not contain a thiolate-imidazolium ion pair in its active site. *J Biol Chem* 278(36):34061–34065
4. Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* 44(33):11188–11200
5. Chen I, Dorr BM, Liu DR (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* 108(28):11399–11404
6. Popp MW, Antos JM, Ploegh HL (2009) Site-specific protein labeling via sortase-mediated transpeptidation. *Curr Protoc Protein Sci* Chapter 15, Unit 15.3
7. Antos JM, Miller GM, Grotenbreg GM, Ploegh HL (2008) Lipid modification of proteins through sortase-catalyzed transpeptidation. *J Am Chem Soc* 130(48):16338–16343
8. Yamamoto T, Nagamune T (2009) Expansion of the sortase-mediated labeling method for site-specific N-terminal labeling of cell surface proteins on living cells. *Chem Commun (Camb)* 9:1022–1024
9. Antos JM, Chew GL, Guimaraes CP, Yoder NC, Grotenbreg GM, Popp MW, Ploegh HL (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* 131(31):10800–10801
10. Guimaraes CP, Carette JE, Varadarajan M, Antos J, Popp MW, Spooner E, Brummelkamp TR, Ploegh HL (2011) Identification of host cell factors required for intoxication through use of modified cholera toxin. *J Cell Biol* 195(5):751–764
11. Popp MW, Artavanis-Tsakonas K, Ploegh HL (2009) Substrate filtering by the active site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations. *J Biol Chem* 284(6):3593–3602
12. Popp MW, Karssemeijer RA, Ploegh HL (2012) Chemoenzymatic site-specific labeling of influenza glycoproteins as a tool to observe virus budding in real time. *PLoS Pathog* 8(3):e1002604
13. Antos JM, Popp MW, Ernst R, Chew GL, Spooner E, Ploegh HL (2009) A straight path to circular proteins. *J Biol Chem* 284(23):16028–16036
14. Claessen JH, Witte MD, Yoder NC, Zhu AY, Spooner E, Ploegh HL (2013) Catch-and-release probes applied to semi-intact cells reveal ubiquitin-specific protease expression in *Chlamydia trachomatis* infection. *ChemBioChem* 14(3):343–352
15. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3(11):707–708
16. Popp MW, Dougan SK, Chuang TY, Spooner E, Ploegh HL (2011) Sortase-catalyzed

- transformations that improve the properties of cytokines. *Proc Natl Acad Sci U S A* 108(8):3169–3174
17. Swee LK, Guimaraes CP, Sehrawat S, Spooner E, Barrasa MI, Ploegh HL (2013) Sortase-mediated modification of alphaDEC205 affords optimization of antigen presentation and immunization against a set of viral epitopes. *Proc Natl Acad Sci U S A* 110(4):1428–1433
 18. Witte MD, Cagnolini JJ, Dougan SK, Yoder NC, Popp MW, Ploegh HL (2012) Preparation of unnatural N-to-N and C-to-C protein fusions. *Proc Natl Acad Sci U S A* 109(30):11993–11998
 19. Tanaka T, Yamamoto T, Tsukiji S, Nagamune T (2008) Site-specific protein modification on living cells catalyzed by Sortase. *Chembiochem* 9(5):802–807
 20. Mao H, Hart SA, Schink A, Pollok BA (2004) Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* 126(9):2670–2671
 21. Hendricks GL, Weirich KL, Viswanathan K, Li J, Shriver ZH, Ashour J, Ploegh HL, Kurt-Jones EA, Fygenson DK, Finberg RW, Comolli JC, Wang JP (2013) Sialylnolacto-N-tetraose c (LSTc)-bearing liposomal decoys capture influenza A virus. *J Biol Chem* 288(12):8061–8073
 22. Tafesse FG, Sanyal S, Ashour J, Guimaraes CP, Hermansson M, Somerharju P, Ploegh HL (2013) Intact sphingomyelin biosynthetic pathway is essential for intracellular transport of influenza virus glycoproteins. *Proc Natl Acad Sci U S A* 110(16):6406–6411
 23. Hess GT, Cagnolini JJ, Popp MW, Allen MA, Dougan SK, Spooner E, Ploegh HL, Belcher AM, Guimaraes CP (2012) M13 bacteriophage display framework that allows sortase-mediated modification of surface-accessible phage proteins. *Bioconjug Chem* 23(7):1478–1487
 24. Esteban A, Popp MW, Vyas VK, Strijbis K, Ploegh HL, Fink GR (2011) Fungal recognition is mediated by the association of dectin-1 and galectin-3 in macrophages. *Proc Natl Acad Sci U S A* 108(34):14270–14275
 25. Hirota N, Yasuda D, Hashidate T, Yamamoto T, Yamaguchi S, Nagamune T, Nagase T, Shimizu T, Nakamura M (2010) Amino acid residues critical for endoplasmic reticulum export and trafficking of platelet-activating factor receptor. *J Biol Chem* 285(8):5931–5940
 26. Wu Z, Guo X, Wang Q, Swarts BM, Guo Z (2010) Sortase A-catalyzed transpeptidation of glycosylphosphatidylinositol derivatives for chemoenzymatic synthesis of GPI-anchored proteins. *J Am Chem Soc* 132(5):1567–1571
 27. Popp MW, Ploegh HL (2011) Making and breaking peptide bonds: protein engineering using sortase. *Angew Chem Int Ed Engl* 50(22):5024–5032
 28. Strijbis K, Spooner E, Ploegh HL (2012) Protein ligation in living cells using sortase. *Traffic* 13(6):780–789
 29. Hendrickx AP, Budzik JM, Oh SY, Schneewind O (2011) Architects at the bacterial surface—sortases and the assembly of pili with isopeptide bonds. *Nat Rev Microbiol* 9(3):166–176
 30. Ling JJ, Policarpo RL, Rabideau AE, Liao X, Pentelute BL (2012) Protein thioester synthesis enabled by sortase. *J Am Chem Soc* 134(26):10749–10752
 31. Pierce NW, Lee JE, Liu X, Sweredoski MJ, Graham RL, Larimore EA, Rome M, Zheng N, Clurman BE, Hess S, Shan SO, Deshaies RJ (2013) Cand1 promotes assembly of new SCF complexes through dynamic exchange of F box proteins. *Cell* 153(1):206–215
 32. Pos W, Sethi DK, Call MJ, Schulze MS, Anders AK, Pyrdol J, Wucherpfennig KW (2012) Crystal structure of the HLA-DM-HLA-DRI complex defines mechanisms for rapid peptide selection. *Cell* 151(7):1557–1568
 33. Bornhorst JA, Falke JJ (2000) Purification of proteins using polyhistidine affinity tags. *Methods Enzymol* 326:245–254
 34. Coin I, Beyermann M, Bienert M (2007) Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat Protoc* 2(12):3247–3256

Chapter 14

BONCAT: Metabolic Labeling, Click Chemistry, and Affinity Purification of Newly Synthesized Proteomes

Peter Landgraf, Elmer R. Antileo, Erin M. Schuman,
and Daniela C. Dieterich

Abstract

Metabolic labeling of proteins using classical radioisotope-labeled amino acids has enabled the analysis and function of protein synthesis for many biological processes but cannot be combined with modern high-throughput mass spectrometry analysis. This chapter describes the unbiased identification of a whole de novo synthesized proteome of cultured cells or of a translationally active subcellular fraction of the mammalian brain. This technique relies on the introduction of a small bioorthogonal reactive group by metabolic labeling accomplished by replacing the amino acid methionine by the azide-bearing methionine surrogate azidohomoalanine (AHA) or the amino acid homopropargylglycine (HPG). Subsequently an alkyne- or azide-bearing affinity tag is covalently attached to the group by “click chemistry”—a copper(I)-catalyzed [3+2] azide-alkyne cycloaddition. Affinity tag-labeled proteins can be analyzed in candidate-based approaches by conventional biochemical methods or with high-throughput mass spectrometry.

Key words Protein synthesis, Click chemistry, Affinity purification, Proteome, Mass spectrometry

1 Introduction

Intracellular and intercellular communication and signaling in every cell mainly rely on proteins and their specific interactions and functions. Being part of a dynamic and functional entity such as a whole cell, a subcellular compartment like, for example, the chemical synapse or a signaling cascade requires that each computational unit is able to adjust to changes in the environment or in the activity pattern on the timescale of seconds to minutes. Thus, understanding cellular protein homeostasis including protein synthesis, protein turnover and degradation, and the various types of post-translational modifications is of pivotal interest to decipher the complex and dynamic processes of, for example, the development of an organism or how the mammalian brain enables lifelong learning and long-lasting memories. In neuroscience, a large body of literature suggests that protein synthesis—besides posttranslational

protein modifications, such as protein phosphorylation, glycosylation, acylation, and others, as well as directed protein turnover—is important for both long-term synaptic plasticity and memory. The detailed molecular and cellular processes, however, by which an intricate cellular network like the brain assembles and the mechanisms underlying the dynamic properties of the adult brain that allow it to engage in complex processes like learning and memory are still not completely resolved. Other examples include organ formation and patterning during development or the dysbalances in protein homeostasis that lead to pathophysiological conditions like cancer or neurodegenerative diseases.

Facing an estimated number of approximately 10,000 different proteins in a single mammalian cell [1], a total number of 1,000–1,500 different/unique proteins per chemical synapse [2], and a turnover rate of 0.7 % per hour of the synaptic protein content [3], in-depth identification of a cell’s entire proteome, let alone the comparison to another proteome, is a major challenge for modern proteomics. While today’s state-of-the-art MS instruments routinely sequence single purified proteins with subfemtomolar sensitivity, the effective identification of low-abundance proteins is orders of magnitude lower in complex mixtures due to limited dynamic range and sequencing speed [4]. The characterization of a proteome is an even more difficult challenge if temporal and spatial aspects of a proteome or a subpopulation of the proteome have to be taken into consideration. The separation and enrichment of the subproteome in question is key for its successful characterization. While posttranslational modifications such as phosphorylation or ubiquitination readily provide a suitable handle for enrichment of the “phosphoproteome” or for proteins destined for degradation, reducing sample complexity by selectively enriching newly synthesized proteins is troublesome, since all proteins—old and new—share the same pool of 20 amino acids.

To specifically label newly synthesized proteins, BONCAT (bioorthogonal noncanonical amino acid tagging) and FUNCAT (fluorescent noncanonical amino acid tagging) were developed [5–8]. These complementary techniques enable one to identify and visualize the subpopulation of newly synthesized protein using bioorthogonal noncanonical amino acids by utilizing the cell’s own translation machinery in a highly specific fashion.

With BONCAT it is possible to focus on one aspect of protein homeostasis, which enables a higher degree of in-depth analysis. In this chapter the application of BONCAT to investigate newly synthesized proteomes is described in detail for HEK293 cells, primary cortical cultures from rats, and synaptoneuroosomes (SNS) prepared from the rat brain.

In the first step of BONCAT, newly synthesized proteins are labeled using either the azide-bearing artificial amino acid azidohomoalanine (AHA) or the alkyne-bearing amino acid

homopropargylglycine (HPG) as surrogates for methionine, thus endowing the proteins with a novel azide or alkyne functionality that serves to distinguish them from the pool of preexisting proteins. These small chemical groups deliver unique chemical functionality to their target molecules, which can subsequently be tagged with exogenously delivered probes for detection or isolation in a highly selective manner. The metabolic labeling step is very similar to classical radioisotope labeling and can be combined or follow pharmacological treatment or classical pulse-chase experiments (Fig. 1). To increase the fraction of substituted methionine, a methionine-depletion step prior to AHA or HPG addition is advisable, and methionine must be absent from the medium during the metabolic labeling reaction. Employing copper-catalyzed azide-alkyne ligation (or “click chemistry”), the reactive azide

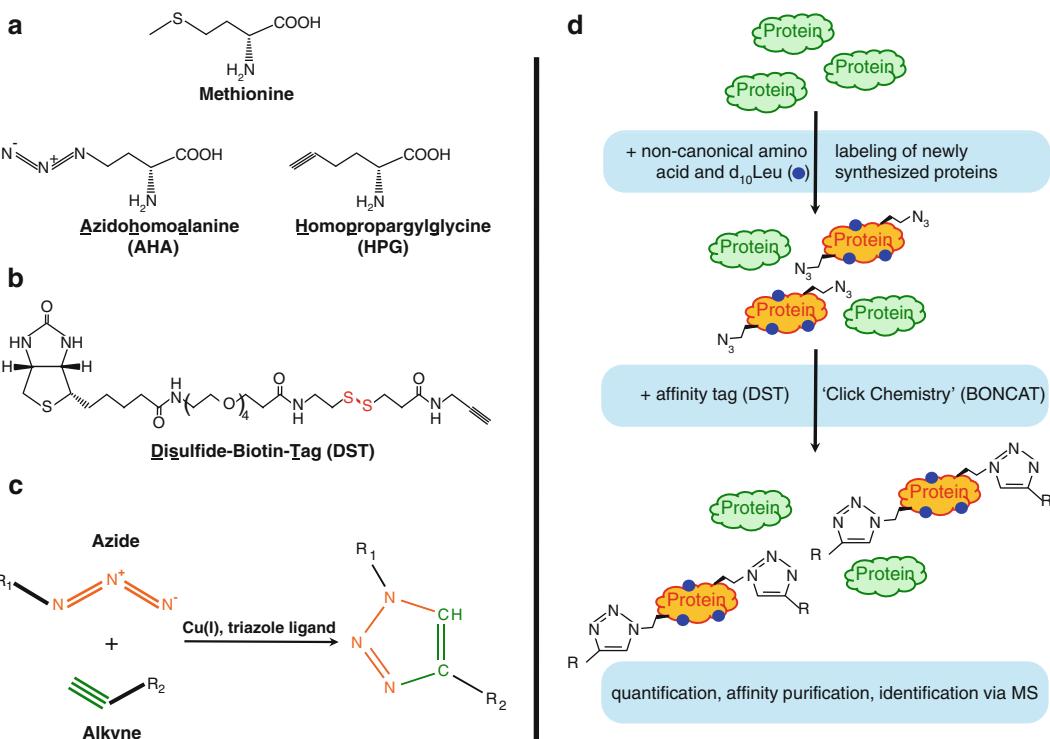


Fig. 1 BONCAT—reagents, principle, and application. (a) Chemical structures of methionine and its noncanonical surrogates azidohomoalanine (AHA) and homopropargylglycine (HPG). (b) DST—a cleavable alkynyl disulfide (red) biotin affinity tag. (c) Reaction scheme of the copper-catalyzed [3+2] azide-alkyne cycloaddition. (d) Cells or translationally active cellular components are supplemented with an azide- or alkyne-bearing noncanonical amino acid in methionine-free media. As an additional validation component, d_{10} L (blue dot) is added. After incorporation into newly synthesized proteins, AHA- or HPG-harboring proteins can be tagged via a “click chemistry” reaction with an affinity tag bearing an alkyne or an azide moiety. In downstream applications newly synthesized, affinity-tagged proteins can be analyzed via Western blot or processed for affinity purification and identification by MS

group of AHA or the alkyne group of HPG is covalently coupled to an alkyne-bearing affinity tag or to an azide-bearing affinity tag, respectively, in the second step (Fig. 1). These tags enable the subsequent detection, affinity purification, and eventually MS identification of AHA- or HPG-labeled proteins.

2 Materials

2.1 Equipment

1. Temperature-controlled shaker for microcentrifuge and 15 mL Falcon tubes (temperature range: room temperature to 100 °C).
2. Microcentrifuge for 1 and 2 mL tubes, cooled.
3. Centrifuge for 15 and 50 mL Falcon tubes, cooled.
4. Vortex.
5. Rotator for microcentrifuge and Falcon tubes.
6. Dot blot apparatus.
7. SDS-PAGE and Western blot unit.
8. Accessory materials: automatic pipettes (e.g., Eppendorf) ranging from 0.1 to 1,000 µL and respective plastic tips, microcentrifuge tubes (1.5 and 2.0 mL), Falcon tubes (15 and 50 mL), petri dishes with a diameter of 60 × 15 mm, microcentrifuge spin columns with polyethylene filter (30 µm pore size, e.g., Pierce), cooling box for crushed ice, glass beakers of different sizes.

2.2 Chemicals

- All chemicals should be of analytical grade unless otherwise noted; specific requirements in quality are indicated in the text.
- Water, molecular biology grade.
- L-Azidohomoalanine (AHA) (keep powder desiccated at room temperature).
- L-Methionine (Met) (nonanimal source, cell culture tested, store desiccated at room temperature).
- L-Leucine-d₁₀ (d₁₀L, d₁₀>98 %, protect from light. Store desiccated at room temperature).
- L-Glutamine, 200 mM (100×) (store in 1 mL aliquots at -20 °C).
- Triazole ligand (store powder desiccated at room temperature).
- Dimethyl sulfoxide (DMSO, ≥99.7 %, packed in 5 or 10 mL flame-sealed vials, store at room temperature) (*see Note 1*).
- Biotin-alkyne disulfide tag (DST) (*see Note 2*).
- Copper(I) bromide, 99.999 % (small portions should be finely ground with a mortar and a pestle and can be kept for 1 month, store powder desiccated at room temperature).

- Immobilized Tris-(2-carboxyethyl)phosphine hydrochloride (immobilized TCEP, keep at 4 °C).
- Iodoacetamide ($\text{C}_2\text{H}_4\text{INO}$, store desiccated at 4 °C).
- *Complete*TM EDTA-free protease inhibitor cocktail (Roche, store at 4 °C).
- Benzonase Nuclease ($\geq 250 \text{ U}/\mu\text{L}$, store at -20 °C).
- RiboLock RNase inhibitor ($40 \text{ U}/\mu\text{L}$, store at -20 °C).
- PD-10 columns (GE Healthcare, store at room temperature).
- Immobilized NeutrAvidin (store at 4 °C).
- Triton X-100 (in best available quality, store at room temperature).
- Igepal CA-630 (chemically not distinguishable from NP-40, store at room temperature).
- 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, store at room temperature).
- Ammonium bicarbonate (CH_5NO_3 , store at room temperature).
- 2-Mercaptoethanol (store at 4 °C).
- Polyclonal rabbit biotin antibody (e.g., Bethyl Laboratories, 1:10,000, store at 4 °C).
- Secondary HRP-conjugated anti-rabbit antibody (keep as recommended by the supplier).
- Neurobasal[®] Medium (Invitrogen, store at 4 °C).
- B-27[®] Serum-Free Supplement (Invitrogen, 50 ×, store at 4 °C).
- Penicillin/Streptomycin (Pen/Strep, 100×, store portioned in 1 mL aliquots at -20 °C).
- GlutamaxTM (Invitrogen, 200 mM, store in 1 mL aliquots at -20 °C).

2.3 Stock Solutions and Buffer

2.3.1 Stock Solutions

The following stock solutions are useful to perform the experiments in an efficient and speedy manner. They should be stored as indicated in the text or mentioned in Subheading 4.

1. Neurobasal[®] Medium + (NB+): Neurobasal[®] Medium supplemented with 0.8 mM L-glutamine and B-27[®] (1:50).
2. 2× HBS (HEPES-buffered saline), pH 7.35: 20 mM HEPES, 238 mM NaCl, 10 mM KCl, 4 mM CaCl₂, 4 mM MgCl₂, 60 mM glucose (store in 50 mL aliquots at -20 °C).
3. 10× PBS (phosphate-buffered saline), pH 7.4 and pH 7.8: 1.37 M NaCl, 27 mM KCl, 43 mM Na₂HPO₄, 14 mM NaH₂PO₄ (see Note 3).
4. 20 % (v/v) Triton X-100 in ultrapure water.
5. 20 % (w/v) SDS in ultrapure water.

6. 100 mM AHA in molecular biology grade water (*see Note 4*).
7. 100 mM L-methionine in molecular biology grade water (*see Note 4*).
8. 100 mM L-leucine-d₁₀ in molecular biology grade water (*see Note 4*).
9. 200 mM triazole ligand in DMSO (keep in 20 µL aliquots at -20 °C).
10. 25 mM biotin-alkyne disulfide tag (DST) in 1× PBS, pH 7.8 (keep in 20 µL aliquots at -20 °C).
11. 4× SDS sample buffer: 1 % (w/v) SDS, 40 % (v/v) glycerol, 20 % (v/v) 2-mercaptoethanol, 0.004 % (w/v) bromophenol blue, 250 mM Tris/HCl, pH 6.8.
12. 4× SDS sample buffer (without 2-mercaptoethanol): 1 % (w/v) SDS, 40 % (v/v) glycerol, 0.004 % (w/v) bromophenol blue, 250 mM Tris/HCl, pH 6.8.
13. 50× PI: *Complete* EDTA-free protease inhibitor dissolved in 1 mL molecular biology grade water (store at -20 °C up to 1 month).

2.3.2 Buffers and Working Solutions

The following buffers and working solutions should be prepared either freshly or kept at the appropriate conditions mentioned in Subheading 4. All buffers and solutions used for click chemistry must be free of EDTA, EGTA, or other chelators as these reagents lead to the inactivation of the copper(I) catalyst.

1. PBS-MC: 1× PBS, pH 7.4, 1 mM MgCl₂, 0.1 mM CaCl₂.
2. PBS-PI: 1× PBS, pH 7.8, supplemented with *Complete* EDTA-free protease inhibitor (*see Note 5*).
3. PD-10 column buffer: 0.5 % (w/v) SDS in 1× PBS, pH 7.8 or pH 7.5 (*see Note 6*).
4. 0.5 M iodoacetamide: Dissolve 92.5 mg in 1 mL molecular biology grade water (*see Note 7*).
5. Cu(I) Br suspension: Suspend 10 mg in 1 mL molecular biology grade water (*see Note 8*).
6. NeutrAvidin-binding buffer: 1 % (v/v) Igepal and 0.5 % (w/v) SDS in 1× PBS, pH 7.5.
7. NeutrAvidin wash buffer A: 1 % (v/v) Igepal in 1× PBS, pH 7.5.
8. NeutrAvidin wash buffer B: 1× PBS, pH 7.5.
9. 50 mM ammonium bicarbonate: Dissolve 200 mg in 50 mL molecular biology grade water and prepare freshly.

2.3.3 Reagent Setup

For metabolic labeling of newly synthesized proteomes of HEK293 cells, primary neuronal cultures, or subcellular fractions, like synaptoneuroosomes (SNS) prepared from rat brains,

we use a final concentration of 4 mM AHA and 4 mM d₁₀L in the respective culture media or labeling buffer; for further details, *see Note 9*. Controls are supplemented with 4 mM L-methionine instead of AHA. Labeling times strongly depend on the protein synthesis capacity of the respective cultures or fraction and should be chosen accordingly to the scientific question (*see Note 10*). For cultures and subcellular fractions, the following media and buffers are used:

HEK293 cells

1× HBS, pH 7.35, alternatively: DMEM (Dulbecco's Modified Eagle Medium) -Met/-L, supplemented with B-27® (1:50), Pen/Strep (1:100), Glutamax™ (1:100).

Primary neuronal cultures

Hibernate -Met/-L: medium without Met and L, supplemented with B-27® (1:50), Pen/Strep (1:100), Glutamax™ (1:100).

Synaptoneuroosomes (SNS)

Preincubation buffer: 25 mM HEPES, pH 7.4, 20 mM glucose, 3.5 mM KCl, 1.2 mM Na₂HPO₄, 2 mM MgCl₂, 129 mM NaCl.

Incubation buffer: 25 mM HEPES, pH 7.4, 20 mM glucose, 3.5 mM KCl, 1.2 mM Na₂HPO₄, 1 mM MgCl₂, 1.8 mM CaCl₂, 129 mM NaCl (*see Note 11*).

3 Methods

3.1 Metabolic

Labeling of HEK293

Cells, Primary

Neuronal Cultures,

and SNS Using AHA

and d₁₀L

Required conditions for a sufficient metabolic labeling with AHA/d₁₀L largely depend on the respective cell type (culture system) or subcellular fraction as well as on the scientific question, i.e., time window, with or without stimulation, etc. For efficient labeling it is important to deplete the respective medium for Met in order to increase the amount of AHA/d₁₀L incorporation into the nascent peptide chains, as AHA is not efficiently charged onto the Met-tRNA in the presence of Met [9]. However, optimal conditions for the maintenance of cell viability and protein homeostasis are desirable at the same time and thus pose limits on the duration of Met depletion. In the following three different, cell type-specific labeling strategies are described (Fig. 2):

HEK293 cells

HEK293 cells are a robust, rapid proliferating cell line, requiring a high amount of de novo synthesized proteins for processes like cell division, growth, and differentiation. Thus, efficient metabolic labeling with AHA/d₁₀L can be realized in relative short time periods (2–4 h) under comparable less demanding conditions. The following labeling and tagging protocol is outlined for one 75 cm² culture flask.

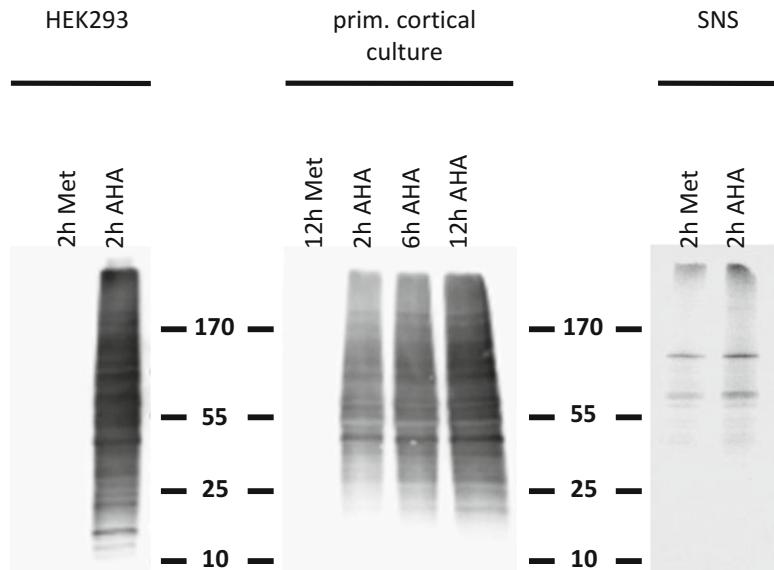


Fig. 2 Western blot analysis of newly synthesized, biotin-tagged proteins from HEK cells, primary neuronal cultures, and translational active synaptoneuroosomes (SNS). Proliferating HEK cells are characterized by high protein turnover rates, whereby efficient protein labeling appears within a short time period (2 h). In contrast, the protein turnover of primary cortical cultures is comparably low, and a longer labeling time is required (2, 6 and 12 h). SNS are a translation-active subcellular fraction with comparably low and time-limited protein synthesis capacity (2 h; note that the blot shown on the right was exposed for a longer time in order to visualize the biotin signal)

1. Cultivate HEK293 cells under appropriate conditions to 80–90 % confluence.
2. Pre-warm a sufficient amount of 1× HBS, pH 7.35 to 37 °C.
3. Rinse cells briefly and gently with 7 mL 1× HBS in the 75 cm² culture flask.
4. Incubate cells with 7 mL 1× HBS for 30 min at 37 °C and 5 % CO₂ in a cell culture incubator (*see Note 12*).
5. Carefully remove 1× HBS and add 7 mL fresh 1× HBS, supplemented with 4 mM AHA and 4 mM d₁₀L (final concentrations); incubate for 2–4 h at 37 °C and 5 % CO₂. For the control group, supplement with 4 mM L-methionine instead of AHA.
6. After incubation, place the culture flasks on ice, carefully remove the AHA/d₁₀L containing HBS, and rinse cells briefly in chilled PBS-MC (*see Note 13*).
7. Harvest cells in 5 mL of chilled PBS-PI and transfer the suspensions to a 15 mL Falcon tube. Spin down for 5 min at 3,000 × g and 4 °C, carefully remove the supernatant from the cell pellets, and continue directly with Subheading 3.2, or freeze the pellets at –80 °C (*see Note 14* for further details).

Primary neuronal cultures

Primary neuronal cells represent a highly sensitive and finicky culture system, characterized by low protein synthesis rates compared to rapidly dividing cells such as HEK293 cells. Therefore, longer time periods are required for sufficient metabolic labeling of newly synthesized proteins with AHA/d₁₀L. Under these circumstances the selection of an appropriate neuronal cell culture media for the maintenance of cell viability, if necessary over days, is of particular importance. For the protocol described below, one 75 cm² culture flask containing 21-day-old (DIV 21) primary cortical cultures from rat brains is used. Culture conditions are 3 × 10⁶ cell/75 cm² culture flask, 10 mL NB+, 37 °C, 5 % CO₂, and saturated humidity [10].

1. Pre-warm a sufficient volume of Hibernate -Met/-L, supplemented with B-27® (1:50), Pen/Strep (1:100), and Glutamax™ (1:100) to 37 °C.
2. Carefully remove NB+ from the cells and gently and briefly rinse with 7 mL pre-warmed Hibernate -Met/-L.
3. Incubate neuronal cells with 7 mL Hibernate -Met/-L for 30 min at 37 °C, 5 % CO₂, and saturated humidity in an incubator (*see Note 11*).
4. Remove the Hibernate medium and add 7 mL fresh, 4 mM AHA/d₁₀L containing Hibernate -Met/-L, supplemented with B-27® (1:50), Pen/Strep (1:100), and Glutamax™ (1:100); incubate for up to 3 days at 37 °C, 5 % CO₂, and saturated humidity (*see Note 15*). For the control group, supplement with 4 mM L-methionine instead of AHA.
5. For finishing the experiment, place the culture flasks on ice, remove carefully the AHA/d₁₀L containing Hibernate, and briefly rinse cells in chilled PBS-MC (*see Note 13*). Repeat this washing step once more.
6. Collect the cells in 500 µL of PBS-PI, pH 7.8, using a cell scraper and transfer the suspension to a 1.5 mL plastic tube. Continue directly with Subheading 3.2, or freeze the suspension at -80 °C (*see Notes 14 and 16*).

Synaptoneuroosomes (SNS)

SNS are a protein translation-active, subcellular biochemical fraction containing post- and presynaptic specializations, distal parts of dendritic and axonal structures, astroglial terminals, and all essential components of the protein synthesis machinery. Therefore, analyzing de novo synthesized proteins from SNS is an essential prerequisite to define the role of local protein synthesis during neuronal development, maintenance, and function. A variety of different preparation techniques for SNS have been described in the literature [11–14]. We prefer a preparation method using initial subfractionation via differential centrifugation, followed by

velocity sedimentation using a Ficoll step gradient (5, 13 and 16 %), leading to a higher enrichment degree and lower contamination levels with intact soma or nuclei. Regardless of the preparation method used, a short preparation time is key for maintaining the protein synthesis capacity of the isolated structures in vitro. The protocol described in the following paragraph is optimized for SNS prepared from one adult rat brain and starts with the respective final SNS fraction from a density step gradient. The obtained amount of SNS is sufficient for 4–5 individual samples that can be tested under different experimental conditions.

1. Harvest the specific SNS band from the density step gradient (ca. 1.5 mL) and transfer to a 15 mL Falcon tube.
2. Adjust to 15 mL with chilled preincubation buffer and spin down for 5 min at $3,000 \times g$ and 4 °C. Remove the supernatant carefully.
3. Resuspend the pellet in 15 mL chilled preincubation buffer by gently pipetting up and down, and spin down for 5 min at $3,000 \times g$ and 4 °C. Remove the supernatant carefully.
4. Add 12 mL chilled incubation buffer and gently resuspend the SNS fraction by pipetting up and down several times (*see Note 17*).
5. Transfer the SNS suspension in 3 mL aliquots to petri dishes with a diameter of 60 × 15 mm and dispense the suspension homogenously in order to achieve a maximal surface for efficient gaseous exchange.
6. Supplement each petri dish with 120 µL AHA and d₁₀L from 100 mM stock solutions in order to reach a final concentration of 4 mM each. For the control group, supplement with 4 L-methionine instead of AHA.
7. Shake the suspension gently 3–4 times and place the dishes for 2 h in an incubator at 37 °C, 5 % CO₂, and saturated humidity.
8. After incubation place the dishes on crushed ice and transfer the suspensions to fresh 15 mL Falcon tubes. Rinse the dishes with 3 mL PBS-MC each (*see Note 13*), and combine these suspensions with the corresponding Falcon tube. Spin down for 5 min at $3,000 \times g$ and 4 °C and discard the supernatants carefully.
9. Adjust each (AHA-incorporated) SNS fraction to 10 mL PBS-PI, pH 7.8, and resuspend gently. Spin down for 5 min at $3,000 \times g$ and 4 °C once more; carefully remove the supernatant and discard it.
10. Proceed directly with Subheading 3.2, or freeze the pellet at -80 °C (*see Note 14*).

3.2 Lysis of AHA-Labeled Cells or Subcellular Protein Fractions

The following protocol section is designed for pellets from one 75 cm² culture flask of 90 % confluent, AHA-labeled HEK293 cells, yielding 5 mL of protein extract. For smaller volumes the particular components must be adjusted accordingly (*see Note 18*).

1. Thaw the frozen cell pellet on ice and add 975 µL of PBS-PI, pH 7.8; homogenize carefully by pipetting up and down or vortex briefly.
2. Add 25 µL 20 % (w/v) SDS and vortex vigorously.
3. Add 1 µL of Benzonase (\geq 500 U) and mix well by pipetting up and down several times.
4. Boil the sample for 10 min at 95–100 °C while gently mixing (700 rpm) in a thermo shaker.
5. Chill on ice and add 3,950 µL PBS-PI, pH 7.8, and 50 µL 20 % (v/v) Triton X-100; mix gently (*see Note 19*).
6. Centrifuge for 5 min at 3,000 $\times g$ and 4 °C.
7. Transfer the remaining supernatant, representing the cell-free protein extract, to a fresh Falcon tube (*see Note 20*).

3.3 Reduction and Alkylation of AHA-Labeled Protein Extracts

1. Wash immobilized TCEP resin: Transfer 1 mL TCEP suspension to a 15 mL Falcon tube and adjust to 10 mL with PBS, pH 7.8. Mix carefully and spin down at 1,000 $\times g$ and 4 °C for 5 min. Discard the supernatant and repeat twice.
2. Add the protein extract from Subheading 3.2, and incubate with the TCEP resin for 75 min under gentle agitation at room temperature in an end-over-end mixer.
3. Add 120 µL freshly prepared 0.5 M iodoacetamide solution and wrap the Falcon tube in aluminum foil to protect it from light. Agitate for 30 min in the end-over-end mixer at room temperature.
4. Spin down the resin at 1,000 $\times g$ and 4 °C for 5 min. Transfer the remaining supernatant to a fresh Falcon tube.
5. If necessary adjust the volume to 5 mL with PBS, pH 7.8, and start immediately with desalting (Subheading 3.4).

3.4 First Desalting with PD-10 Columns

At this point a first desalting step is necessary to remove any excess of TCEP and iodoacetamide. You will need one PD-10 column for 2.5 mL of protein extract and two columns for 5 mL protein extract. If you have smaller volumes, either adjust your sample to 2.5 mL or use smaller-sized columns (*see Note 21*).

1. Equilibration of PD-10 column: Cut off the tips from the column, open top lid, and attach the columns in an appropriate holder. Let the preserving buffer flow through by gravity. Equilibrate the column by adding five times 5 mL 0.05 % SDS in PBS, pH 7.8.

2. Load 2.5 mL of the reduced and alkylated protein extract from Subheading 3.3 on one PD-10 column and discard the flow-through.
3. Elute proteins from the column by adding 3.5 mL 0.05 % SDS in PBS, pH 7.8, and collect eluate in a fresh 15 mL Falcon tube. Combine the eluates from one sample (7 mL in total).
4. Add 140 μ L of 50× PI, mix carefully, and proceed immediately with the tagging reaction.

3.5 Tagging Reaction (Click Chemistry)

For each 1 mL of the eluate from Subheading 3.4, add 1 μ L triazole ligand, 1 μ L DST tag, and 10 μ L freshly prepared copper(I) bromide suspension (in the mentioned order). The copper(I) bromide suspension must be prepared immediately before starting the tagging reaction (*see Note 8*). The steps in the following section are calculated for 7 mL of desalted protein extract from Subheading 3.4.

1. Add 7 μ L triazole ligand directly into the sample and vortex for 10 s (*see Note 22*).
2. Add 7 μ L DST tag and vortex for 10 s.
3. Add 70 μ L copper(I) bromide suspension and vortex finally for at least 30 s (*see Note 23*).
4. Incubate at 4 °C over night or at least for 6 h at room temperature with agitation in an end-over-end mixer.
5. Spin down at 3,000 $\times g$ at 4 °C for 5 min. Transfer the supernatant to a fresh Falcon tube, and adjust to 7.5 mL with PBS, pH 7.8 (*see Note 24*). Proceed immediately with the second desalting step.

3.6 Second Desalting with PD-10 Columns

A second desalting step is required to remove any excessive, unreacted DST tag, copper, and triazole ligand. The procedure follows exactly the description in Subheading 3.4, except that the elution buffer is 0.05 % SDS in PBS, pH 7.5. For desalting 7.5 mL of tagged sample and three fresh PD-10 columns are required (*see Note 25*).

3.7 Affinity Purification of DST-Tagged Proteins Using NeutrAvidin® Agarose

1. Wash NeutrAvidin resin three times with ten bed volumes of NeutrAvidin-binding buffer (agitate briefly so the resin becomes homogenously distributed, centrifuge at 3,000 $\times g$ for 5 min at 4 °C, discard supernatant, and repeat this washing step twice).
2. Add Igepal to a final concentration of 1 % (v/v) to the sample from Subheading 3.6 and agitate for 20 min at room temperature in the end-over-end mixer until the detergent is completely dissolved (*see Note 26*).
3. Add the tagged sample to the NeutrAvidin Agarose from point 1 and agitate the mix constantly in an end-over-end mixer at 4 °C over night (12–16 h) to ensure proper binding of the biotin moiety to the NeutrAvidin resin (*see Note 27*).

4. Spin down the resin at $3,000 \times g$ for 5 min at 4 °C; carefully remove the supernatant (unbound fraction) to a fresh tube (*see Notes 26 and 28*).
5. Wash the resin (bound fraction) three times with ten bed volumes of NeutrAvidin washing buffer A (agitate constantly for 10 min at room temperature, spin down at $3,000 \times g$ for 5 min at 4 °C, discard the supernatant, and repeat procedure twice).
6. Wash the resin three times with ten bed volumes of NeutrAvidin washing buffer B as described in **step 5**.
7. Wash the resin in ten bed volumes of 50 mM ammonium bicarbonate for 10 min at room temperature.
8. Spin down the resin at $3,000 \times g$ for 5 min at 4 °C and remove the supernatant except for 1.5 mL resin/ammonium bicarbonate slurry.
9. Transfer the slurry to a 2 mL centrifuge tube and spin the resin at $3,000 \times g$ for 5 min at 4 °C.
10. Carefully remove all of the supernatant and adjust to 1 mL with 50 mM ammonium bicarbonate (*see Note 26*).
11. Add 50 µL 2-mercaptoethanol to the sample and agitate for 30 min at room temperature in an end-over-end mixer to cleave the tagged proteins from the resin (*see Note 29*).
12. Spin down at $3,000 \times g$ for 5 min at 4 °C and collect the supernatant (eluate 1) containing newly synthesized proteins in a fresh 2 mL centrifuge tube.
13. Adjust the resin to 1 mL with 5 % (v/v) 2-mercaptoethanol/ammonium bicarbonate solution, agitate for 30 min at room temperature in an end-over-end mixer, spin down at $3,000 \times g$ for 5 min at 4 °C, collect the supernatant (eluate 2), and combine with eluate 1.
14. Repeat **step 13** yielding eluate 3 and combine with eluates 1 and 2 (*see Notes 26, 30, and 31*).
15. Finally pass the combined eluates through a microcentrifuge spin column containing a polyethylene filter (30 µm pore size) according to the manufacturer's protocol and dry the proteins in a lyophilizer (*see Notes 26, 31, and 32*).

4 Notes

1. DMSO is very hygroscopic. Therefore, avoid the exposure to air and water. For the preparation of the triazole ligand solution, the DMSO must be water-free. Use only freshly opened flame-sealed vials. Keep the triazole ligand solutions in 25 µL aliquots at -20 °C.
2. The DST (biotin-alkyne disulfide tag) is synthesized according to a published protocol described in [15].

3. 10× PBS is prepared in double-distilled water and can be stored at room temperature up to 6 months. The respective pH of 7.4, 7.5, or 7.8 is adjusted with a few drops of 1 N NaOH. 1× PBS should be prepared freshly for each experiment and the respective pH adjusted again, if necessary.
4. 100 mM amino acid stock solutions in molecular biology grade water can be stored for up to 1 month at 4 °C.
5. PBS-PI should be prepared freshly! At 4 °C the buffer can be stored up to 1 week. If other protease inhibitors as recommended are used, make sure that they are free of EDTA, EGTA, or other chelators.
6. PD-10 columns should be equilibrated with the same buffer that is used for the subsequent elution of the newly synthesized proteins and depends from the experimental requirements of the following step.
7. Iodoacetamide solution has to be prepared freshly and kept protected from light.
8. The catalytically active Cu(I) ion of the copper bromide suspension reacts in the presence of water very rapidly to Cu(0) and Cu(II). Therefore, prepare the copper bromide suspension immediately before starting with click chemistry. In case of numerous samples, prepare multiple suspensions, for 3–5 reactions each. Vortex the suspension very carefully for 15–20 s.
9. The average content of Met in proteins is 1.7 %. Thus, if Met is surrogated by its noncanonical sibling AHA, a comparable small part of newly synthesized proteins carries the metabolic label. In consequence, fewer proteins can be positively validated after MS analysis by means of identifiable metabolically incorporated noncanonical amino acids. For this reason, the simultaneous metabolic labeling with a second, modified amino acid is recommended. In this protocol, tenfold deuterated L-leucine ($d_{10}L$) is used, because leucine is the most abundant essential amino acid (7.5 % of all amino acid residues, in contrast, for instance, arginine with only 4.7 % of all amino acid residues [16]). An alternative method, especially if quantification of identified proteins is a concern, was published very recently [17] and is based on the combination of BONCAT and stable isotope labeling of amino acids (SILAC).
10. Rapidly proliferating cell lines are characterized by a high capacity for the de novo synthesis of proteins, consistent with the requirements of fast dividing, growing, and differentiating cells. Thus, a sufficient amount of newly synthesized proteins can be obtained after a comparable short labeling time with AHA/ $d_{10}L$ (1–2 h). In contrast, differentiated primary cell cultures with low or, as in the case of neurons, no cell division

require longer labeling times with AHA/d₁₀L due to lower protein turnover rates (>12 h). Under these conditions, the selected labeling time additionally depends on the turnover rates of interested proteins. Proteins with high turnover rates can be labeled within relative short time periods, whereas proteins with low turnover rates require longer labeling times for subsequent detection of candidates.

11. The composition of the preincubation and the incubation buffer used in this protocol has been modified from a buffer system described in Meffert, M., Premack, B. A., and Schulman [18] accordingly. For labeling of SNS with AHA/d₁₀L, other click chemistry compatible buffer systems might be used.
12. This step allows the depletion of essential amino acids including methionine and leucine.
13. MgCl₂ and CaCl₂ of the PBS-MC preserve membrane integrity by maintaining proper ionic strength in the buffer system.
14. After flash freezing in liquid nitrogen cell pellets can be stored at -80 °C for several months without any effect on click chemistry efficiency.
15. Hibernate medium was prepared according to the protocol from Brewer and Price, 1996 [19], whereby methionine and leucine were omitted. Additional supplementation with B27®, Pen/Strep, and Glutamax facilitates long-lasting labeling experiments in primary neuronal cultures with AHA/d₁₀L, while simultaneously a maximum of cellular viability is maintained.
16. Neuronal cells adhere very tightly at the bottom of culture flasks. Harvesting the cells with a scraper can lead to the damage of cells and the consequent release of cytosol. In order to keep cytosolic proteins, cells are harvested in the required reaction buffer for the subsequent lysis and click chemistry.
17. Take care to obtain a real homogenous suspension in order to realize comparable conditions between different samples from one preparation.
18. Proteins of neuronal cultures from one 75 cm² culture flask (3 Mio cells, DIV 21) are solubilized in 0.5 mL PBS-PI, pH 7.8. For the pellets from one SNS sample, 1 mL PBS-PI, pH 7.8, is used.
19. The final SDS concentration in the reaction mix should be below 0.2 %, because higher concentrations decrease the efficiency of the tagging reaction. After adding the Triton X-100, avoid foam formation by intense mixing.
20. The protein extract can be frozen at -80 °C for a longer time. If the procedure is continued, thaw the extract on ice and heat for 5 min at 95 °C; chill on ice.

21. Follow exactly the manufacturer's protocol for the PD-10 columns! Differing volumes for loaded samples or elution buffers lead to the loss of proteins and/or unwanted contamination.
22. After adding the triazole ligand, the sample appears slightly milky, which vanishes after adding DST tag and copper bromide suspension.
23. For an efficient click chemistry reaction, it is essential to keep the described order without any delay in order to avoid triazole ligand precipitation and disproportionation of the Cu(I) ion. Thus, the reaction mix has to be set up separately for each sample! If more than two samples are processed, use multiple fresh copper bromide suspensions, prepared immediately before starting these reactions.
24. Adjusting the volume to 7, 5 ml is required for proper gravity flow using the PD-10 columns.
25. PD-10 columns can be regenerated several times, in particular within one tagging and affinity purification experiment. For this purpose rinse the columns four times with 25 mL ddH₂O. Finally equilibrate with 25 mL PBS, pH 7.5.
26. In order to monitor successful tagging reaction and affinity purification, take a 30 µL aliquot, solubilize with 10 µL 4× SDS sample buffer (without 2-mercaptoethanol), and boil for 5 min at 95 °C. Analyze on SDS-PAGE and Western blot or freeze temporary at -20 °C. Attention: If other sample buffers are used, they should not contain 2-mercaptoethanol or other reducing compounds. Otherwise the disulfide bound of the DST tag will be cleaved, leading to the release of the biotin tag.
27. The overnight agitation with NeutrAvidin can be used as a pause point in the protocol.
28. The unbound fraction should be kept at -20 °C. In case of inefficient binding or excess of tagged proteins, affinity purification can be repeated.
29. Skin contact and inhalation of 2-mercaptoethanol is highly toxic. For handling the samples, take care for self-protection and use a fume hood.
30. At this point, the effective elution of proteins cleaved from the DST tag can be analyzed by solubilizing an aliquot of the remaining NeutrAvidin matrix with 4× SDS sample buffer.
31. Cleavage of the DST tag leads releases the biotin! Thus, no biotin signal is visible in Western blots. The relevant lanes appear empty but can be analyzed in the SDS-PAGE.
32. Filtration of the combined eluates removes contamination from the NeutrAvidin matrix in the sample that can be stored at -80 °C till lyophilizing.

Acknowledgments

This work has received funding from the Deutsche Forschungsgemeinschaft (DI1512/1-1 and DI1512/1-2), the DIP (Deutsch-Israelische-Projektkooperation) German-Israeli Project Cooperation foundation, and the CBBS, Magdeburg, Germany, to DCD.

References

1. Pandey A, Mann M (2000) Proteomics to study genes and genomes. *Nature* 405: 837–846
2. Pielot R, Smalla KH, Müller A et al (2012) SynProt: a database for proteins of detergent-resistant synaptic protein preparations. *Front Synaptic Neurosci* 10(3389)
3. Cohen LD, Zuchman R, Sorokina O et al (2013) Metabolic turnover of synaptic proteins: kinetics, interdependencies and implications for synaptic maintenance. *PLoS One* 8:e63191
4. de Godoy LM, Olsen JV, de Souza GA et al (2006) Status of complete proteome analysis by mass spectrometry: SILAC labelled yeast as a model system. *Genome Biol* 7:R50
5. Dieterich DC, Link AJ, Graumann J et al (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* 103:9482–9487
6. Dieterich DC, Lee JJ, Link AJ et al (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal noncanonical amino acid tagging. *Nat Protoc* 2:532–540
7. Dieterich DC, Hodas JJ, Gouzer G et al (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat Neurosci* 13:897–905
8. Taylor AM, Dieterich DC, Ito HT et al (2010) Microfluidic local perfusion chambers for the visualization and manipulation of synapses. *Neuron* 66:57–68
9. Link AJ, Vink MKS, Tirell DA (2007) Preparation of the functionalizable methionine surrogate azidohomoalanine via copper-catalyzed diazo transfer. *Nat Protoc* 2: 1879–1883
10. Goslin K, Asmussen H, Banker G (1989) Rat hippocampal neurons in low-density culture. In: Banker G, Goslin K (eds) *Culturing nerve cells*. MIT Press, Cambridge, pp 339–370
11. Kiebler MA, Lopez-Garcia JC, Leopold PL (1999) Purification and characterization of rat hippocampal CA3-dendritic spines associated with mossy fiber terminals. *FEBS Lett* 445: 80–86
12. Rao A, Steward O (1991) Evidence that protein constituents of postsynaptic membrane specialisations are locally synthesized: analysis of proteins synthesized within synaptosomes. *J Neurosci* 11:2881–2895
13. Williams C, Shai RM, Wu Y et al (2009) Transcriptome analysis of synaptoneuroosomes identifies neuroplasticity genes overexpressed in incipient Alzheimer disease. *PLoS One* 4:e4936. doi:[10.1371/journal.pone.0004936](https://doi.org/10.1371/journal.pone.0004936)
14. Troca-Marin JA, Alves-Sampaio A, Tejedor FJ et al (2010) Local translation of dendritic RhoA revealed by an improved synaptoneurosome preparation. *Mol Cell Neurosci* 43:308–314
15. Szychowski J, Mahdavi A, Hodas JJ et al (2010) Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J Am Chem Soc* 132:18351–18360
16. Bruice PY (2004) *Organic chemistry*, 4th edn. Pearson Education Inc, New Delhi, pp 960–962
17. Howden AJM, Geoghegan V, Katsch K et al (2013) QuaNCAT: quantitating proteome dynamics in primary cells. *Nat Methods* 10:343–346
18. Meffert MK, Premack BA, Schulman H (1994) Nitric oxide stimulates Ca^{2+} -independent synaptic vesicle release. *Neuron* 12:1235–1244
19. Brewer GJ, Price PJ (1996) Viable cultured neurons in ambient hibernate storage for a month. *Neuroreport* 7:1509–1512

Chapter 15

Genetic Encoding of Unnatural Amino Acids for Labeling Proteins

Kathrin Lang, Lloyd Davis, and Jason W. Chin

Abstract

The site-specific incorporation of bioorthogonal groups via genetic code expansion provides a powerful general strategy for site-specifically labeling proteins with any probe. Here we describe the genetic encoding of dienophile-bearing unnatural amino acids into proteins expressed in *Escherichia coli* and mammalian cells using the pyrrolysyl-tRNA synthetase/tRNA_{CUA} pair and its variants. We describe the rapid fluorogenic labeling of proteins containing these unnatural amino acids in vitro, in *E. coli*, and in live mammalian cells with tetrazine–fluorophore conjugates in a bioorthogonal Diels–Alder reaction with inverse electron demand. These approaches have been extended to site-specific protein labeling in animals, and we anticipate that they will have a broad impact on the labeling and imaging field.

Key words Genetic encoding, Unnatural amino acid, Bioorthogonal reaction, Diels–Alder cycloaddition, Site-specific protein labeling, Tetrazine–fluorophores, Dienophiles

1 Introduction

Developments in biomolecule labeling and microscopy are necessary to improve the scope of biological imaging and have historically been tightly coupled. The ability to label any specific protein at a single position in the polypeptide with any desired small molecule fluorophore or probe would provide the ultimate labeling method for super-resolution microscopy, Fluorescence resonance energy transfer (FRET), and single-molecule imaging [1, 2]. This, however, still represents an outstanding challenge for chemical biologists. Enzyme-mediated labeling approaches may allow for rapid labeling, but require the use of protein or peptide fusions that introduce perturbations into the protein under study and limit the sites that can be labeled [3–6]. In principle the genetically encoded, site-specific incorporation of custom-synthesized unnatural amino acids bearing a bioorthogonal functional group and their rapid and highly specific labeling with a chemical reporter with tailored physical and biological properties could fulfill many of the requirements for an ideal

protein labeling strategy (Fig. 1). The incorporation of designer amino acids can be achieved using “orthogonal” aminoacyl-tRNA synthetase/tRNA pairs that direct amino acid incorporation in response to an amber stop codon (UAG) placed in a gene of interest [7–17]. The reaction between the unnatural amino acid and the chemical reporter has to proceed under biologically compatible conditions, in which the components react quickly with each other to form a stable product but do not react with other chemical groups found within the cellular context, a property referred to as bioorthogonality. Many bioorthogonal reactions for which a component can be genetically encoded, however, are too slow to affect quantitative site-specific labeling of proteins on a time scale that is useful for studying many biological processes. Building on our previous work, we have recently discovered pyrrolysyl-tRNA synthetase/tRNA pairs for the efficient site-specific incorporation of several unnatural amino acids useful for site-specific protein labeling, including dienophile-bearing amino acids Ne -*trans*-cyclooctene-4-oxycarbonyl-L-lysine, Ne -(bicyclo[6.1.0]non-4-yn-9-yl-methoxy)carbonyl-L-lysine, and Ne -5-norbornene-2-yloxycarbonyl-L-lysine (**TCOK**, **BCNK**, and **NorK**, respectively, Fig. 2a), into proteins expressed in *Escherichia coli* and mammalian cells. These modified proteins can be site-specifically labeled with tetrazine-fluorophore conjugates in a Diels–Alder reaction with inverse electron demand (Fig. 2b) [18, 19]. These reactions are very specific and very rapid, and the tetrazine fluorophores, which are initially weakly fluorescent, become strongly fluorescent once attached to the protein via the chemical reaction, making the signal to noise of this labeling approach superior.

Recent work has also demonstrated the viability of unnatural amino acid incorporation in *C. elegans* and *D. melanogaster*, and it is hoped that this technology can be further expanded into other organisms.



Fig. 1 Bioorthogonal chemical reactions allow labeling of an incorporated unnatural amino acid bearing a bioorthogonal functional group (*yellow circle*) with a chemical reporter (*red star*) under biologically compatible conditions

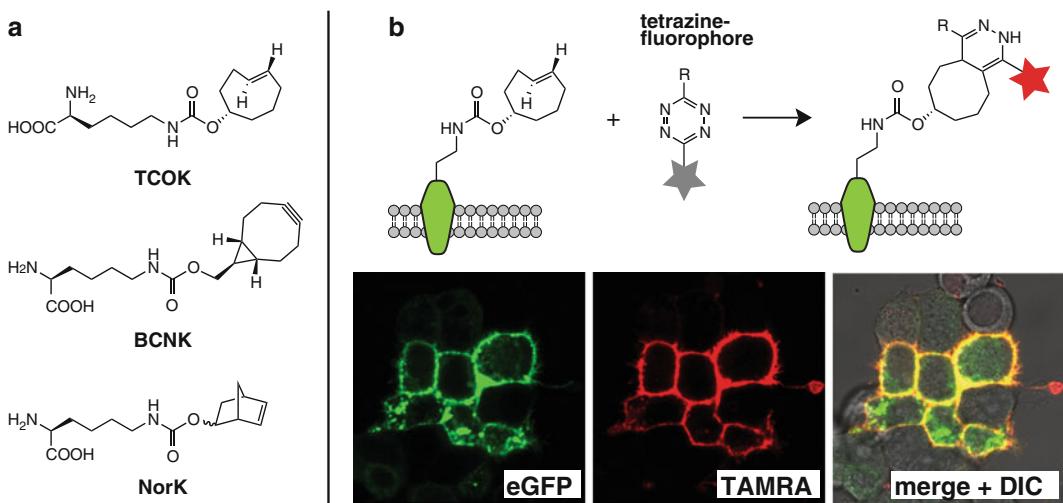


Fig. 2 Site-specific labeling of a cell surface receptor in live mammalian cells. (a) Amino acid structures of trans-cyclooctene-, bicyclo[6.1.0]nonyne-, and norbornene-bearing amino acids (**TCOK**, **BCNK**, and **NorK**). (b) Site-specific incorporation of **TCOK** into the extracellular domain of an epidermal growth factor (EGFR)-eGFP fusion and its rapid, fluorogenic, and specific labeling with a tetrazine–tetramethylrhodamine (TAMRA) conjugate. Note that despite endogenous amber codons, we see minimal background labeling of nontarget proteins, either on the cell surface or intracellularly

Compared to other methods for fluorescently labeling proteins, the site-specific incorporation of unnatural amino acids, bearing bioorthogonal groups, and their subsequent chemoselective reaction with customized probes introduces minimal perturbations to proteins. Furthermore, this approach to labeling proteins is highly flexible and allows labeling of proteins with a variety of probes including a range of fluorophores, cross-linking agents, and cytotoxic molecules. It is hoped that this method will expand the range of available cellular labeling targets, thus facilitating previously unobtainable imaging of protein dynamics, structure, and function in living cells and animals and helping to answer important biological questions that cannot be addressed by other labeling approaches.

In the following we describe the site-specific labeling of proteins *in vitro*, in *E. coli*, and in live mammalian cells (both on the surface and intracellularly) via incorporation of the unnatural amino acids **TCOK**, **BCNK**, and **NorK** and their subsequent chemoselective reaction with tetrazine–fluorophore conjugates.

For clarity reasons we exemplify in the following Subheadings 2 and 3 the incorporation of TCOX and labeling of proteins thereof. Slight modifications in the protocol for incorporation of BCNK and NorK can be found in the Subheading 4.

2 Materials

2.1 Expression and Purification of Proteins with Site-Specifically Incorporated TCOK in *E. coli*

1. Chemically competent *E. Coli* DH10B cells (homemade or commercial).
2. Plasmid construct that encodes for the mutant TCOK-incorporating *Methanosaarcina barkeri* pyrrolysyl-tRNA synthetase (TCORS) (see Note 1).
3. Plasmid construct that encodes for the wild-type pyrrolysyl-tRNA_{CUA} and a C-terminally hexahistidine-tagged protein of interest containing an in-frame amber codon in the open reading frame at the site at which it is to be labeled (see Notes 2–4).
4. Stock solutions of unnatural amino acids were prepared as follows: TCOK, 50 mM in 0.1 M NaOH; BCNK, 50 mM in 0.1 M NaOH; NorK, 100 mM in DMSO (see Note 5).
5. SOB medium supplemented with 0.2 % glucose.
6. LB medium.
7. 1,000-fold antibiotic stock solutions were prepared as follows: ampicillin, 100 mg/mL (in water); tetracycline, 25 mg/mL (in 7/3 ethanol/water).
8. L-(+)-arabinose: 20 % w/v in water.
9. Lysis buffer: 10 mM Tris–HCl, 20 mM imidazole, 200 mM NaCl, pH 8, 1 mM phenylmethanesulfonyl fluoride, 1 mg/mL lysozyme, 100 µg/mL DNase A, Roche protease inhibitor (1 tablet/50 mL buffer).
10. Ni²⁺-NTA beads (QIAGEN).
11. Wash buffer: 20 mM Tris–HCl, 30 mM imidazole, 300 mM NaCl, pH 8.
12. Elution buffer: 20 mM Tris–HCl, 300 mM imidazole, 300 mM NaCl, pH 8.
13. Buffer A: 20 mM Tris–HCl, 100 mM NaCl, 2 mM EDTA, pH 7.4.
14. HiLoad 16/60 Superdex 75 prep grade column.
15. Amicon Ultra-15 3 kDa MWCO centrifugal filter device (Millipore).
16. NuPAGE LDS sample buffer (Invitrogen).
17. SDS-PAGE running buffer (1× NuPAGE MES SDS running buffer, Invitrogen).

2.2 *E. coli* Protein Labeling with Tetrazine–Dye Probes

1. Tetrazine–fluorophore conjugates: stock solutions, 2 mM in DMSO (see Note 6).
2. Phosphate buffered saline (PBS buffer).

3. NuPAGE LDS sample buffer (Invitrogen).
4. SDS-PAGE running buffer (1× NuPAGE MES SDS running buffer, Invitrogen).

2.3 Incorporation of TCOK in Mammalian Cell Proteins and Site-Specific Protein Labeling

1. Imaging dishes (35 mm μ-dishes, Ibidi) or 25 mm coverslips in a 6-well cell culture plate.
2. 0.01 % poly-L-lysine (Sigma).
3. Phosphate buffered saline (PBS buffer).
4. HEK 293 cells.
5. Plasmid encoding for a protein of interest containing an in-frame amber codon at the site at which it is to be labeled (*see Note 2*) as well as for the mutant TCOK-incorporating *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (*see Notes 1 and 7*).
6. Plasmid encoding for the wild-type pyrrolysyl-tRNA_{CUA} (*see Note 8*).
7. Cell culture medium: DMEM (Dulbecco's modified eagle medium) containing 10 % FBS (fetal bovine serum) and penicillin/streptomycin.
8. Lipofectamine 2000 (Life Technologies).
9. Opti-MEM serum-free medium (Life Technologies).
10. TCOK.
11. Tetrazine-dyes: 2 mM in DMSO (*see Note 6*).

3 Methods

3.1 Expression and Purification of Proteins with Site-Specifically Incorporated TCOK in *E. coli*

Following this protocol we expressed and purified many different C-terminally His-tagged recombinant cytosolic proteins (10–30 kDa) like T4 lysozyme, sfGFP, myoglobin, and ubiquitin. Changes in the purification protocol might be necessary for larger or insoluble proteins.

1. Mix 50–100 ng of the plasmid that encodes for TCORS and the plasmid that encodes for the tRNA_{CUA} and a C-terminally hexahistidine-tagged protein of interest containing an in-frame amber codon in the open reading frame at the site at which it is to be labeled (**items 2 and 3** of Subheading **2.1**) with 50 μL of homemade chemically competent DH10B cells. Incubate for 15 min on ice, heat-shock for 60 s at 42 °C, and incubate for 5 min on ice. Recover cells in 1 mL of SOB medium (supplemented with 0.3 % glucose) for 1 h at 37 °C.
2. Incubate recovered cells in 100 mL of LB containing 100 μg/mL ampicillin and 25 μg/mL tetracycline at 37 °C overnight.
3. Use 20 mL of this overnight culture to inoculate 1 L of LB supplemented with 50 μg/mL ampicillin and 12.5 μg/mL tetracycline and incubate at 37 °C.

4. At OD₆₀₀ 0.3–0.5, add the TCOK stock solution (*see item 4* of Subheading 2.1) to a final concentration of 2 mM (*see Note 9*).
5. After 30 min induce protein expression by the addition of arabinose to a final concentration of 0.2 %.
6. After 3.5 h of induction harvest cells by centrifugation. At this point the cell pellet can be frozen at –80 °C until required.
7. Thaw cell pellet on ice and suspend in 30 mL of lysis buffer.
8. Lyse cells by sonication at 4 °C.
9. Clarify the extract by centrifugation (20 min, 21,000 ×*g*, 4 °C). Remove the supernatant to a new container and discard the pellet.
10. Add 1 mL of Ni²⁺-NTA beads to the extract and incubate the mixture with agitation for 1 h at 4 °C.
11. Collect beads by centrifugation (10 min, 1,000 ×*g*).
12. Wash beads three times with 30 mL wash buffer and collect by centrifugation (10 min, 1,000 ×*g*).
13. Subsequently resuspend beads in 10 mL of wash buffer and transfer them to a column.
14. Elute protein with 3 mL of elution buffer.
15. Check purity of protein by SDS-PAGE (*see Note 10*).
16. If necessary purify protein further by chromatography. Perform size-exclusion chromatography employing a HiLoad 16/60 Superdex 75 prep grade column at a flow rate of 1 mL/min with buffer A.
17. Pool fractions containing the protein and concentrate the protein. Use an Amicon Ultra-15 3 kDa MWCO centrifugal filter device.
18. Confirm the mass of the protein and the incorporation of TCOK by mass spectrometry (*see Note 11*).

3.2 Protein Labeling with Tetrazine-Dye Probes

3.2.1 In Vitro Labeling of Purified Proteins with Tetrazine-Dye Conjugates

1. Incubate 0.2–2 nmol of purified protein with incorporated TCOK (in buffer A, *see Note 12*) with ten equivalents of tetrazine-fluorophore conjugate (stock solution, 2 mM in DMSO, *see Note 6*).
2. Incubate for 2–15 min.
3. Analyze by 4–12 % SDS-PAGE and scan gel with a Typhoon imager to make fluorescent bands visible. Alternatively analyze by ESI-MS (*see Note 11*).

*3.2.2 Labeling of the Whole *E. coli* Proteome with Tetrazine-Dye Conjugates*

*To demonstrate that the reaction between TCOK and tetrazine-fluorophore conjugates is highly selective within a cellular context, the labeling reaction can be performed on *E. coli* expressing a protein incorporating TCOK.*

1. After growing bacteria in the presence of TCOK for 3.5 h (**step 6** of Subheading 3.1), collect cell pellet by centrifugation ($16,000 \times g$, 2 min) of 1 mL of cell suspension.
2. Suspend cell pellet in 500 μ L PBS buffer, spin down ($16,000 \times g$, 5 min), and discard supernatant (*see Note 13*).
3. Repeat **step 2** twice.
4. Resuspend cell pellet in 100 μ L PBS and incubate with 3 μ L of tetrazine-dye conjugate (2 mM in DMSO) for 2–15 min at 37 °C.
5. Collect cell pellet by centrifugation ($16,000 \times g$, 5 min) and discard supernatant.
6. Resuspend cells and wash cell pellet twice with 500 μ L PBS buffer (*see Note 14*).
7. Suspend cell pellet in 100 μ L NuPAGE LDS sample buffer supplemented with 5 % β -mercaptoethanol.
8. Heat the mixture to 90 °C for 10 min and centrifuge at $16,000 \times g$ for 10 min.
9. Analyze the crude cell lysate by 4–12 % SDS-PAGE to assess protein levels.
10. Scan gel with a Typhoon imager to make fluorescent bands visible (*see Note 15*).

3.3 Incorporation of TCOK in Mammalian Cell Proteins and Site-Specific Protein Labeling

1. Incubate the imaging dishes or glass coverslips with 0.01 % poly-L-lysine solution for 30 min. Rinse the dishes/coverslips twice with PBS buffer.
2. Plate cells onto the dishes/coverslips in cell culture medium and incubate overnight (37 °C, 10 % CO₂). Allow the cells to grow to 70–90 % confluence.
3. For transfection mix 10 μ L of Lipofectamine 2000 transfection reagent with 250 μ L of Opti-MEM serum-free medium. Mix gently and incubate for 5 min at room temperature.
4. Mix 2 μ g of each plasmid (**items 5** and **6** of Subheading 2.3) with 250 μ L of Opti-MEM serum-free medium. Incubate for 5 min at room temperature. (Note: proceed to **step 5** within 20 min.)
5. After 5 min incubation combine the diluted DNA (**step 4**) with diluted Lipofectamine 2000 transfection reagent (**step 3**) (total volume = 500 μ L). Mix gently and incubate for 20 min at room temperature.
6. Add the transfection mixture from **step 5** to the imaging dishes containing cells and medium by gently dropping it onto the surface. Mix gently by rocking the dish back and forth.
Incubate cells for at least 4 h in the transfection mixture. This incubation should be conducted in the same medium conditions as used for growing cells overnight.

7. To dissolve TCOK in cell culture medium, add 1 mL of cell culture medium to the appropriate amount of TCOK and then add 100 μ L of 1 M NaOH solution to dissolve the amino acid. Add medium to the final volume (final concentration of TCOK 1 mM) and neutralize by adding 100 μ L of 1 M HCl. Before adding to cells, filter this solution through a 0.2 μ m membrane sterile syringe filter (*see Note 16*).
8. Pipette off the transfection medium and cover cells with 2 ml of cell culture medium containing TCOK (**step 7**). Mix gently by rocking the dish back and forth.
9. Incubate cells overnight at 37 °C in 10 % CO₂.
10. Pipette off the TCOK-containing medium and add fresh cell culture medium. Rinse the cells twice more with fresh cell culture medium. Then leave cells for at least 1 h in fresh medium at 37 °C in 10 % CO₂.
11. After this incubation, exchange the medium once more before adding the labeling medium containing tetrazine–fluorophore conjugate.
12. Prepare cell culture medium containing 200 nM tetrazine–fluorophore conjugate by diluting a 2 mM stock solution of the corresponding tetrazine–fluorophore conjugate in fresh cell culture medium. Add 2 mL of this solution directly to imaging dishes to cover cells.
13. Incubate cells in the presence of tetrazine–fluorophore conjugate for 2–15 min at 37 °C in 10 % CO₂.
14. Remove the fluorophore-containing medium and rinse the cells three times with fresh cell culture medium.
15. At this point labeled cell surface proteins can be imaged by fluorescence microscopy.
16. For internal proteins it may be necessary to perform additional wash steps with incubations in order to clear dye within cells (*see Note 17*).

4 Notes

1. The TCORS is a *Methanosarcina barkeri* wild-type pyrrolysyl-tRNA synthetase variant containing the mutations Y271A, L274M, and C313A (TCORS). For incorporation of BCNK we used a *Methanosarcina barkeri* wild-type pyrrolysyl-tRNA synthetase variant containing the mutations Y271M, L274G, and C313A (BCNRS). Incorporation of NorK was achieved using the *Methanosarcina barkeri* wild-type pyrrolysyl-tRNA synthetase. (The plasmid that encodes TCORS has ampicillin resistance in our experiments.)

2. There are several considerations to be made when selecting a labeling site. The site of labeling must be sufficiently surface exposed to be accessible to the tetrazine–fluorophore during labeling. Furthermore, this residue must be one that can be mutated without disruption to folding of the protein. Furthermore, as a considerable proportion of the protein will terminate at this point, it is important to take care that the truncated protein, N-terminal of the amber codon, does not result in a toxic translation product.
3. We used an approach where the TCORS is encoded by one plasmid, while reporter and wild-type pyrrolysyl-tRNA_{CUA} were cloned into a second plasmid. Thereby the tRNA is constitutively expressed, while expression of the reporter (i.e., protein of interest) is under control of an inducible promoter. (We used an arabinose-inducible PBAD promoter; this plasmid has tetracycline resistance).
4. Alternative expression systems, where both the pyrrolysyl-tRNA synthetase and the tRNA_{CUA} are cloned into the same plasmid, while the protein of interest with the introduced in-frame amber codon is expressed by a second plasmid, have been described. This option has the obvious advantage of higher flexibility, since the protein of interest resides on a separate plasmid; however, in our hands, it suffered from inferior yields compared to the option used here.
5. Be careful, TCOK and BCNK are not stable toward strong acids below pH < 5.
6. Tetrazines with a primary amino group as a handle were synthesized as in ref. 18. These tetrazines can be readily reacted with various commercially available amine-reactive fluorophores. We mainly coupled them to commercially available *N*-hydroxysuccinimidyl ester or isothiocyanate derivatives of fluorophores. An overview of synthesized and tested tetrazine–fluorophore conjugates can be found in refs. 18, 19.
7. In our experiments the reporter was cloned into the same construct as the TCORS.
8. A single copy of this tRNA may be insufficient to provide good reporter expression. We used a four-copy tRNA construct in which four copies of tRNA_{CUA} are expressed behind a U6 promoter with a CMV enhancer [20].
9. Before expressing a protein on larger scale, we suggest to optimize expression and amber suppression by altering TCOK concentrations (from 0.5 mM to 5 mM) as well as expression times and temperatures. Protein levels under different conditions were checked by Western blots with antibodies against the hexahistidine tag.

10. Depending on what the site-specifically modified protein is needed for, some further chromatography-based purification might be necessary.
11. We analyzed protein total mass using an Agilent 1200 LC-MS system; ESI-MS was carried out with a 6130 Quadrupole spectrometer. The solvent system consisted of 0.2 % formic acid in H₂O as buffer A and 0.2 % formic acid in acetonitrile as buffer B. LC-ESI-MS on proteins was carried out using a Phenomenex Jupiter C4 column (150×2 mm, 5 µm), and samples were analyzed in the positive mode, following protein UV absorbance at 214 and 280 nm. Total protein masses were calculated by deconvolution within the MS ChemStation software (Agilent Technologies). Additionally, protein total mass was determined on an LCT time-of-flight mass spectrometer with electrospray ionization (ESI, Micromass). Proteins were rebuffered in 20 mM of ammonium bicarbonate and mixed 1:1 with acetonitrile, containing 1 % formic acid. Alternatively samples were prepared with a C4 ZipTip (Millipore) and infused directly in 50 % aqueous acetonitrile containing 1 % formic acid. Samples were injected at 10 µL min⁻¹, and calibration was performed in positive ion mode using horse heart myoglobin. 30 scans were averaged and molecular masses obtained by maximum entropy deconvolution with MassLynx version 4.1 (Micromass). Theoretical masses of wild-type proteins were calculated using ProtParam (<http://web.expasy.org/protparam>), and theoretical masses for unnatural amino acid containing proteins were adjusted manually. To confirm the incorporation of TCOK at the right site, tryptic digestion followed by LC-MS/MS might be carried out.
12. Labeling reactions were carried out in buffer A, PBS, water, or LB medium. The reaction is compatible with most buffer and cell medium conditions. Using buffer systems with pH < 5 might lead to some degradation of TCOK.
13. Washing the cell pellet after protein expression is necessary to get rid of excess TCOK before the tetrazine–fluorophore conjugate is added to the reaction.
14. If the samples are not to be analyzed on an SDS-PAGE gel but are intended to be imaged directly by fluorescence microscopy, we recommend washing the cell pellet thoroughly with PBS buffer or LB medium containing up to 10 % DMSO to clear dye that attaches nonspecifically to bacteria.
15. In our *E. coli* proteome labeling experiments, we controlled the level of recombinant protein expression so that it was equal to or less than that of many endogenous proteins by modulating the concentration of TCOK added to cells. This ensures that any specific labeling of the target protein versus native

proteins was not an artifact of the abundance of the target protein. We normally also conduct one control experiment, where we incorporate a tetrazine-unreactive unnatural amino acid at the position of TCOK to show that the labeling is specific. Our standard control consists in incorporating Ne-*tert*-butyloxycarbonyl-L-lysine, which is a good substrate for the *Methanosarcina barkeri* wild-type pyrrolysyl-tRNA synthetase and does not react with tetrazine-fluorophore conjugates.

16. For incorporation of BCNK a 100 mM stock solution in DMSO (solution might be cloudy) is prepared and then diluted in cell culture medium to make up a 0.5 mM solution, which is also filtered through a 0.2 µm membrane sterile syringe filter before being added to cells. We were careful not to expose cells to a final concentration of more than 0.5 % DMSO. Cell culture medium containing 1 mM NorK is prepared in a similar fashion to TCOK.
17. Internal proteins were imaged by labeling cells with a cell-permeable diacetyl fluorescein dye coupled to a tetrazine (*see* ref. 19). After rinsing the dye out of the medium, we incubate cells in fresh cell culture medium three times for 30 min. These incubations are helpful in reducing background, but the requirements may vary depending on the dye used.

References

1. Fernandez-Suarez M, Ting AY (2008) Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Biol* 9:929–943
2. Zhang J, Campbell RE, Ting AY, Tsien RY (2002) Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Biol* 3:906–918
3. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3:373–382
4. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat Biotechnol* 21:86–89
5. Fernandez-Suarez M, Baruah H, Martinez-Hernandez L, Xie KT, Baskin JM, Bertozzi CR, Ting AY (2007) Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* 25:1483–1487
6. Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21:766–776
7. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc* 124:9026–9027
8. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG (2003) An expanded eukaryotic genetic code. *Science* 301:964–967
9. Zhang Z, Wang L, Brock A, Schultz PG (2002) The selective incorporation of alkenes into proteins in *Escherichia coli*. *Angew Chem Int Ed Engl* 41:2840–2842
10. Wang L, Zhang Z, Brock A, Schultz PG (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* 100:56–61
11. Fekner T, Li X, Lee MM, Chan MK (2009) A pyrrolysine analogue for protein click chemistry. *Angew Chem Int Ed Engl* 48:1633–1635
12. Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J Am Chem Soc* 131:8720–8721
13. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW (2011) Genetically encoded 1,2-Aminothiols facilitate rapid and site-specific

- protein labeling via a bio-orthogonal cyano-benzothiazole condensation. *J Am Chem Soc* 133:11418–11421
- 14. Wang Y, Song W, Hu WJ, Lin Q (2009) Fast alkene functionalization *in vivo* by Photoclick chemistry: HOMO lifting of nitrile imine dipoles. *Angew Chem Int Ed Engl* 48: 5330–5333
 - 15. Hancock SM, Uprety R, Deiters A, Chin JW (2010) Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J Am Chem Soc* 132:14819–14824
 - 16. Greiss S, Chin JW (2011) Expanding the genetic code of an animal. *J Am Chem Soc* 133:14196–14199
 - 17. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW (2012) Expanding the genetic code of *Drosophila melanogaster*. *Nat Chem Biol* 8:748–750
 - 18. Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* 4:298–304
 - 19. Lang K, Davis L, Wallace S, Mahesh M, Cox DJ, Blackman ML, Fox JM, Chin JW (2012) Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling *in vitro* and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *J Am Chem Soc* 134:10317–10320
 - 20. Gautier A, Nguyen DP, Lusic H, An W, Deiters A, Chin JW (2010) Genetically encoded photocontrol of protein localization in mammalian cells. *J Am Chem Soc* 132:4086–4088

Chapter 16

Labeling Proteins by Affinity-Guided DMAP Chemistry

Tomonori Tamura and Itaru Hamachi

Abstract

Catalysts have long played an essential role in organic synthesis and thus hold potential as tools for chemical protein modification. However, there are only a few examples of catalyst-mediated protein labeling under biological conditions because of the difficulty of designing molecular catalysts that work in aqueous environments with high target selectivity and reaction efficiency. To overcome this situation, we have previously developed a new catalyst-based method, termed affinity-guided DMAP (4-dimethylaminopyridine) (AGD) chemistry, for site-specific protein labeling in a target-selective manner using an acyl transfer reaction. More recently, we discovered that the labeling rate and efficiency can be greatly enhanced by using “multivalent” DMAP groups. Here, we describe the principle of the multivalent AGD chemistry and the protocol for chemical labeling of FK506-binding protein 12 (FKBP12) in test tubes. In this method, the FKBP12 labeling is completed within 30 min and occurs site specifically at the vicinity of the ligand-binding pocket of the protein.

Key words Protein labeling, FKBP12, DMAP catalyst, Acyl transfer reaction, AGD chemistry

1 Introduction

Site-specific chemical modification of proteins should be useful not only for investigating the biological roles of proteins but also for providing new analytical tools for detecting biochemical events, such as ligand binding or protein–protein interactions [1, 2]. To date, the most popular strategy for protein labeling relies on the use of peptide or enzyme tags [3, 4]. In this approach, target proteins are expressed as a fusion with a tag sequence in cells and then labeled using a chemical or enzymatic reaction with a designed probe. Although undoubtedly valuable, this tag-based approach is restricted to recombinant proteins. On the other hand, chemical strategies that can be used for site-specific native protein labeling are still limited.

Our group has been developing several organic chemistry-based methods for site-specific native protein labeling [5]. In particular, the previously developed affinity-guided DMAP (4-dimethylaminopyridine) (AGD) catalysts consisting of the DMAP group and a protein

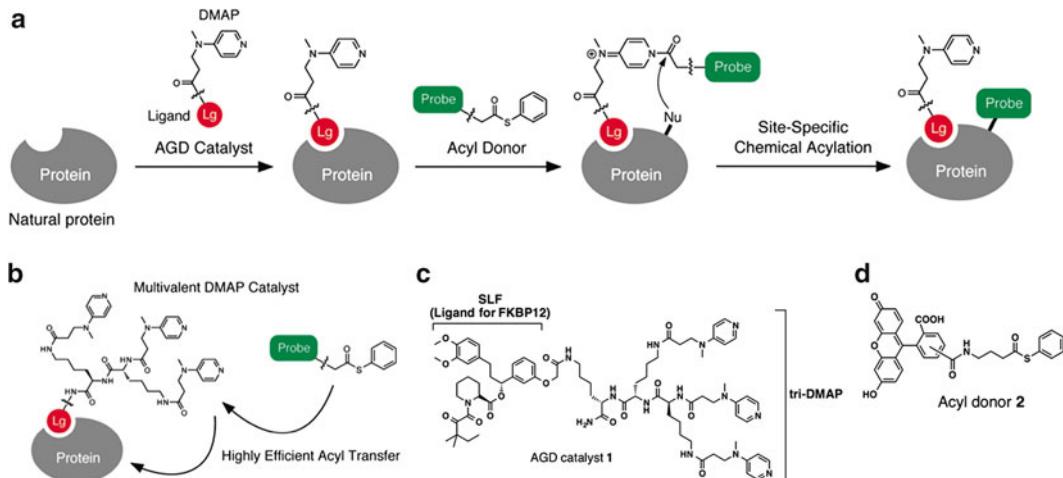


Fig. 1 AGD catalyst-mediated selective chemical protein labeling. Schematic illustrations of (a) the basic strategy and (b) the multivalent DMAP system. Lg, protein ligand; Nu, nucleophilic amino acid. (c, d) Molecular structure of (c) AGD catalyst **1** for FKBP12 labeling and (d) acyl donor **2** containing fluorescein

ligand allowed us to introduce diverse synthetic probes at the vicinity of the ligand-binding pocket of target proteins with high labeling efficiency and target selectivity [6–8]. The strategy of the AGD chemistry is shown in Fig. 1a. DMAP is a well-established acyl transfer catalyst, which can activate an acyl ester for transfer to a nucleophilic residue. The AGD catalyst selectively binds to the target protein, and the DMAP moiety mediates the transfer of the acyl group from the thiophenyl ester type of acyl donor to a nucleophilic amino acid located near the ligand-binding pocket. Importantly, in this system we can easily alter the target protein by switching the ligand moiety, while the acyl donor can be commonly used for various proteins. By using this strategy, we have previously demonstrated the selective chemical acylation of various native proteins *in vitro*, in a bacterial cell lysate, and in an animal tissue extract (Table 1). More recently, we discovered that the labeling rate and efficiency can be greatly enhanced by increasing the number of DMAP groups in the AGD catalysts with high site specificity (Fig. 1b) [7]. This improvement facilitates rapid and selective chemical labeling of bradykinin B₂ receptor (B₂R), G protein-coupled receptor, on the live cell surface (85 % yield within 30 min). Although AGD-mediated protein labeling does not currently work well inside cells, this approach is superior in terms of the labeling efficiency and the ease of synthesis of labeling reagents to other protein-labeling methods developed in our group, such as ligand-directed tosyl (LDT) chemistry [9–11] or acyl imidazole (LDAI) chemistry [12, 13].

Here, we describe the standard protocols of site-specific FKBP12 labeling using AGD chemistry *in vitro*. The synthesis of AGD catalyst **1** is outlined in Fig. 2. The Michael reaction of

Table 1
Current examples of proteins labeled by the AGD chemistry

Target protein	Ligand	Probe	Context	Labeled amino acid(s)	Ref.
Congerin II	Lactose	Fluorescein Coumarin Biotin ¹⁹ F probe	Test tube <i>E. coli</i> lysate Mucus tissue	Tyr	[6-8]
Concanavalin A	Mannose	Fluorescein Coumarin	Test tube	Not determined	[6]
Wheat germ agglutinin	N-Acetylglucosamine	Fluorescein Coumarin	Test tube	Not determined	[6]
SH2 domain	Phosphotyrosine peptide	Fluorescein	Test tube	Lys	[7]
FKBP12	SLF	Fluorescein	Test tube	Lys	[7]
B ₂ R	B ₂ R-selective antagonist	Fluorescein biotin	HEK293 cell surface	Not determined	[7]

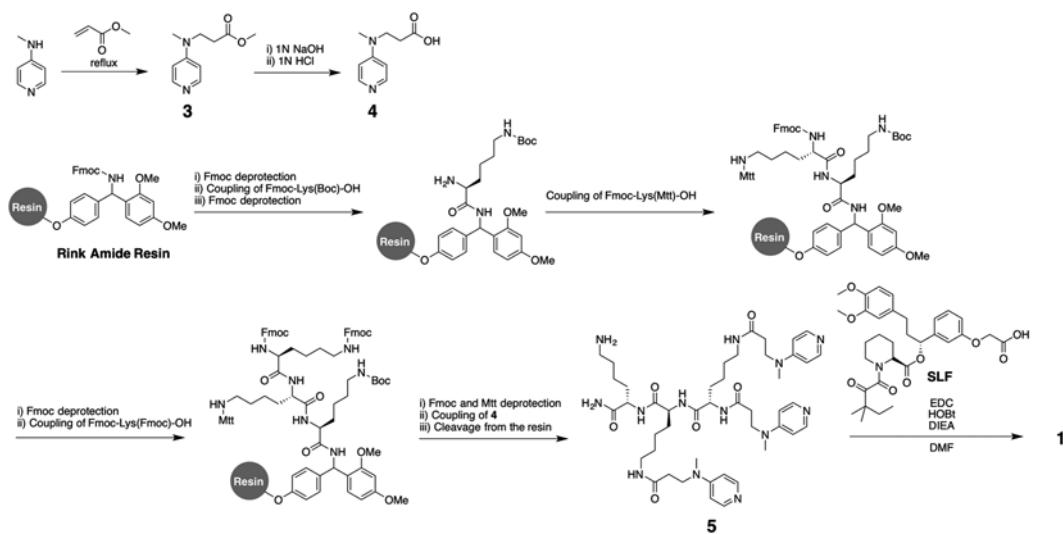


Fig. 2 Synthetic scheme of AGD catalyst 1

4-(methylamino)pyridine with methyl acrylate gives **3**, which is converted to **4** by alkaline hydrolysis. Tri-DMAP moiety **5** is prepared by solid phase synthesis using Fmoc-amino acids and **4**. The synthetic ligand of FKBP12 (SLF) [14] is conjugated with **5** to generate the AGD catalyst **1** for FKBP12 labeling. The synthesis of acyl donor **2** is outlined in Fig. 3. 5,6-Carboxyfluorescein is coupled with γ -aminobutyric acid *tert*-butyl ester. After deprotection of the *tert*-butyl ester, thiophenol is condensed to give acyl donor **2**.

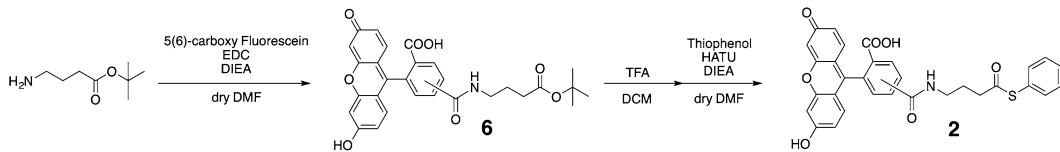


Fig. 3 Synthetic scheme of acyl donor **2**

Native FKBP12 labeling is carried out in vitro by incubating AGD catalyst **1** and acyl donor **2** with recombinant FKBP12. The labeling yields and the labeling site are evaluated by MALDI-TOF MS and MS/MS spectrometry.

2 Materials

Standard laboratory glassware and equipment are required for the organic syntheses, including, but not limited to, round-bottom flasks, syringes, chromatography columns, funnels, mechanical stirrers, and evaporators.

2.1 Synthesis of SLF-Tethered AGD Catalysts **1**

1. 4-(Methylamino)pyridine: 98 % purity.
2. Methyl acrylate: > 99.0 % purity.
3. Methanol: > 99.5 % purity.
4. Chloroform: > 99 % purity.
5. 1 N sodium hydroxide (NaOH).
6. 1 N hydrochloride (HCl).
7. Rink amide resin (Novabiochem).
8. Fmoc-Lys(Boc)-OH.
9. Fmoc-Lys(Mtt)-OH.
10. Fmoc-Lys(Fmoc)-OH.
11. *N*-Methylpyrrolidone (NMP): > 99.0 % purity.
12. Piperidine: > 99.0 % purity.
13. Trifluoroacetic acid (TFA): > 99.0 % purity.
14. Triisopropylsilane (TIS): > 98.0 % purity.
15. 1,2-Ethanedithiol: > 98.0 % purity.
16. 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium-hexafluorophosphate (HBTU).
17. 1-Hydroxybenzotriazole, monohydrate (HOBr): > 97.0 % purity.
18. *N,N*-Diisopropylethylamine (DIEA): > 98.0 % purity.
19. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC): > 98 % purity.
20. *N,N*-Dimethylformamide (DMF), anhydrous: > 99.8 % purity.
21. Synthetic ligand of FKBP (SLF) (*see* Ref.10).

22. Dichloromethane (DCM), anhydrous: >99.5 % purity.
23. Ammonia solution (28 % in water).
24. Thin-layer chromatography (TLC).
25. UV handheld lamp.
26. Silica gel: e.g., silica gel 60 N (spherical, neutral) 40–50 µm.
27. C18 reversed-phase (RP)-HPLC, e.g., Hitachi LaChrom L-7100 system equipped with a LaChrom L-7400 UV detector, and a YMC-Pack ODS-A column (5 µm, 250×10 mm).
28. Acetonitrile, HPLC grade: >99.93 % purity.

2.2 Synthesis of Acyl Donor 2

1. 5,6-Carboxyfluorescein.
2. γ -Aminobutyric acid: >99 % purity.
3. EDC.
4. HOBr.
5. DIEA.
6. DMF.
7. Ethyl acetate: >99 % purity.
8. Citric acid: >99.5 % purity.
9. Sodium chloride: >99.0 % purity.
10. Magnesium sulfate, anhydrous: 99.0 % purity.
11. TFA.
12. DCM.
13. Thiophenol: >98.0 % purity.
14. 2-(1*H*-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate methanaminium (HATU): >99 % purity.
15. Acetic acid: >98 % purity.
16. Thin-layer chromatography (TLC).
17. UV handy lamp.
18. Silica gel.

2.3 FKBP12 Labeling

1. Recombinant FKBP12 (*see* ref. 14).
2. SLF-tethered AGD catalysts **1** (prepared via synthesis).
3. Acyl donor **2** (prepared via synthesis).
4. FK506.
5. 50 mM Tris-HCl, pH 8.0.
6. ZipTip C4 resin.
7. Matrix for MALDI-TOF MS: 10 mg/mL of α -cyano-4-hydroxycinnamic acid (CHCA) in 50 % water, 50 % acetonitrile, 0.1 % TFA.
8. MALDI-TOF MS, e.g., Autoflex-III (Bruker).

2.4 Identification of the Labeling Site of FKBP12

1. Dialysis membrane (MWCO, 3,000).
2. Centrifugal filter device, e.g., Centricon Ultracel YM-3 (Millipore).
3. 50 mM HEPES, pH 8.0.
4. Urea, proteomics grade.
5. Lysyl endopeptidase (LEP), proteomics grade.
6. C18 reversed-phase (RP)-HPLC, e.g., Hitachi LaChrom L-7100 system equipped with a LaChrom L-7400 UV detector, and a YMC-Pack ODS-A column (5 µm, 250×10 mm).
7. Acetonitrile, HPLC grade: >99.93 % purity.
8. MALDI-TOF MS and MS/MS spectrometer, e.g., Autoflex-III (Bruker).

3 Methods

3.1 Synthesis of SLF-Tethered AGD Catalysts 1 (See Note 1)

3.1.1 Synthesis of Compound 3 (Fig. 4)

1. Charge a round-bottom flask with 4-(methylamino)pyridine (3.45 g, 31.9 mmol) and dissolve in 25 mL of methyl acrylate.
2. Reflux the mixture for 20 h.
3. Evaporate the excess methyl acrylate under vacuum.
4. Purify the residue by column chromatography on silica gel ($\text{CHCl}_3\text{-MeOH-28 \% NH}_3\text{ aq} = 200:10:1$).
5. Evaporate the collected fraction to afford compound **3** as a yellow oil (5.50 g, 28.3 mmol, 89 %).

^1H NMR (400 MHz, CDCl_3), δ/ppm 8.24 (d, 2H, $J=7.2$ Hz), 6.50 (d, 2H, $J=7.2$ Hz), 3.75 (t, 2H, $J=6.8$ Hz), 3.70 (s, 3H), 3.00 (s, 3H), 2.60 (t, 2H, $J=6.8$ Hz).

3.1.2 Synthesis of Compound 4 (Fig. 5)

1. Charge a round-bottom flask with compound **3** (5.5 g, 28.3 mmol) and dissolve in MeOH (40 mL).
2. Add 1 N NaOH (40 mL) and stir the mixture for 18 h at room temperature.
3. Neutralize with 1 N HCl solution on ice bath (see Note 2).
4. Lyophilize the solution.

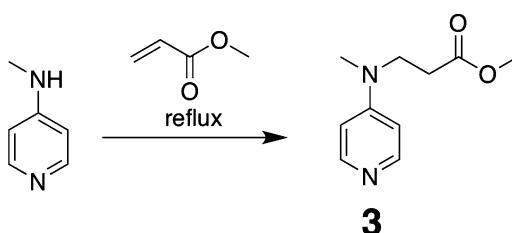


Fig. 4 Synthesis of compound **3**

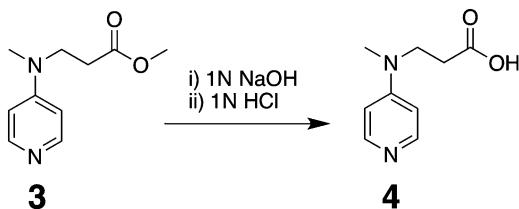


Fig. 5 Synthesis of compound 4

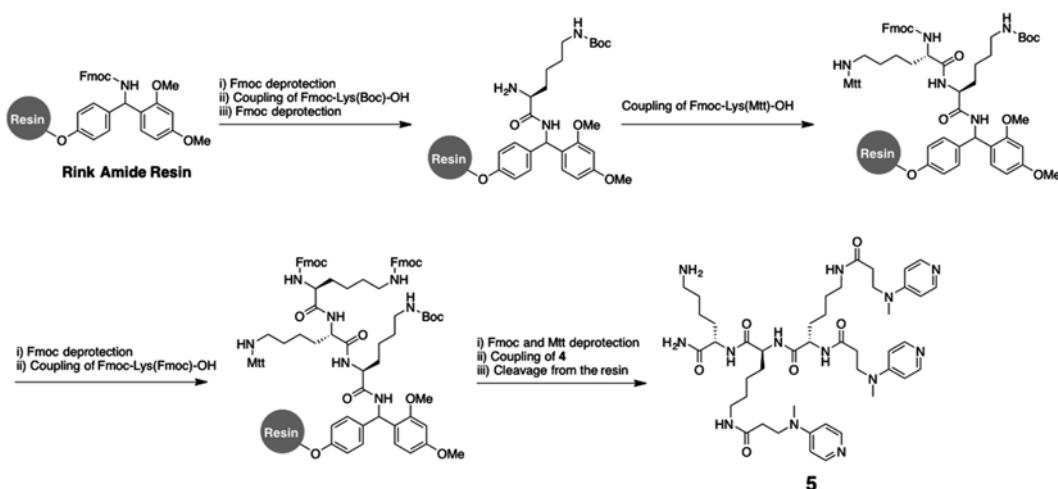


Fig. 6 Synthesis of compound 5

5. Add MeOH (30 mL) and CHCl₃ (10 mL) and remove the insoluble salts by filtration.
 6. Evaporate the filtrate to afford compound **4** as a white solid (5.0 g, 27.7 mmol, 98 %).

¹H NMR (400 MHz, CD₃OD), δ /ppm 8.07 (t, 2H, J =7.2 Hz), 7.00 (d, 2H, J =7.2 Hz), 3.84 (t, 2H, J =7.2 Hz), 3.20 (s, 3H), 2.49 (t, 2H, J =7.2 Hz).

3.1.3 Solid-Phase Synthesis of Compound 5 (Fig. 6)

Solid-phase synthesis is carried out using Rink amide resin. Fmoc-Lys(Boc)-OH, Fmoc-Lys(Mtt)-OH, and Fmoc-Lys(Fmoc)-OH are used as building blocks. All coupling and deprotection steps are monitored by the Kaiser test (*see* ref. 15).

1. Perform each coupling step with a mixture of carboxylic acid substrate (4 eq.), HBTU (4 eq.), HOEt (4 eq.), and DIEA (8 eq.) in NMP at room temperature.
 2. Carry out Fmoc deprotection with 20 % piperidine in N-methylpyrrolidone (NMP) for 20 min at room temperature.
 3. Carry out 4-methyltrityl (Mtt) deprotection as follows: Treat the resin with DCM containing 1 % TFA and 5 % TIS. The beads are stirred for 5 min, filtered, washed with DCM, and

immediately resuspended in DCM containing 1 % TFA and 5 % TIS. The procedure is repeated three times.

4. Following assembly, perform global deprotection and cleavage from the resin with TFA containing 1 % TIS, 2.5 % ethanedi-thiol, and 2.5 % H₂O.
5. Remove the solvent in vacuo.
6. Purify the residue by C18 reversed-phase (RP)-HPLC (CH₃CN/H₂O/0.1 % TFA) and lyophilize to afford compound **5**.

¹H NMR (400 MHz; CD₃OD), δ/ ppm 8.12 (d, 6H, *J*=6.0 Hz), 7.06 (d, 6H, *J*=63.2 Hz), 4.35–4.17 (m, 3H), 3.95–3.85 (m, 6H), 3.21 (s, 9H), 3.18–3.05 (m, 4H), 3.00–2.90 (m, 2H), 2.68–2.50 (m, 6H), 1.90–1.25 (m, 18H).

MALDI-TOF MS (CHCA), calcd for [M+H]⁺=888.5567, obsd 888.5589.

3.1.4 Synthesis of AGD Catalyst **1** (Fig. 7)

1. Charge a round-bottom flask with SLF (3 mg, 5 μmol) and dissolve in anhydrous DMF (0.5 mL) (*see Notes 3 and 4*).
2. Add compound **5** (3 mg, 3.4 μmol), HOEt (1.0 mg, 6.7 μmol), EDC (1.3 mg, 6.7 μmol), and DIEA (2.4 μL, 13.5 μmol).
3. Stir the reaction mixture at room temperature overnight.
4. Purify the crude solution by RP-HPLC (CH₃CN/H₂O/0.1 % TFA) and lyophilize to afford AGD catalyst **1** as a yellow solid (1.7 mg, 35 %).

¹H NMR (400 MHz; CD₃OD), δ/ ppm 8.10 (d, 6H, *J*=6.4 Hz), 7.30 (t, 1H, *J*=8.0 Hz), 7.00–6.93 (m, 9H), 6.84 (d, 1H, *J*=8.0 Hz), 6.77 (d, 1H, *J*=8.0 Hz), 6.70 (d, 1H, *J*=8.0 Hz), 5.74–5.71 (m, 1H), 5.30 (m, 1H), 4.52 (s, 2H), 4.29–4.23

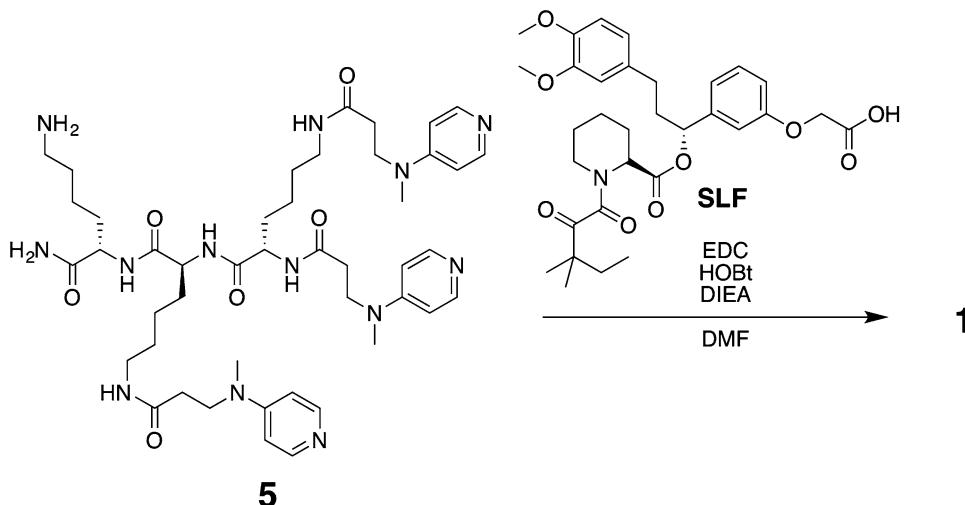


Fig. 7 Synthesis of compound **1**

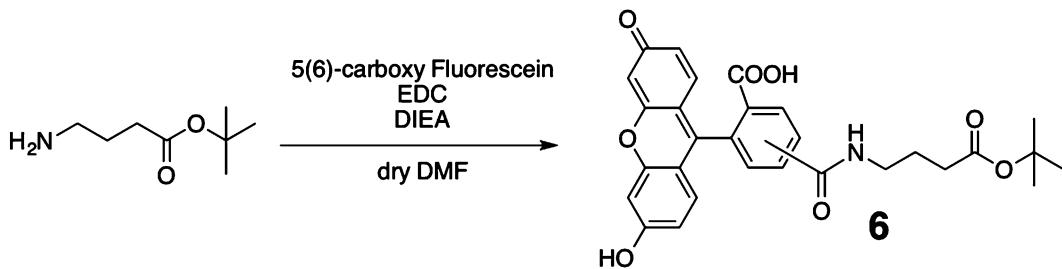


Fig. 8 Synthesis of compound 6

(m, 3H), 3.92–3.85 (m, 6H), 3.80 (s, 3H), 3.79 (s, 3H), 3.20 (s, 3H), 3.18 (s, 3H), 3.18 (s, 3H), 3.16–3.14 (m, 2H), 3.12–3.05 (m, 6H), 2.65–2.50 (m, 8H), 2.35 (d, 2H, $J=14.4$ Hz), 2.03 (m, 2H), 1.85–1.30 (m, 24H), 1.19 (s, 6H), 0.86 (t, 3H, $J=8.0$ Hz).

MALDI-TOF MS (CHCA), calcd for $[M+]^+=1,453.8242$, obsd 1,453.8224.

3.2 Synthesis of Acyl Donor 2 (See Note 1)

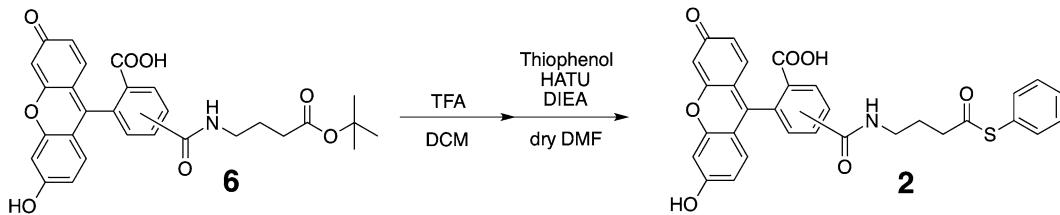
3.2.1 Synthesis of Compound 6 (Fig. 8)

1. Charge a round-bottom flask with 5,6-carboxyfluorescein (100 mg, 266 μ mol) and dissolve in DMF (5 mL).
2. Add EDC (78.6 mg, 399 μ mol), DIEA (153 μ L, 931 μ mol), and γ -aminobutyric acid *tert*-butyl ester (104 mg, 532 μ mol).
3. Stir the reaction mixture for 3 h at room temperature.
4. Evaporate the solvent and dissolve the residue in ethyl acetate (200 mL).
5. Wash the organic layer twice with 5 % aqueous citric acid (50 mL) and once with brine (50 mL).
6. Dry the organic layer over $MgSO_4$, filter, and evaporate.
7. Purify the residue by column chromatography on silica gel ($CHCl_3$ –MeOH– $CH_3COOH=200:10:1$) to afford compound 6 as a yellow solid (72 mg, 139 μ mol, 59 %). Compound 6 is obtained as a mixture of 5' and 6' isomers.

1H NMR (400 MHz, CD_3OD), δ /ppm 1.37 (s, 9H(6')), 1.44 (s, 9H(5')), 1.79 (quintet, $J=7.2$ Hz, 2H(6')), 1.91 (quintet, $J=7.2$ Hz, 2H(5')), 2.25 (t, $J=7.2$ Hz, 2H(6')), 2.35 (t, $J=7.2$ Hz, 2H(5')), 3.33 (t, $J=7.2$ Hz, 2H(6')), 3.46 (t, $J=7.2$ Hz, 2H(5')), 6.51–6.61 (m, 4H(5', 6')), 6.68–6.69 (m, 2H(5', 6')), 7.29 (d, $J=8.0$ Hz, 1H(5')), 7.60 (d, $J=2.0$ Hz, 1H(6')), 8.07 (d, $J=8.0$ Hz, 1H(6')), 8.12 (d, $J=8.0$ Hz, 1H(6')), 8.18 (d, $J=8.0$, 1H(5')), 8.42 (s, 1H (5')).

3.2.2 Synthesis of Acyl Donor 2 (Fig. 9)

1. Charge a round-bottom flask with compound 6 (72 mg, 139 μ mol) and dissolve in DCM (2 mL) and TFA (0.8 mL).
2. Stir for 1 h at room temperature.
3. Evaporate the solvent and dissolve the residue in DMF (2 mL).

**Fig. 9** Synthesis of compound **2**

4. Add HATU (52.8 mg, 139 μ mol), DIEA (121 μ L, 695 μ mol), and thiophenol (17 μ L, 167 μ mol).
5. Stir for 1 h at room temperature.
6. Evaporate the solvent and dissolve the residue in ethyl acetate (200 mL).
7. Wash the organic layer twice with 5 % aqueous citric acid (50 mL) and once with brine (50 mL).
8. Dry the organic layer over $MgSO_4$, filter, and evaporate.
9. Purify the residue by column chromatography on silica gel ($CHCl_3$ -MeOH- CH_3COOH = 300:10:1) to afford compound **2** as a yellow solid (25 mg, 45.9 μ mol, 33 %). Acyl donor **2** is obtained as a mixture of 5' and 6' isomers.

1H NMR (400 MHz, CD_3OD), δ /ppm 1.91 (quintet, J =7.2 Hz, 2H(6')), 2.06 (quintet, J =7.2 Hz, 2H(5')), 2.69 (t, J =7.2 Hz, 2H(6')), 2.81 (t, J =7.2 Hz, 2H(5')), 3.36 (t, J =7.2 Hz, 2H(6')), 3.49 (t, J =7.2 Hz, 2H(5')), 6.50–6.70 (m, 4H(5', 6')), 6.68–6.69 (m, 2H(5', 6')), 7.26–7.40 (m, 5H(5', 6')+1H(5')), 7.61 (d, J =2.0 Hz, 1H(6')), 8.06 (d, J =8.0 Hz, 1H(6')), 8.12 (d, J =8.0 Hz, 1H(6')), 8.18 (J =8.0 Hz, 1H(5')), 8.41 (s, 1H(5')).

HRMS (FAB) calcd for [M+H $^+$] ($C_{31}H_{24}NO_7S$), 554.1268; found, 554.1285.

3.3 General Protocol for FKBP12 Labeling by AGD Chemistry

Final concentrations of reaction components are the following: 10 μ M AGD catalyst **1**, 50 μ M acyl donor **2**, 100 μ M FK506, 10 μ M FKBP12.

1. Prepare a 1 mM stock solution of AGD catalyst **1** in DMSO (*see Note 5*).
2. Prepare a 5 mM stock solution of acyl donor **2** in anhydrous DMSO (*see Note 6*).
3. Add 1 μ L of AGD catalysts **1** (1 mM stock solution in DMSO) and 1 μ L of acyl donor **2** (5 mM stock solution in DMSO) to a solution of FKBP12 (12 μ g) in 50 mM Tris buffer (100 μ L, pH 8.0) (*see Notes 7 and 8*). As a negative control, FK506 (100 μ M) is added at the first step to the reaction mixture (*see Note 9*).

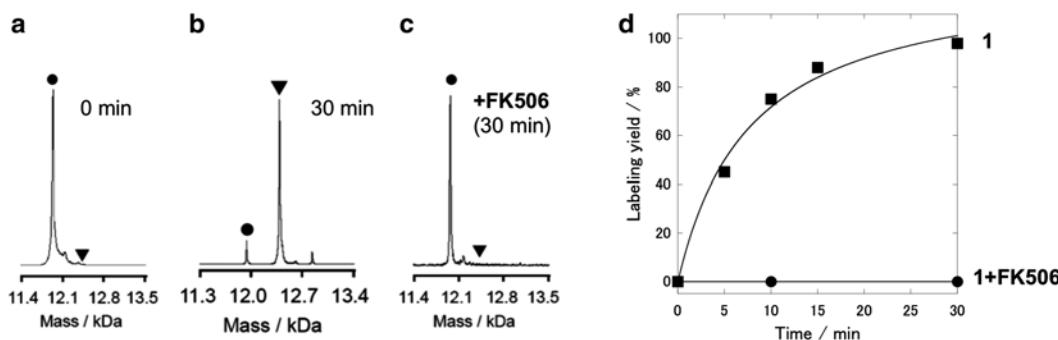


Fig. 10 (a–c) MALDI-TOF MS analysis of FKBP12 labeling using AGD catalyst **1** and acyl donor **2** in (a and b) the absence or (c) presence of FK506. After incubation at 37 °C for 0 or 30 min, the reaction mixtures are analyzed by MALDI-TOF MS (filled circle), unmodified FKBP12 (filled inverted triangle), fluorescein-labeled FKBP12. (d) Reaction kinetics of FKBP12 labeling using AGD catalyst **1** and acyl donor **2** in (filled square) the absence or (filled circle) presence of FK506

4. Mix the solution thoroughly by pipetting, and incubate at 37 °C (*see Note 10*).
5. Take the samples (5 µL) at different time points.
6. Purify and concentrate the protein with ZipTip C4 resin.
7. Elute the protein with MALDI-TOF MS matrix (CHCA) solution and spot on MALDI plate and dry for 10 min.
8. Determine the labeling yields with MALDI-TOF MS spectra (Fig. 10) (*see Note 11*).

3.4 General Protocol for Identification of the Labeling Site of FKBP12

1. Labeling is performed as described above under the following condition: 20 µM FKBP12, 20 µM AGD catalyst **1**, 40 µM fluorescein acyl donor in 50 mM Tris-HCl (pH 8.0) at 37 °C for 10 min.
2. Purify the labeled FKBP12 with RP-HPLC (a gradient of 5–55 % of solvent A (CH₃CN) over 100 min, UV detection at 220 nm, fluorescence detection at 515 nm (excitation at 480 nm)) (*see Note 12*).
3. Collect the fraction containing the labeled FKBP12 and dialyze against 50 mM HEPES (pH 8.0) with a Spectra/Por dialysis membrane (MWCO, 3,000).
4. Concentrate the resulting solution using a Centricon Ultracel YM-3 (Millipore) to afford 0.3 µM of the labeled FKBP12.
5. Add urea (at a final concentration of 2 M) and lysyl endopeptidase (LEP) (LEP/substrate ratio = 1/10 (w/w)) to this solution (*see Notes 13 and 14*).

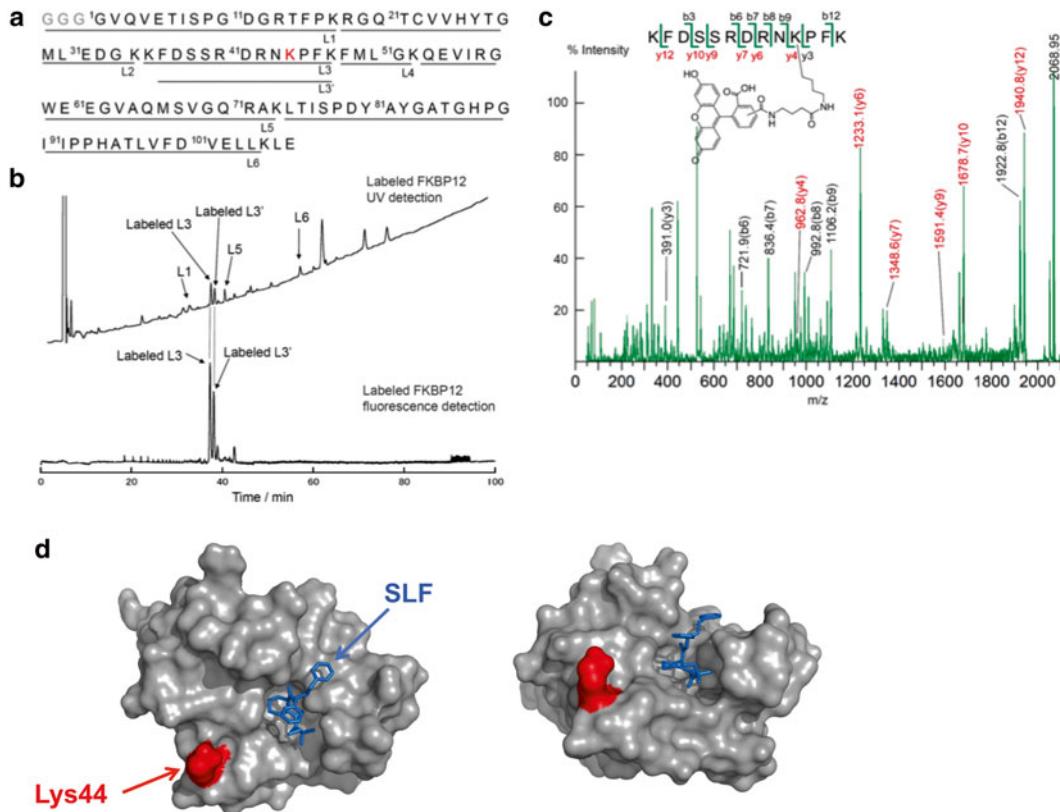


Fig. 11 Identification of the labeling site of FKBP12 treated with AGD catalyst **1** and acyl donor **2**. (a) The primary sequence of FKBP12 and the assignment of each fragment generated by lysyl endopeptidase (LEP) digestion. The extra amino acids shown in gray (GGG) are derived from the expression vector. (b) Reversed-phase HPLC analysis of LEP-digested, fluorescein-labeled FKBP12. A gradient of 5–55 % of CH_3CN over 100 min is used with UV detection at 220 nm (*top*) and fluorescence detection at 515 nm (excitation at 480 nm, *bottom*). The labeled L3 fragment is characterized by MALDI-TOF MS. MALDI-TOF MS (CHCA), calcd for $[\text{M}+\text{H}]^+=2,068.92$, obsd = 2,068.95. The second major peak could be assigned to labeled L3' fragment (see (a)). (c) MALDI-TOF MS/MS analysis of the fluorescein-labeled L3 fragment. (d) The crystal structure of the FKBP12–SLF complex (PDB ID: 1FKG)

6. After incubation at 37 °C for 30 h, separate the digested peptides with analytical RP-HPLC (a gradient of 5–55 % of solvent A (CH_3CN) over 100 min, UV detection at 220 nm, and fluorescence detection at 515 nm (excitation at 480 nm)) (Fig. 11b).
7. Collect the fluorescence fractions and analyze with MALDI-TOF MS (matrix: CHCA).
8. Carry out the further characterization by MALDI-TOF MS/MS analysis (Fig. 11c) (see Note 15).

4 Notes

1. AGD reagents are not commercialized at the moment. We will provide AGD catalysts, acyl donors, or those intermediates we have published upon request.
2. Be careful not to excessively add 1 N HCl. If the solution becomes acidic, the deprotected carboxylic acid is sensitive to esterization by MeOH.
3. Protein ligands containing primary or secondary amino groups or a phenolic hydroxyl group may not be used as a ligand for the AGD catalyst because they can react with the *N*-acyl-pyridinium intermediate in the labeling reaction.
4. In general, the affinity of the ligand moiety determines the specificity and the selectivity of labeling reaction. We usually use a ligand having at least a micromolar dissociation constant, but the required affinity is expected to depend on the concentration of the target protein in the context of interest.
5. The DMSO stock solution of AGD catalyst **1** is stable at least for 1 year at -30 °C. The concentration of AGD catalyst **1** is determined by measuring the absorbance at 280 nm using a molar absorption coefficient of 54,000 cm⁻¹ M⁻¹ in MeOH.
6. To prevent hydrolysis of acyl donor **2**, we prepared a stock solution using anhydrous DMSO in a vial and sealed it with parafilm after use. The DMSO stock solution of acyl donor **2** is stable at least for 1 year at -30 °C. The concentration of acyl donor **2** is determined by measuring the absorbance at 494 nm using a molar absorption coefficient of 68,000 cm⁻¹ M⁻¹ in pH 8.0 Tris buffer.
7. The labeling reaction is sensitive to pH. Our studies indicate that the optimal pH is generally 8.0.
8. DMSO contents in the reaction buffer should be kept below 5 % to avoid protein denaturation.
9. FK506 inhibits the labeling reaction by competitively binding to the same site that SLF ligand binds to.
10. The labeling reaction of FKBP12 with the AGD catalyst **1** is very rapid and generally completed within 30 min at 37 °C.
11. If the labeling yield is insufficient, we recommend optimizing the structure of the AGD catalyst by changing the length (and/or rigidity) of the linker between the ligand and the tri-DMAP moiety.
12. Size-exclusion chromatography is also available for purification of the labeled FKBP12.
13. A proteomic grade of protease should be used for peptide mapping.

14. If the protein digestion with LEP is insufficient, trypsin (trypsin/substrate ratio = 1/30 (w/w)) may be added to the solution.
15. Our previous studies have demonstrated so far that lysine and tyrosine residues can be modified by AGD chemistry.

Acknowledgement

We thank Dr. Yoichiro Koshi and Dr. Hangxiang Wang for their great contribution to the development and application of the AGD chemistry. T.T. acknowledges the JSPS Research Fellowships for Young Scientists.

References

1. Hellinga HW, Marvin JS (1998) Protein engineering and the development of generic biosensors. *Trends Biotechnol* 16:183–189
2. Nalbant P, Hodgson L, Kraynov V, Toutchkine A, Hahn KM (2004) Activation of endogenous Cdc42 visualized in living cells. *Science* 305: 1615–1619
3. Griffin BA, Adams SR, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281: 269–272
4. Kepler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* 21:86–89
5. Hayashi T, Hamachi I (2012) Traceless affinity labeling of endogenous proteins for functional analysis in living cells. *Acc Chem Res* 45: 1460–1469
6. Koshi Y, Nakata E, Miyagawa M, Tsukiji S, Ogawa T, Hamachi I (2008) Target-specific chemical acylation of lectins by ligand-tethered DMAP catalysts. *J Am Chem Soc* 130: 245–251
7. Wang H, Koshi Y, Minato D, Nonaka H, Kiyonaka S, Mori Y, Tsukiji S, Hamachi I (2011) Chemical cell-surface receptor engineering using affinity-guided, multivalent organocatalysts. *J Am Chem Soc* 133:12220–12228
8. Sun Y, Takaoka Y, Tsukiji S, Narazaki M, Matsuda T, Hamachi I (2011) Construction of a ¹⁹F-lectin biosensor for glycoprotein imaging by using affinity-guided DMAP chemistry. *Bioorg Med Chem Lett* 21:4393–4396
9. Tsukiji S, Miyagawa M, Takaoka Y, Tamura T, Hamachi I (2009) Ligand-directed tosyl chemistry for protein labeling in vivo. *Nat Chem Biol* 5:341–343
10. Tamura T, Tsukiji S, Hamachi I (2012) Native FKBP12 engineering by ligand-directed tosyl chemistry: labeling properties and application to photo-cross-linking of protein complexes in vitro and in living cells. *J Am Chem Soc* 134:2216–2226
11. Tamura T, Kioi Y, Miki T, Tsukiji S, Hamachi I (2013) Fluorophore labeling of native FKBP12 by ligand-directed tosyl chemistry allows detection of its molecular interactions in vitro and in living cells. *J Am Chem Soc* 135:6782–6785
12. Fujishima S, Yasui R, Miki T, Ojida A, Hamachi I (2012) Ligand-directed acyl imidazole chemistry for labeling of membrane-bound proteins on live cells. *J Am Chem Soc* 134:3961–3964
13. Matsuo K, Kioi Y, Yasui R, Takaoka Y, Miki T, Fujishima S, Hamachi I (2013) One-step construction of caged carbonic anhydrase I using a ligand-directed acyl imidazole-based protein labeling method. *Chem Sci* 4:2573–2580
14. Clackson T, Yang W, Rozamus LW, Hatada M, Amara JF, Rollins CT, Stevenson LF, Magari SR, Wood SA, Courage NL, Lu X, Cerasoli F, Gilman M, Holt DA (1998) Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc Natl Acad Sci U S A* 95:10437–10442
15. Kaiser E, Colescott RL, Bossinger CD, Cook PI (1970) Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem* 34:595–598

Chapter 17

Ligand-Directed Tosyl Chemistry for Selective Native Protein Labeling In Vitro, In Cells, and In Vivo

Shinya Tsukiji and Itaru Hamachi

Abstract

Introducing nongenetically encoded, synthetic probes into specific proteins is now recognized as a key component in chemical biology. In particular, the ability to chemically modify specific “native” proteins in various contexts from *in vitro* to cellular systems is of fundamental importance to study biological systems. We developed a protein-labeling technique termed ligand-directed tosyl (LDT) chemistry for this purpose. This method is capable of labeling specific native proteins with diverse synthetic probes with high site specificity and target selectivity without compromising protein function. Here we describe the principle of the LDT chemistry and the protocol for selective chemical labeling of native carbonic anhydrase *in vitro*, in blood cells (*ex vivo*), and in living mice (*in vivo*).

Key words Chemical protein modification, Affinity labeling, Tosyl chemistry, Carbonic anhydrase

1 Introduction

The chemical modification (labeling) of proteins is a powerful approach in chemical biology. This technique allows diverse synthetic molecules, such as fluorescent dyes, affinity labels, or NMR/MRI probes, to be introduced to proteins. The resulting chemically modified proteins serve as useful tools for investigating biological processes. During the past few decades, various protein modification methods have been reported [1–3]. However, most of them are applicable only to purified proteins in test tubes. Since there has been a growing need to study the behavior and function of proteins in their native environment, the development of new techniques for the chemical labeling of target proteins in the context of living cells or even in living animal bodies is now strongly demanded [4–8]. The most popular approach is the use of peptide or enzyme tags [9–16]. In this method, target proteins are expressed as a fusion with a tag sequence in cells, and then the tag is labeled by a chemical or enzymatic reaction with a designed probe. Although undoubtedly valuable, this tag-based technique is

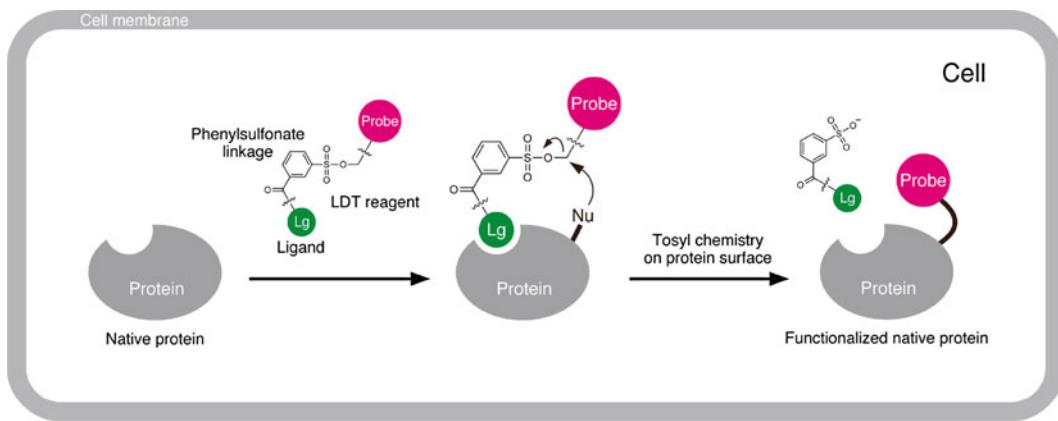


Fig. 1 Schematic of the principle of LDT chemistry. The LDT-based protein labeling method allows the attachment of a probe of interest to a specific target endogenous protein in cells. *Lg* protein ligand, *Nu* nucleophilic amino acid

restricted to recombinant proteins. The ability to chemically modify nontagged “endogenous” (native) proteins of interest in living systems should greatly facilitate the functional analysis of proteins in their physiological context.

To address this challenge, we recently developed a novel protein-labeling methodology termed ligand-directed tosyl (LDT) chemistry (Fig. 1) [17]. This chemistry is based on the principle of affinity labeling [18, 19] and involves the use of labeling reagents in which a protein ligand and a synthetic probe of interest are connected by an electrophilic phenyl sulfonate (tosylate) ester group. As shown in Fig. 1, the LDT reagent selectively binds to the target protein through the specific protein-ligand interaction, covalently transferring the synthetic probe to a nucleophilic amino acid residue near the ligand-binding pocket. Because the ligand moiety is cleaved off concomitantly with the labeling process, the function of the labeled protein is preserved. This method provides a powerful way to introduce diverse synthetic probes to target proteins in a traceless manner with high site specificity and target selectivity. The utmost appeal of this chemistry is its applicability to native proteins in living systems. Indeed, this technique has so far been successfully employed by us [17, 20–23] and others [24–26] to label various native proteins *in vitro*, in various types of mammalian cells, and even *in vivo*. The previous applications of the LDT method are summarized in Table 1.

As representative examples, here we show the application of the LDT chemistry for selective chemical modification of native carbonic anhydrase (CA) in various contexts. The CA-targeted LDT reagents are shown in Fig. 2 together with their control compounds. Using LDT reagent 1, we were able to introduce a

Table 1
Current examples of proteins labeled by the LDT chemistry

Target protein	Ligand	Probe	Context	Labeled amino acid(s)	Ref.
Carbonic anhydrase	Benzenesulfonamide	Coumarin Biotin ¹⁹ F probe	Test tube Red blood cells In vivo (mouse)	His	[17, 20, 21]
FKBP12	SLF	Biotin Diazirine Oregon Green	Test tube Culture cells (Jurkat, HeLa, A549)	His, Tyr, Glu	[22, 23]
Congerin II	Lactose	Coumarin	Mucus tissue	Not determined	[17]
SH2 domain	Phosphotyrosine peptide	Coumarin	Test tube <i>E. coli</i> lysate	His	[20]
14-3-3	Fusicoccin analogue (ISIR-042)	Dansyl BODIPY	Test tube <i>E. coli</i> lysate Culture cells (U937)	His	[24]
Hsc70	Apotozole	Coumarin	Test tube	His, Asp	[25]
<i>E. coli</i> DHFR mutant	Trimethoprim	Atto520	Test tube Culture cells (HEK293T, MEF)	Cys	[26]

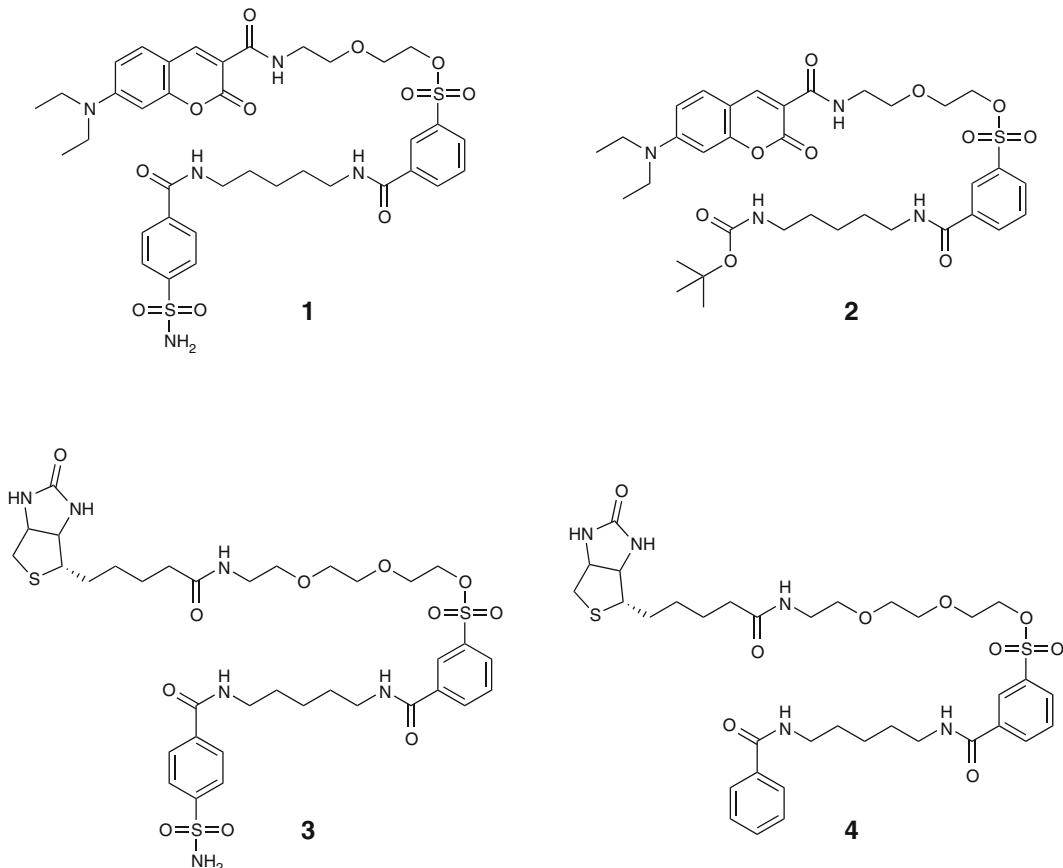


Fig. 2 Structures of LDT reagents **1** and **3** for carbonic anhydrase and their control compounds **2** and **4**. Reagent **1** contains the benzenesulfonamide ligand and the 7-diethylaminocoumarin (DEAC) probe. Reagent **3** contains the same ligand and the biotin tag

diethylaminocoumarin (DEAC) fluorophore to native (purified) CA in vitro (Fig. 3) [17]. When LDT reagent **3** was used, endogenously expressed CA in blood cells (ex vivo) and in living mice (in vivo) were selectively modified with a biotin affinity tag (Fig. 4). In the following, we provide protocols for the chemical synthesis of these LDT reagents and their use for CA labeling. In addition, general guidelines on the design and synthesis of LDT reagents are included in Note 1. For details on labeling experiments, please refer to the figure legends.

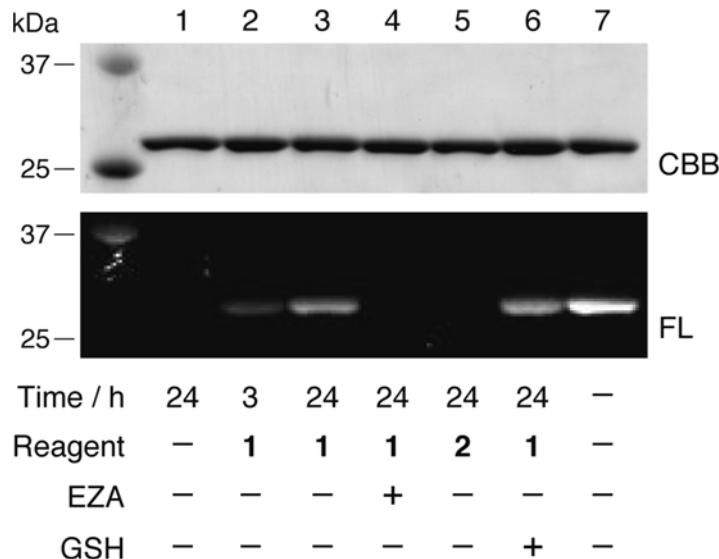


Fig. 3 Carbonic anhydrase II (CAlI) labeling in vitro. Purified CAlI (40 µM) was incubated with **1** or **2** (80 µM) in buffer (pH 7.2) with or without either ethoxzolamide (EZA) (400 µM) or reduced glutathione (GSH) (10 mM) at 37 °C. In lane 7, a 1:1 conjugate of CAlI and DEAC dye was used as a standard marker to determine the labeling yields. Samples were subjected to SDS-PAGE and analyzed by in-gel fluorescence imaging (FL) and Coomassie Brilliant Blue (CBB) staining

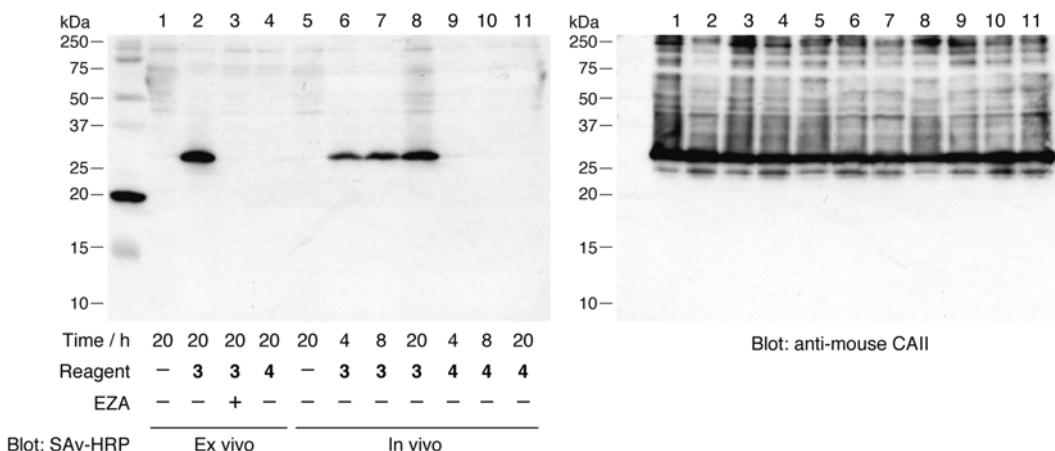


Fig. 4 Carbonic anhydrase-selective labeling ex vivo and in vivo. For ex vivo labeling, blood withdrawn from a mouse was incubated with reagent **3** or **4** (5 µM) in buffered saline (pH 7.4) with or without EZA (500 µM) at 35 °C for 20 h. For in vivo labeling, mice were intravenously injected with **3** or **4** (100 µM in 0.5 mL of buffered saline). Blood was taken from the tail vein and analyzed by immunoblot using streptavidin-horseradish peroxidase conjugate (SAv-HRP) (*left*) and anti-mouse CAlI antibody (*right*)

2 Materials

2.1 Synthesis of LDT Reagent 1

Standard laboratory glassware and equipments are required for the organic syntheses, including, but not limited to round-bottom flasks, syringes, chromatography columns, funnels, mechanical stirrers, and evaporators.

2.1.1 Reactants

1. 7-Diethylaminocoumarin-3-carboxylic acid: We have synthesized this compound according to a previous report [27].
2. 1-Hydroxybenzotriazole monohydrate (HOBr): >97 % purity.
3. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC): >98 % purity.
4. 2-(2-Aminoethoxy)ethanol: >98 % purity.
5. Magnesium sulfate anhydrous (MgSO_4): >98 % purity.
6. 3-(Chlorosulfonyl)benzoyl chloride: 98 % purity.
7. *N*-(*tert*-Butoxycarbonyl)-1,5-diaminopentane: >98 % purity.
8. *N,N*-Diisopropylethylamine (DIEA): >98 % purity.
9. 4-Dimethylaminopyridine (DMAP): >99 % purity.
10. Trifluoroacetic acid (TFA): >99 % purity.
11. 4-Sulfamoylbenzoic acid: >95 % purity.
12. *N*-Hydroxysuccinimide: >98 % purity.
13. Biotin: >98 % purity.
14. 2-(2-(2-Azidoethoxy)ethoxy)ethanol: This compound is not commercially available and thus needs to be synthesized according to a previous report [28].
15. Palladium carbon (Pd-C): 10 % on carbon.
16. Hydrogen gas: >99.9 % purity.

2.1.2 Solvents and Solutions

1. *N,N*-Dimethylformamide (DMF): anhydrous, >99.5 % purity.
2. Chloroform (CHCl_3): >99 % purity.
3. Saturated sodium bicarbonate (NaHCO_3) solution.
4. Diethyl ether (Et_2O): >99.5 % purity.
5. Dichloromethane (CH_2Cl_2): anhydrous, >99.5 % purity.
6. Methanol (MeOH) (for column chromatography): >99.5 % purity.
7. Toluene: >99 % purity.
8. Ethyl acetate (AcOEt): >99 % purity.
9. Brine (saturated sodium chloride solution).
10. Ethanol (EtOH): >99.5 % purity.
11. Acetic acid (AcOH): >99 % purity.
12. Methanol (for reaction): anhydrous, >99.5 % purity.

2.1.3 Devices

1. Thin layer chromatography (TLC): e.g., TLC silica gel 60 F₂₅₄ aluminum sheets (Merck).
2. UV handy handheld lamp.
3. Ninhydrin spray (0.5 % in 1-Butanol) for TLC staining.
4. Silica gel: e.g., silica gel 60 N (spherical, neutral) 40–50 µm (Merck).
5. Balloon for hydrogen gas: Use an appropriate connector to attach H₂-filled balloon to a reaction flask.

2.2 Chemical Labeling Experiments

2.2.1 In Vitro Labeling

Prepare all solutions using ultrapure water (e.g., Milli-Q water) and analytical grade reagents. Standard biological equipments are required, including, but not limited to, automatic pipettes, pipette tips, plastic tubes, and glass bottles.

1. LDT reagent **1**: For synthesis, *see* Subheading [3.1](#).
2. Dimethyl sulfoxide (DMSO): >99 % purity.
3. Purified human carbonic anhydrase isozyme II (Sigma, C6165): >80 % purity.
4. HEPES buffer: 50 mM HEPES, pH 7.2.

2.2.2 In Cells (Ex Vivo) and In Vivo Labeling

1. LDT reagent **3**: For synthesis, *see* Subheading [3.2](#).
2. Dimethyl sulfoxide (DMSO): >99 % purity.
3. Experimental mice: e.g., Slc:ICR mice, specific pathogen-free grade, male, ca. 8-week-old.
4. Polypropylene container.
5. Diethyl ether (Et₂O): >98 % purity.
6. General mouse surgical instruments: surgical table, dissecting scissor, forceps, etc.
7. Disposable sterile 1 mL syringes.
8. Disposable sterile needles: 27 G.
9. Ethylenediaminetetraacetic acid (EDTA) solution: 0.5 M, pH 8.0.
10. Phosphate-buffered saline (PBS): 1.5 mM KH₂PO₄, 8.1 mM NaHPO₄, 137 mM NaCl, 2.7 mM KCl, pH 7.4.
11. NP-40^{EZA} lysis buffer: 50 mM Tris-HCl, 150 mM NaCl, 2.5 mM EDTA, 1 % Nonidet P-40, 0.5 mM ethoxzolamide (EZA), pH 7.4.
12. Mouse holder.

2.2.3 SDS-PAGE and Band Detection

1. 2× Laemmli buffer: 125 mM Tris-HCl, 4 % SDS, 20 % glycerol, 10 % 2-mercaptoethanol, 0.004 % bromophenol blue, pH 6.8.
2. Standard SDS-polyacrylamide gel electrophoresis (SDS-PAGE) apparatus: e.g., Mini-PROTEAN Tetra cell system (Bio-Rad).

3. Standard Western blotting apparatus: e.g., Mini Trans-Blot cell system (Bio-Rad).
4. 12.5 and 15 % polyacrylamide gel.
5. SDS-PAGE running buffer: 25 mM Tris-HCl, 192 mM glycine, 0.1 % SDS, pH 8.1–8.5.
6. Fluorescence imager: e.g., ChemiDoc XRS system (Bio-Rad) equipped with UV light illuminator and 480BP70 filter.
7. PVDF or nitrocellulose membranes: e.g., Immun-Blot PVDF membrane (Bio-Rad).
8. Western blot transfer buffer: 25 mM Tris-HCl, 192 mM glycine, pH 8.1–8.5.
9. Tris-buffered saline (TBS): 10 mM Tris-HCl, 150 mM NaCl, pH 7.4.
10. TBS containing 0.05 % Tween-20 (TBST).
11. Blocking solution: 5 % skim milk in TBST.
12. Streptavidin-horseradish peroxidase conjugate (SAv-HRP) (Invitrogen).
13. Anti-mouse carbonic anhydrase II (CAII) antibody and anti-goat IgG-HRP conjugate (Santa Cruz Biotechnology).
14. Chemiluminescence reagent: e.g., Chemi-Lumi One (Nacalai Tesque).
15. Chemiluminescence imager: e.g., ImageQuant LAS 4000 (GE Healthcare).

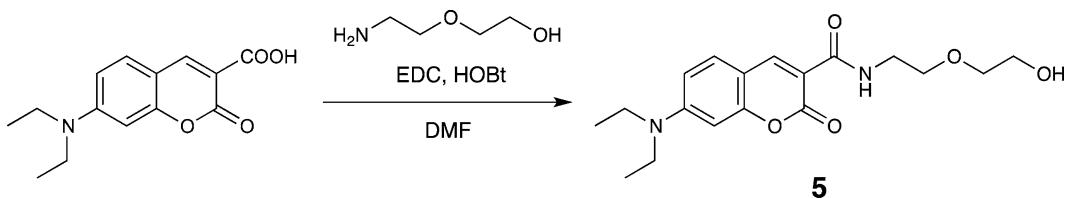
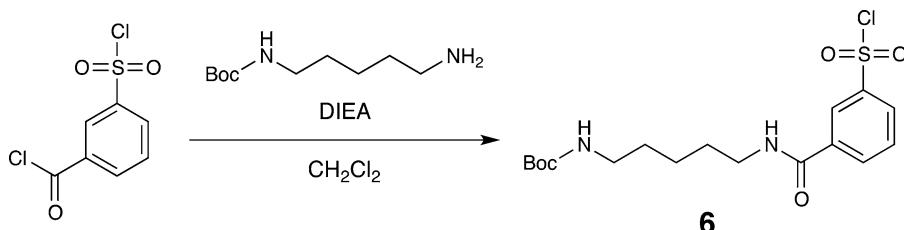
3 Methods

Unless otherwise specified, all reactions are performed at room temperature under argon atmosphere. For the synthesis of control compound **4**, *see ref. 17*.

3.1 Synthesis of LDT Reagent 1

3.1.1 Synthesis of Compound 5 (Fig. 5)

1. Charge a round-bottom flask with 7-diethylaminocoumarin-3-carboxylic acid (2.0 g, 7.65 mmol) and dissolve in anhydrous DMF (15 mL).
2. Add HOBr (1.50 g, 9.79 mmol) and EDC (1.94 g, 10.12 mmol).
3. Stir the reaction mixture for 5 min.
4. Add 2-(2-aminoethoxy)ethanol (990 µL, 9.98 mmol).
5. Stir the reaction mixture overnight (*see Note 2*).
6. Evaporate the solvent and dissolve the residue in CHCl₃ (30 mL).
7. Wash the organic layer three times with saturated NaHCO₃ solution (20 mL).

**Fig. 5** Synthesis of compound 5**Fig. 6** Synthesis of compound 6

8. Dry the organic layer over MgSO₄, filter, and evaporate.
9. Triturate the residue with Et₂O.
10. Collect the solid by filtration and dry it in a vacuum desiccator to afford compound 5 as a yellow solid (2.35 g, 88 %). ¹H NMR (500 MHz; CDCl₃): δ 9.11 (br s, 1H), 8.70 (s, 1H), 7.43 (d, 1H, *J*=10.0 Hz), 6.65 (dd, 1H, *J*=10.0, 5.0 Hz), 6.50 (d, 1H, *J*=5.0 Hz), 3.77 (m, 2H), 3.69–3.63 (m, 6H), 3.46 (q, 4H, *J*=5.0 Hz), 1.24 (t, 6H, *J*=5.0 Hz).

3.1.2 Synthesis of Compound 6 (Fig. 6)

1. Charge a round-bottom flask with 3-(chlorosulfonyl)benzoyl chloride (3.0 mL, 18.82 mmol) and dissolve in anhydrous CH₂Cl₂ (10 mL).
2. Add dropwisely a solution of *N*-(*tert*-butoxycarbonyl)-1,5-diaminopentane (1.88 g, 9.29 mmol) and DIEA (3.93 mL, 23.11 mmol) in anhydrous CH₂Cl₂ (8 mL) at 4 °C.
3. Stir the reaction mixture at 4 °C for 15 min and then at room temperature for 1 h (*see Note 2*).
4. Evaporate the solvent.
5. Purify the residue by column chromatography on silica gel (CHCl₃).
6. Evaporate the collected fraction to afford compound 6 as a pale yellow sticky solid (3.11 g, 83 %). ¹H NMR (500 MHz; CDCl₃): δ 8.43 (s, 1H), 8.22 (d, 1H, *J*=7.9 Hz), 8.15 (d, 1H, *J*=8.2 Hz), 7.71 (dd, 1H, *J*=7.9, 7.9 Hz), 6.83 (br s, 1H), 4.66 (br s, 1H), 3.49 (m, 2H), 3.13 (m, 2H), 1.68 (m, 2H), 1.54 (m, 2H), 1.40 (m+s, 2+9H).

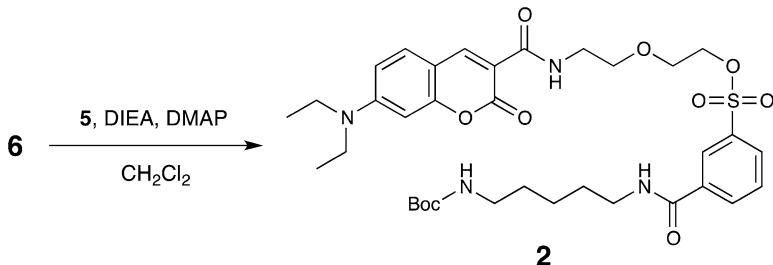


Fig. 7 Synthesis of compound 2

3.1.3 Synthesis of Compound 2 (Fig. 7)

1. Charge a round-bottom flask with compound **6** (3.0 g, 7.41 mmol) and dissolve in anhydrous CH_2Cl_2 (10 mL).
2. Add compound **5** (1.29 g, 3.70 mmol), DIEA (3.13 mL, 18.41 mmol), and DMAP (45 mg, 0.37 mmol).
3. Stir the reaction mixture for 6 h (*see Note 2*).
4. Evaporate the solvent (*see Note 3*).
5. Purify the residue by column chromatography on silica gel ($\text{CHCl}_3/\text{MeOH}$, 100:1).
6. Evaporate the collected fraction to afford compound **2** as a yellow solid (1.60 g, 60 %) (*see Note 4*). ^1H NMR (500 MHz; CDCl_3): δ 8.91 (br s, 1H), 8.68 (s, 1H), 8.40 (s, 1H), 8.19 (d, 1H, $J=7.9$ Hz), 8.04 (d, 1H, $J=7.6$ Hz), 7.62 (dd, 1H, $J=7.9$, 7.9 Hz), 7.44 (d, 1H, $J=9.2$ Hz), 6.66 (dd, 1H, $J=8.9$, 2.5 Hz), 6.48 (d, 1H, $J=2.1$ Hz), 4.59 (br s, 1H), 4.23 (m, 2H), 3.68 (m, 2H), 3.55–3.44 (m, 10 H), 3.09 (m, 2H), 1.66 (m, 4H), 1.51 (m, 2H), 1.41 (s, 9H), 1.25 (t, 6H, $J=7.2$ Hz). ^{13}C NMR (100 MHz; CDCl_3): δ 165.68, 163.44, 162.92, 157.73, 156.11, 152.75, 148.82, 148.33, 136.40, 136.24, 133.26, 131.30, 130.49, 129.51, 126.11, 110.17, 110.08, 108.44, 96.55, 69.95, 69.72, 68.37, 45.17, 40.39, 40.17, 39.26, 29.80, 29.29, 28.47, 24.18, 12.51. HR-FAB MS (3-nitrobenzyl alcohol): calculated for $\text{C}_{35}\text{H}_{48}\text{N}_4\text{O}_{10}\text{S} [\text{M}]^+$ = 716.3091; found 716.3083.

3.1.4 Synthesis of LDT Reagent 1 (Fig. 8)

1. Charge a round-bottom flask with compound **2** (205 mg, 0.29 mmol) and dissolve in anhydrous CH_2Cl_2 (8 mL).
2. Add TFA (1.6 mL).
3. Stir the reaction mixture for 30 min (*see Note 2*).
4. Add toluene (8 mL) and evaporate the solution. Repeat this once again.
5. Dissolve the residue in anhydrous DMF (2 mL) to prepare a solution of deprotected **2**.
6. Charge another round-bottom flask with 4-sulfamoylbenzoic acid (86 mg, 0.43 mmol) and dissolve in anhydrous DMF (2 mL).
7. Add DIEA (241 μL , 1.42 mmol) and EDC (108 mg, 0.56 mmol).

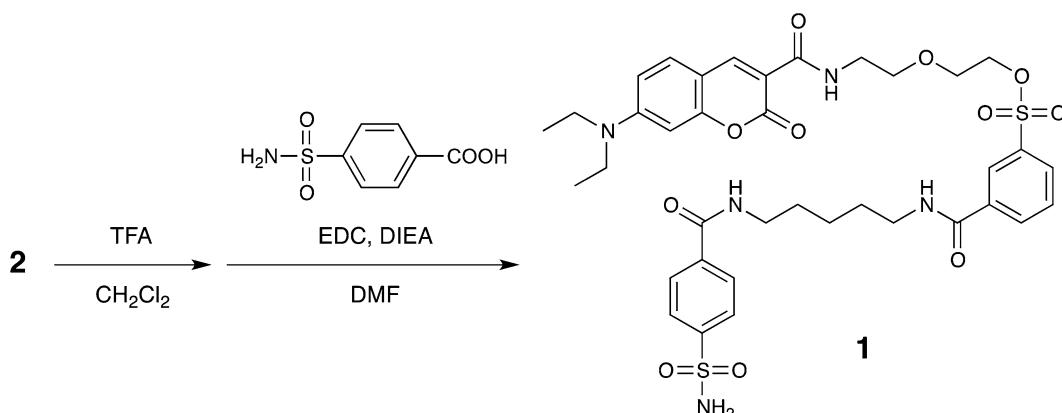


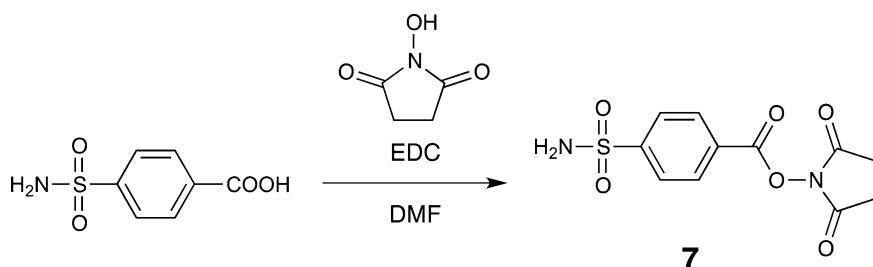
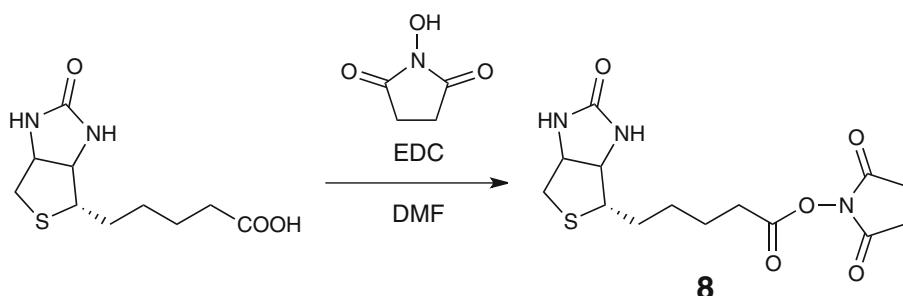
Fig. 8 Synthesis of LDT reagent **1**

8. Stir the reaction mixture for 5 min.
9. Add the solution of deprotected **2** prepared above.
10. Stir the reaction mixture for 3 h (*see Note 2*).
11. Evaporate the solvent and dissolve the residue in AcOEt (10 mL).
12. Wash the organic layer twice with saturated NaHCO₃ solution (5 mL) and once with brine (5 mL).
13. Dry the organic layer over MgSO₄, filter, and evaporate.
14. Purify the residue by column chromatography on silica gel (CHCl₃/MeOH, 25:1).
15. Evaporate the collected fraction and triturate the residue with Et₂O.
16. Collect the solid by filtration and dry it in a vacuum desiccator to afford LDT reagent **1** as a yellow solid (60 mg, 26 %) (*see Note 4*). ¹H NMR (500 MHz; CDCl₃): δ 8.87 (br t, 1H, *J*=5.3 Hz), 8.59 (s, 1H), 8.34 (s, 1H), 8.05 (d, 1H, *J*=7.9 Hz), 7.97 (d, 1H, *J*=8.0 Hz), 7.75 (s, 4H), 7.54 (dd, 1H, *J*=7.9, 7.9 Hz), 7.39 (d, 1H, *J*=9.2 Hz), 6.64 (dd, 1H, *J*=9.2, 2.5 Hz), 6.45 (d, 1H, *J*=2.5 Hz), 5.99 (s, 2H), 4.19 (m, 2H), 3.63 (m, 2H), 3.49–3.39 (m, 12H), 1.65 (m, 4H), 1.42 (m, 2H), 1.23 (t, 6H, *J*=7.2 Hz). ¹³C NMR (100 MHz; CDCl₃): δ 166.67, 166.19, 163.63, 162.93, 157.72, 152.91, 148.39, 144.84, 138.24, 136.30, 136.06, 132.98, 131.41, 130.51, 129.64, 127.88, 126.36, 110.31, 109.59, 108.38, 96.52, 70.07, 69.81, 68.35, 45.21, 39.94, 39.72, 39.30, 29.03, 28.46, 22.95, 12.52. HR-FAB MS (3-nitrobenzyl alcohol): calculated for C₃₇H₄₅N₅O₁₁S₂ [M]⁺=799.2557; found 799.2559.

3.2 Synthesis of LDT Reagent 3

3.2.1 Synthesis of Compound 7 (Fig. 9)

1. Charge a round-bottom flask with 4-sulfamoylbenzoic acid (500 mg, 2.49 mmol) and dissolve in anhydrous DMF (20 mL).
2. Add DIEA (1.3 mL, 7.64 mmol) and EDC (720 mg, 3.76 mmol).

**Fig. 9** Synthesis of compound 7**Fig. 10** Synthesis of compound 8

3. Stir the reaction mixture for 5 min.
4. Add *N*-hydroxysuccinimide (370 mg, 3.21 mmol).
5. Stir the reaction mixture for 4 h (*see Note 2*).
6. Evaporate the solvent and dissolve the residue in AcOEt (20 mL).
7. Wash the organic layer three times with saturated NaHCO₃ solution (5 mL) and once with brine (5 mL).
8. Dry the organic layer over MgSO₄, filter, and evaporate to afford compound 7 as a white solid (650 mg, 88 %). ¹H NMR (400 MHz; CD₃OD) δ 8.28 (d, 2H, *J*=8.8 Hz), 8.10 (d, 2H, *J*=8.8 Hz), 2.92 (s, 4H).

3.2.2 Synthesis of Compound 8 (Fig. 10)

1. Charge a round-bottom flask with biotin (1.18 g, 4.83 mmol) and dissolve in anhydrous DMF (8 mL).
2. Add *N*-hydroxysuccinimide (667 mg, 5.80 mmol) and EDC (1.11 g, 5.79 mmol).
3. Stir the reaction mixture for 10 h (*see Note 2*).
4. Evaporate the solvent.
5. Purify the residue by recrystallization from EtOH/AcOH/H₂O (95:1:4).

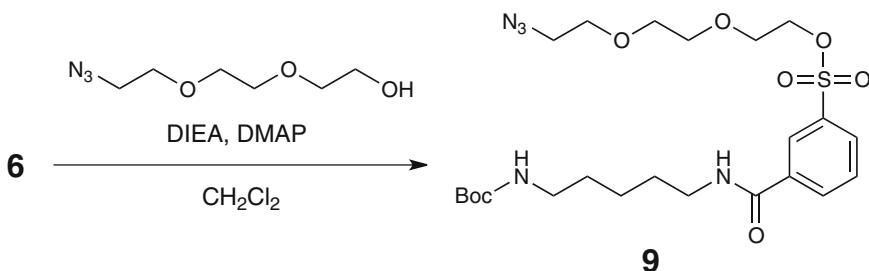


Fig. 11 Synthesis of compound 9

- 3.2.3 Synthesis of Compound 9 (Fig. 11)**

6. Collect the precipitate by filtration and dry under vacuum to afford compound **8** as a white solid (1.30 g, 79 %). ^1H NMR (400 MHz; $\text{CDCl}_3/\text{CD}_3\text{OD}$, 1:1): δ 4.49 (m, 1H), 4.31 (dd, 1H, J =7.6, 4.4 Hz), 3.17 (m, 1H), 2.91 (dd, 1H, J =12.8, 5.2 Hz), 2.84 (s, 4H), 2.71 (d, 1H, J =12.8 Hz), 2.63 (dt, 2H, J =7.2, 2.0 Hz), 1.83–1.51 (m, 6H).

3.2.4 Synthesis of Compound 10 (Fig. 12)

1. Charge a round-bottom flask with compound **6** (900 mg, 2.22 mmol) and dissolve in anhydrous CH_2Cl_2 (5 mL).

2. Add 2-(2-(2-azidoethoxy)ethoxy)ethanol (300 mg, 1.71 mmol), DIEA (872 μL , 5.13 mmol), and DMAP (63 mg, 0.52 mmol).

3. Stir the reaction mixture overnight (*see Note 2*).

4. Evaporate the solvent (*see Note 3*).

5. Purify the residue by column chromatography on silica gel ($\text{CHCl}_3/\text{MeOH}$, 30:1).

6. Evaporate the collected fraction to afford compound **9** as a colorless oil (570 mg, 61 %) (*see Note 4*). ^1H NMR (400 MHz; CDCl_3) δ 8.26 (s, 1H), 8.12 (d, 1H, J =7.6 Hz), 8.04 (d, 1H, J =8.0 Hz), 7.64 (dd, 1H, J =8.0, 7.6 Hz), 6.50 (br s, 1H), 4.59 (br s, 1H), 4.23 (m, 2H), 3.71 (m, 4H), 3.63 (t, 2H, J =5.0 Hz), 3.59 (s, 4H), 3.47 (m, 2H), 3.36 (t, 2H, J =5.0 Hz), 3.13 (m, 2H), 1.66 (m, 4H), 1.53 (m, 2H), 1.41 (s, 9H).

1. Charge a round-bottom flask with compound **9** (420 mg, 0.77 mmol) and dissolve in anhydrous CH_2Cl_2 (10 mL).

2. Add TFA (4 mL).

3. Stir the reaction mixture for 15 min (*see Note 2*).

4. Add toluene (4 mL) and evaporate the solution. Repeat this once again.

5. Dissolve the residue (deprotected **9**) in anhydrous DMF (2 mL).

6. Add compound **7** (299 mg, 1.00 mmol) and DIEA (390 μL , 2.29 mmol).

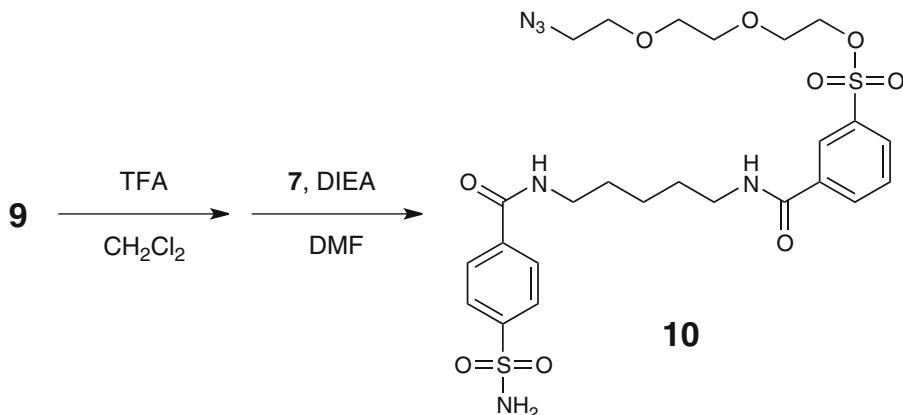


Fig. 12 Synthesis of compound **10**

7. Stir the reaction mixture for 3 h (*see Note 2*).
 8. Evaporate the solvent and dissolve the residue in AcOEt (5 mL).
 9. Wash the organic layer twice with saturated NaHCO₃ solution (5 mL) and once with brine (1 mL), dry over MgSO₄, filter, and evaporate.
 10. Purify the residue by column chromatography on silica gel (CHCl₃/MeOH, 20:1 to 10:1).
 11. Evaporate the collected fraction to afford compound **10** as a white solid (180 mg, 37 %) (*see Note 4*). ¹H NMR (400 MHz; CD₃OD) δ 8.35 (s, 1H), 8.13 (d, 1H, *J*=8.0 Hz), 8.07 (d, 1H, *J*=8.0 Hz), 7.95 (d, 2H, *J*=8.4 Hz), 7.92 (d, 2H, *J*=8.8 Hz), 7.71 (dd, 1H, *J*=8.0, 8.0 Hz), 4.22 (m, 2H), 3.67 (m, 2H), 3.60 (t, 2H, *J*=5.0 Hz), 3.54 (s, 4H), 3.42 (m, 4H), 3.31 (m, 2H), 1.70 (m, 4H), 1.48 (m, 2H).
- 3.2.5 Synthesis of LDT Reagent 3 (Fig. 13)**
1. Charge a round-bottom flask with compound **10** (6.5 mg, 10.4 μmol) and dissolve in anhydrous MeOH (2 mL).
 2. Add Pd-C (10 %, 10 mg) (*see Note 5*).
 3. Replace the argon gas in the flask with H₂ (*see Note 6*).
 4. Stir the reaction mixture under H₂ atmosphere for 1 h (*see Notes 2 and 7*).
 5. Change the H₂ gas in the flask back to argon.
 6. Add compound **8** (4.5 mg, 13.2 μmol) and DIEA (3.6 μL, 21.2 μmol).
 7. Stir the reaction mixture for 1 h (*see Note 2*).
 8. Remove the Pd-C by filtration and evaporate the filtrate.

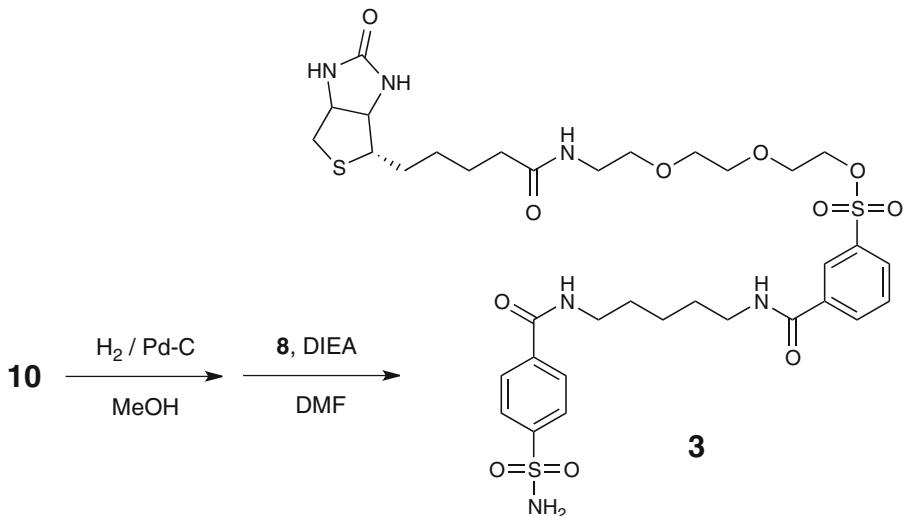


Fig. 13 Synthesis of LDT reagent **3**

9. Purify the residue by column chromatography on silica gel ($\text{CHCl}_3/\text{MeOH}$, 5:1).
10. Evaporate the collected fraction to afford LDT reagent **3** as a colorless oil (6.4 mg, 75 %) (*see Note 4*). ^1H NMR (400 MHz; $\text{CDCl}_3/\text{CD}_3\text{OD}$, 1:1) δ 8.35 (s, 1H), 8.11 (d, 1H, $J=7.6$ Hz), 8.02 (d, 1H, $J=8.0$ Hz), 7.92 (d, 2H, $J=8.4$ Hz), 7.89 (d, 2H, $J=8.0$ Hz), 7.65 (dd, 1H, $J=8.0, 7.6$ Hz), 4.47 (m, 1H), 4.28 (dd, 1H, $J=7.8, 4.6$ Hz), 4.21 (m, 2H), 3.68 (m, 2H), 3.62–3.48 (m, 8H) 3.42–3.33 (m, 4H), 3.14 (m, 1H), 2.88 (dd, 1H, $J=12.8, 4.4$ Hz), 2.69 (d, 1H, $J=12.8$ Hz), 2.18 (t, 2H, $J=7.2$ Hz), 1.69–1.56 (m, 8H), 1.47–1.38 (m, 4H). ^{13}C NMR (100 MHz; CD_3OD) δ 176.41, 168.78, 167.87, 166.09, 147.61, 139.19, 138.11, 137.23, 133.59, 131.63, 131.00, 128.94, 127.81, 127.32, 71.56, 71.52, 71.18, 70.61, 69.74, 63.36, 61.62, 56.99, 41.05, 41.04, 40.96, 40.33, 36.75, 30.07, 29.72, 29.47, 26.84, 25.36. HR-FAB MS (3-nitrobenzyl alcohol): calculated for $\text{C}_{35}\text{H}_{50}\text{N}_6\text{O}_{11}\text{S}_3$ $[\text{M}]^+=826.2700$; found 826.2672.

3.3 *In Vitro Carbonic Anhydrase II Labeling*

1. Prepare a stock solution of 10 mM DEAC-tethered LDT reagent **1** in DMSO.
2. Prepare a solution of 40 μM purified carbonic anhydrase II (CAII) in HEPES buffer (50 mM, pH 7.2) (*Note 8*).
3. Transfer 100 μL of the CAII solution in a 1.5 mL tube (*see Note 9*).
4. Add 0.8 μL of LDT reagent **1** to 100 μL (final concentration, 80 μM) and mix well by gently pipetting it up and down.

5. Incubate the reaction mixture at 37 °C.
6. At 3 and 24 h, transfer 40 µL of the reaction solution into a new 1.5 mL tube and mix with 40 µL of 2× Laemmli buffer.
7. Boil the sample at 95 °C for 5 min.
8. Go to Subheading 3.4 for in-gel fluorescence analysis of DEAC-labeled CAII.

3.4 In-Gel Fluorescence Analysis of DEAC-Labeled Carbonic Anhydrase II

The following SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and in-gel fluorescence detection can be performed using a standard procedure:

1. Load samples onto 12.5 or 15 % polyacrylamide gel followed by SDS-PAGE.
2. Place the gel on a stage of a Bio-Rad ChemiDoc XRS system.
3. Detect DEAC-labeled CAII by UV light excitation at 302 nm with a 480BP70 filter.
4. After fluorescence imaging, stain the gel with CBB.

3.5 Ex Vivo Chemical Biotinylation of Endogenous Carbonic Anhydrase in Living Blood Cells

All animal experiments must be performed in accordance with institutional guidelines for the care and use of animals in research:

1. Prepare a stock solution of 1 mM LDT reagent **3** in DMSO.
2. Anesthetize an Slc:ICR mouse with Et₂O in a polypropylene container.
3. Fix the mouse on a surgical table on its back.
4. After opening the abdomen by a standard animal surgical procedure, withdraw blood (0.4 mL) from the abdominal aorta using a disposable syringe and mix with 2 µL of 0.5 M EDTA solution (pH 8.0) in a 1.5 mL tube (*see Note 10*).
5. Dilute the blood sample two-fold with phosphate-buffered saline (PBS, pH 7.4).
6. Transfer 100 µL of the blood cell suspension in a 1.5 mL tube (*see Note 9*).
7. Add 0.5 µL of LDT reagent **3** (final concentration, 5 µM) and mix well by gently pipetting it up and down.
8. Incubate the reaction mixture at 35 °C for 20 h (*see Note 11*).
9. Lyse the cells by adding 30 µL of NP-40^{EZA} lysis buffer (*see Note 12*).
10. Mix the cell lysate with 130 µL of 2× Laemmli buffer, and boil the sample at 95 °C for 5 min.
11. Go to Subheading 3.7 for immunoblot analysis of biotin-labeled carbonic anhydrase.

3.6 In Vivo Chemical Biotinylation of Endogenous Carbonic Anhydrase in Living Mice

1. Prepare a stock solution of 10 mM LDT reagent **3** in DMSO.
2. Add 5 µL of LDT reagent **3** into 0.5 mL of PBS solution (final concentration, 100 µM) and mix well by pipetting it up and down.
3. Fix an Slc:ICR mouse in a mouse holder (no anesthesia) and secure the tail.
4. Inject the 0.5 mL PBS solution containing **1** through the tail vein.
5. Release the mouse in a cage.
6. At 4, 8, and 20 h after injection, fix the mouse again in a mouse holder.
7. Make a small cut in the tail vein with a surgical knife (*see Note 13*) and take blood (10 µL) using an automatic pipette.
8. Mix the blood immediately with 40 µL of NP-40^{EZA} lysis buffer in a 1.5 mL tube (*see Notes 10 and 14*).
9. Mix the blood lysate with 50 µL of 2× Laemmli buffer, and boil the sample at 95 °C for 5 min.
10. Go to Subheading **3.7** for immunoblot analysis of biotin-labeled carbonic anhydrase.

3.7 Immunoblotting of Biotin-Labeled Endogenous Carbonic Anhydrase

The following SDS-PAGE and immunoblotting can be performed using a standard procedure:

1. Load samples onto 15 % polyacrylamide gel followed by SDS-PAGE.
2. Electrotransfer the proteins onto an Immun-Blot PVDF membrane.
3. Block and wash the membrane using blocking solution and TBST described in Subheading **2.2**, respectively.
4. Detect biotinylated protein(s) with SAv-HRP and Chemi-Lumi One using an ImageQuant LAS 4000 imager.
5. For immunodetection of endogenous carbonic anhydrase, use anti-mouse CAII antibody and anti-goat IgG-horseradish peroxidase conjugate.

4 Notes

1. As shown in Fig. **14**, LDT reagents consist of five modules: (**a**) a specific ligand for a target protein, (**b**) a linker, (**c**) the phenyl sulfonate ester (tosylate) group, (**d**) a linker, and (**e**) a chemical probe that is desired to be introduced to the target protein. The ligand molecule **a** needs to be attached via the linker **b** to the phenyl ring of the tosylate group **c**, while the chemical probe **e** is linked via the linker **d** to the tosylate group **c** through

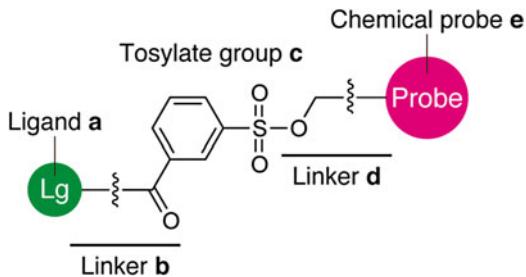


Fig. 14 Basic design configuration of LDT reagent

the sulfonate ester bond. This configuration allows the cleavage of the ligand moiety concomitantly with the affinity-induced labeling reaction. In addition to this basic design, the following considerations need to be made to generate LDT reagents that can be used for selective native protein labeling:

- (a) Ligand **a**: The choice of the ligand molecule is of primary importance to determine the selectivity and labeling efficiency of LDT reagents to their target protein. So far, we have obtained successful results by using highly specific ligands with a K_d below μM [17, 20–23], but the required affinity is expected to depend on the concentration of the target protein in the context of interest. It should also be noted that the use of too-high-affinity ligands may perturb the function of the labeled protein due to the difficulty of release of the cleaved ligand fragment from the ligand-binding pocket of the protein after labeling. Thus, depending on applications, several trials to test ligand candidates with different affinities (if available) might be required. In some cases, the cleaved ligand fragment can be removed from the cells (1) by simple washing procedures if the fragment is cell permeable or (2) with the help of cellular organic anion transporter systems that exclude anionic organic compounds, such as phenyl sulfonate anion derivatives, from the interior to the exterior of cells [21].
- (b) Linker **b**: In general, the efficiency of affinity labeling is governed by the proximity between the reactive (electrophilic) group of the labeling reagent and the nucleophilic amino acid of the target protein when they are complexed [22]. Thus, in the design of LDT reagents, the choice of linker **b** is a critical factor that controls the location of the tosylate group around the ligand-binding pocket. As shown in Table 1, we now know that the tosylate electrophile is reactive toward His, Tyr, Glu, Asp, and Cys residues [29]. Therefore, when the crystal structure of the target protein is available, it is a clever way to look for these amino acids near the ligand-binding pocket and adjust the linker length so that the tosylate group can be brought

close to the targeted amino acid(s). In some cases, the use of rigid structures, such as cyclic compounds, as linker **b** is effective to enhance labeling efficiency and specificity [22]. Nevertheless, it is still difficult to predict the best linker structure just from crystal structural information because of protein dynamics and/or the lack of tools to estimate the nucleophilicity of amino acids on protein surfaces. We thus recommend to test several types of linker to achieve better labeling efficiency. Such an effort, i.e., testing a set of labeling reagents with various linkers, is crucial when no structural information of the target protein is available.

- (c) Linker **d**: Linker **d** is a part of the tosylate ester, and thus its structure affects the chemical reactivity of the tosylate group itself. For example, when we attempted to use a benzyl alcohol derivative as linker **d**, the resulting benzyl phenyl sulfonate compound was too reactive to be isolated [30]. So far, we have obtained satisfactory results by using ethylene glycol-based linkers. Ethylene glycol-based linkers are also advantageous in terms of their flexibility and water solubility.
 - (d) Chemical Probe **e**: In principle, a variety of chemical probes can be used in the LDT chemistry. However, when applying this method to intracellular proteins, the LDT reagent must be cell permeable. Therefore, in such cases, the use of cell-impermeable probes might not be suited. In this regard, it should be noted that cell permeability of LDT reagents is not determined only by the probe moiety but is controlled by the whole structure of the reagent. For example, when a cell-impermeable fluorescein dye was introduced as a chemical probe into an LDT reagent targeting FKBP12, the resulting reagent was cell permeable owing to the high cell permeability of the ligand used [23]. It is also important that the chemical probe of interest does not interact nonspecifically with cellular components such as (nontarget) proteins or lipids, because such undesired interactions will cause off-target labeling.
 - (e) Synthetic aspect: LDT reagents are not commercialized at the moment. Therefore, researchers need to synthesize their own LDT reagents by standard organic synthesis techniques. For this purpose, various synthetic routes can be considered. We usually adopt synthetic schemes in which ligands and/or chemical probes are attached at late steps in order to facilitate the exchange of these moieties. We are willing to provide LDT reagents or their intermediates that we have published upon request.
2. Regardless of the indicated reaction time, we recommend to monitor the reaction by TLC. TLC should be visualized by fluorescence quenching using a handheld lamp or by ninhydrin staining.

3. We recommend to purify the product immediately after the evaporation. The tosylate ester tends to decompose gradually when it is left as the crude residue.
4. Tosylate ester compounds should be kept at -20 °C.
5. Caution: Pd-C should be added when the flask is filled with argon. Be also careful to put all the Pd-C into the solvent. Pd-C may ignite if it is exposed to H₂ gas in a dry state.
6. We attach a H₂-filled balloon to the flask using a three-way cock.
7. In monitoring the reaction by TLC, change the H₂ gas in the flask back to argon first.
8. The concentration of CAII can be determined by measuring the absorbance at 280 nm using a molar absorption coefficient of 54,000 M⁻¹ cm⁻¹ [31].
9. Control samples are prepared in the same way.
10. This step should be performed quickly to avoid blood clotting.
11. We usually do not shake the tube even though cells gradually settle down during incubation.
12. Any other lysis buffer, such as RIPA buffer, can be used. However, we recommend to add EZA, a high-affinity inhibitor of carbonic anhydrase, to the lysis buffer in order to avoid potential progress of the labeling reaction during the lysis process.
13. A very small cut is sufficient to take 10 µL of blood for analysis.
14. We usually prepare 40 µL of NP-40^{EZA} lysis buffer in a 1.5 mL tube first and add the withdrawn blood to the tube.

Acknowledgment

We thank Masayoshi Miyagawa, Dr. Tomonori Tamura, Dr. Yousuke Takaoka, Dr. Hangxiang Wang, Yoshiyuki Kioi, and Takayuki Miki for their great contribution to the development and application of the LDT chemistry.

References

1. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* 72:249–289
2. Wang L, Schultz PG (2005) Expanding the genetic code. *Angew Chem Int Ed* 44:34–66
3. Chalker JM, Bernardes GJL, Lin YA, Davis BG (2009) Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem Asian J* 4:630–640
4. Prescher JA, Bertozzi CR (2005) Chemistry in living systems. *Nat Chem Biol* 1:13–21
5. Saghatelyan A, Cravatt BF (2005) Assignment of protein function in the postgenomic era. *Nat Chem Biol* 1:130–142
6. Johnsson K (2009) Visualizing biochemical activities in living cells. *Nat Chem Biol* 5:63–65
7. Sletten EM, Bertozzi CR (2009) Bioorthogonal chemistry: fishing for selectivity in a sea of

- functionality. *Angew Chem Int Ed* 48: 6974–6998
8. Takaoka Y, Ojida A, Hamachi I (2013) Protein organic chemistry and applications for labeling and engineering in live-cell systems. *Angew Chem Int Ed* 52:4088–4106
 9. Griffin BA, Adams SR, Tsien RY (1998) Selective covalent labeling of recombinant protein molecules inside live cells. *Science* 281: 269–272
 10. Kepler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat Biotechnol* 21:86–89
 11. George N, Pick H, Vogel H, Johnsson N, Johnsson K (2004) Specific labeling of cell surface proteins with chemically diverse compounds. *J Am Chem Soc* 126:8896–8897
 12. Chen I, Howarth M, Lin WY, Ting AY (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* 2:99–104
 13. Fernández-Suárez M, Baruah H, Martínez-Hernández L, Xie KT, Baskin JM, Bertozzi CR, Ting AY (2007) Redirecting lipoic acid ligase for cell surface protein labeling with small-molecules probes. *Nat Biotechnol* 25: 1483–1487
 14. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3:707–708
 15. Los GV et al (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3:373–382
 16. Nonaka H, Fujishima S, Uchinomiya S, Ojida A, Hamachi I (2010) Selective covalent labeling of tag-fused GPCR proteins on live cell surface with a synthetic probe for their functional analysis. *J Am Chem Soc* 132:9301–9309
 17. Tsukiji S, Miyagawa M, Takaoka Y, Tamura T, Hamachi I (2009) Ligand-directed tosyl chemistry for protein labeling *in vivo*. *Nat Chem Biol* 5:341–343
 18. Wold F (1977) Affinity labeling—an overview. *Methods Enzymol* 46:3–14
 19. Hayashi T, Hamachi I (2012) Traceless affinity labeling of endogenous proteins for functional analysis in living cells. *Acc Chem Res* 45: 1460–1469
 20. Tsukiji S, Wang H, Miyagawa M, Tamura T, Takaoka Y, Hamachi I (2009) Quenched ligand-directed tosylate reagents for one-step construction of turn-on fluorescent biosensors. *J Am Chem Soc* 131:9046–9054
 21. Takaoka Y, Sun Y, Tsukiji S, Hamachi I (2011) Mechanisms of chemical protein ¹⁹F-labeling and NMR-based biosensor construction *in vitro* and in cells using self-assembling ligand-directed tosylate compounds. *Chem Sci* 2: 511–520
 22. Tamura T, Tsukiji S, Hamachi I (2012) Native FKBP12 engineering by ligand-directed tosyl chemistry: labeling properties and application to photo-cross-linking of protein complexes *in vitro* and in living cells. *J Am Chem Soc* 134:2216–2226
 23. Tamura T, Kioi Y, Miki T, Tsukiji S, Hamachi I (2013) Fluorophore labeling of native FKBP12 by ligand-directed tosyl chemistry allows detection of its molecular interactions *in vitro* and in living cells. *J Am Chem Soc* 135:6782–6785
 24. Takahashi M, Kawamura A, Kato N, Nishi T, Hamachi I, Ohkanda J (2012) Phosphopeptide-dependent labeling of 14-3-3 ζ proteins by fusicoccin-based fluorescent probes. *Angew Chem Int Ed* 51:509–512
 25. Cho HJ, Gee HY, Baek K-H, Ko S-K, Park J-M, Lee H, Kim N-D, Lee MG, Shin I (2011) A small molecule that binds to an ATPase domain of Hsc70 promotes membrane trafficking of mutant cystic fibrosis transmembrane conductance regulator. *J Am Chem Soc* 133: 20267–20276
 26. Jing C, Cornish VW (2013) A fluorogenic TMP-tag for high signal-to-background intracellular live cell imaging. *ACS Chem Biol* 8: 1704–1712
 27. Song A, Wang X, Lam KS (2003) A convenient synthesis of coumarin-3-carboxylic acid via Knoevenagel condensation of Meldrum's acid with *ortho*-hydroxyaryl aldehydes or ketones. *Tetrahedron Lett* 44:1755–1758
 28. Fernandez-Megia E, Correa J, Rodríguez-Meizoso I, Riguera R (2006) A click approach to unprotected glycodendrimers. *Macromolecules* 39:2113–2120
 29. Weerapana E, Simon GM, Cravatt BF (2008) Disparate proteome reactivity profiles of carbon electrophiles. *Nat Chem Biol* 4:405–407
 30. Tsukiji S., Hamachi, I. Unpublished results
 31. Supuran CT, Briganti F, Tilli S, Chegwidden WR, Scozzafava A (2001) Carbonic anhydrase inhibitors: sulfonamides as antitumor agents? *Bioorg Med Chem* 9:703–714

INDEX

A

- Acceptor peptide (AP) 171. *See also* AviTag peptide
AeL-TerL intein 132, 140, 141
AcpS phosphopanthetheinyl transferase (PPTase) 162
ACP-tag 56
Acylation of N-terminal amine group 148, 157
Acyl transfer reaction 230
ADC. *See* Antibody drug conjugates (ADC)
Affinity-guided DMAP (4-dimethylaminopyridine) (AGD) chemistry 229–242
Affinity labeling 244, 260
Affinity probes 163
Affinity purification 186, 199–214
Affinity-tag 201
AGD chemistry. *See* Affinity-guided DMAP
 (4-dimethylaminopyridine) (AGD) chemistry
AHA. *See* Azidohomoalanine (AHA)
Aldehyde tag 11
Alkyne-bearing affinity tag 202
Amber stop codon 218
Amine-tag 148, 149, 151, 152, 154, 155
1,2-Aminothiol 83, 84, 91
Antibody drug conjugates (ADC) 8, 15, 16, 22, 145
Avidin 9, 174, 181, 182
Avi-tag 8–11, 16–21, 23
AviTag peptide 175
Azide-bearing affinity tag 202
Azidohomoalanine (AHA) 200–202, 204–208,
 212, 214
p-Azidophenylalanine (AZF) 149, 151, 152, 157

B

- Benzylchloropyrimidine derivatives 10
Benzylcytosine derivatives 70
Benzylguanine derivatives 70
Beta-adrenoceptor G-protein coupled receptors 96
Biarsenical compound 3
Ne-(Bicyclo[6.1.0]non-4-yn-9-yl-methoxy)carbonyl-L-lysine (BCNK) 218–220, 224, 225, 227
Bioconjugation reaction 81, 130, 146, 148
Bioorthogonal Diels–Alder reaction
 with inverse electron demand 218

- Bioorthogonal non-canonical amino acid tagging (BONCAT) 199–214
Bioorthogonal reactions 146, 157, 218
Biopharmaceutical 8, 11, 14–16, 23
Biophysical methods 7–10, 22
Biosensor 65–66
Biotin-CoA conjugate 162–164, 166
Biotin labeling 162, 163, 167–169, 258, 259
Biotin ligase 8, 17, 146, 172, 173, 181
Biotinylation 9, 10, 16–19, 23, 171–182, 258, 259
BirA 8, 17, 171–182
BL-tag 55, 146

C

- Carbonic anhydrase (CA) 66, 67, 98, 99, 103,
 203, 208, 244–247, 249, 250, 257–259, 262
Cell-surface labeling 71, 120
Cellular lysates 122
Chemical crosslinker 64
Chemical protein modification 243
Chemical synthesis 72, 189, 246
Chemical-tag labeling 145–157
Chemoenzymatic labeling 185–197
Chem-tag 130, 148
Click chemistry 10, 13, 15, 21, 82, 199–227
Click-tag 148, 149, 151, 152, 154–157
CLIP-tag 55–60, 64, 65, 108, 146
Coenzyme A (CoA) 8, 22, 29, 161–164, 166
Copper-catalyzed alkyne–azide cycloaddition
 (CuAAC) 146, 149, 156
Covalent labeling 107, 116, 119–127
2-Cyanobenzothiazole (CBT) condensation 81–91
Cysteine 1, 2, 5, 12, 15, 16, 20, 57, 81–91,
 94, 95, 99, 100, 103, 129, 141, 145, 147, 148, 156,
 185, 186, 193, 196
Cysteine-reactive probes 103

D

- De novo synthesized proteome 205, 207
Design of labeling molecules 30
Dienophiles 218
DMAP catalyst 229, 230
Drug discovery 7–23

E

- E. coli.* See *Escherichia coli (E. coli)*
 Environmentally-sensitive fluorophores 94, 95
 Enzyme-mediated labeling of tags 217
Escherichia coli (E. coli) 12, 17, 84, 86, 87, 115, 133, 136, 150, 152, 156, 162, 166–168, 171, 173, 175–177, 179, 182, 219–223, 226, 231, 245
 1,2-Ethanedithiol (EDT) 4, 133, 232
 Exteins 130, 141, 147, 152, 156

F

- FK506-binding protein 12
 (FKBP12) 230–234, 238–241, 245, 261
 Fluorescein Arsenic Helix-binder
 (FIAsh) 1–5, 13, 21, 146
 Fluorescence energy transfer
 (FRET) 10, 58, 63–66, 94, 162, 217
 Fluorescence labeling 93–104
 Fluorescent gel scan 122
 Fluorescent probes 59, 61
 Fluorogenic labeling 219

G

- GABA_A receptor channels 96
 Genetic code expansion 218, 219
 Genetic encoding of unnatural amino acids 217–227
 G protein-coupled receptor
 (GPCR) 12, 20, 43, 58, 230
 Green fluorescent protein
 (GFP) 1, 13, 21, 22, 43, 119, 146

H

- HaloTag 55, 56, 119–127
 Homopropargylglycine (HPG) 201, 202

I

- Intracellular protein labeling 40
 In vitro cell-free expression systems 122
 In-vitro imaging 222, 247
 In vivo animal imaging 123
 In-vivo imaging 13–14, 107, 123
 Ion channels 12, 36, 93–104

L

- Labeling specificity 56, 103
 Ligand-directed tosyl chemistry 243–262
 Lipoic acid ligase 8, 13, 55, 146

M

- Mass spectrometry 74, 90, 156, 195, 222
 Membrane permeation 32–34, 39, 41
 Metabolic labeling 12, 199–214

- Methionine surrogates 201
Mxe GyrA intein 147

N

- Nanoparticle 44, 58, 68, 69, 163, 173
 Native chemical ligation
 (NCL) 82, 83, 86, 129, 130, 146
Ne-5-norbornene-2-yloxy carbonyl-L-lysine
 (NorK) 218–220, 224, 227
 Neurophysiology 93
 Neutravidin-agarose 210
 Non-canonical amino acid 200, 201, 212
 No-wash labeling 122, 125, 127
Npu DnaE intein 147–149, 156

O

- O⁶-alkylguanine-DNA alkyltransferase*
 (hAGT) 57, 107

P

- Patch clamp fluorimetry (PCF) 93–104
 Peptide ligation 129–141
 Peptidyl carrier protein (PCP)
 Permeability 30, 32–37, 39–45, 57, 59, 121, 261
 Phage display 16, 162, 163, 169, 171
 Phosphopantetheinyl transferase
 (PPTase) 8, 22, 55, 161–169
 Phosphoproteome 200
 Photoactivatable molecules 58
 PhotoActivated Localization Microscopy
 (PALM) 60

- Post-translational modification 14, 146, 161, 199
PPTase. See *Phosphopantetheinyl transferase (PPTase)*

Protein

- degradation 122
 engineering 82, 163, 189
 half-life 57, 108
 homeostasis 199, 200, 205
 labeling 7–23, 58, 82, 86, 108, 113, 121–127, 131, 147, 157, 161–169, 186, 206, 218, 220–223, 229, 230, 243–262
 native 229, 243–262
 recombinant 145–157
 modification 10, 15, 86, 146, 171, 200, 243
 semi-synthesis 130
 splicing 83, 156
 synthesis 38, 122, 133, 199, 205–208
 trafficking 56, 103, 122, 126
 turn-over 76, 199, 200, 206, 213

- Protein-drug interactions 11, 108
 Protein-protein interactions 10, 11, 19–21, 56, 58, 63–65, 107, 174, 229
 Proteomic profiling 201–214
Psp-GDB Pol intein 147

- Pulse-chase labeling.....122, 124, 126
 Pyrrolysyl tRNA synthetase/tRNA_{CUA} pair
 and variant83, 218, 225
- R**
- Radiolabeling.....86
 Reactive labeling.....21
 Resorufin Arsenic Helix-binder (ReAsH).....5
- S**
- Selective cross-linking (S-CROSS).....57, 109
 Self-labeling protein.....11, 55, 120
 Self-labeling tag.....11, 21, 107, 119
 Semisynthetic protein trans-splicing129–141, 149
Sfp phosphopanthetheinyl transferase (PPTase)162
 Site-specific labeling.....11, 14, 16, 21, 23, 56, 60,
 64, 65, 81–91, 185–197, 218, 219
 SNAP-tag.....13, 14, 20, 21, 41, 55–68, 71,
 73–76, 107–117
 SNS. *See* Synaptoneuroosomes (SNS)
 Solid-phase peptide synthesis.....130, 132–134,
 146, 149, 192
 Sortase.....8, 55, 146, 174, 185–197
 Split intein.....130–132, 138, 145–157, 174
Ssp DnaB intein.....131, 134, 141, 147
 Stimulated emission depletion
 (STED)60, 112–113, 117
 Stochastic optical reconstruction microscopy
 (STORM)60
 Streptavidin9–11, 19, 20, 155, 162,
 163, 167–169, 171–175, 179–182, 247, 250
 Streptavidin gel-shift.....179–181
 Structural dynamics.....261
 Subproteome200
 Super-resolution microscopy60, 217
 Synaptoneuroosomes (SNS).....200, 204–208, 213
 Synthetic peptide.....129–141

T

- TAMRA carboxytetramethylrhodamine97
 Targeted protein inactivation.....107
 Terminal cysteine.....81–91
 Tetracysteine Tag56
 Tetrazine-fluorophore
 conjugates218–220, 222, 225, 227
 Thiol-linked fluorophore.....99
 TMP-tag55, 56
 Tosyl chemistry.....243–262
 Traceless bioconjugation.....146
Nε trans-cyclooctene-4-oxy carbonyl-L-lysine
 (TCOK)218–227
 Transglutaminase labeling8, 20
 Transmembrane receptors.....93
 Transpeptidation186–188, 195–197
 Traptavidin173

U

- Ubiquitin activating enzyme
 (UAE)21, 162–165, 168, 169
 Unnatural amino acids.....11, 129, 146, 149,
 157, 217–227

V

- Voltage clamp fluorometry (VCF).....94–96
 Voltage-gated ion channels.....96
 Voltage-gated proton channel Hv1.....96, 97

X

- Xenopus laevis oocytes96, 99
Y
 Yeast cell surface display163