

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie

Labor für Synthetische Biochemie

**Genetically encoding unnatural amino acids: Novel tools for protein labelling and chemical stabilisation of low-affinity protein complexes**

**Marko Cigler**

Vollständiger Abdruck der, von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Angela Casini

Prüfende/-r der Dissertation: 1. Prof. Dr. Kathrin Lang  
2. Junior Group Leader Dr. Stephan M. Hacker

Die Dissertation wurde am 26.09.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 21.10.2019 angenommen.



*The work presented within this thesis was carried out in the Lab for Synthetic Biochemistry, Technical University Munich, from January 2015 to September 2019, under the supervision of Prof. Dr. Kathrin Lang*

*Furthermore, this work was supported by the two year Kekulé Mobility Fellowship (Kekulé-Stipendium) awarded by the Fonds der Chemischen Industrie (FCI).*



## Acknowledgements

*To whom it may concern,*

*I would like to use the space within the next page or two to express my gratitude towards quite a few people who have accompanied me on this rocky road called PhD. I really wish I could say “I did it all by myself!” but, even though I am quite good (talking about being modest...), this is just not how life works. In the last (almost!) 5 years I came across several wonderful people (and a couple of less-wonderful, to remain polite) and I want to acknowledge them properly. Before we start though, please keep in mind that I’m an Ausländer, so don’t be surprised if you suddenly do not understand the text you are reading – it might be in a different language. Also, apologies to people I might have forgotten but definitely feel deserving of my gratitude. If you are one of them, know that I will find you and I will hug you (and I am not joking!)*

*First and foremost, I must thank the person without whom this thesis would definitely not exist. Thank you Kathrin very much for giving me the opportunity to work in your lab, learn a bunch of new stuff from you and develop cool new methods. An enormous “Thank you” for encouraging me to think both critically and creatively of my work but also for pushing me when I needed to be pushed. Thank you for, and I think I speak for all of us at #KLangLab, always having our back. This has definitely been a transformative experience that has made me a better scientist and for this you will always have my gratitude. I hope you didn’t regret your decision of hiring “a timid little boy from a backward little town”; I sure do not regret of doing my PhD in your lab!*

*I would, furthermore, like to thank my second examiner Stephan Hacker for his time and effort in evaluating the work from the past 4.5 years as well as Fonds der Chemischen Industrie for their financial support during my time as PhD.*

*To my family – mom, dad & Luka – I want to thank you guys straight away so that my appreciation and gratitude doesn’t get lost somewhere at the end. A colossal thank you for your unconditional love and support. And Luka, Chemistry is still way cooler than Math!*

*À ma chère Ex-Maman, je suis si heureux d’avoir passé tant de temps merveilleux avec toi! Merci beaucoup pour tous les souvenirs et.... Mets une banane dans ton oreille! Many, many thanks go to the coolest Chipmunks in the world as well as to Mr. Billy Stinkwater. I had a blast with you guys! To the incomparable Herzig group - you rock! #YOLO. Hülya, Schluggi, the best (and only) Aussie I know, das Kind, René and all of the other Gulder people; thank you for making my PhD time great. Nikako ne mogu zaboraviti na ekipu iz Zagreba i okolice (nazovimo Milano okolicom)... Kaj da vam velim? Jel se mi zbilja još uvijek družimo s Cindrom?*

*This acknowledgement is also nothing without mentioning a lot of both past and current members of #KLangLab! From Thorsten “Thorstibär” Mülleimer and Mathias, Dominik (Essig), Daniel Ghetko, Patrickant, Tanja, Ángela, Carina and Caro all the way to den jungen, hübschen und knackigen Maxl and AberDomi.*

*And finally, finally before all of this gets really repetitive and boring, I would like to dedicate a tiny bit of space to each of my colleagues, my peers, my equals, who have made*

coming to work for the past 4.5 years a joy! In what I hope is supposed to be chronological order, my deepest gratitude from the bottom of my heart and the bottom of my bottom goes to:

Susanne; We've been through a lot you and I, haven't we? Thank you for your friendship, for all the help and fruitful discussions (and we had lots of those, every time something didn't come out of the column and the contents of a flask unexpectedly turned black) and basically for a great time inside and outside the lab! Can't wait to brunch with you as soon as I'm jobless!

Max; I can still remember the first days of August 2015. Unusually cold for that time of year, but with a dash of unexplainable mystery in the air. But all was finally clear to me as I came to the lab that morning and gazed upon your beautiful hazel eyes for the first time... (at least before that silly accident with Schleppis and almost ending up with a #bloodynose #thathurted). 4 words: You're the man! (And I secretly still hope to get at least a TA position at #AKFottner one day! I'd also apply for the butler position or the Eppi-shaker, FYI).

Meine kleine Mariechen und für immer mein Fujikaiten "partner in crime", du bist die putzigste Person die ich kenne! Und auch wenn ich dir manchmal Hamsterbäckchen sage, du bist immer für mich da und ich habe sehr viel Spaß mit dir! Ich freue mich schon riesig für dich alle Garnelen zu schnappen!

The schmockastic Anh from the god-forgotten Schmockistan, I knew you were trouble when you walked in! Danke dir für alle Tage wo du Toni & Azo fertig gemacht hast... jeden Tag deine unglaublichen Sprüche zu hören war einfach der Wahnsinn. Kudos for the Tropic music as well. I really hope you become a president of something one day... You definitely have my vote!

Kristina, X-Tina, Augustina, AzoK, AlcoK, KaKa, Alkoxide, Kazo and dozens of other nicknames I cannot remember right now. I have so many things to say to you and so little space left... Will you be happy with three little words? (And no... they are not "Ich will schokiii"). In all seriousness, thank you for everything you did and for all the fun times we had! My only regret is that I didn't see you clean the pumps more often.

Die geile Lenny Lennox Lenovo MJ! (Another one with great nicknames!) Ich mag dich sehr, das weißt du hoffentlich schon! Bleib so wie du bist und ich freue mich abartig auf unser Pilzdoktor Abend; versprochen ist versprochen!

Toni... I've given a lot of thought what to say to you (that I already haven't said in your face). I heard some whispers that you are a good chemist so I am very happy for you. Keep doing what you are doing and don't think of Erdbeeralehyd!

And last but not least, Vera die Beschte – it has been the privilege and the honour of my life to Schlenk proteins with you! Your spirit is full of life and these past weeks where you've been gone were terrible! Don't you ever change Wankstar!  
#SchlenkBuddies4Life #We'llAlwaysHaveTheSchlenkLine

To sum it all up, It's been a pleasure to work alongside all of you guys, Kathrin included!  
I will miss you all!

M, signing out.  
(KLangLab 2015-2019)

PS. Is there a mike I can drop? This feels like a cool moment to do it



## List of Publications

*Parts of this thesis have been published in peer-reviewed journals:*

T. A. Nguyen, **M. Cigler**, K. Lang, *Expanding the Genetic Code to Study Protein-Protein Interactions*, *Angew. Chem. Int. Ed.* **57** (2018), 14350-14361.

**M. Cigler**, T. G. Müller, D. Horn-Ghetko, M.-K. von Wrisberg, M. Fottner, R. S. Goody, A. Itzen, M. P. Müller, K. Lang, *Proximity-triggered covalent stabilization of low-affinity protein complexes in vitro and in vivo*, *Angew. Chem. Int. Ed.* **56** (2017), 15737-15741.

*Parts of this thesis have been published in book chapters:*

**M. Cigler**, T. A. Nguyen, K. Lang, *Genetic code expansion approaches to introduce artificial covalent bonds into proteins in vivo*. In *Oxidative folding of proteins: Basic principles, cellular recognition and engineering* (pp. 399-420), Matthias J. Feige (Ed), The Royal Society of Chemistry, 2018.

*Parts of this thesis have been presented at scientific conferences:*

**M. Cigler**, T. G. Müller, D. Horn-Ghetko, M.-K. von Wrisberg, M. Fottner, R. S. Goody, A. Itzen, M. P. Müller, K. Lang, *Proximity-triggered covalent stabilization of low-affinity protein complexes in vitro and in vivo*, poster presentation, 2018 Genetic Code Expansion Conference, 9.-12.8.2018, Corvallis, Oregon, USA

**M. Cigler**, K. Lang, *Stabilization of low-affinity protein-protein interactions by site-specific incorporation of unnatural amino acids*, oral contribution, 25<sup>th</sup> Croatian Meeting of Chemists and Chemical Engineers with International Participation (25HSKIKI), 19-22.4.2017, Poreč, Croatia

**M. Cigler**, T. G. Müller, M. Müller, A. Itzen, K. Lang, *Chemical stabilization of low-affinity protein complexes*, poster presentation, EMBO Conference: Chemical Biology 2016, 31.8-3.9.2016, Heidelberg, Germany



## Table of Contents

<b>ABSTRACT</b> .....	<b>XIII</b>
<b>ZUSAMMENFASSUNG</b> .....	<b>XV</b>
<b>§ 1. INTRODUCTION</b> .....	<b>1</b>
<b>1.1. Visualising proteins inside living cells: An overview</b> .....	<b>1</b>
<i>1.1.1. ... And then there was GFP</i> .....	<i>1</i>
<i>1.1.2. Genetic tags for labelling</i> .....	<i>3</i>
<b>1.2. “The stuff that proteins are made of”</b> .....	<b>5</b>
<b>1.3. Expanding the genetic code: Suppression of the amber codon</b> .....	<b>8</b>
<i>1.3.1. Applications and efforts within the field</i> .....	<i>11</i>
<b>1.4. Bioorthogonal reactions in chemical biology</b> .....	<b>13</b>
<i>1.4.1. Hydrazone/oxime ligations</i> .....	<i>13</i>
<i>1.4.2. Azides as bioorthogonal handles</i> .....	<i>14</i>
<i>1.4.3. Inverse electron-demand Diels-Alder cycloadditions</i> .....	<i>18</i>
<i>1.4.4. Light-induced reactions in chemical biology</i> .....	<i>21</i>
<i>1.4.4.1. 1,3-dipolar cycloadditions</i> .....	<i>22</i>
<i>1.4.4.2. Light-induced Diels-Alder reactions</i> .....	<i>24</i>
<i>1.4.4.3. Thiol-ene and thiol-yne click reactions</i> .....	<i>24</i>
<b>1.5. Proximity-triggered protein crosslinking</b> .....	<b>26</b>
<i>1.5.1. Overview of UAAs able to undergo proximity-enabled reactions</i> .....	<i>28</i>
<i>1.5.1.1. Reactive halides</i> .....	<i>28</i>
<i>1.5.1.2. Michael acceptors</i> .....	<i>30</i>
<i>1.5.1.3. Other reactive UAAs</i> .....	<i>31</i>
<b>§ 2. DEVELOPMENT OF A NOVEL LIGHT-INDUCED BIOORTHOGONAL REACTION FOR SITE-SPECIFIC PROTEIN LABELLING</b> .....	<b>33</b>
<b>2.1. Quinone methides (QMs)</b> .....	<b>34</b>
<b>2.2. Light-induced hetero-Diels-Alder reaction between <i>o</i>NQMs and vinyl ethers</b> .....	<b>35</b>
<b>2.3. Building the initial setup</b> .....	<b>38</b>
<b>2.4. The rocky road towards obtaining <i>o</i>NQM-based UAAs</b> .....	<b>41</b>
<b>2.5. Incorporation of vinyl ether and vinyl sulfide UAAs and subsequent labelling experiments</b>	<b>46</b>
<b>2.6. The final experiment</b> .....	<b>53</b>
<b>2.7. Pursuing alternative bioorthogonal reactions for UAAs containing vinyl ether and vinyl sulfide moieties</b> .....	<b>55</b>
<i>2.7.1. ortho-Quinoline quinone methides (<i>o</i>QQMs)</i> .....	<i>56</i>

---

2.7.2. 9,10-phenanthrenequinones (PQs).....	59
2.8. Conclusion and outlook .....	65
<b>§ 3. PROXIMITY-TRIGGERED COVALENT STABILISATION OF LOW-AFFINITY PROTEIN COMPLEXES IN VITRO AND IN VIVO.....</b>	<b>67</b>
3.1. Synthesis and site-specific incorporation .....	67
3.2. Choosing the ideal proof-of-principle system.....	69
3.3. <i>In vitro</i> crosslinking experiments.....	71
3.4. <i>In vivo</i> crosslinking experiments in <i>E. coli</i> and mammalian cells.....	74
3.5. Expanding the crosslinking approach towards MS/MS-cleavable UAAs.....	78
3.5.1. Design and site-specific incorporation of MS <sup>2</sup> -cleavable UAAs .....	81
3.6. Conclusion and outlook .....	87
<b>§ 4. EXPERIMENTAL PART .....</b>	<b>89</b>
4.1. General experimental details – Chemistry .....	89
4.1.1. General procedure for preparation of protected amino acids .....	90
4.2. General experimental details – Biology .....	92
4.2.1. Transformation of competent <i>E. coli</i> cells.....	93
4.2.2. QuikChange <sup>TM</sup> site-directed mutagenesis .....	93
4.2.3. Synthetase screening for site-specific incorporation of UAAs into sfGFP .....	94
4.2.4. General protocol for expression and purification of UAA-containing proteins .....	95
4.3. Supporting information for Chapter 2 .....	96
4.3.1. Photochemistry equipment .....	96
4.3.2. Small molecule photochemistry experiments .....	99
4.3.2.1. Small molecule photolysis experiments with oNQMPs.....	99
4.3.3.2. Small molecule photochemistry experiments with PQs .....	100
4.3.3. Chemical synthesis .....	101
4.3.3.1. ortho-Naphthoquinone methide (oNQM) derivatives and related compounds .....	101
4.3.3.2. ortho-Quinolinone quinone methide (oQQM) derivatives and related compounds .....	125
4.3.3.3. 9,10-phenanthrenequinones (PQs) derivatives and related compounds .....	129
4.3.4. Biological methods and labelling experiments .....	133
4.4. Supporting information for Chapter 3 .....	136
4.4.1. Chemical synthesis .....	136
4.4.1.1. Bromoalkyl-bearing UAAs.....	136
4.4.1.2. Sulfoxide-containing UAAs .....	140
4.4.1.3. Bromoalkyl-bearing urea derivatives .....	143
4.4.2. Biological methods .....	149
<b>§ 5. REFERENCES .....</b>	<b>157</b>

---

---

**§ 6. LIST OF ABBREVIATIONS .....173**

**§ 7. APPENDIX .....CLXXVII**

**7.1. Supporting Figures – Chapter 2 ..... clxxvii**

**7.2. Supporting Figures – Chapter 3 ..... cxci**

**§ 8. EIDESSTATTLICHE ERKLÄRUNG .....CXCVIII**

---



## ABSTRACT

The genetic code is a nearly universal set of rules by which living cells translate the information encoded in their genomes into proteins. Commonly composed of 20 natural amino acids, proteins play a notable part in essentially all biological processes necessary for normal functioning of cells and organisms. Considering the limited diversity of functionalities throughout the natural amino acids, recent years have seen an increased focus on artificially expanding the genetic code to include unnatural amino acids (UAAs) adorned with unique biochemical handles. Reassignment of the amber stop codon (also known as amber suppression) definitely stands out as a powerful tool which allows site-specific incorporation of UAA into proteins *in vivo* in both prokaryotes and eukaryotes. Key requirements necessary for this endeavour include an orthogonal aminoacyl-tRNA synthetase (aaRS) that specifically recognizes and charges the UAA exclusively onto its cognate tRNA<sup>CUA</sup> and not onto endogenous host tRNAs. Likewise, the tRNA<sup>CUA</sup> is not a substrate for endogenous aaRSs and consequently directs the incorporation of the UAA in response to a premature amber codon (UAG) within the mRNA during translation. By altering the specificity of orthogonal aaRS through directed evolution approaches, more than 250 UAAs have already been site-specifically incorporated into proteins of interest. The chemical variety of incorporated UAAs, ranging from bioorthogonal handles and posttranslational modifications to photo- and chemical crosslinkers, has enabled multiple opportunities to site-specifically label proteins *in vivo* and modulate their function, map transient protein-protein interactions as well as elucidate diverse biological processes as they proceed in space and time.

Within the scope of this thesis we aspired to employ site-specific incorporation of novel UAAs for two conceptually different *in vitro* and *in vivo* applications: (i) site-specific labelling of proteins *via* a novel light-induced bioorthogonal reaction and (ii) proximity-triggered covalent stabilisation and characterisation of low-affinity protein complexes.

The development of fast, chemoselective and high-yielding bioorthogonal reactions between UAAs and externally added chemical probes has become an indispensable tool for labelling individual proteins and proteomes. Recent years have especially witnessed a focus on light-induced bioorthogonal reactions as these are distinguished by elegant reaction control with exceptional spatial and temporal resolution. Herein we attempted to exploit the distinctive reactivity of photochemically generated *ortho*-naphthoquinone methides (*o*NQMs) towards vinyl ethers and sulfides in order to develop a novel and exceptionally fast hetero-Diels-Alder reaction (second-order rate constants  $\sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ) for site-specific labelling of proteins. We devised and simultaneously pursued two labelling strategies, depending on which reaction partner would be incorporated as an UAA and which utilised as an externally added labelling reagent. While the synthesis of UAAs bearing *o*NQM precursors (*o*NQMPs) was challenging and their site-specific incorporation into proteins requires directed evolution,

we managed to efficiently introduce vinyl ether- and sulfide-containing UAAs into sfGFP in both *E. coli* and mammalian cells. Nevertheless, in all of our labelling attempts on purified proteins, we were confronted with the undesired reactivity of *o*NQMs with nucleophilic amino acid residues. Unfortunately, all of our efforts to circumvent this problem were unsuccessful and we opted not to pursue the reaction for labelling purposes any further. This thesis furthermore describes several alternative reactions for vinyl ether/sulfide-bearing UAAs, notably the visible-light-mediated [4+2] cycloaddition with 9,10-phenanthrenequinones, which ultimately suffered the same fate as *o*NQMs.

Formation of low-affinity and transient protein-protein interactions is indispensable in signalling and regulatory pathways in the cell. Due to their highly dynamic nature, however, high-resolution structural studies of low-affinity protein complexes are often elusive as they frequently require the generation of stable complexes. For this purpose, we sought out to link low-affinity protein complexes in their native conformation by incorporating flexible bromoalkyl-bearing UAAs that are inert under physiological conditions but can react with nucleophilic amino acid such as cysteines upon increasing their local concentration through proximity. Consequently, we were able to covalently crosslink positions brought into proximity during the formation of the protein complex, both *in vitro* and in living bacteria and mammalian cells. Moreover, we chemically stabilized and determined the crystal structure of a transient GDP-bound complex between a small human GTPase Rab1b and the GEF domain of its nucleotide exchange factor DrrA from *Legionella pneumophila*. We are currently attempting to expand this approach to UAAs containing an additional MS/MS cleavable moiety which would not only help entrap but also facilitate the characterisation of hitherto unmapped protein interactions.

---

## ZUSAMMENFASSUNG

Der genetische Code ist eine nahezu universelle Reihe von Regeln, die lebende Zellen nutzen, um die in ihren Genomen gespeicherte Information in Proteine zu übersetzen. Proteine, welche üblicherweise aus 20 natürlichen Aminosäuren bestehen, spielen eine wichtige Rolle im Grunde aller biologischen Prozesse, die notwendig sind für die normale Funktion von Zellen und Organismen. Betrachtet man die eingeschränkte Vielfalt von chemischen Gruppen unter den natürlichen Aminosäuren, dann konnte man in den letzten Jahren einen erhöhten Fokus auf die künstliche Erweiterung des genetischen Codes mit unnatürlichen Aminosäuren (UAAs) beobachten, die einzigartige biochemische Gruppen tragen. Die Neuordnung des Amber-Stopp-Codons (auch bekannt als Amber-Suppression) ist ein wirkungsvolles Werkzeug für den positionsspezifischen Einbau von UAAs in ein Protein *in vivo* in Prokaryoten und Eukaryoten. Hauptanforderungen sind dabei ein orthogonales Aminoacyl-tRNA<sub>CUA</sub> Synthetase/tRNA<sub>CUA</sub> Paar, das die UAA spezifisch erkennt und verarbeitet, ohne Kreuzreaktivitäten gegenüber dem Wirt zu zeigen. Mehr als 200 UAAs konnten bereits positionsspezifisch in Zielproteine eingebaut werden, indem die Spezifität der orthogonalen Synthetase durch gerichtete Evolution verändert wurde. Die biochemische Vielfalt der eingebauten UAAs reicht von bioorthogonalen Gruppen, die über spezielle chemische Reaktionen weiter funktionalisiert werden können, und posttranslationalen Modifikationen zu Photo- und chemischen Crosslinkern. Diese haben die positionsspezifische Markierung von Proteinen *in vivo* ermöglicht, sowie die Regulierung ihrer Funktion, die Abbildung von kurzlebigen Protein-Protein-Interaktionen und die Aufklärung von verschiedenen biologischen Prozessen in zeitlicher und räumlicher Auflösung.

Im Rahmen dieser Thesis wollten wir den positionsspezifischen Einbau von neuen UAAs mit orthogonalen tRNA-Synthetase/tRNA<sub>CUA</sub> Paaren für zwei konzeptionell verschiedene *in vitro* und *in vivo* Anwendungen einsetzen: (i) für die positionsspezifische Markierung von Proteinen über eine neue lichtinduzierte bioorthogonale Reaktion und (ii) für eine Proximitäts-vermittelte kovalente Stabilisierung und Charakterisierung niedrig-affiner Proteinkomplexe.

Die Entwicklung von schnellen, chemoselektiven bioorthogonalen Reaktionen mit hohen Ausbeuten zwischen UAAs und extern hinzugefügten chemischen Sonden wird allmählich ein unentbehrliches Werkzeug für die Markierung von einzelnen Proteinen und Proteomen. In den letzten Jahren konnte ein spezieller Fokus auf lichtinduzierten bioorthogonalen Reaktionen beobachtet werden, da diese eine elegante Reaktionskontrolle aufweisen mit außergewöhnlicher räumlicher und zeitlicher Kontrolle. In dieser Arbeit wollten wir uns die charakteristische Reaktivität von photochemisch-generierten *ortho*-Naphthoquinonmethiden (*o*NQMs) gegenüber Vinylethern und -sulfiden zu Nutze machen, um eine neue und sehr schnelle hetero-Diels-Alder Reaktion (mit Ratenkonstanten zweiter

---

Ordnung  $\sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ) für die positionsspezifische Markierung von Proteinen zu entwickeln. Wir haben dazu zwei Strategien für die Markierung entworfen und verfolgt. Während die Synthese von UAAs mit *o*NQM Vorläufern (*o*NQMPs) herausfordernd war und der Einbau der entsprechenden UAAs in Proteinen gerichtete Evolution erfordert, konnten wir Vinylether- und -sulfid-tragende UAAs effektiv in sfGFP einbringen, sowohl in *E. coli* und Säugetierzellen. Die Markierungsversuche an aufgereinigten Proteinen zeigten jedoch ungewünschte Reaktivität der *o*NQM gegenüber nukleophilen Aminosäureseitenketten. Unglücklicherweise waren alle Bemühungen dieses Problem zu lösen nicht erfolgreich, weswegen die Reaktion zur Markierung nicht weiterverfolgt wurde. Weiterhin sind in dieser Thesis mehrere alternative Reaktionen für Vinylether/sulfid-beinhaltende UAAs beschrieben, insbesondere die durch sichtbares Licht vermittelte [4+2] Cycloaddition mit 9,10-Phenanthrenquinonen, welche jedoch dieselben Nebenreaktionen zeigt wie die Reaktion mit *o*NQMs.

Die Bildung von niedrig-affinen und kurzlebigen Protein-Protein-Interaktionen ist unabdingbar in der Signalweiterleitung und regulatorischen Wegen in der Zelle. Aufgrund ihrer dynamischen Art ist die Untersuchung der hochaufgelösten Struktur niedrig-affiner Proteinkomplexe jedoch schwierig, da hierfür die Bildung stabiler Komplexe notwendig ist. Aus diesem Grund wollten wir niedrig-affine Proteinkomplexe in ihrer nativen Struktur über eine eingebaute, flexible, Bromoalkyl-basierte UAA verknüpfen. Diese ist unter physiologischen Bedingungen inert, kann jedoch mit nukleophilen Aminosäuren wie Cysteinen bei erhöhter lokaler Konzentration durch Proximität reagieren. Dies hat uns ermöglicht, Positionen, welche durch Bildung des Proteinkomplexes in eine gewisse Nähe zueinander gebracht wurden, *in vitro*, sowie in lebenden Bakterien und Säugetierzellen kovalent zu verknüpfen. Weiterhin konnten wir die Struktur des kurzlebigen GDP-gebundenen Komplexes zwischen der kleinen humanen GTPase Rab1b und der GEF Domäne ihres Nukleotidaustauschfaktors DrrA aus *Legionella pneumophila* chemisch stabilisieren und die Kristallstruktur davon bestimmen. Derzeit versuchen wir diesen Ansatz auf UAAs auszuweiten, die ein zusätzliches, spaltbares MS/MS Motiv beinhalten, welche nicht nur dabei helfen würden unbekannte Protein-Interaktionen einzufangen, sondern auch deren Charakterisierung erleichtern würden.

---



## § 1. INTRODUCTION

The fundamentals of every living system lie on the shoulders of cells. Though tiny in size (~ 1-100  $\mu\text{m}$ , depending on their type), these “working units of life” are run by complex and sophisticated interactions of their numerous constituents, such as nucleic acids, proteins, small molecules and ions. Over the course of past decades, our understanding of the molecular basis of life has increased tremendously by utilizing a vast plethora of biochemical and biophysical techniques, ranging from X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy to whole-genome sequencing. Nevertheless, the available toolbox is often limited to *in vitro* studies on confined biomolecules, overlooking their complex and dynamic behaviour in the living cell. Finding new ways to investigate biomolecular interactions *in vivo* as they proceed in space and time is, thus, of special importance in modern molecular biology.

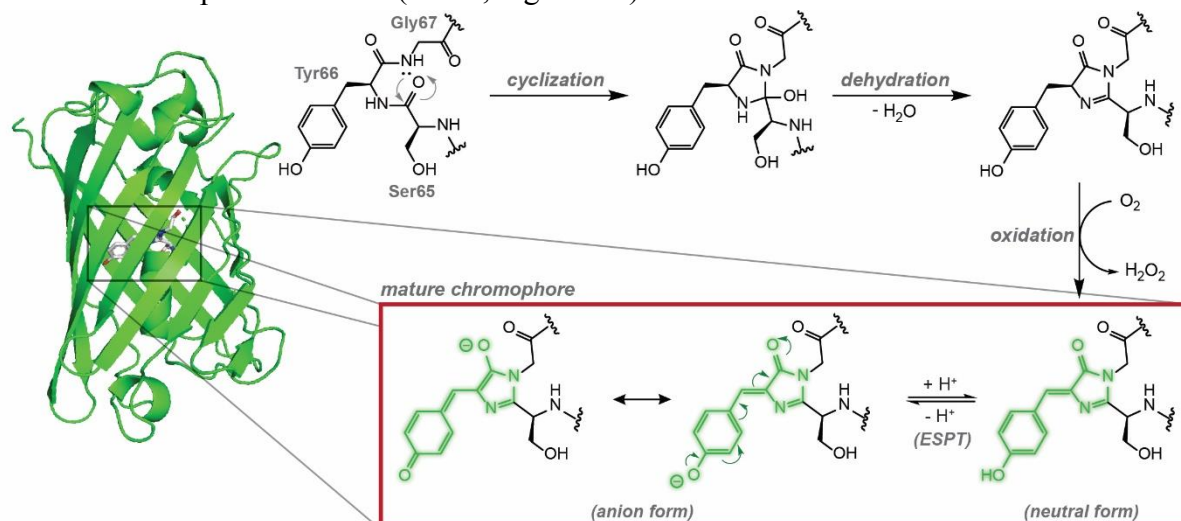
This thesis is comprised of two major parts; the first concentrates on efforts to develop a novel light-induced bioorthogonal reaction intended for site-specific labelling of proteins, both *in vitro* and *in vivo*. The second part details the development of a crosslinking tool used for stabilising low-affinity protein-protein interactions in living cells. Both of these topics make use of an elegant approach that emerged out of interdisciplinary fields of synthetic and chemical biology – the genetic code expansion – which allows site-specific incorporation of unnatural amino acids (UAAs) to modulate and study proteins of interest.

### 1.1. Visualising proteins inside living cells: An overview

#### 1.1.1. ... And then there was GFP

The year 1962 saw The Beatles releasing their first hit, Spider-Man making his first appearance in comic books and Marilyn Monroe singing “Happy Birthday” to the US president John F. Kennedy, months before her sudden passing. Remarkably, this same year at the Princeton University, Osamu Shimomura, intrigued by the bioluminescence of the crystal jellyfish *Aequorea victoria*, isolated a protein which would decades later become a revolutionary and indispensable imaging tool – the green fluorescent protein (GFP).<sup>[1]</sup> Consisting of 238 amino acid residues (26.9 kDa), GFP adapts a beta barrel (also referred to as “beta-can”) structure composed of eleven  $\beta$ -strands with a central alpha helix running along its axis. When exposed to light in the blue to ultraviolet range, GFP exhibits a bright green fluorescence ( $\lambda_{\text{em}} = 509 \text{ nm}$ ) caused by the posttranslationally formed 4-(*p*-hydroxybenzylidene)imidazolidine-5-one (HBI) chromophore (Figure 1).<sup>[2]</sup> Covalently bound to the interior helix, the HBI chromophore forms spontaneously from the tripeptide Ser65-Tyr66-Gly67 during maturation of the GFP scaffold. Mechanistically speaking, a nucleophilic attack of the Gly67 amide nitrogen atom on the carbonyl group of Ser65 and subsequent dehydration first create a heterocyclic imidazolone. The rate-limiting dehydrogenation of the  $\alpha$ - $\beta$  bond of Tyr66 by molecular oxygen, finally leads to conjugation of the aromatic Tyr rest with the imidazolone.<sup>[3]</sup> Depending on the hydrogen-bond network around the Tyr66, the HBI

chromophore can either be protonated (neutral form) or deprotonated (anionic form) with both forms exhibiting different spectral properties ( $\lambda_{\text{abs}}(\text{anionic}) = 475 \text{ nm}$ ;  $\lambda_{\text{abs}}(\text{neutral}) = 395 \text{ nm}$ ). Upon excitation, however, the neutral form undergoes rapid conversion into the anionic form *via* excited state proton transfer (ESPT, Figure 1.1).<sup>[4]</sup>



**Figure 1.1.** The chromophore formation of green fluorescent protein (GFP, PDB: 4KW4).

Even though the chemical and optical nature of wild-type GFP had been described by the end of 1970s<sup>[2a]</sup>, the realization of its utility began in the early 1990s with the cloning of the GFP gene<sup>[5]</sup> and its recombinant expression in *Escherichia coli* and *Caenorhabditis elegans*.<sup>[6]</sup> The latter was achieved by the lab of Martin Chalfie, who realized that GFP could be expressed and used as a genetically encoded fluorescence marker in almost every organism, due to the spontaneous formation of the HBI chromophore. Shortly after, the report of the 3D structure of GFP provided crucial information on the chromophore formation and the interactions between the surrounding amino acid residues.<sup>[2b, 2c]</sup> This led to the seminal work of Roger Tsien and his colleagues who engineered a large variety of GFP and GFP-like variants with improved spectral properties such as increased photostability and fluorescence (with diminished absorbance at 395 nm) as well as different colour such as blue (BFP), cyan (CFP) and yellow (YFP).<sup>[4]</sup>

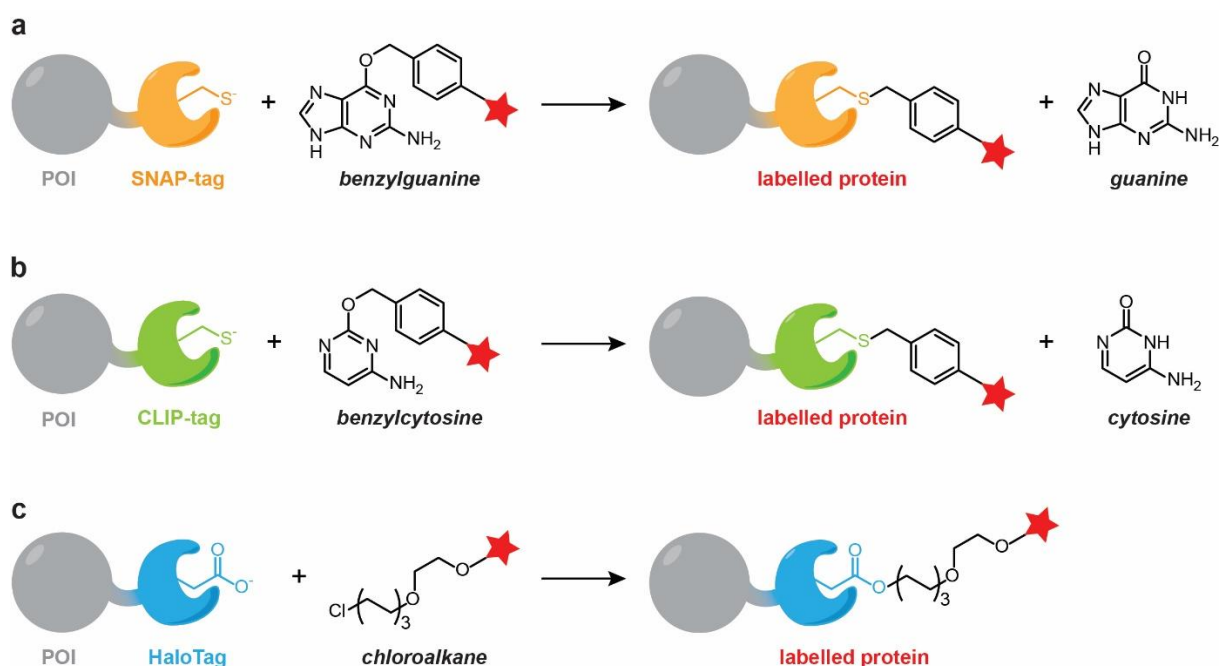
Today, the development of GFP and its various genetically encoded variants has had a tremendous impact on the understanding of cellular processes and protein transport.<sup>[7]</sup> When fused to the N or C terminus of the protein of interest, GFP can act as a reporter of protein subcellular localization. Furthermore, employment of fluorescent protein fusions has given insight into protein trafficking and interaction pathways, and at the same time has facilitated their purification for molecular analysis. For the discovery and development of GFP technology and its substantial impact on life sciences, in 2008, O. Shimomura, M. Chalfie and R. Tsien were acknowledged with the Nobel Prize in Chemistry.<sup>[8]</sup>

### 1.1.2. Genetic tags for labelling

The usage of protein fusions for labelling purposes, however, is not always ideal. Introduction of an additional sequence into the protein of interest may perturb its structure and hinder its function. The relatively large size of all GFP variants limits their fusion to solely N and C terminus of proteins, hindering this way their placement at any position in the protein of interest. Furthermore, their photophysical properties are not always as good as those of organic dyes (e.g. lower fluorescence quantum yields or occurrence of photobleaching upon denaturation), making them generally not suitable for single-molecule spectroscopy. Therefore, several alternative methods for labelling proteins have been developed. Among them, self-labelling enzyme-based fusion tags, which catalyse the covalent attachment of an exogenously added synthetic substrate (i.e. an organic fluorophore), should be noted.<sup>[9]</sup> As opposed to GFP and other fluorescent proteins, self-labelling enzymes only become fluorescent when the cells expressing them are exposed to the added fluorophore (Figure 1.2). This enables a spatiotemporal control of the labelling, as well as an opportunity to sequentially label proteins with different dyes. Human *O*<sup>6</sup>-alkylguanine-DNA-alkyltransferase (SNAP-tag) was the first self-labelling enzyme used to covalently attach a fluorophore onto a protein of interest *in vivo*.<sup>[9a]</sup> This 20 kDa DNA repair protein was engineered using directed evolution to react specifically with *O*<sup>6</sup>-alkyl- and *O*<sup>6</sup>-benzylguanine derivatives, through a cysteine residue in its active site (Figure 1.2a). Moreover, another variant of this enzyme has been further evolved to react specifically with *O*<sup>2</sup>-benzylcytosine derivatives, instead of *O*<sup>6</sup>-benzylguanine (CLIP-tag, Figure 1.2b).<sup>[9b]</sup> SNAP- and CLIP-tag are orthogonal to each other and can therefore be used to label two proteins with two different fluorophores within the same cell. Similarly, a 33 kDa bacterial enzyme haloalkane dehalogenase has been developed as a self-labelling fusion tag (HaloTag, Figure 1.2c).<sup>[9c]</sup> This hydrolase reacts with substrates bearing primary alkyl halides, during which a stable ester bond is formed between the substrate and an Asp residue in the active site. A conserved His residue in the wild-type enzyme, which in fact hydrolyses the alkyl-enzyme intermediate and thus regenerates the Asp nucleophile, has been mutated out of the sequence in order to preserve the stable covalent alkyl-enzyme adduct.

SNAP/CLIP-tags and HaloTag have been used extensively to investigate protein localization and trafficking in living cells.<sup>[9b, 10]</sup> The benefits of using these self-labelling enzyme tags include the fact that the attachment of the functionalized substrate is highly specific and irreversible. Importantly, the tag can be genetically encoded in both prokaryotes and eukaryotes, while the synthetic substrates can be coupled to different fluorescent dyes and affinity handles. The reactions occur rapidly at physiological conditions in living cells, but can also be extended to chemically fixed cells, if the cells are labelled before fixation. This is especially important when using HaloTags, where the labelling is inhibited by both fixation and detergents. Even though they provide higher labelling specificity, self-labelling enzyme tags exhibit similar problems as fluorescent proteins, especially regarding their size. In order to tackle the problems that arise from the molecular size of these tags, two further labelling approaches have been developed.

---

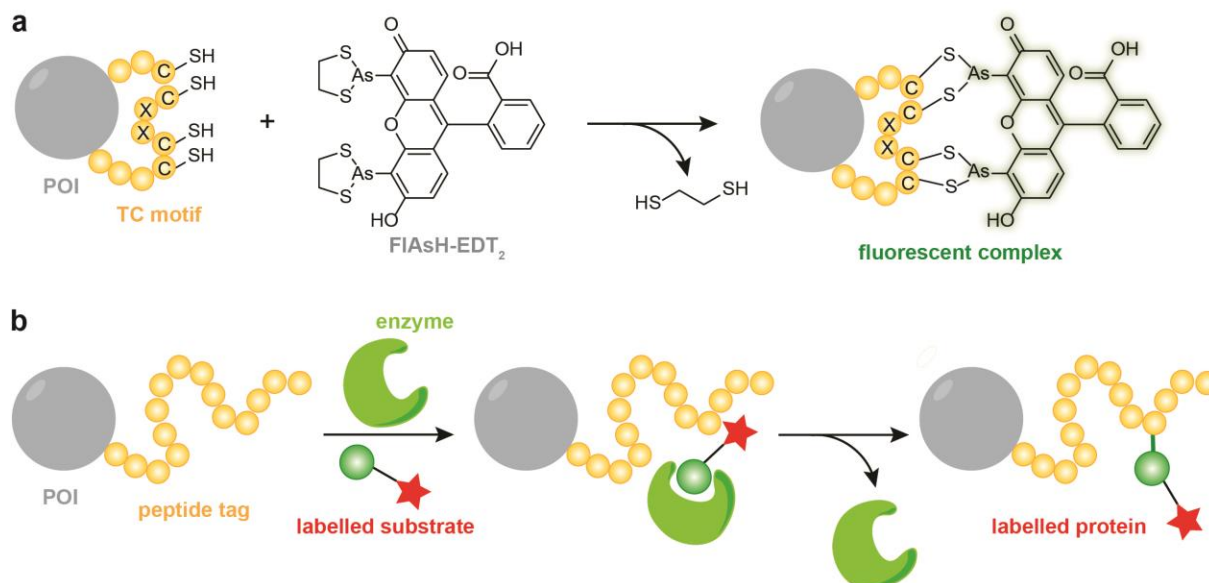


**Figure 1.2.** Self-labelling enzymes. (a) The SNAP-tag specifically reacts with externally added small molecules bearing a  $O^6$ -benzylguanine moiety. (b) CLIP-tag specifically reacts with  $O^2$ -benzylcytosine derivatives. (c) HaloTag undergoes nucleophilic displacement reactions with substrates bearing primary alkyl halides.

The first approach has its roots in the coordination chemistry of arsenic(III). Developed by Roger Tsien and his colleagues, cell-permeable biarsenical dyes bind with high affinity to a genetically encodable tetracysteine (TC) motif (-Cys-Cys-X-X-Cys-Cys-, where X refers to any amino acid other than cysteine).<sup>[11]</sup> The commonly used dye is the green-emitting FAsH-EDT<sub>2</sub> (fluorescein arsenical hairpin binder-ethanedithiol) containing fluorescein in its core with two 1,3,2-dithioarsolane moieties. FAsH-EDT<sub>2</sub> is essentially non-fluorescent and only becomes fluorescent upon binding to the TC motive, with absorbance and emission maxima at 508 nm and 528 nm respectively (Figure 1.3a). An optimised recognition motif (-Cys-Cys-Pro-Gly-Cys-Cys-) provides even higher affinity binding with dissociation constant of 10 pM, enabling quantitative labelling of the protein of interest.<sup>[12]</sup> While several cell-permeable dyes with different spectral properties are known,<sup>[13]</sup> the usage of biarsenical dyes is not without its problems.<sup>[14]</sup> Apart from overall toxicity of arsenic compounds to living cells, background labelling due to non-specific binding of the dyes to proteins rich in cysteines remains a problem, as it limits detection of poorly expressed proteins. As with fluorescent proteins and self-labelling enzymes, binding of biarsenical dyes to target proteins can also have an effect on their function and localization.

The second approach utilizes enzymes that catalyse covalent attachment of a labelling probe to a specific peptide sequence in living cells (Figure 1.3b). For example, a bacterial lipoic acid ligase (LplA) has been engineered to efficiently attach substrates containing different functionalities (azides, alkynes, *trans*-cyclooctenes, reactive halides and fluorophores such as 7-hydroxycoumarin, amongst others) to a LplA acceptor peptide (LAP) tag consisting of 13 amino acids.<sup>[15]</sup> While both the ligase and the LAP-bearing protein have to be co-

expressed within the same cell to ensure specific labelling, the variety of different functionalities an engineered ligase can recognise allows their usage in combination with bioorthogonal chemistries (see Chapter 1.4). Still, the development of new labelling substrates requires the engineering of the LplA active site.



**Figure 1.3.** Self-labelling tags and enzyme-mediated labelling. (a) Biarsenical dyes such as FIAsh-EDT<sub>2</sub> recognise genetically encoded tetracysteine (TF) motifs, CCXXCC, and specifically react with them affording fluorescently labelled proteins. (b) Engineered enzymes such as liponic acid ligase specifically recognise peptide tags fused to the protein of interest and efficiently attach labelled substrates to it.

Considering the described limitations of genetic tagging technologies, the development of general methods to selectively and site-specifically label proteins remains a topic of extensive research. An ideal method for protein labelling would encompass several important features. First, such a method would allow facile placement of a wide range of probes at any position in any protein expressed in live cells. Furthermore, the labelling would be quantitative, rapid and highly specific for an arbitrarily defined site in a protein. No toxic reagents would be employed and no toxic by-products generated with such a method and, importantly, it would display “turn-on” fluorescence as a mean of site-specific protein labelling detection with minimal background labelling. Considering the aforementioned features, site-specific incorporation of unnatural amino acids bearing bioorthogonal reactive groups *via* genetic code expansion allows chemoselective labelling of proteins at defined sites with essentially any probe.<sup>[16]</sup>

## 1.2. “The stuff that proteins are made of”

Proteins are commonly assembled from 20 natural (or proteinogenic) amino acids using information encoded in genes. More precisely, protein-coding genes composed of triplet codons are transcribed into messenger RNA (mRNA) which serves as a template for protein synthesis. Each of these nucleotide triplets encodes a specific amino acid. Translation of genetic information contained in the mRNA is accomplished by the ribosome, a multi-subunit cellular machine made out of ribosomal RNA (rRNA) as well as a variety of ribosomal

proteins, which complements each codon with its matching aminoacyl-tRNA. At the basis of this pairing lie specific hydrogen bond patterns formed between the codon and its complementary anticodon located on the tRNA. Aminoacyl-tRNAs are composed of specific amino acids covalently attached to 3' ends of their cognate tRNAs. This esterification reaction, also known as “tRNA charging”, is catalysed by aminoacyl-tRNA synthetases (aaRS), highly specific enzymes for both their cognate amino acid and tRNA.<sup>[17]</sup> The high specificity of aaRS must be noted since it maintains the fidelity of protein translation. Some aaRS have evolved different editing mechanisms in order to prevent proteome-wide statistical mutations that would arise from erroneous mischarging of the tRNA and which would have catastrophic consequences for the cell.<sup>[18]</sup> Mistranslation typically occurs at very low levels in most organisms (1 error per  $10^3$ - $10^4$  translated codons<sup>[19]</sup>) although, under certain circumstances (e.g. oxidative stress), higher tolerance towards errors during translation is beneficial for some species.<sup>[20]</sup>

Prokaryotic translation is initiated by binding of the small ribosomal subunit 30S to the purine-rich Shine-Dalgarno sequence located upstream of the initiation codon AUG on the mRNA. After formation of the initiation complex, formylmethionyl-tRNA<sup>fMet</sup><sup>1</sup> binds to the AUG codon, accompanied by binding of the larger ribosomal subunit 50S to form the mature ribosome 70S which contains three tRNA-binding sites: the aminoacyl site (A-site), the peptidyl site (P-site) and the exit site (E-site). After the initiator tRNA<sup>fMet</sup> is positioned in the P-site, the second aminoacyl-tRNA, brought to the ribosome by elongation factor EF-Tu, binds to the A-site. Subsequent nucleophilic substitution between the amine group of the aminoacyl-tRNA (A-site) and the carbonyl group of the ester fMet-tRNA<sup>fMet</sup> (P-site) results in a dipeptide bound to the tRNA in the A-site. After translocation of the now uncharged tRNA<sup>fMet</sup> from the P-site to the E-site and the dipeptidyl-tRNA from the A-site to the P-site, the next aminoacyl-tRNA can bind to the A-site while the tRNA<sup>fMet</sup> dissociates from the ribosome.<sup>[21]</sup> The elongation of the polypeptide chain essentially follows the same concept, with the ribosome working at the speed of 12-21 amino acids per second in *E. coli*.<sup>[22]</sup> Compared to bacterial DNA replication rates of approx. 600 base-pairs per second,<sup>[23]</sup> protein translation is quite sluggish.

With exception of selenocysteine and pyrrolysine, all of the proteinogenic amino acids are coded for by a codon. Nevertheless, there are 64 possible codons which makes the genetic code redundant or degenerate. Degeneracy is exhibited as multiplicity of codon combinations that code for a specific amino acid (e.g. GCU, GCC, GCA and GCG all code for alanine). Three of the 64 codons do not code for an amino acid; instead, the protein translation, as described above, continues until one of the three codons is recognized by a release factor

---

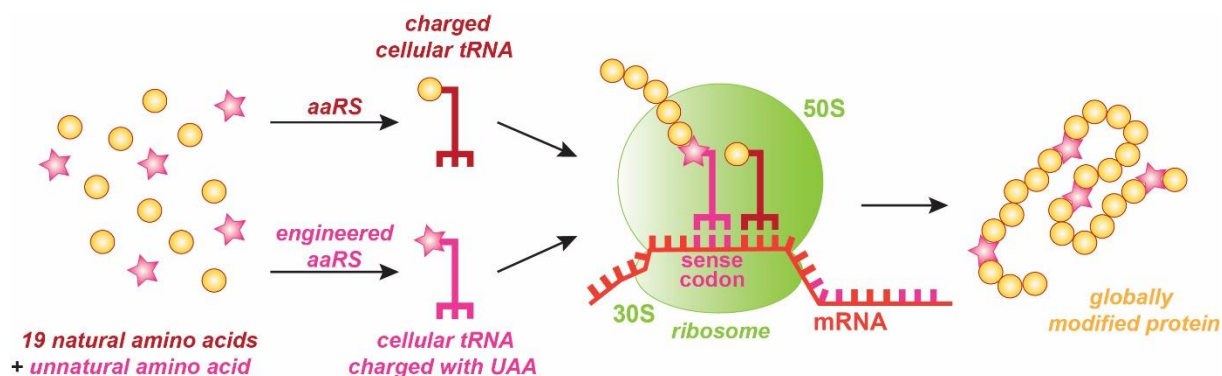
<sup>1</sup> In bacteria, and sometimes even in mitochondria and chloroplasts, *N*-formylmethionine (fMet) is the starting amino acid of the nascent polypeptide. It is charged onto a specific initiator tRNA<sup>fMet</sup> in form of Met, which then gets *N*-formylated by methionyl-tRNA formyltransferase. Even though both tRNA<sup>fMet</sup> and tRNA<sup>Met</sup> have an anticodon specific towards AUG codon, fMet is only incorporated at the beginning, while Met is incorporated at every other position of the protein sequence. *N*-formylation is later on removed from the majority of mature proteins by a specific hydrolase.

---

which causes disassociation of the ribosomal subunits and termination of protein synthesis. These stop codons are the amber (UAG), ochre (UAA) and opal (UGA) codons.

Considering the limited diversity of functionalities within the 20 proteinogenic amino acids, decoration of proteins with unique biochemical handles in form of unnatural (or non-canonical) amino acids (UAAs) has become a long-lasting desire within the scientific community. Various early methods that enable introduction of amino acids bearing novel functionalities into proteins include solid-phase synthesis<sup>[24]</sup>, chemical or ribozyme-mediated aminoacylation of tRNAs employed in *in vitro* translation systems<sup>[25]</sup> and native chemical ligation.<sup>[26]</sup> Nonetheless, these methods are limited to *in vitro* applications and production of smaller proteins. On the contrary, incorporation of UAAs into proteins *in vivo* can be achieved through the use of the translational machinery of the cell in two ways: by residue-specific incorporation *via* selective pressure (Figure 1.4) or by expanding the genetic code for site-specific incorporation (Figure 1.5).<sup>[16a, 27]</sup>

In residue-specific incorporation methods, the UAA partially replaces one of the 20 natural amino acids by competing for its aaRS. Such reassignment of one of the codons is achieved in *E. coli* by expressing the proteins with the UAA in auxotrophic strains and growing them in media depleted with the respective natural amino acid. In this fashion, the selective pressure allows the incorporation of the UAA into positions that are normally occupied by the replaced natural amino acid, not only in the overexpressed protein of interest but proteome-wide as well. Naturally, the aaRS of the replaced natural amino acid must accept the UAA as its substrate. This has prominently been achieved with the methionine codon AUG and MetRS (Figure 1.4).<sup>[28]</sup> Since endogenous aaRS can only incorporate a finite amount of UAAs, introduction of mutant aaRS has expanded the range of substrates the endogenous aaRS did not recognize.



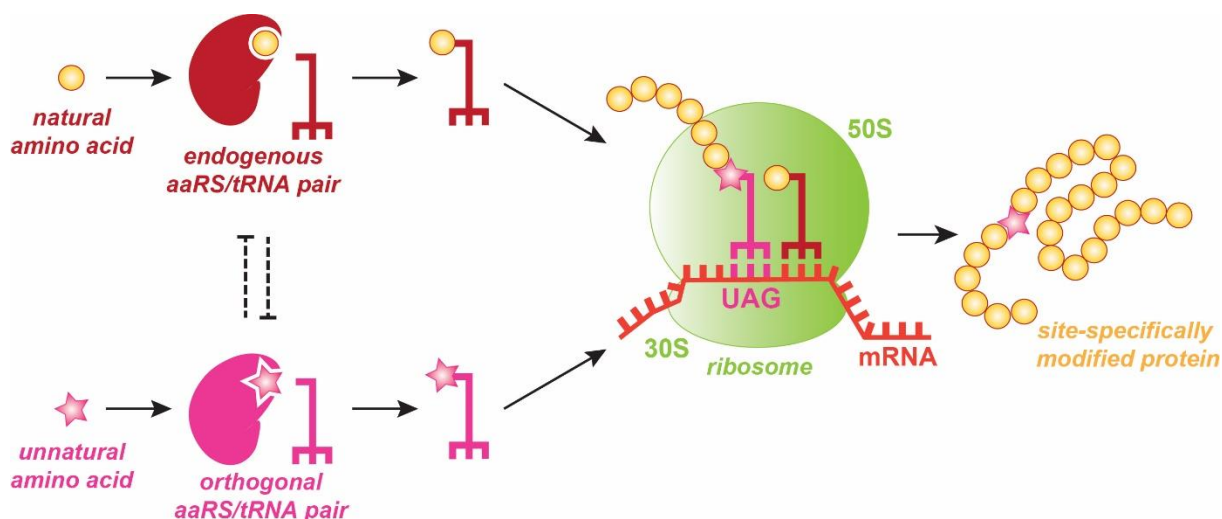
**Figure 1.4.** Principle of residue-specific incorporation of unnatural amino acids into proteins. One of the 20 natural amino acids (e.g. methionine) is depleted from the cell culture medium of an auxotrophic strain. The selective pressure allows the UAA to be recognised by endogenous MetRS which charges it onto cellular tRNA, Ultimately, the UAA is incorporated globally into the proteome in place of methionine.

As mentioned before, three of the 64 possible codons signal for termination of translation. Taking into account that termination only requires one stop codon, the remaining two could be reassigned to a new, unnatural amino acid.<sup>[29]</sup> Interestingly, the amber stop codon is the least used in *E. coli* (9 %) and rarely terminates essential genes.<sup>[30]</sup> Furthermore, certain species do not use the amber codon as a stop codon at all, but rather use it to introduce

an amino acid, such as the methanogenic archaeon *Methanocaldococcus jannaschii* (formerly *Methanococcus jannaschii*) which introduces a tyrosine at a UAG codon.<sup>[31]</sup> Pioneered by Peter Schultz and others, the reassignment of the amber codon (also known as amber suppression) is a powerful tool that has opened up multiple possibilities to study proteins through site-specific incorporation of UAAs, not only in bacteria<sup>[16a, 27c]</sup> but also in yeast (*Saccharomyces cerevisiae*),<sup>[32]</sup> mammalian cells,<sup>[33]</sup> plants<sup>[34]</sup> and certain animals as well.<sup>[35]</sup>

### 1.3. Expanding the genetic code: Suppression of the amber codon

Several key requirements need to be fulfilled in order to expand the genetic code. These include: (i) a blank codon that needs to be reassigned, (ii) a tRNA that can recognize this codon, (iii) an aaRS specific for that tRNA only, and finally (iv) the UAA of interest. To clarify, amber codon suppression relies on a complementary amber tRNA<sup>CUA</sup> that is specifically aminoacylated by an orthogonal aaRS, evolved to accept an UAA of choice but none of the 20 natural amino acids (Figure 1.5). The tRNA<sup>CUA</sup> must be a poor substrate for any of the endogenous aaRS and is thus able to direct the incorporation of the UAA in response to an amber codon on the mRNA during translation.



**Figure 1.5.** Site-specific incorporation of unnatural amino acids into proteins *via* genetic code expansion. An unnatural amino acid (pink star) is specifically recognised by an orthogonal aminoacyl-tRNA synthetase (aaRS) and charged exclusively onto an orthogonal amber suppressor tRNA (pink) and not onto endogenous host tRNAs (red). Likewise, the amber suppressor tRNA is not a substrate for the endogenous aaRSs. Consequently, the UAA is co-translationally incorporated into a protein of interest in response to a premature amber codon (UAG) introduced at a specific site in the gene of interest.

The first orthogonal aaRS/tRNA pair was discovered by importing the tyrosyl tRNA and synthetase from *M. jannaschii* (*Mj*) into *E. coli*.<sup>[31]</sup> The *Mj*tRNA<sup>CUA</sup> primarily differs from its *E. coli* counterpart in the nucleotide content of its acceptor stem, while the respective *Mj*TyrRS has structural features that allow minimal recognition of the anticodon of *Mj*tRNA<sup>CUA</sup>. Furthermore, *Mj*TyrRS doesn't possess any editing mechanisms to proofread its amino acid substrates. By mutating the amino acid binding pocket of *Mj*TyrRS, the aforementioned features enabled site-specific incorporation of *p*-methoxyphenylalanine into dihydrofolate reductase (DHFR).<sup>[31]</sup> Other commonly used aaRS/tRNA pairs include the

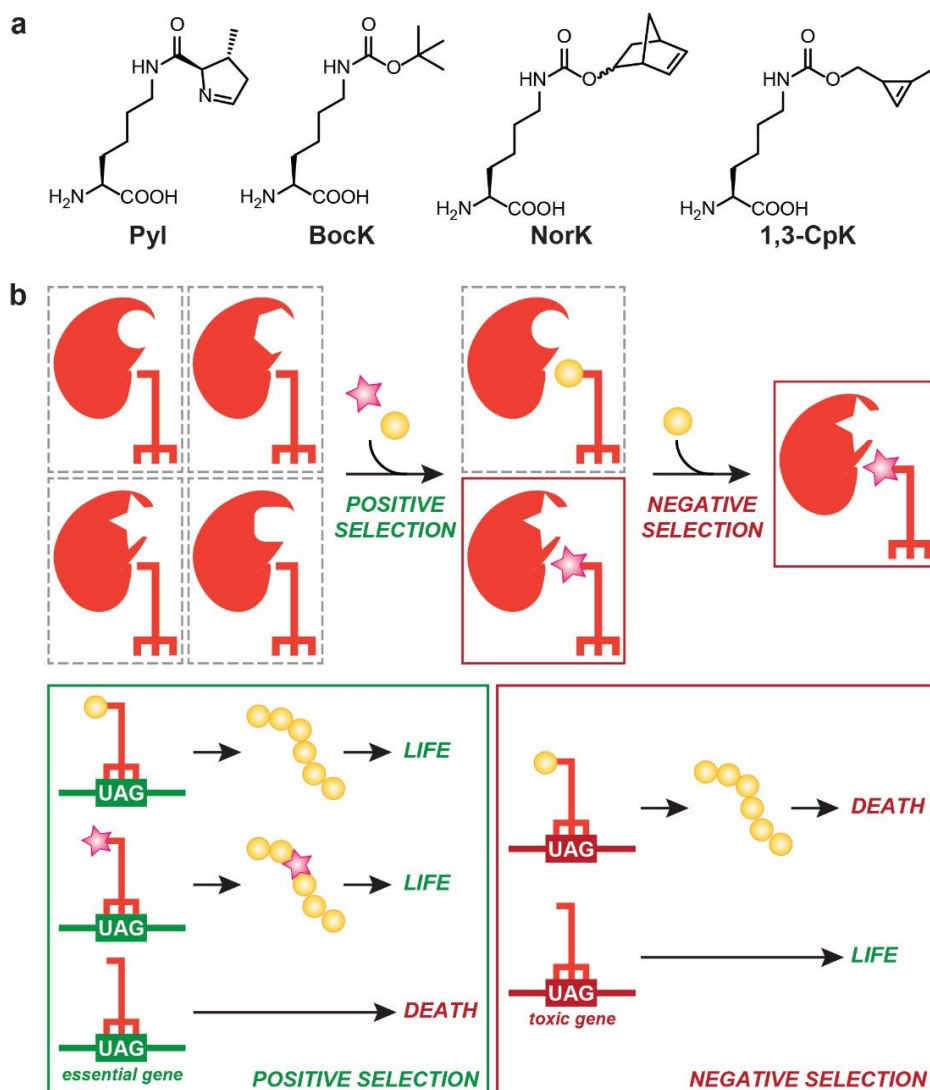


tyrosyl-tRNA synthetase (*EcTyrRS*)/tRNA<sup>CUA</sup> from *E. coli*, leucyl-tRNA synthetase (*EcLeuRS*)/tRNA<sup>CUA</sup> from *E. coli* and pyrrolysyl-tRNA synthetase (*PylRS*)/tRNA<sup>CUA</sup> from *Methanosarcinae*. The orthogonality of an aaRS/tRNA pair is reflected in their interaction with the translational machinery (i.e. endogenous aaRS/tRNA pairs) of the host organism. Due to inherent differences between different hosts, however, the orthogonality of an individual synthetase-tRNA pair in one organism does not necessarily guarantee orthogonality in another. For example, *MjTyrRS*/tRNA<sup>CUA</sup> pair is widely used for genetic code expansion in *E. coli*, nonetheless it is not orthogonal in eukaryotic cells. On the other hand, *EcTyrRS*/tRNA<sup>CUA</sup> as well as *EcLeuRS*/tRNA<sup>CUA</sup> pairs are orthogonal only in eukaryotic cells, while the *PylRS*/tRNA<sup>CUA</sup> pair is orthogonal in both hosts. The latter pair is considered particularly useful for incorporation of a variety of different UAAs as it normally only recognizes pyrrolysine (Pyl), the so-called 22<sup>nd</sup> proteinogenic amino acid, identified only in methanogenic archaea and a few microbes.

Wild-type *PylRS* from organisms like *Methanosarcina mazei* (*Mm*) and *Methanosarcina barkeri* (*Mb*) contains a C-terminal catalytic domain and an N-terminal domain variable in length, which was thought for a long time to be important for tRNA recognition and aminoacylation.<sup>[36]</sup> Nevertheless, a novel *PylRS*/tRNA<sup>CUA</sup> pair lacking the N-terminal domain was very recently discovered from *Methanomethylophilus alvus* (*Ma*).<sup>[37]</sup> The poor solubility of the N-terminal domain has so-far prevented efforts to crystalize the full-length *PylRS* from *Methanosarcinae*<sup>[38]</sup>, although crystal structures of the truncated C-terminal domains have been reported.<sup>[33a, 39]</sup> These structures elucidated seven amino acid residues in the active site of the *MmPylRS* which form a hydrophobic pocket for the binding of Pyl. Although this binding is comparable to other aaRSs, the hydrophobic interactions are non-specific as the *PylRS* is highly promiscuous towards variations of the side chains of its substrates (Figure 1.6a).<sup>[36]</sup> The wild-type enzyme for example readily incorporates UAAs that are, structurally speaking, completely different from Pyl such as *N*<sup>6</sup>-(*tert*-butoxycarbonyl)-L-lysine (BocK).<sup>[33a]</sup> The lack of an editing domain (that a lot of aaRSs contain in order to maintain the high fidelity of translation) also contributes to the unique promiscuity of *PylRS*. Another striking feature of *PylRS* is its low selectivity towards the tRNA anticodon. While the majority of aaRSs interact with anticodons in order to pair the tRNAs with the corresponding amino acids, the aminoacylation activity of *PylRS* was remarkably not abolished by mutating the anticodon of tRNA<sup>Pyl</sup>.<sup>[39b, 40]</sup> The exceptional features of wild-type *PylRS*/tRNA<sup>CUA</sup> pair allowed incorporation of a variety of UAAs into proteins for different purposes (Figure 1.6a). The pair's scope for new amino acids, however, can be further expanded by directed evolution (Figure 1.6b).<sup>[16a, 27c, 41]</sup>

Directed evolution is a two-step selection procedure that allows finding an active-site mutant of an orthogonal *PylRS* specific for an UAA of interest, without recognizing any of the 20 natural amino acids (Figure 1.6b). The procedure starts with the creation of a library that normally contains around 10<sup>9</sup> randomized active-site mutants of the orthogonal aaRS. The library is then transformed into *E. coli* and subjected to positive selection for aminoacylation activity in the presence of both natural amino acids and the particular UAA. This selection step singles out aaRS/tRNA pairs on their ability to suppress the amber codon

introduced into an essential gene. The most commonly used positive selection marker is chloramphenicol acetyltransferase; incorporation of either the UAA or one of the natural amino acids in response to the stop codon leads to survival of cells grown on chloramphenicol, whereas termination of translation leads to cell death. To eliminate the false positives obtained in the first selection step, the survived synthetase variants are screened for their ability to suppress the amber codon introduced into a toxic gene in the absence of the UAA. The negative selection marker of choice is the ribonuclease barnase (usually carrying two amber codons to tighten the selection pressure) which is expressed if the synthetase variants accept natural amino acids as substrates.<sup>[41b]</sup> Ideally, this leaves only the aaRS mutants specific for the desired UAA (Figure 1,6b). The two-step selection process is, however, repeated several times to afford the synthetase mutants with the best specificity towards the desired UAA.



**Figure 1.6.** Evolving aaRS-tRNA pairs specific for a new unnatural amino acid. (a) The wild-type PylRS is highly promiscuous and, together with its orthogonal tRNA<sup>CUA</sup> pair can incorporate several different UAAs, some of which are displayed in this Figure. (b) The scope of the PylRS/tRNA<sup>CUA</sup> pair can be expanded by directed evolution approach. Consisting of positive and negative selection steps, directed evolution is employed for finding new orthogonal aaRS-tRNA<sup>CUA</sup> pairs specific for an unnatural amino acid of interest (pink star).

### 1.3.1. Applications and efforts within the field

The utility of evolved PylRS mutants can be illustrated best with more than 200 different UAAs that have so far been incorporated site-specifically into proteins. The introduction of novel functionalities into proteins has opened up multiple possibilities to investigate and elucidate diverse biological processes with impressive precision as they proceed in space and time. The applications of UAAs have been highlighted in several excellent reviews,<sup>[16]</sup> nevertheless I would like to single out a few. UAAs modified with biophysical probes such as spin-labels,<sup>[42]</sup> fluorescent probes,<sup>[43]</sup> NMR-active isotopes (<sup>19</sup>F, <sup>15</sup>N, <sup>13</sup>C)<sup>[44]</sup> and IR probes<sup>[45]</sup> provides new insights into protein structure and chemical environment as well as dynamics of protein activation. Posttranslational modifications including methylation,<sup>[46]</sup> acetylation,<sup>[47]</sup> phosphorylation<sup>[48]</sup> and ubiquitylation<sup>[49]</sup> have been successfully introduced into recombinant proteins in order to study their function and impact on other proteins. Photocaged amino acids are used for manipulating protein functions by light as well as for elucidating signalling processes in the cell.<sup>[33b, 50]</sup> With the development of new bioorthogonal chemistries, incorporation of UAAs bearing bioorthogonal handles that are inert under physiological conditions but react selectively and rapidly with externally added chemical probes not only allow site-specific labelling of proteins *in vitro* and *in vivo* with reporters such as fluorophores and biotin, but covalent attachment of diverse biochemical and biophysical probes as well.<sup>[16b]</sup> Some of these bioorthogonal functionalities include azides, alkenes, alkynes, ketones, tetrazoles and tetrazines. Another way of introducing artificial covalent linkages into proteins but for slightly different purposes is the incorporation of UAAs containing photo- and chemical crosslinking moieties.<sup>[51]</sup> Photocrosslinking UAAs allow mapping of transient protein-protein interactions,<sup>[52]</sup> while introduction of chemical crosslinkers enhances protein stability and stabilises low-affinity protein complexes in their native environment.<sup>[53]</sup> The latter can be achieved through site-specific incorporation of electrophilic UAAs whose reactivity is triggered by the presence of nucleophilic residues in their proximity. The following chapters 1.4. and 1.5. examine the versatility and applications of bioorthogonal chemistries and proximity-triggered crosslinking approaches respectively.

Although exceptionally useful, genetic code expansion is not without its limitations and overcoming them is a subject of extensive research. Some of the technical issues include lower expression yields compared to the wild-type proteins, which highly depend on the suppression efficiency of the used aaRS mutant. Even though an UAA can in theory be introduced at any desired position in a protein, the suppression efficiencies are also position-dependent, as lower incorporation efficiencies have been observed when an amber codon is introduced at the beginning of a gene. This, however, is also related to the types of proteins of interest, with stable and easily overexpressed proteins (such as GFP or ubiquitin) typically showing good UAA incorporation. In addition, the extent of incorporation is also affected by competition of endogenous release factors towards the stop codon; binding of the release factor terminates the translation affording truncated variants of the proteins of interest. Therefore, affinity tags such as His- or Strep-tag used for enrichment of recombinant proteins are typically introduced at the C-terminus of the protein of interest. This way only full-length proteins that incorporate UAAs bind to, for example, affinity columns while the truncated

---

variants are eluted out. Several *E. coli* strains where the release factor RF-1 was knocked out showed to circumvent the competition problem.<sup>[54]</sup> RF-1 recognises UAG and UAA codons<sup>[55]</sup> and its removal requires genome-wide replacement of all the endogenous UAG codons (321 known positions in *E. coli* strain MG1655) with UAA, allowing RF-2 (which recognizes UAA and UGA codons<sup>[55]</sup>) to overtake its function. Nevertheless, the abolishment of RF-1 comes with a price, as the knock-out strains grow at a much slower rate.

It also should be noted that directed evolution of new specific aaRS is quite laborious and can take months before finding a positive hit, if any, as several cycles of positive and negative selection need to be performed and different mutant libraries tested. Properties of the UAAs such as their structure, solubility, cell-permeability and potential toxicity are further aspects impeding the discovery of new synthetases.

In spite of its relevance, incorporation of a single UAA into a protein is often not sufficient for certain biological applications (e.g. elucidation of protein dynamics). Considering that the amber codon can only direct the incorporation of one kind of amino acid, parallel incorporation of multiple different UAAs into the same protein requires different aaRS/tRNA pairs as well as unique codons for each amino acid. Importantly, the used pairs must be mutually orthogonal (each pair recognizes its respective codon and amino acid). The fact that PylRS does not recognize the CUA anticodon of tRNA<sup>Pyl</sup> was conveniently used to reassign the synthetase to other codons such as ochre (UAA) and opal (UGA) simply by mutating the tRNA<sup>Pyl</sup> anticodon.<sup>[56]</sup> The same pair was repurposed to rare sense codons such as AGG (one of the six codons for arginine) as well.<sup>[57]</sup> Taking into account the mutually orthogonality of *MjTyrRS* and *MmPylRS*, the pairs have been used in *E. coli* for simultaneous incorporation of two different UAAs in response to amber and ochre (UAA)<sup>[58]</sup> or even a quadruplet codon (AGGA).<sup>[59]</sup> The quadruplet codons were decoded by means of engineered orthogonal ribosomes and tRNAs with extended anticodons.<sup>[59-60]</sup> With respect to site-specific dual incorporation of UAAs in mammalian cells, the strategies developed hitherto struggle with significant problems. Although a *EcTyr*<sup>UAG</sup>/*Pyl*<sup>UAA</sup> pair was initially used for dual amber and ochre suppression<sup>[61]</sup>, the codon combination suffered from misincorporation due to competition of the UAG suppressor tRNA for the UAA codon.<sup>[62]</sup> Better alternatives were developed in forms of *EcTyr*<sup>UGA</sup>/*Pyl*<sup>UAG</sup> and *EcLeu*<sup>UAG</sup>/*Pyl*<sup>UGA</sup> pairs, with the former pair allowing site-specific dual incorporation of azide-/alkyne- and cyclopropene-/azide-bearing UAAs.<sup>[62]</sup> These suppression systems, however, exhibit low efficiencies due to weak activities of UGA suppressor tRNAs. Substantial lack of diversity in the UAA repertoire that the *EcTyrRS* and *EcLeuRS* can incorporate is another hindering factor. Remarkably, several proof-of-principle experiments were able to show the mutual orthogonality of the recently discovered PylRS/tRNA<sup>CUA</sup> pair from *M. alvus* and the *MmPylRS*/tRNA<sup>UGA</sup> pair in both *E. coli* and mammalian cells.<sup>[37]</sup> The substrate scope of the novel synthetase was, furthermore, broadened by transplanting active site mutations from *MmPylRS*, allowing double incorporation of distinct UAAs in response to amber and opal stop codons. Being the first mutually orthogonal pairs in both bacteria and eukaryotic systems, *MmPylRS*/tRNA and *MaPylRS*/tRNA open up exciting opportunities for simultaneous incorporation of multiple (thus far unavailable) functionalities into proteins in mammalian cells.

---

## 1.4. Bioorthogonal reactions in chemical biology

While labelling reactions between unnatural amino acids and the externally added chemical probe offer a unique opportunity for imaging individual proteins and labelling proteomes, their success highly depends on properties of the chemical reactions used. In every bioorthogonal reaction, naturally, both the chemical functionality and the external probe must react specifically with one another under physiological conditions (i.e. aqueous medium, moderate temperatures, neutral pH, high salt concentrations) while simultaneously avoiding nonspecific reactions with a broad range of functionalities present within complex biological systems. However, the chemoselectivity of these reactions is only one of the necessary requirements that make the development and implementation of novel bioorthogonal reactions quite challenging. The reactions need to be robust, with high yields and fast rates at relatively low concentrations of reactants. The bioorthogonal moieties must also be kinetically, thermodynamically and metabolically stable in aqueous environment and react readily to give non-toxic products and only inoffensive by-products.<sup>[16b, 63]</sup>

Despite the numerous constraints and challenges, a number of chemoselective reactions have been developed that show good biocompatibility and selectivity in living systems (Sections 1.4.1 to 1.4.4). These reactions vary widely in their rates and selectivity and some are limited only to *in vitro* labelling of biomolecules. Most bioorthogonal reactions follow second-order kinetics and their rates depend directly on the intrinsic second-order rate constant ( $k_2$ ) as well as concentrations of both reaction partners. Therefore, rapid reactions that exhibit high  $k_2$  are favourable for labelling of low abundance proteins or for labelling on a very short time scale. Low abundance proteins can be labelled by using a large excess of labelling reagents, however, such strategies are largely burdened by low-solubility, off-target reactions (either covalent or non-covalent) and toxicity of labelling reagents at high concentrations. Second-order rate constants for bioorthogonal reactions, for which one of the partners can be genetically encoded, varied until recently between  $10^{-4}$ - $10 \text{ M}^{-1} \text{ s}^{-1}$ . These sluggish rates are 8 to 4 orders of magnitude slower than enzymatic labelling approaches such as SNAP- or CLIP-tag, making bioorthogonal reactions in many cases sufficient for labelling proteins *in vitro* or on the cell surface. Considering, however, that biological processes can be very rapid, fast and quantitative reactions with high second-order rate constants are needed to facilitate protein labelling both *in vitro* and in living cells.

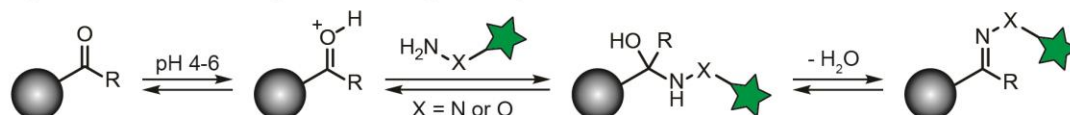
### 1.4.1. Hydrazone/oxime ligations

The earliest attempts to find suitable biorthogonal reagents relied on the reactivity of aldehydes and ketones which form hydrazones and oximes in reactions with hydrazines and alkoxyamines respectively.<sup>[64]</sup> These reactions, however, typically exhibit sluggish kinetics ( $k_2$  ranges from  $10^{-4}$  to  $10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ )<sup>[65]</sup> and require acidic conditions (optimal pH 4-6) in order to activate the carbonyl group of aldehydes and ketones (Figure 1.7a). Furthermore, both hydrazines and alkoxyamines are nucleophilic enough to react with a variety of other functionalities present inside the cell. Therefore, hydrazone/oxime ligation reactions as such are limited to *in vitro* and cell-surface labelling, where high concentrations of labelling

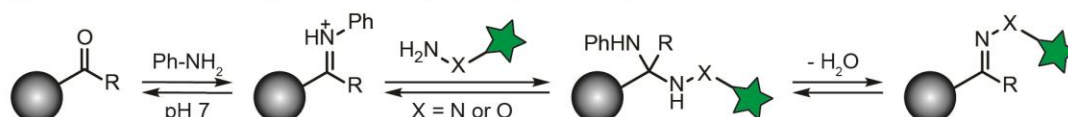
---

reagents are employed and optimal pH can easily be adjusted.<sup>[27c, 32, 65-66]</sup> Nevertheless, presence of aniline as organocatalyst has been shown to circumvent aforementioned problems to an extent (Figure 1.7b). Aniline forms a highly reactive iminium intermediate with the carbonyl group, which is readily trapped by nitrogen nucleophiles at neutral pH, boosting the rate constants up to  $10^1$ - $10^3$   $M^{-1} s^{-1}$ .<sup>[67]</sup> Aniline-catalysed oxime ligations were successfully applied for cell-surface labelling of glycoproteins as well as for live cell labelling of intracellular bacterial receptors.<sup>[68]</sup> Several other more efficient organocatalysts have been reported in recent literature as well.<sup>[69]</sup>

**a** Hydrazone/oxime ligation catalysed by acid



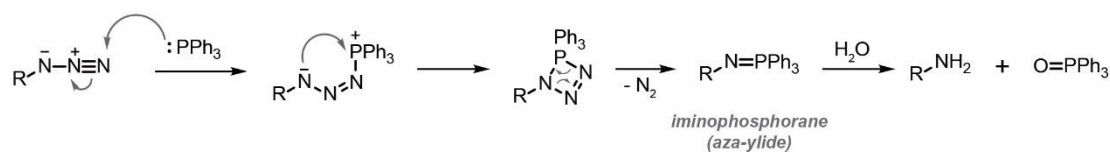
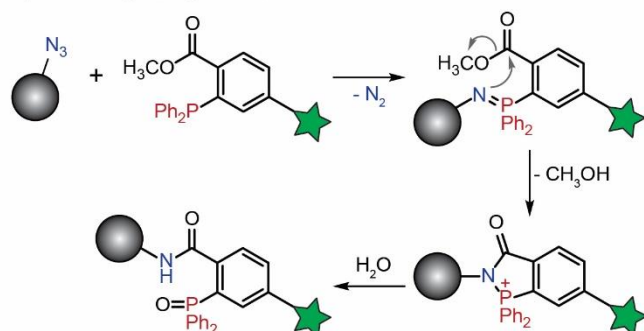
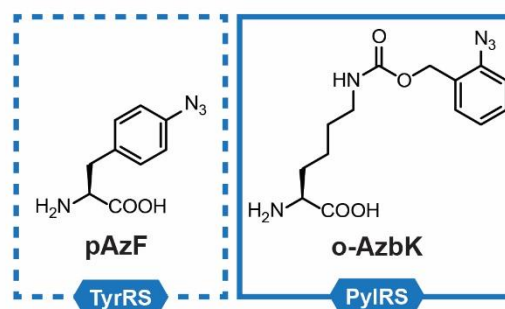
**b** Hydrazone/oxime ligation catalysed by aniline



**Figure 1.7.** Hydrazone and oxime ligations. (a) Aldehydes and ketones react with hydrazines and alkoxyamines under acidic conditions (pH 4-6) to afford hydrazones and oximes, respectively. (b) Aniline-catalysed hydrazone/oxime ligation at neutral pH.

### 1.4.2. Azides as bioorthogonal handles

In the last decade, a set of azide-, alkyne- and alkene-functionalized UAAs have been prepared (see Figures 1.8 and 1.9e). The azide group is absent in biomolecules and is stable at physiological conditions. Moreover, it is small in size which makes azide group incorporation into proteins, as well as other biomolecules, relatively straightforward.<sup>[28c, 70]</sup> One of the classic reactions involving organic azides is the Staudinger reaction with phosphines, in which azides are reduced to primary amines (the reaction is also known as Staudinger reduction, Figure 1.8a).<sup>[71]</sup> A reactive iminophosphorane intermediate is initially formed through nucleophilic addition of the phosphine at the terminal nitrogen atom of the organic azide moiety, setting free molecular nitrogen as the non-toxic by-product. In the presence of water, the aza-ylide then spontaneously hydrolyses to afford a primary amine and a phosphine oxide. The Staudinger reaction was further modified with the design of a phosphine reagent that has an ester group adjacent to the phosphorus atom.<sup>[72]</sup> This way, the aza-ylide reacts with an electrophilic carbonyl group to form a stable amide bond, making the reaction, now known as Staudinger ligation, suited for protein labelling (Figure 1.8b). However, slow kinetics ( $k_2 \sim 10^{-3} M^{-1} s^{-1}$ ) of Staudinger ligations necessitate the use of high concentrations of phosphine reagents and thus limit its imaging applications. Although azide-bearing UAAs (Figure 1.8c) are essentially stable, it has been observed that some of them can get reduced *in cellulo* to corresponding amines, diminishing the labelling efficiencies.<sup>[28c]</sup> Many phosphines can also get easily inactivated through oxidation by air.

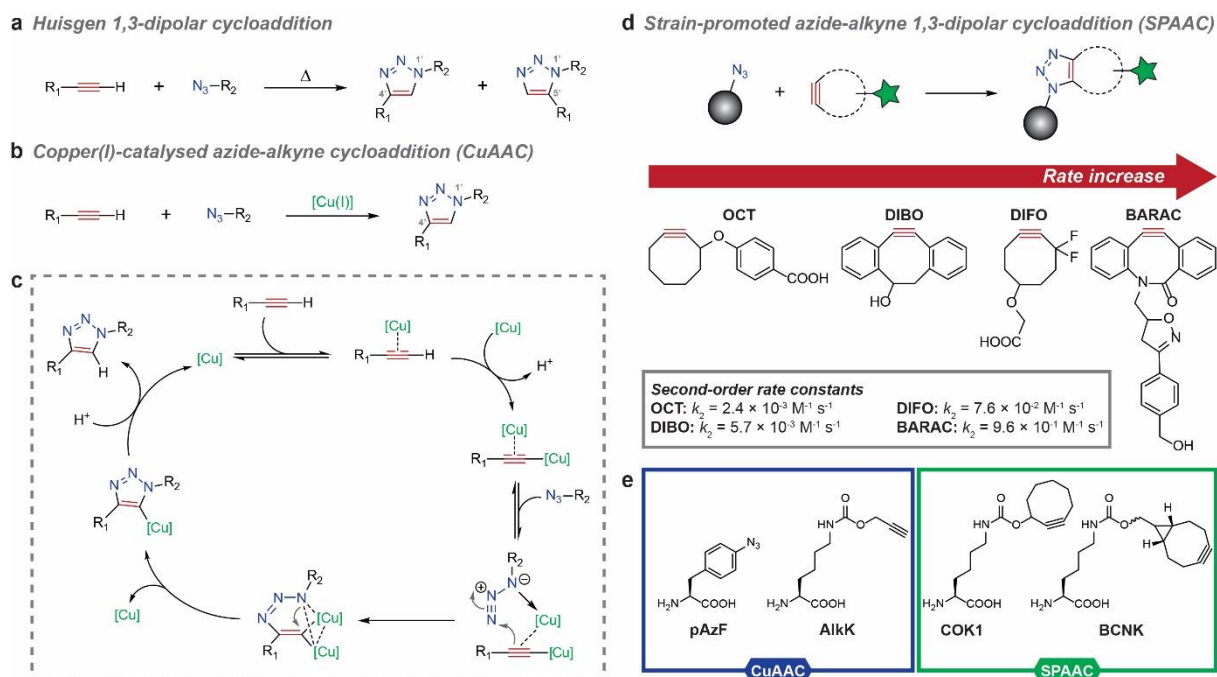
**a** *Staudinger reduction***b** *Staudinger ligation***c** *Azide-containing UAAs*

**Figure 1.8.** Staudinger reaction and chemoselective labelling *via* Staudinger ligation. (a) Mechanism of Staudinger reduction of azides into amines mediated by aryl phosphines. (b) A modification of the Staudinger reduction, the Staudinger ligation can be employed for selective labelling of proteins. (c) Examples of azide-containing UAAs that have been site-specifically incorporated into proteins and used for protein labelling *via* Staudinger ligation.

Azides can also react with alkynes in 1,3-dipolar cycloadditions to yield stable 1,2,3-triazoles. First described by Rolf Huisgen more than four decades ago, the reaction requires high temperatures and pressure to form the 1,2,3-triazole products (usually a mixture of 1,4- and 1,5-regioisomers) in acceptable yields (Figure 1.9a).<sup>[73]</sup> However, an impressive rate acceleration was achieved when catalytic amounts of copper(I) salts were used (Figure 1.9b).<sup>[74]</sup> First reported independently by Morten Meldal and K. Barry Sharpless, the Cu(I)-catalysed variant of cycloaddition between azides and terminal alkynes (CuAAC), widely known as “click reaction”, proceeds efficiently at physiological conditions (aqueous solutions, room temperature and neutral pH), yielding only the 1,4-regioisomer. The Cu(I) catalyst is either added to the reaction in form of a salt (e.g. copper(I) iodide or bromide) or, more often, generated *in situ* by using a mixture of copper(II) sulfate and sodium ascorbate as the reducing agent. Stabilizing ligands such as TBTA are also used to improve reaction yields, due to poor stability of Cu(I) in aqueous media.<sup>[75]</sup> Mechanistically speaking, CuAAC is no longer a formal 1,3-dipolar cycloaddition as it doesn’t occur in a concerted fashion. Instead, it is thought to involve two chemically equivalent copper atoms: one  $\sigma$ -bound copper acetylide bearing a second  $\pi$ -bound copper atom which reversibly coordinates the organic azide. Following the nucleophilic attack of the acetylide onto the azide, a cyclic intermediate is formed which gives the 1,4-triazole product after ring contraction and protonolysis (Figure 1.9c).<sup>[76]</sup> An analogous ruthenium-catalysed variant (RuAAC) yields regioselectively only the 1,5-triazole.<sup>[77]</sup>

The impact of click chemistry has substantially grown over the years with CuAAC being described as “the cream of the crop”.<sup>[84]</sup> Indeed, CuAAC has been broadly applicable in synthesis of small molecules as well as polymers and dendrimers,<sup>[85]</sup> macrocyclizations,<sup>[86]</sup> functionalization of surfaces,<sup>[87]</sup> natural product and drug discovery<sup>[88]</sup> and bioconjugations.<sup>[89]</sup>

Even though it has been employed in various biological studies and proceeds at least 25 times faster than Staudinger ligation,<sup>[90]</sup> CuAAC *per se* is not truly biorthogonal and is unsuitable for labelling applications in live cells due to the toxicity of Cu(I) ions.<sup>[91]</sup> Ligand coordination sometimes helps in reducing the concentrations of toxic ions, however the toxicity of copper complexes highly depends on the ligand environment, and some complexes can still be taken up by the cell, directly affecting cellular metabolism and function.<sup>[92]</sup> Therefore, several metal-free azide-alkyne cycloadditions have been successfully developed.<sup>[93]</sup>



**Figure 1.9.** Azide-alkyne cycloaddition reactions. (a) Huisgen 1,3-dipolar cycloaddition. (b) Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC). (c) Mechanism of CuAAC involving two chemically equivalent copper atoms. (d) Strain-promoted azide-alkyne cycloaddition (SPAAC). The labelling probe is represented by the green star and the biomolecule by the grey sphere. The reactivity of the strained alkynes can be enhanced either by introducing electron-withdrawing groups (DIFO)<sup>[78]</sup> or by modulating the ring strain (DIBO and BARAC). All second-order rate constants were measured for reactions with benzyl azide by NMR in CD<sub>3</sub>CN except for the reaction with DIBO which was measured in MeOD. (e) Examples of UAAs employed for CuAAC (blue) and SPAAC (green). pAzF is incorporated in *E. coli* by a *MjTyrRS* mutant<sup>[79]</sup> and in eukaryotes by an evolved *EcTyrRS* variant.<sup>[32, 80]</sup> Alkyne-bearing AlkK<sup>[81]</sup>, COK1<sup>[82]</sup> and BCNK<sup>[83]</sup> are incorporated by respective PylRS variants.

The need for Cu(I) catalyst in azide-alkyne cycloaddition reactions was effectively eliminated by introducing ring strain into the alkyne moiety (Figure 1.9d).<sup>[94]</sup> In its first version, the reaction known as strain-promoted [3+2] azide-alkyne cycloaddition (SPAAC) was shown to occur with cyclooctyne (smallest stable cycloalkyne) derivatives at rates comparable to Staudinger ligations (OCT, Figure 1.9d).<sup>[90]</sup> Albeit slower than CuAAC, the introduction of the ring-strain is a significant improvement over Huisgen's 1,3-dipolar cycloaddition. The deformation of the triple bond angle in the strained alkyne, significantly lowers the activation energy of cyclooctynes as opposed to terminal alkynes.<sup>[95]</sup> Therefore, cyclooctynes give 1,2,3-



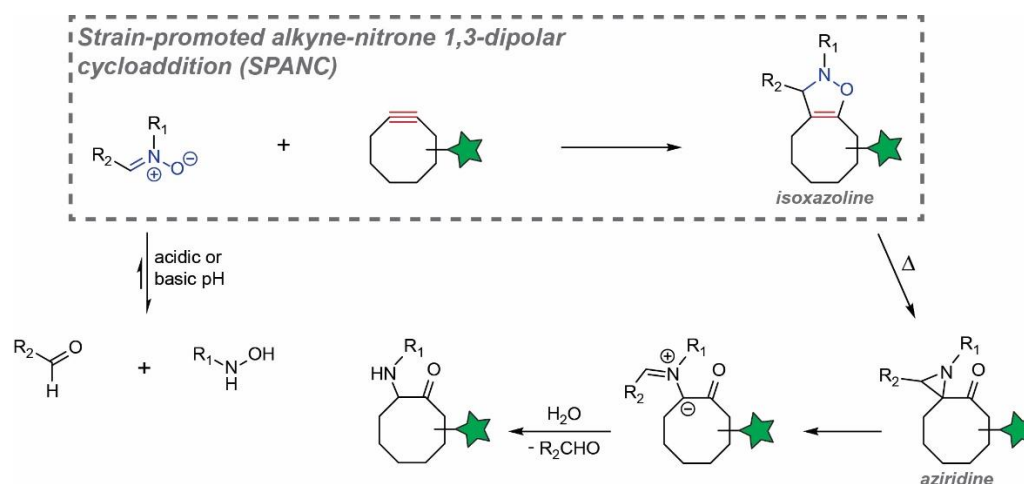
triazoles<sup>2</sup> with azides at physiological conditions without the use of heat or a catalyst. Improvements over the reaction rates and solubility of the cyclooctyne reagents generated a series of functionalized derivatives including: difluorinated cyclooctynes (DIFO),<sup>[78]</sup> dibenzocyclooctynes (DIBO)<sup>[97]</sup> and biarylazacyclooctynones (BARAC).<sup>[98]</sup> The fusion of two aryl rings to the cyclooctyne moiety results in a higher ring strain, which is even further enhanced by incorporation of an amide bond into the ring skeleton, while improving the solubility of the reagent at the same time. BARAC exhibits the fastest kinetics for a SPAAC reaction thus far (up to  $1 \text{ M}^{-1} \text{ s}^{-1}$ , Figure 1.9d), however its improved reactivity comes with the cost of lower stability, as it is also sensitive to oxygen and light and can undergo acid catalysed intramolecular rearrangements.<sup>[99]</sup> Additionally, cyclooctynes are susceptible to the nucleophilic attack of glutathione present in living cells.<sup>[100]</sup> Nevertheless, SPAAC has enabled labelling of abundant biomolecules within mammalian cells and *C. elegans*, as well as live zebrafish and mice.<sup>[78, 100b, 101]</sup> UAAs containing strained alkynes have also been successfully employed for labelling via SPAAC. (Figure 1.9e)<sup>[82, 83b]</sup>. A light-induced version of SPAAC has been developed as well and used as a tool for temporally-controlled labelling, which will be described in detail later on.<sup>[102]</sup>

In general, the synthesis of cyclooctyne-bearing reagents has been shown to be often tedious and low-yielding.<sup>[78, 103]</sup> Nevertheless, their applicability extends the SPAAC reaction; strained cyclooctynes can also react with nitrones in [3+2] cycloadditions (strain-promoted 1,3-dipolar azide-nitron cycloaddition or SPANC, Figure 1.10) which are up to 30 times faster than the corresponding reactions with azides.<sup>[104]</sup> Notable labelling experiments have been performed on the cell surface through chemical introduction of a nitron moiety into proteins; either by oxidation of an N-terminal serine residue with sodium periodate and subsequent condensation of the formed aldehyde with hydroxylamine<sup>[104a]</sup> or by labelling of surface-exposed lysine residues with activated NHS-esters of cyclic nitrones.<sup>[105]</sup> Nitrones, however, are prone to hydrolysis, while isoxazoline products may undergo Baldwin rearrangements at physiological conditions which limits their application (Figure 1.10).<sup>[106]</sup>

---

<sup>2</sup> Since SPAAC follows the same concerted mechanism of Huisgen's 1,3-dipolar cycloaddition, the 1,2,3-triazole product is a mixture of regioisomers. For labelling purposes this is generally not a major issue, however the regioselectivity problem limits applications of SPAAC in areas where homogenous adducts are indispensable (e.g. drug design). Nevertheless, symmetrical cyclooctynes such as BCN have been reported to form a single regioisomer upon cycloaddition with azides.<sup>[96]</sup>

---

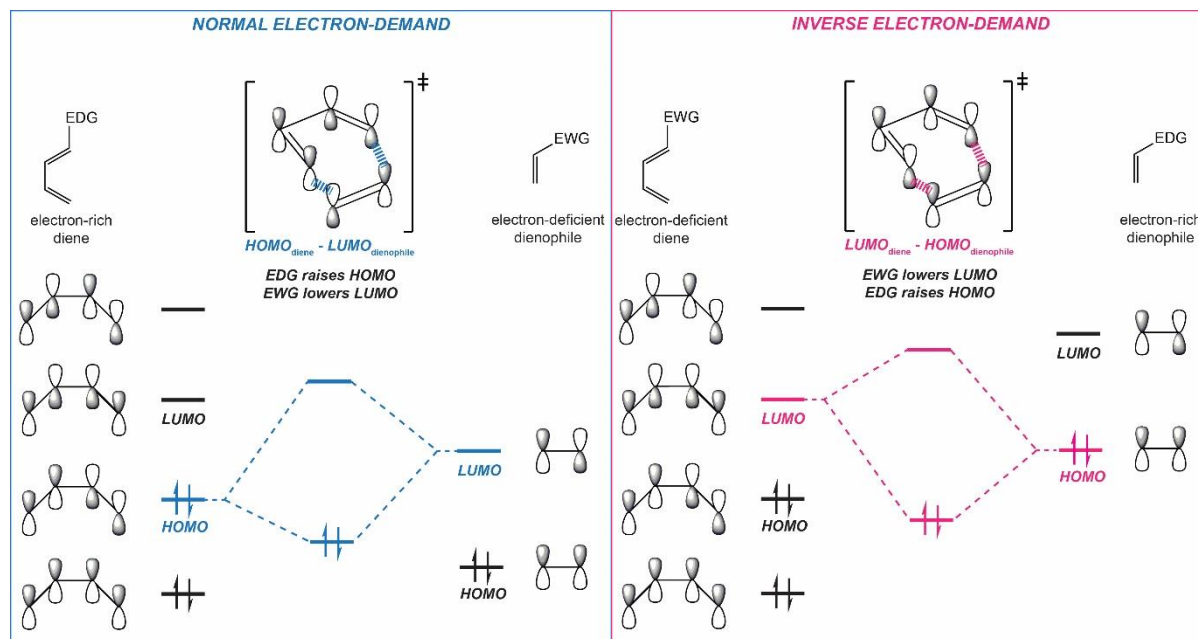


**Figure 1.10.** Strain-promoted alkyne-nitron 1,3-dipolar cycloaddition (SPANC) and possible side-reactions of its reactants and products. Nitrones are prone to hydrolysis under acidic and basic conditions to corresponding aldehydes and hydroxylamines. Isoxazoline products can undergo thermal Baldwin rearrangement to give substituted aziridines. Aziridine ring-opening followed by hydrolysis yields a  $\beta$ -aminoketone. The stability of both nitrones and isoxazoline products largely depends on substituents  $R_1$  and  $R_2$ ; cyclic nitrones exhibit greater stability towards hydrolysis while electron-withdrawing groups can minimize the isoxazoline rearrangement. Green star represents the labelling probe.

### 1.4.3. Inverse electron-demand Diels-Alder cycloadditions

Bioorthogonal reactions based on Diels-Alder (DA) cycloaddition have proven to be quite popular since they can proceed in high yields under physiological conditions and don't produce any by-products.<sup>[107]</sup> First described by Otto Diels and Kurt Alder in 1928, Diels-Alder cycloaddition is a pericyclic reaction between a conjugated diene and a dienophile yielding a six-membered ring.<sup>[108]</sup> The reaction proceeds through a concerted suprafacial/suprafacial interaction of  $4\pi$ -electrons of the diene and  $2\pi$ -electrons of the dienophile and is, therefore, classified as a thermally allowed  $[4+2]$  cycloaddition or, according to Woodward-Hoffman rules, a  $[\pi4s + \pi2s]$  cycloaddition (where “s” indicates the suprafacial interaction between the  $\pi$ -electrons). Highly regarded as one of the most important carbon-carbon bond forming reactions with good control over the regio- and stereochemical outcomes, the DA cycloaddition has found broad applications especially in the synthesis of natural products and new materials.<sup>[109]</sup> Depending on the electronic effects of the substituents on the diene and the dienophile, DA cycloaddition is classified into more common “normal” electron-demand DA and inverse electron-demand DA reaction (iEDDA) (Figure 1.11). The “normal” electron-demand DA typically occurs between an electron-rich diene and an electron deficient dienophile through the strong interactions between the highest occupied molecular orbital (HOMO) of the diene and the lowest unoccupied molecular orbital (LUMO) orbital of the dienophile. The energy gap between the HOMO and LUMO frontier orbitals, however, is low enough that, by changing the electronic properties of the substituents on both reaction partners, their roles can essentially be reversed. Consequently, electron-withdrawing groups (EWGs) on the diene lower its LUMO, while electron-donating groups (EDGs) on the dienophile raise the energy of its HOMO. The energetically favourable bond formation in an inverse-electron demand DA ergo comes from the interaction between the LUMO of the diene

and the HOMO of the dienophile. Lowering the energy gap between  $\text{HOMO}_{\text{dienophile}}$  and  $\text{LUMO}_{\text{diene}}$  through modulation of bound substituents can accelerate the rates of iEDDA. Even though it has been observed that DA cycloadditions exhibit a notable rate enhancement when carried out in polar solvent systems, even in water,<sup>[110]</sup> their biological application is limited due to possible cross-reactions of electron-rich dienes with multiple nucleophilic functionalities within the cell.<sup>[111]</sup>

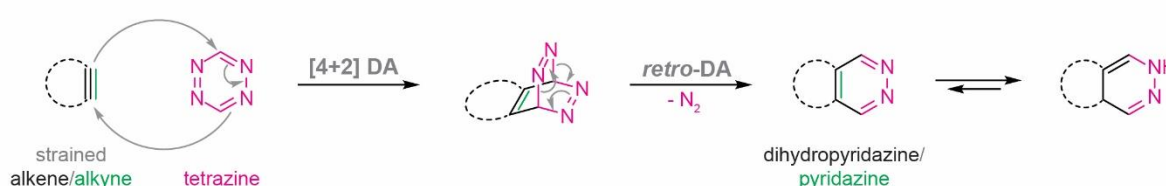


**Figure 1.11.** Normal and inverse electron-demand Diels-Alder cycloaddition. The energy gap between the frontier molecular orbitals and thus the kinetics of the cycloaddition reaction can be affected by the electronic properties of the substituted (either electron-donating (EDG) or electron-withdrawing (EWG) groups) reactants.

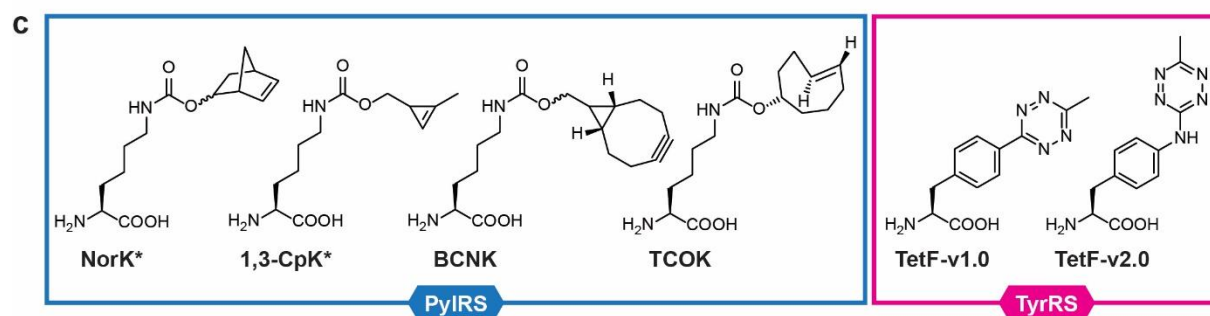
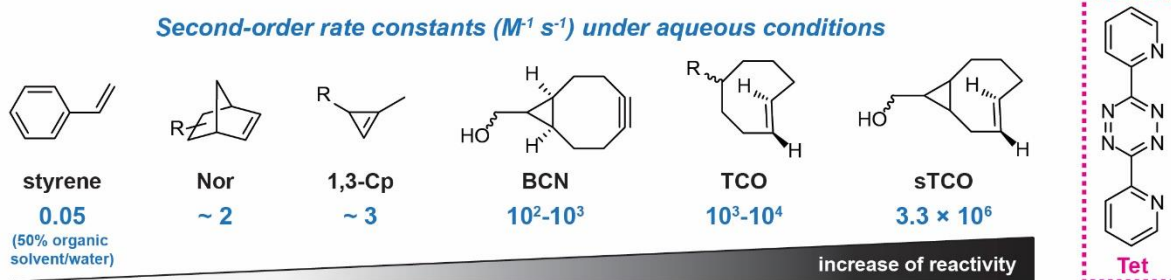
Recently, an iEEDA reaction between strained alkenes and alkynes and 1,2,4,5-tetrazines has been rediscovered and repurposed for selective labelling of biomolecules.<sup>[112]</sup> Defined by high regio-, stereo- and chemoselectivity, these reactions yield dihydropyridazines or pyridazines alongside molecular nitrogen as the only non-toxic by-product (Figure 1.12a). The extrusion of nitrogen is the result of a retro-DA reaction which immediately follows the initial iEEDA cycloaddition between an electron deficient 1,2,4,5-tetrazine and the electron rich strained alkene or alkyne. Apart from their high specificity, these reactions are highlighted by their impressively fast kinetics, with rates comparable to kinetics of biological processes ( $10^4$ - $10^5 \text{ M}^{-1} \text{ s}^{-1}$ , Figure 1.12b). Depending on which reaction partners are used, the reaction can achieve full conversion within seconds. Indeed, several UAAs bearing strained alkene and alkyne functionalities, including norbornenes (NorK),<sup>[33c, 113]</sup> 1,3-disubstituted cyclopropenes (1,3-CpK),<sup>[114]</sup> bicyclo[6.1.0]nonynes (BCNK)<sup>[83]</sup> and *trans*-cyclooctenes (TCOK),<sup>[83a, 113b]</sup> have successfully been incorporated site-specifically into proteins *via* engineering respective PylRS/tRNA<sup>CUA</sup> pairs (Figure 1.12c). Orthogonal PylRS variants allowed site-specific incorporation of these moieties in bacteria as well as in mammalian cells and subsequent labelling with functionalized tetrazine probes, with NorK exhibiting lowest reaction rates ( $\sim 1$ - $10 \text{ M}^{-1} \text{ s}^{-1}$ ) and TCOK the fastest ( $\sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$ , Figure 1.12b).<sup>[33c, 83a, 115]</sup>

When coupled to a series of green- and red-emitting fluorophores, tetrazines (which are chromophores themselves, absorbing light between 500-530 nm) have interestingly showed to quench their fluorescence. Once covalently attached to a protein of interest bearing a strained alkene/alkyne, the fluorescence of the cycloaddition adducts dramatically increased.<sup>[83a, 116]</sup> Further fine-tuning of the reaction, by introducing ring-strain and various substituents to the reaction partners, resulted with the fastest reported reaction with  $k_2$  of  $3.3 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$  (between dipyridyltetrazine and a strained *trans*-cyclooctene, sTCO, Figure 1.12b).<sup>[117]</sup> It must be noted that tetrazine-bearing UAAs have been prepared as well, however their applications have so far been limited to *E. coli*, as they are recognized by *Mj*TyrRS variants (Figure 1.12c).<sup>[118]</sup>

**a** Inverse electron-demand Diels-Alder cycloaddition



**b** Reactivities of unstrained and strained dienophiles with dipyridyltetrazine (Tet)



**Figure 1.12.** Inverse electron-demand Diels-Alder (iEDDA) cycloaddition between tetrazines and strained alkenes/alkynes. (a) Mechanism of the iEDDA reaction. Dihydropyridazine or pyridazine products are obtained through a DA/retro-DA sequence with the extrusion of molecular nitrogen as the only non-toxic by-product. (b) Examples of relative reactivities of unstrained and strained dienophiles with dipyridyltetrazine and their second-order rate constants ( $\text{M}^{-1} \text{s}^{-1}$ ). (c) Examples of UAAs containing strained alkenes/alkynes and tetrazines incorporated by corresponding PylRS and TyrRS variants respectively. Asterisk (\*) denotes UAAs that are accepted by the wild-type PylRS.

Together with the fast kinetics and superior signal-to-noise ratio, iEDDA between tetrazines and strained alkenes/alkynes has been the first choice for many spatiotemporal labelling applications *in vitro* and *in vivo*, both on the cell surface and within the cells. For

example, CpK in combination with fluorophore-conjugates of tetrazines was used for statistical labelling of *E. coli* and mammalian cell proteomes as well as for cell-specific proteomics in animals such as *D. melanogaster*.<sup>[114, 119]</sup> Another example involves site-specific incorporation of BCNK into a specific kinase, which permitted the control of enzyme's activity through tethering of a tetrazine-containing inhibitor. Furthermore, introduction of a photoswitchable linker between the tetrazine and inhibitor moieties enabled light-dependent and reversible toggling of enzyme's function in live mammalian cells.<sup>[120]</sup>

#### 1.4.4. Light-induced reactions in chemical biology

Employment of light-induced reactions in biological systems has attracted considerable attention as these reactions are distinguished by superior spatial and temporal resolution and elegant reaction control.<sup>[121]</sup> Such control is achieved through the use of light sources with suitable wavelengths and light intensities and, naturally, through overall irradiation time. Apart from their operational simplicity, importantly, light-induced reactions often proceed without the need for ligands or potentially toxic metal catalysts making them compatible with living biological systems. The underlying principle of photo-triggered reactions consists of *in situ* generation of highly reactive and short-living intermediates *via* irradiation, which subsequently react with their specific counterparts present in the biological system (e.g. through a cycloaddition reaction). The photo-generation of reactive intermediates is typically irreversible and their stability is directly influenced by functional groups present on the molecule. Depending on what types of reactive intermediates are formed, the mechanisms of their formation and their subsequent reactivities, light-induced reactions in chemical biology include:

- (a) Reactions where the reactive intermediates are formed through irreversible extrusion of CO (e.g. from cyclopropanones<sup>[122]</sup>) or N<sub>2</sub> (e.g. from tetrazoles or diazirines).
- (b) Reactions where the reactive intermediates are formed through rearrangements or photoelimination reactions (e.g. *o*-quinone methides, hydroxy-*o*-quinodimethanes, nitrile ylides from 2*H*-azirines).
- (c) Reactions involving free radicals whose generation is often mediated by irradiation of a photoinitiator. A prime example is the formation of thiyl radicals for thiol-ene and thiol-yne click reactions.

While the intended use of aforementioned reaction intermediates is light-induced formation of covalent linkages within or between biomolecules (e.g. for labelling or crosslinking purposes; see Chapter 1.5 for the latter), other strategies that involve photosensitive reagents which give a certain response upon irradiation with light have found broad applications in chemical biology, especially in light-dependent modulation of protein structure and function. These strategies include: (d) irreversibly triggered, photolabile protection groups (photocages) as reagents that by structure-guided placement temporarily mask ("cage") a certain functionality in a biologically active molecule (e.g. coumarins and nitroaryls); and (e) reversibly switchable compounds, photoswitches, which can be switched between two thermodynamically stable

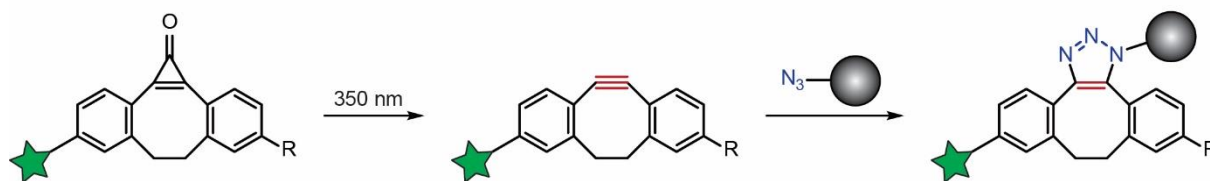
---

forms (e.g. azobenzenes). The photophysical and photochemical properties as well as various applications of these types of compounds have been extensively covered in several reviews<sup>[123]</sup> and are not going to be covered within the scope of this chapter.

#### 1.4.4.1. 1,3-dipolar cycloadditions

As already mentioned in the previous chapter, several efforts have been made to address the toxicity of Cu(I) catalysts when implementing CuAAC reaction for live-cell labelling. After the breakthrough achieved by introducing a ring strain into the alkyne, eliminating the need for metal catalysis, Popik and co-workers have masked the triple bond of dibenzocyclooctyne (DIBO) as cyclopropenone (Figure 1.13).<sup>[102]</sup> Cyclopropenones are thermally stable in aqueous media, however, they readily undergo CO extrusion after short irradiation with wavelengths around 350 nm (one-photon excitation) or 800 nm (two-photon excitation) affording DIBO that can selectively react with organic azides. The “unmasking” reaction is reported to be quantitative and proceed with high quantum yields and was utilized for labelling of living cells expressing azide-modified glycoproteins on their surface.<sup>[102]</sup> More recently, the strategy was expanded towards labelling of nascent RNA in living cells<sup>[124]</sup> and tissue labelling *via* multi-photon activation.<sup>[125]</sup> Nevertheless, the photo-SPAAC strategy is currently still limited to relatively large DIBO derivatives, while one-photon uncaging still depends on usage of UV light.

#### Photo-SPAAC

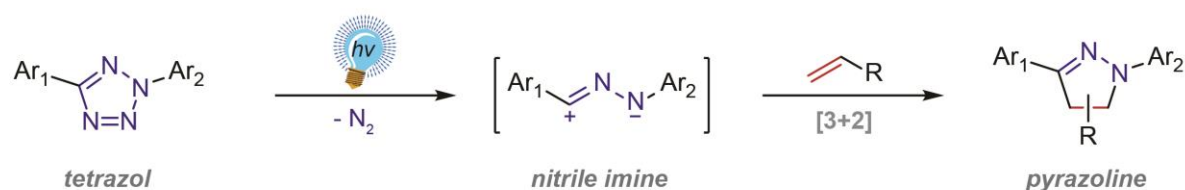


**Figure 1.13.** The light-mediated “unmasking” of cyclopropenones and subsequent photo SPAAC reaction with azides.

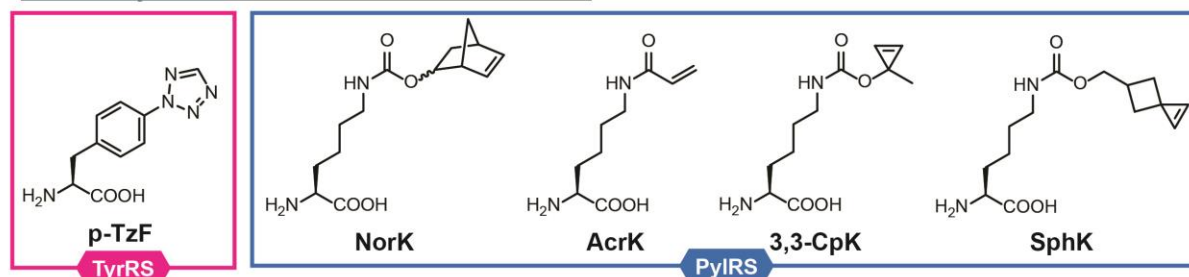
In the 1960s, Rolf Huisgen described another 1,3-dipolar cycloaddition that has found its way towards biology – the reaction between methyl crotonate and the nitrile imine formed by photolysis (or thermolysis at temperatures around 200 °C) of 2,5-diphenyltetrazole, affording a 3:1 mixture of stable pyrazolines.<sup>[126]</sup> After irradiation, the diaryltetrazole undergoes rapid cycloreversion accompanied by extrusion of molecular nitrogen to irreversibly generate the nitrile imine *in situ*, a transient, highly reactive species which tends to dimerize into dihydrotetrazines or bis-diazoethylenes. In the presence of either electron-rich or electron-deficient dipolarophiles (e.g. terminal alkenes), however, nitrile imines spontaneously undergo [3+2] cycloaddition (Figure 1.14a). Initially, the photolysis of diaryltetrazoles was performed using an intense immersion lamp with broad-spectrum UV irradiation. However, Lin and co-workers have reported that the extrusion of nitrogen can be accomplished with simple hand-held UV lamps (302 nm) generally used for thin layer chromatography (TLC) monitoring.<sup>[127]</sup> On small molecule reactions, the formation of nitrile imines typically occurs within a couple of minutes, making the following 1,3-dipolar cycloaddition with terminal alkenes the limiting step of this so-called “photoclick” reaction.

Due to the phototoxicity of the UV light at 302 nm, a series of substituted diaryltetrazols have been made with the goal to shift the absorption maxima towards the long-wavelength region. Indeed, introduction of electron-donating amino-, dimethylamino- and methoxy- groups as well as extension of the conjugated system have resulted with a bathochromic shift towards 365 nm.<sup>[128]</sup> Furthermore, electron-donating groups raise the energy of the HOMO of the nitrile amine, making the HOMO<sub>dipole</sub>-LUMO<sub>dipolarophile</sub> energy gap smaller and thus accelerating the photoclick reaction. Both tetrazole and alkene groups have been incorporated site-specifically into proteins in the form of UAAs (Figure 1.14b).<sup>[34, 113a, 129]</sup> The tetrazole UAA *p*-(2-tetrazole)phenylalanine (*p*-TzF, Figure 1.14b), incorporated by an evolved *Mj*TyrRS mutant into proteins in *E. coli*, only showed sluggish reactivity towards dimethyl fumarate ( $k_2 \sim 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$ ) when irradiated with 254 nm due to its shortened  $\pi$ -conjugation system.<sup>[129b]</sup> Therefore, incorporation of UAAs bearing strained (NorK, 3,3-CpK, SphK)<sup>[113a, 129c, 129d]</sup> or electron-deficient alkenes (AcrK)<sup>[34, 130]</sup> has received more attention as it allowed more freedom in designing tetrazole-based reaction partners that would exhibit desired reactivities (Figure 1.14b).

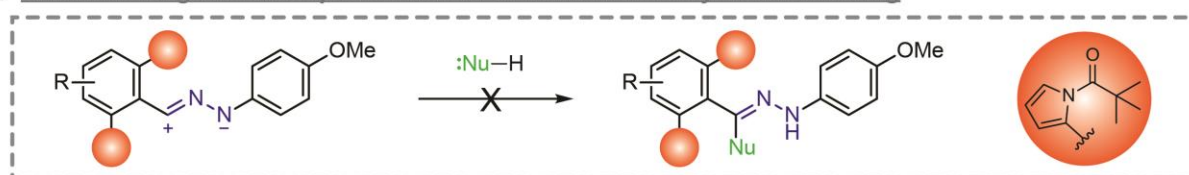
### a Tetrazole “photoclick” reaction



### b Genetically encodable tetrazole and alkene UAAs



### c Circumventing the nucleophilic addition to nitrile iminies by steric shielding



**Figure 1.14.** Tetrazole “photoclick” reaction. (a) Upon irradiation, diaryltetrazoles afford nitrile imines which spontaneously undergo [3+2] cycloaddition reactions with terminal alkenes. (b) Genetically encodable tetrazole and alkene UAAs. Incorporation of alkene-containing UAAs has received more attention as it allows the employment of a variety of tetrazols with different spectral properties. (c) Nitrile imines show a lack of bioorthogonality as they can also react with biological nucleophiles. Nevertheless, this side reaction can be circumvented by sterically shielding the diaryltetrazole.

Typical second-order rate constants for “photoclick” reactions were reported to range from  $32 \text{ M}^{-1} \text{ s}^{-1}$  (NorK) and up to  $60 \text{ M}^{-1} \text{ s}^{-1}$  (3,3-CpK), making “photoclick” reactions

significantly faster than SPAAC.<sup>[129c]</sup> Further increase of the cyclopropene reactivity was achieved by incorporating spiro[2.3]hex-1-ene modified lysine (SphK) into sfGFP using wild-type PylRS, which showed impressively fast kinetics ( $k_2 = 1.0 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  in PBS) with a water-soluble tetrazole, comparable to the fastest tetrazine ligation reactions.<sup>[129d, 131]</sup> Another important feature of the photoclick reaction is the fluorescence that the resulting pyrazolines exhibit, making tetrazoles potentially useful as fluorogenic probes for live cell imaging studies.<sup>[132]</sup> As an illustration, a class of bithiophene-substituted tetrazoles activated by 405 nm laser light irradiation was recently reported to exhibit fast kinetics with diethyl fumarate ( $k_2$  up to  $3960 \text{ M}^{-1} \text{ s}^{-1}$  in 50 % ACN/PBS), affording pyrazoline products that fluoresce in the red channel (575-644 nm in 50 % ACN/PBS).<sup>[133]</sup> Furthermore, two-photon excitation of tetrazoles was shown on a naphthalene-based derivative with a 700 nm pulsed laser, enabling spatio-temporally controlled imaging of microtubules treated with fumarate-modified chemotherapy drug docetaxel *in vivo*.<sup>[134]</sup> A significant drawback of “photoclick” reactions is that they are not strictly bioorthogonal, as the highly-reactive nitrile imine can react with endogenous nucleophiles such as carboxylic acids, unless the diaryltetrazoles are sterically shielded (Figure 1.14c).<sup>[135]</sup> This lack of bioorthogonality has caused background signals amid incubation of native proteins with tetrazoles.

#### 1.4.4.2. Light-induced Diels-Alder reactions

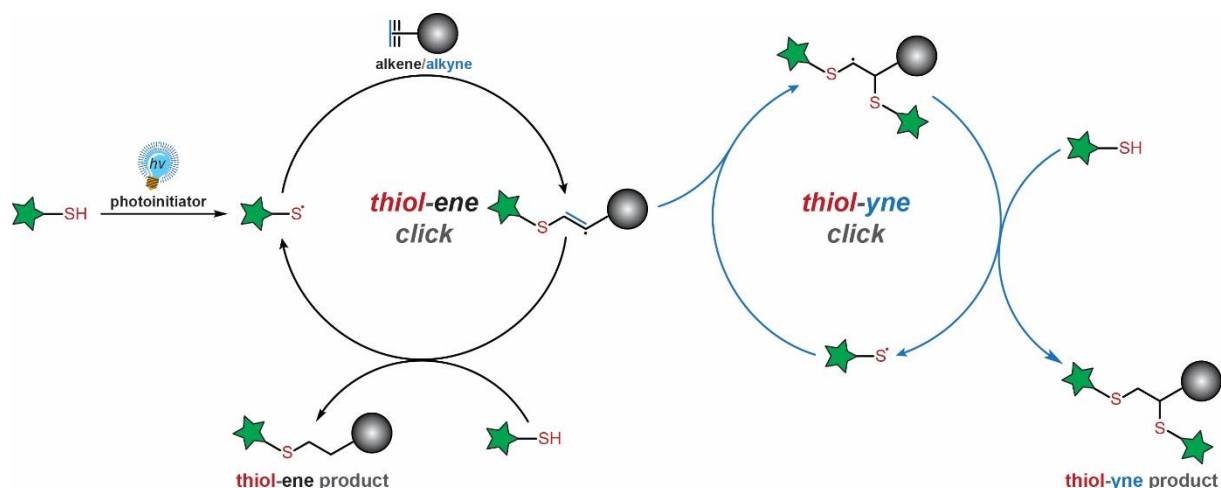
As already mentioned, the attractive features of employing DA cycloadditions in biological systems include the facts that the reactions are high-yielding at physiological conditions, they generally do not produce by-products and are regio-, stereo- and chemoselective. The most popular of these reactions is undoubtedly the rapid iEDDA between tetrazines and strained alkenes and alkynes (see Chapter 1.4.3.). Apart from iEDDA, several light-induced DA reactions have been explored for potential biological use as well. These include: (a) DA reactions of hydroxy-*o*-quinodimethanes with activated olefins<sup>[136]</sup>, (b) hetero-DA reactions of *ortho*-quinone methides with vinyl ethers<sup>[137]</sup>; and (c) visible-light-mediated cycloadditions of phenanthrenequinones and vinyl ethers<sup>[138]</sup>. Since the groundwork of this thesis lies on the shoulders of the latter two reactions, especially on the reactivity of *ortho*-quinone methides, they are going to be thoroughly covered in Chapter 2.

#### 1.4.4.3. Thiol-ene and thiol-yne click reactions

The remarkable reactivity of thiols towards reactive carbon-carbon double bonds has been known since the early 1900s.<sup>[139]</sup> However, the reaction rose to prominence almost a century later as an important example of click chemistry.<sup>[140]</sup> Today, the alkene hydrothiolation reaction is known to proceed either through a catalysed Michael addition to electron-deficient alkenes (termed thiol Michael addition)<sup>[141]</sup> or *via* free-radical addition to electron-rich and electron-deficient alkenes (termed thiol-ene or thiol-ene click reaction, Figure 1.15).<sup>[140b]</sup> The free-radical version of hydrothiolation has also been expanded towards alkynes (thiol-yne reaction, Figure 1.15).<sup>[142]</sup> Thiol-ene click reaction can be initiated thermally or photochemically (UV light) by means of radical initiators. It is regarded to be fast, high-



yielding, regioselective (following exclusively anti-Markovnikov addition) and insensitive to oxygen and water. The UV-light (generally 365 nm) promoted thiol-ene version has found broad applications in syntheses of materials and polymers as well as photopatterning.<sup>[140, 143]</sup> As for biological applications, important examples include thiol glycosylation of olefinic proteins (with residue-specific incorporation of an alkene UAA)<sup>[144]</sup>, introduction of a posttranslational modification mimic *via* site-specific protein acetylation of cysteine with *N*-vinylacetamide<sup>[145]</sup> as well as site-specific protein labelling of genetically-encoded alkenes<sup>[146]</sup> and alkynes.<sup>[147]</sup> Reactions in aqueous media depend on the complexity of the system and required irradiation times typically vary from 15-30 minutes up to 2-5 hours. The use of thiol-ene click chemistry is mostly restricted to *in vitro* biological applications as the major problem of these reactions is their cross-reactivity with thiols which are ubiquitously present in the cell. Furthermore, they rely on addition of free-radical photoinitiators. Several examples of visible-light induced thiol-ene reactions catalysed by ruthenium-based photocatalysts have been reported as well, however they are often low-yielding in aqueous solutions and require long irradiation times (> 10 hours).<sup>[148]</sup>



**Figure 1.15.** Mechanism of the light-induced thiol-ene and thiol-yne click reactions. The formation of reactive thiyl radicals is mediated by photoinitiators.

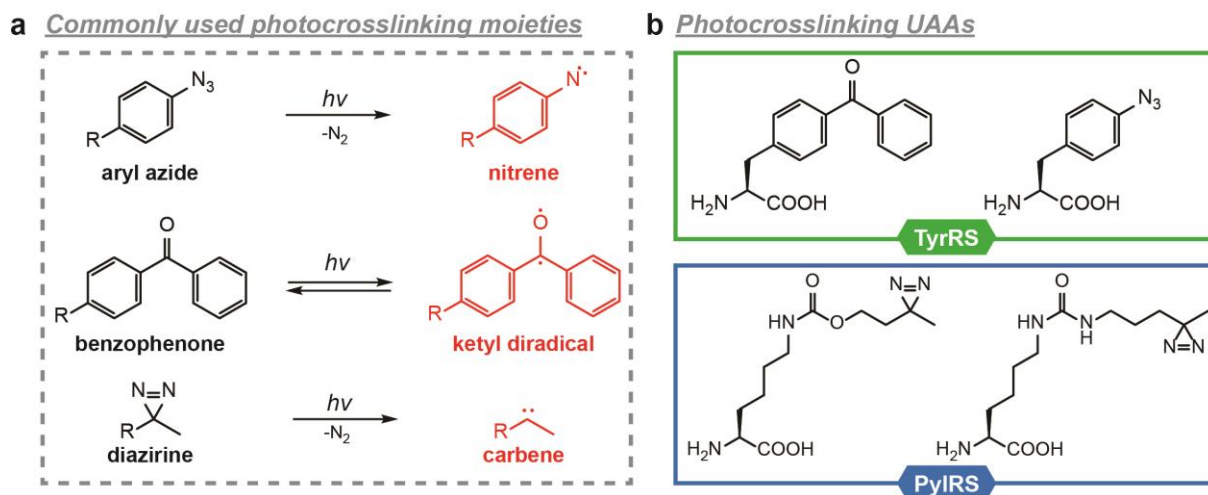
## 1.5. Proximity-triggered protein crosslinking

The structure and function of proteins in living cells are determined by covalent peptide bonds in the backbone as well as multiple noncovalent interactions between amino acid side chains such as hydrogen bonds, ion bridges and hydrophobic interactions. Other covalent linkages innate to different types of proteins in bacteria and eukaryotes come in form of disulfide bonds (or disulfide bridges) between two cysteine residues. The presence of either inter- or intramolecular disulfide bonds in proteins such as antibodies, enzymes and membrane receptors has an important role in their folding and stability.<sup>[149]</sup> For example, aggregation of misfolded proteins is usually caused by formation of non-native disulfide bonds during the protein folding process. In certain bacterial cell-surface proteins, lysine residues can also form intramolecular isopeptide bonds with asparagine or aspartate.<sup>[150]</sup> These isopeptide bonds are formed autocatalytically during protein folding when the respective reacting groups are brought together in a hydrophobic environment. Moreover, the formation of the isopeptide bond requires a nearby glutamate or aspartate as intramolecular catalysts.

Apart from these examples, however, formation of irreversible covalent linkages in proteins is generally not encountered and scientists have over the years pursued tools and approaches that would cause autocatalytic formation of additional types of covalent linkages within and/or between proteins. Such tools would not only enable the design and modulation of novel protein features but allow studies of protein-protein interaction networks as well. The latter is of extreme importance as many signalling and regulatory pathways in the cell depend on formation of transient and low-affinity protein interactions. The inherently dynamic nature of these interactions, however, often hinders high-resolution structural studies, which frequently involve the generation of stable protein complexes. Traditional methods of studying protein-protein interactions include immunoprecipitation<sup>[151]</sup>, yeast two-hybrid screenings<sup>[152]</sup> and FRET-based assays.<sup>[153]</sup> Although powerful, these approaches suffer from many limitations.<sup>[154]</sup> Immunoprecipitation, for example, is unsuitable for characterizing transient interactions<sup>[151a]</sup> while two-hybrid systems typically exhibit high false positive identification rates as they are performed in non-native environments.<sup>[155]</sup> Another method relies on the engineering of disulfide bonds by introducing cysteine residues at specific positions within the interface of two interacting protein partners.<sup>[156]</sup> This method of generating covalent binary complexes (disulfide trapping) is, however, limited by the relative distances that can be spanned by a disulfide bond (5 Å) as well as by rigidity and susceptibility of disulfides towards reduction. Instead, new crosslinking approaches that have recently emerged rely on the site-specific incorporation of two conceptually different but complementary types of UAAs.<sup>[51]</sup>

On the one hand, UAAs bearing photocrosslinking moieties such as benzophenones<sup>[157]</sup>, aryl azides<sup>[79a]</sup> and diazirines<sup>[158]</sup> can form covalent bonds with nearby biomolecules upon UV irradiation (Figure 1.16).<sup>[159]</sup> When irradiated at 250 nm, aryl azides form highly reactive nitrene species that readily insert into C-H and heteroatom-H bonds. Benzophenones generate ketyl diradicals at 350-365 nm while irradiation of diazirines at similar wavelengths leads to formation of reactive carbenes through extrusion of molecular

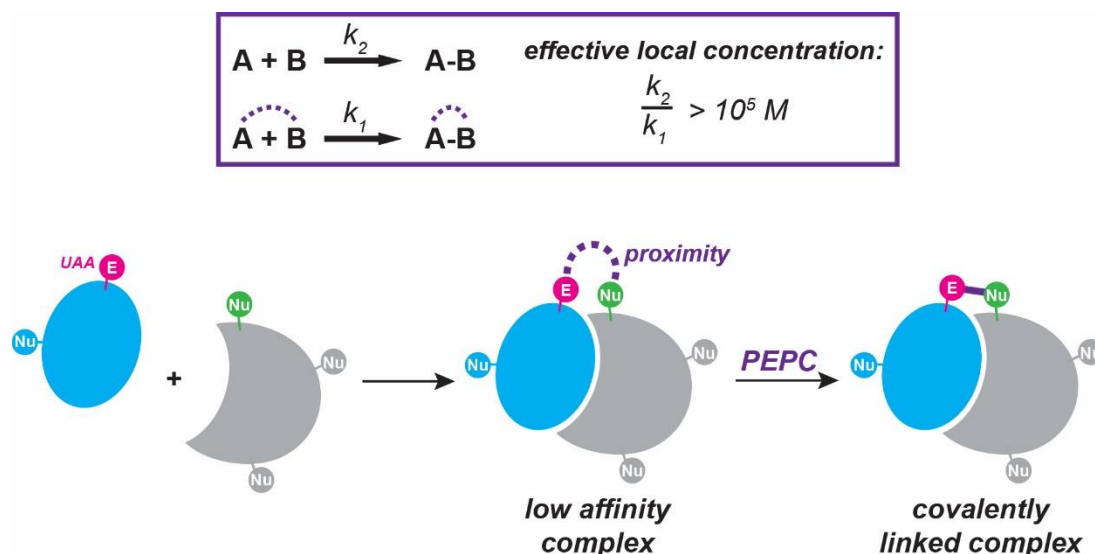
nitrogen. Both ketyl diradicals and carbenes readily insert into C-H bonds of nearby biomolecules, with ketyl diradicals showing tendencies towards  $\gamma$ -CH bonds of methionine.<sup>[159]</sup> Despite the fact that photocrosslinking UAAs have allowed elucidation of weak and transient protein-protein interactions<sup>[51b]</sup>, the respective reactive intermediates are unselective and are easily quenched by water (which, when compared with all biomolecules in the cell, is present in excess) provoking poor crosslinking efficiencies. Furthermore, prolonged and iterative UV irradiations that these molecules often require are harmful to living cells, making them not ideal for homogeneously crosslinking low-affinity complexes.



**Figure 1.16.** Commonly used photocrosslinking UAAs. (a) Light-mediated generation of reactive crosslinking intermediates: nitrenes from aryl azides, ketyl diradicals from benzophenones and carbenes from diazirines. (b) Structures of genetically encoded photocrosslinking UAAs.

The other type of UAAs bear electrophilic moieties that are able to undergo proximity-mediated reactions with adjacent nucleophilic residues. Contrarily to the previously described UAAs used for bioorthogonal labelling strategies (Chapter 1.4), the design of functionalities which are supposed to react specifically with nucleophilic amino acids is vastly demanding as it requires precise fine-tuning of the reactive group in order to ensure unspecific binding to the translational machinery of the cell. In order to achieve high specificity of the covalent linkage formation, the electrophilicity of the reactive group has to be fine-tuned to be chemically inert under physiological conditions, but to react with nucleophilic amino acid side chains when drastically increasing their effective local concentration through proximity.<sup>[160]</sup> The proximity effect can easily be understood by looking at the association between two proteins A and B (affording the complex A-B) as if it were a bimolecular and a unimolecular reactions (Figure 1.17). In a bimolecular reaction, the reaction rate depends on random collision between A and B that would bring the two substrates close together for them to react (in other words, the rate depends directly on the concentrations of both A and B). If the two crosslinking moieties A and B are brought into close proximity, their effective local concentration and, consequently, the reaction rate, rapidly increases with values easily higher than  $10^5$  M.<sup>[161]</sup> To put this number into perspective of reaction rates, the concentrations of A and B in a bimolecular reaction would have to be inconceivably higher than  $10^5$  M in order for the reaction to

proceed at the same rate as the unimolecular one. The expansion of this concept to site-specific introduction of bioreactive amino acids has been dubbed proximity-enabled protein crosslinking (PEPC, Figure 1.17).<sup>[162]</sup> Figure 1.18 depicts an overview of electrophilic UAAs developed for proximity-triggered crosslinking. The representative electrophiles include reactive halides<sup>[53, 162-163]</sup>, Michael acceptors<sup>[164]</sup>, aryl isocyanates<sup>[165]</sup>, aryl carbamates<sup>[166]</sup>, aryl fluorosulfates<sup>[167]</sup> and, very recently, quinone methides.<sup>[168]</sup> Their reactivities towards adjacent nucleophilic amino acids and respective applications are discussed in the following sections.



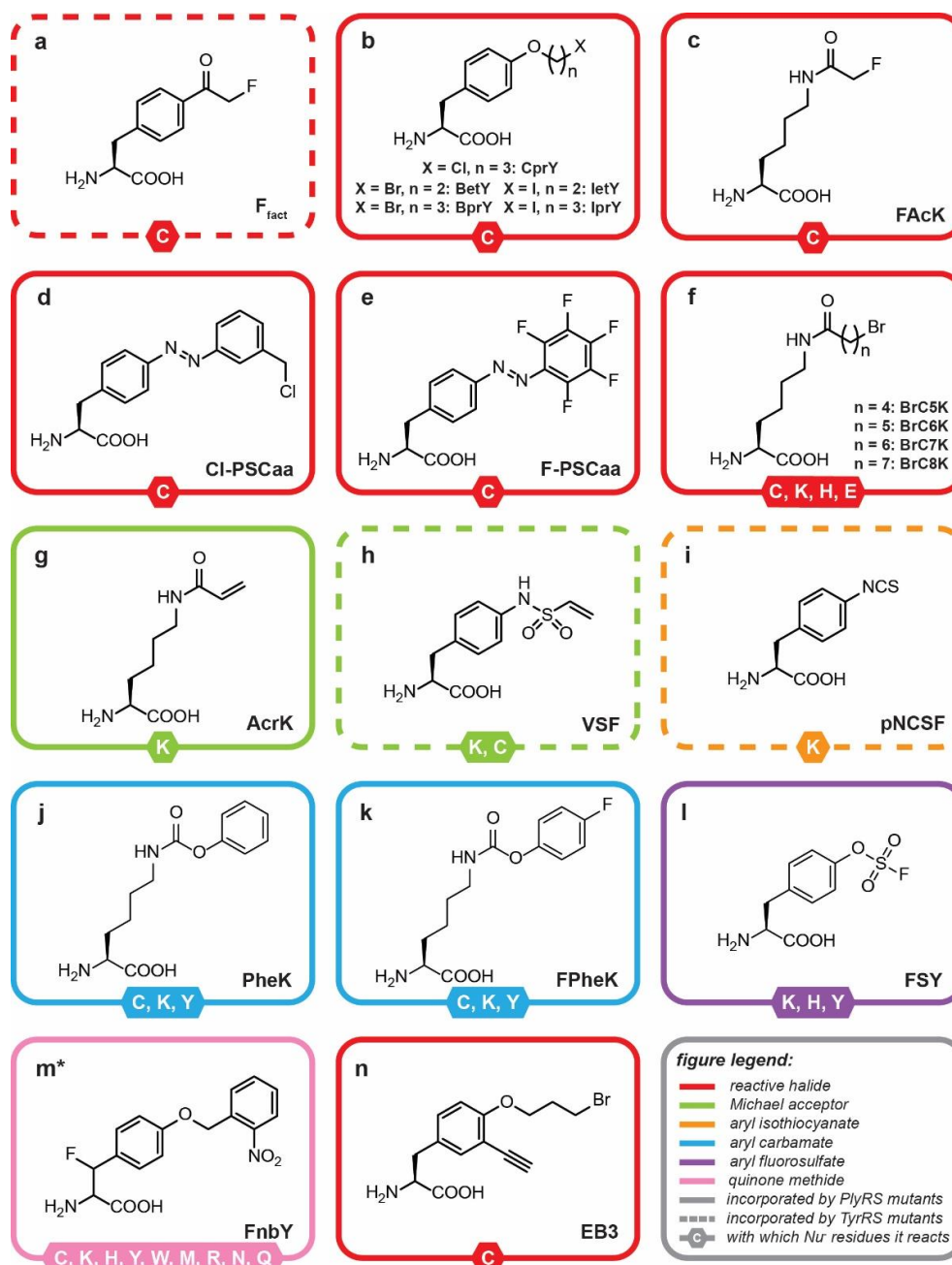
**Figure 1.17.** Proximity-triggered crosslinking of low-affinity protein complexes. A site-specifically incorporated electrophilic UAA (pink circle) is inert under physiological conditions but can react with nucleophilic amino acid residues (green circle) upon increasing their effective local concentration, which is mediated by proximity. The effective local concentration values easily surpass  $10^5$  M and thereby substantially increase the kinetics of the crosslinking reaction.

### 1.5.1. Overview of UAAs able to undergo proximity-enabled reactions

#### 1.5.1.1. Reactive halides

The first UAA successfully applied for proximity-enabled reactions was *p*-2'-fluoroacetylphenylalanine ( $F_{\text{fact}}$ , Figure 1.18a).<sup>[162]</sup> The  $\alpha$ -fluoroketone-bearing  $F_{\text{fact}}$  was designed so that the presence of an adjacent carbonyl group, makes the otherwise inert C-F bond weakly electrophilic. Consequently, the fluoride ion can be substituted in an  $S_N2$ -manner with nearby cysteine residues. Incorporated by using an *Mj*TyrRS mutant,  $F_{\text{fact}}$  was used in several different studies including covalent entrapment of an affibody to its substrate Z protein *in vitro* and enhancement of the photostability of the red fluorescent protein mPlum by covalently linking both ends of its beta-barrel. Both experiment required strategic placement of a cysteine residue in close proximity to the UAA through mutagenesis. Furthermore,  $F_{\text{fact}}$  was used in combination with the photocrosslinker UAA pAzF (Figure 1.18a) for mapping the interaction between the corticotropin-releasing factor receptor type 1 (CRF1R) and its ligand urocortin 1 (Ucn1) in HEK293 cells.<sup>[52]</sup> pAzF was first incorporated

at different positions in the transmembrane domain of CRF1R and irradiated in the presence of Ucn1 in order to unveil their interaction area. F<sub>fact</sub> was then incorporated into different position in the receptor and crosslinked with corresponding cysteine mutants of the ligand which helped elucidate the binding orientation. Combined together, these results provided a complete conformational model of the CRF1R:Ucn 1 complex and insights into the mechanism of receptor activation and function. At the same time this study underlined the mutual complementarity of the two conceptually different crosslinking UAAs.



**Figure 1.18.** An overview of successfully incorporated UAAs that undergo proximity-triggered crosslinking reactions with nearby nucleophilic residues. \*The amino acid FnbY was synthesized as a racemic mixture and used as such throughout the incorporation experiments.

Various other UAAs that target proximal cysteine residues have been designed as well. These include a series of chloro-, bromo- and iodoalkyl tyrosine derivatives with varying alkyl linker lengths<sup>[163a]</sup> and a fluorinated lysine derivative (Figure 1.18b,c).<sup>[163e]</sup> The crosslinking efficiencies of these UAAs were shown in *in vitro* proof-of-principle experiments on the affibody-Z protein complex ( $K_d = 6 \mu\text{M}$ ), based on the known 3D structure. Pairwise substitution of the affibody with UAAs, incorporated site-specifically using different PylRS mutants, and Z-protein with cysteines afforded covalently crosslinked complexes at physiological pH. The crosslinking efficiencies of the tyrosine derivatives displayed a trend (IprY > BprY > CprY) consistent with the order of leaving abilities of halide groups in  $S_N2$  reactions.

Cysteine residues were also targeted as a way of introducing photoswitchable thioether bridges through site-specific introduction of azobenzene-containing UAAs Cl-PSCaa and F-PSCaa (Figure 1.18d,e).<sup>[163c, 163d]</sup> Both of these UAAs undergo a nucleophilic aromatic substitution reaction with cysteines and were incorporated using the same *MmPylRS* mutant in *E. coli* and mammalian cells. Cl-PSCaa was introduced into the central helix of calmodulin (CaM), while a cysteine residue was introduced seven positions further. The intramolecular thioether bridge, whose formation occurred spontaneously during protein expression, was photoisomerized upon exposure to UV light (365 nm), altering the conformation of CaM.<sup>[163c]</sup> F-PSCaa showed similar results, where irradiation with green light (540 nm) drives the *trans*-to-*cis* isomerisation of the photoswitch, while the blue light (405 nm) triggers the isomerisation back to the more stable *trans* isomer.<sup>[163d]</sup>

Considering that the UAAs described so far rely on their reactivity with the strong nucleophilic sulfhydryl group of cysteines, their use requires introduction of cysteines *via* site-directed mutagenesis which are rarely exposed on protein surfaces. An UAA containing a long, linear and flexible bromoalkyl side chain was therefore conceived (BrC6K) with its electrophilicity adjusted to target proximal lysine and histidine residues apart from cysteines (Figure 1.18f,  $n = 5$ ). Efficient inter- and intramolecular crosslinking between BrC6K and cysteines was shown at physiological pH, however crosslinking reactions with lysines and histidines required basic conditions (pH 8.8) and typically yielded only 10-20 % of crosslinked complexes. Apart from several proof-of-principle crosslinking experiments, BrC6K was used to irreversibly bind an affibody to its native human epidermal growth factor receptor 2 (HER2) on the surface of breast cancer cells.<sup>[163b]</sup>

#### 1.5.1.2. Michael acceptors

A couple of UAAs containing Michael acceptors were designed in order to target other nucleophilic amino acid residues besides cysteines (Figure 1.18g,h). When incorporated, UAAs *N*<sup>ε</sup>-acryloyl-(*S*)-lysine (AcrK) and *p*-vinylsulfonamido-(*S*)-phenylalanine (VSF), have shown reactivity towards  $\epsilon$ -amino groups of proximal lysine residues.<sup>[164]</sup> The benefit of these acrylamide- and vinyl sulfonamide-containing UAAs lies in their relatively low nonspecific reactivity towards other cellular proteins. VSF, the more reactive amino acid among the two, was introduced site-specifically into the antigen-binding fragment (Fab) of Herceptin. This

allowed covalent crosslinking to its native HER2 receptor *in vitro* and on cell surfaces. Although the crosslinking efficiency with lysine was reported to be 86 % at pH 7.4 *in vitro*, VSF showed certain reactivity towards glutathione (GSH) during expression of Fab which can limit its applicability.

#### 1.5.1.3. Other reactive UAAs

Other UAAs bearing moderate electrophiles such as aryl isothiocyanates<sup>[165]</sup>, aryl carbamates<sup>[166]</sup> and aryl fluorosulfates<sup>[167]</sup> have been reported to form stable intra- and intermolecular protein crosslinks (Figure 1.18i-l). While aryl isothiocyanate-bearing UAA (pNCSF) shows selectivity towards primary amine residues, genetically encoded aryl carbamate UAAs (PheK and FPheK) exhibit crosslinking reactivity towards proximal lysine, cysteine and tyrosine residues, forming stable intramolecular urea, thiocarbamate and carbamate linkages, respectively.<sup>[166]</sup> The reaction efficiencies of these UAAs, however, are highly pH-dependent with pH 8.5 needed for decent crosslink formation. A genetically encoded aryl fluorosulfate UAA (FSY) similarly showed inter- and intramolecular crosslinking reactivity towards lysine, histidine and tyrosine residues *via* proximity-enabled sulfur-fluoride (SuFEx) reaction, although under physiological conditions.<sup>[167]</sup> The crosslinking properties of FSY were established in several proof-of-principle experiments *in vitro* and *in vivo*.

The proximity-enabled crosslinking strategy was very recently combined with the unmasking properties of photocaging functional groups in form of FnbY, an UAA able to target multiple nucleophilic residues (Figure 1.18m).<sup>[168]</sup> The monofluorinated FnbY contains an *o*-nitrobenzyl photocage moiety which gets cleaved off upon UV irradiation (302 or 365 nm) to spontaneously afford a reactive *p*-quinone methide intermediate after displacement of the fluoride ion (for more information on quinone methides see Chapter 2). The *in situ* generated quinone methide readily undergoes Michael addition reactions with nearby nucleophilic residues including cysteine, lysine, histidine, tyrosine, tryptophan, methionine, arginine, asparagine and glutamine. FnbY was successfully incorporated by a *MbPylRS* mutant into *E. coli* and mammalian cells and showed higher crosslinking efficiencies *in vitro* and *in vivo* (examined *via* dimerization of glutathione S-transferase (GST)) than standard photocrosslinking UAAs (Figure 1.16b).

To conclude this general overview, it is important to mention that the last couple of years have seen a rapid expansion of the toolbox of UAAs that can undergo proximity-enabled crosslinking reactions within and between proteins. The majority of these UAAs, however, were established through *in vitro* crosslinking experiments of either easily overexpressed proteins (e.g. intramolecular crosslinking of fluorescent proteins or dimerization of GST) or well-established stable complexes (e.g. affibody-Z protein complex), often at elevated pH. Only recently, has this approach been expanded by us and others towards *in vivo* entrapment of biologically relevant protein-protein interactions.<sup>[53, 163f]</sup> As described in this thesis, we successfully employed BrC6K and its analogues (Figure 1.18f) for structural elucidation of previously inaccessible low-affinity protein complex.<sup>[53]</sup> More about

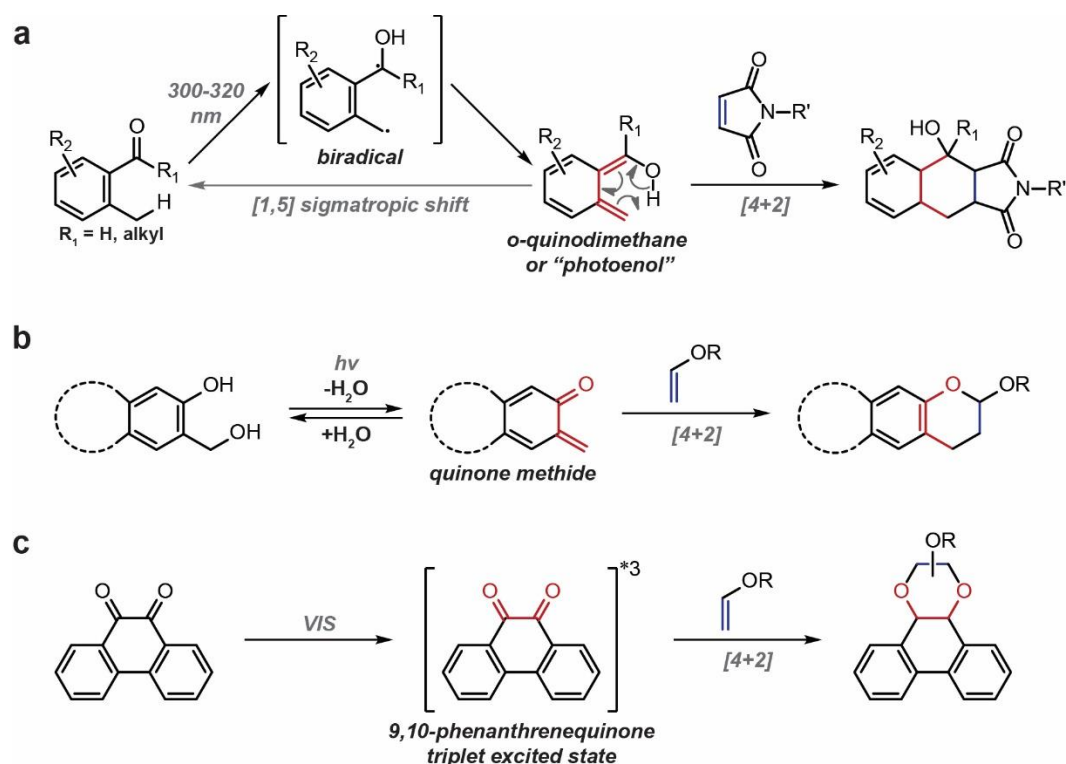
the rationale and the experimental design of this specific application can be found in Chapter 3. A similar approach was used by Lei Wang and co-workers for identification of protein-protein interactions in *E. coli* via mass spectrometry.<sup>[163f]</sup> The bifunctional UAA EB3 (Figure 1.18n), bearing a reactive bromide and an alkyne moiety, can be captured in a covalently linked complexes by proximal cysteine residues of unknown interacting proteins. The alkyne moiety allows specific biotinylation of the crosslinked protein partners via CuAAC and, after protease digestion, enrichment of crosslinked peptides. Finally, mass spectrometry was used for successive analysis and identification of interacting proteins. The aforementioned examples opened up possibilities for practical implementation of crosslinking UAAs in living systems and under native conditions. Combined with high-resolution analysis methods, these are likely to become important tools in synthetic and structural biology.

---



## § 2. DEVELOPMENT OF A NOVEL LIGHT-INDUCED BIOORTHOGONAL REACTION FOR SITE-SPECIFIC PROTEIN LABELLING

The general appeal of using light for inducing bioorthogonal reactivity emanates from the convenient spatiotemporal control over the reaction initiation. While light has played a central role as an external stimulus in ligations based on tetrazoles and cyclopropanones, discussed in Chapter I, light-induced Diels-Alder cycloadditions have mostly been overshadowed by the fastest bioorthogonal reaction to date – the tetrazine ligation – whose light-inducible versions have recently been developed as well.<sup>[169]</sup> Considering the prominence of photoinduced [4+2] cycloadditions in organic synthesis (especially the synthesis of natural products), a few examples of water-compatible reactions have emerged, showing potential for biological applications (Scheme 2.1).



**Scheme 2.1.** Overview of light-induced Diels-Alder cycloaddition reactions. (a) Irradiation of *o*-alkyl phenyl aldehydes or ketones affords *o*-quinodimethanes which either rapidly revert back *via* 1,5-sigmatropic shift or undergo [4+2] cycloaddition with electron-deficient olefins such as maleimides. (b) [4+2] cycloaddition between photochemically generated *o*-quinone methides and vinyl ethers. (c) Visible light (VIS)-mediated [4+2] cycloaddition between 9,10-phenanthrenequinones and vinyl ethers.

One example rests on the irradiation of *o*-alkyl phenyl ketones and aldehydes with UV light (300-320 nm) which generates biradicals that rearrange to hydroxy-*o*-quinodimethanes

or “photoenols”.<sup>[136]</sup> These highly reactive dienes can subsequently react with electron-deficient olefins (e.g. maleimides) to form stable cycloaddition products (Scheme 2.1a).<sup>[170]</sup> Due to the susceptibility of photoenols towards nucleophilic attacks, however, only preliminary studies have so far been conducted including the condensation of maleimide-oligopeptide conjugate with 2-formyl-3-methylphenoxy-derived photoenol in PBS/ACN mixture, under irradiation at 320 nm.<sup>[171]</sup> The other two examples of light-triggered DA reactions embrace the reactivity of electron-rich vinyl ethers with photochemically generated *ortho*-quinone methides (*o*QMs, Scheme 2.1b)<sup>[137]</sup> or with 9,10-phenanthrenequinones (PQs) mediated by visible light (Scheme 2.1c).<sup>[138]</sup> The unique features that these two reactions exhibit, the chemical nature of the reactive intermediates and the scope of their applicability to biological systems play an important role in the making of this thesis.

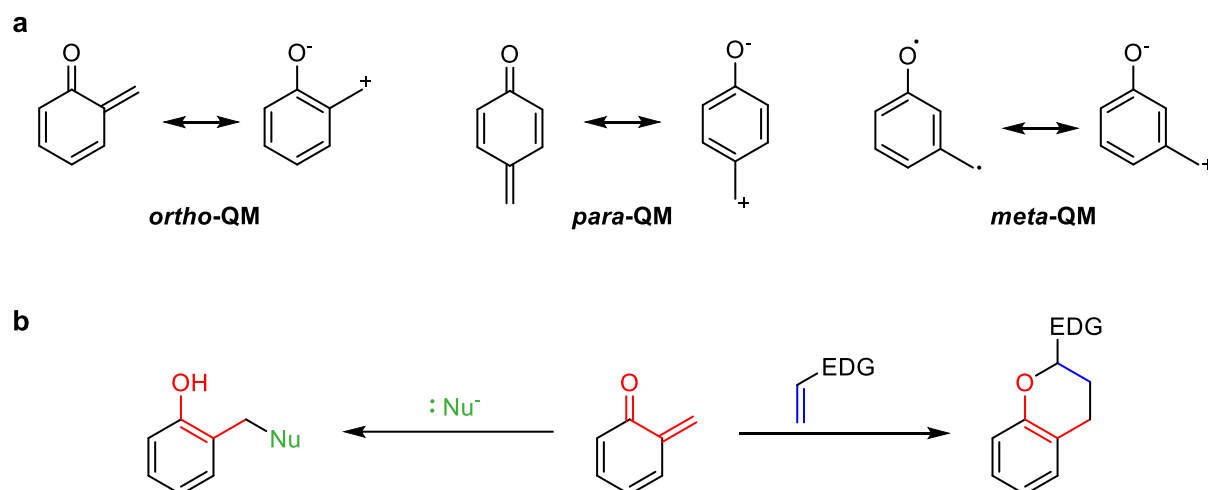
## 2.1. Quinone methides (QMs)

Quinone methides (QMs) are a class of reactive intermediates widely encountered in the chemistry of phenols and related compounds that have received much attention due to their pronounced biological activity.<sup>[172]</sup> Formation of QMs is, for example, deemed to be a crucial step in the mechanism of action of several antitumor antibiotics.<sup>[173]</sup> Structurally related to quinones, the most common QM isomers involve *ortho*- and *para*- orientation of the carbonyl and the methylene group (Scheme 2.2a). Although formally neutral, these molecules are short lived species that exhibit distinctive reactivities as a result of enhanced polarization between the carbonyl and the methylene group. Accordingly, *ortho*- and *para*-QMs can be essentially viewed as benzylic carbocations that are resonantly stabilized by the electron-donating phenoxide ion. Non-Kekulé *meta*-isomers on the other hand are less stable and often transient or difficult to generate as they do not possess direct orbital interactions between the oxygen and the methylene group and are thus represented by a biradical structure (Scheme 2.2a).<sup>[174]</sup> As a result of their unique chemical nature, QMs are susceptible to Michael additions at the electrophilic methylene group, while *ortho*-isomers can undergo [4+2] cycloaddition reactions with electron-rich dienophiles affording chromanes as products.<sup>[172b]</sup> The driving force for both of these types of reactions is aromatisation (Scheme 2.2b).

Generally speaking, QMs can be prepared by various synthetic methods including thermal dehydration or extrusion of small molecules,<sup>[175]</sup> oxidation of phenols,<sup>[176]</sup> reductive elimination of quinones<sup>[177]</sup> and photolysis.<sup>[178]</sup> *o*QMs are readily prepared with the latter method through photochemical elimination of *o*-hydroxybenzyl alcohols and analogues such as ethers and Mannich bases at 254 nm.<sup>[172b, 179]</sup> The mechanism of this reaction in aqueous media was extensively studied by nanosecond laser flash photolysis (LFP) and is thought to involve an excited state intramolecular proton transfer (ESIPT) of a phenolic proton to the oxygen atom (or heteroatom) in the benzylic position, followed by heterolysis of the C-O bond.<sup>[172a, 179d, 180]</sup> The higher acidity of phenols in the excited state ( $pK_a \sim 4$ ) as opposed to the ground state ( $pK_a \sim 10$ ) expedites the proton transfer. Although these analogues have been developed as photolabile protection groups,<sup>[181]</sup> reliance on UV-C light limits their overall applicability. 3-hydroxymethyl-2-naphthols and their derivatives, on the other hand, generate

---

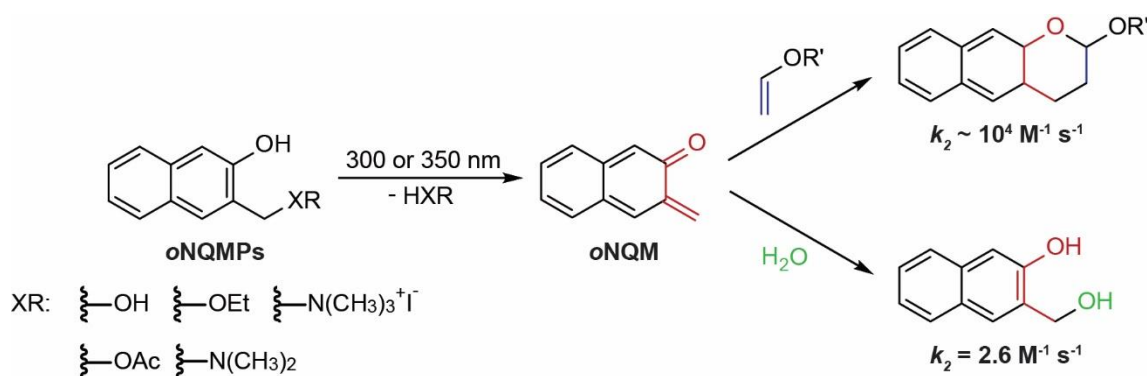
*o*-naphthoquinone methides (*o*NQMs) by employing light higher or equal to 300 nm, making them attractive for potential biological applications.<sup>[137b, 182]</sup> Furthermore, the ESIPT is facilitated by the higher excited state acidity of *o*-naphthols. Indeed, Popik and co-workers have successfully established the photochemical generation of *o*NQMs and their subsequent trapping by electron-rich alkenes as a bioconjugation reaction<sup>[137a, 183]</sup> that we set out to expand towards site-specific labelling of proteins *via* incorporation of corresponding UAAs.



**Scheme 2.2.** Structures and reactivity of quinone methides (QMs). (a) *Ortho*-, *para*- and *meta*-isomers of QMs. (b) Due to polarisation, QMs readily react with nucleophiles while *ortho*-isomers also engage in [4+2] cycloadditions with electron-rich alkenes.

## 2.2. Light-induced hetero-Diels-Alder reaction between *o*NQMs and vinyl ethers

As already mentioned, *o*NQMs can easily be prepared *in situ* by irradiation of aqueous solutions of 3-hydroxymethyl-2-naphthol and its derivatives (*o*NQM precursors or *o*NQMPPs) at 300 or 350 nm within minutes. The lifetime of *o*NQM in aqueous solutions is very short (only ~ 7 ns) as it rapidly and quantitatively reacts with water, regenerating thereby the starting material ( $k_2 = 2.6 \text{ M}^{-1} \text{ s}^{-1}$ ).<sup>[137b]</sup> Nevertheless, in the presence of electron-rich polarized alkenes such as vinyl ethers, *o*NQM readily undergoes a hetero-Diels-Alder reaction with inverse-electron demand to afford substituted 2-alkoxy-3,4-dihydro-2*H*-naphtho[2,3-*b*]pyrans as the only products (Scheme 2.3). The reported rate of this reaction is in the order of  $10^4 \text{ M}^{-1} \text{ s}^{-1}$ , which is exceptionally fast (comparable to iEDDA reactions between tetrazines and strained alkenes/alkynes) and, importantly, outcompetes the hydrolysis reaction. Furthermore, the naphthopyrane products are stable towards prolonged irradiations and only undergo slow hydrolysis at pH 1.<sup>[137a]</sup>



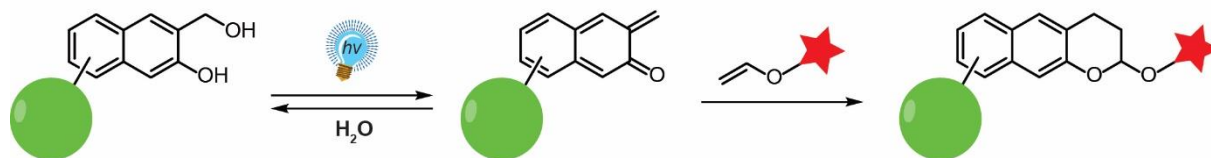
**Scheme 2.3.** Formation of *o*NQMs via photoelimination of corresponding precursors (*o*NQMPs) and their subsequent reactivity with vinyl ethers and water.

Susceptibility of QMs towards nucleophilic attacks under aqueous, physiological conditions has been well documented with the lifetime of QMs and, hence, their reactivities determined by the rates of the hydration reaction. Different QMs have been shown to efficiently react with nucleophilic amino acids such as cysteines, lysines and tyrosines as well as N-terminal residues, either as free amino acids<sup>[184]</sup> or in peptides and proteins<sup>[185]</sup> with little or no competition from QM hydration. Furthermore, QMs are capable of alkylating nucleosides<sup>[186]</sup> and have displayed DNA alkylating and crosslinking actions as well.<sup>[187]</sup> Similarly, Popik and co-workers have shown that *o*NQMs readily undergo Michael additions with thiols ( $k_2 = 2.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ) and azides ( $k_2 = 2.0 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ) which are either faster or slightly slower than the corresponding photocycloaddition reaction with vinyl ethers ( $k_2 = 4 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ).<sup>[137]</sup> Considering that thiols are important, ubiquitously present biological nucleophiles (e.g. glutathione), this cross-reactivity would severely obstruct the development of selective labelling approaches based on *o*NQMs. Nevertheless, irradiation of respective *o*NQM-thioether adducts at 300 or 350 nm regenerates reactive *o*NQMs which can then be irreversibly trapped by vinyl ethers, making the thiol addition reaction *de facto* reversible.<sup>[137a]</sup> Indeed, this feature has been used for selective and reversible derivatisation of cysteines in peptides and bovine serum albumin (BSA).<sup>[183b]</sup> Apart from selective labelling of BSA containing chemically introduced vinyl ether moieties,<sup>[183b]</sup> the light-induced hetero-Diels-Alder cycloaddition of *o*NQMs was furthermore used for spatiotemporally controlled derivatisation of surfaces immobilised with vinyl ether moieties.<sup>[183a]</sup> The orthogonality of the photo-reversible *o*NQM-thiol as well as the *o*NQM-vinyl ether photocycloaddition to SPAAC or CuAAC was also demonstrated on sequential labelling of patterned surfaces.<sup>[188]</sup>

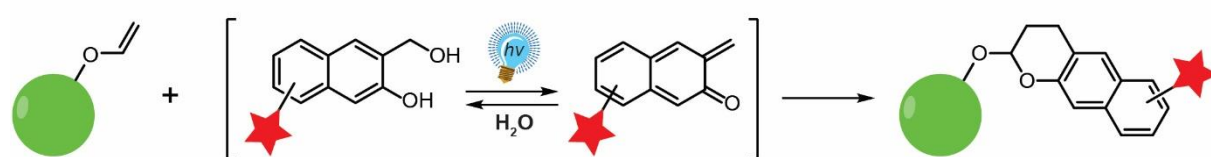
Due to promising reports in the past, we set out to adapt the light-induced hetero iEDDA reaction between *o*NQMs and vinyl ethers first as a proof-of-principle *in vitro* reaction on purified proteins with the potential for expansion towards live cell labelling in both *E. coli* and mammalian cells later on. We, therefore, opted for the design of lysine-based UAAs that would be incorporated by respective PylRS/tRNA<sup>CUA</sup> pairs, as this pair is orthogonal in both prokaryotic and eukaryotic systems. We approached the UAA incorporation conundrum by envisioning two different strategies, depending on which

reaction partner would be incorporated as an UAA and which utilized as an externally added probe (e.g. fluorophore or biotin derivative) (Figure 2.1).

### Labelling strategy I



### Labelling strategy II



**Figure 2.1.** Devised strategies for site-specific labelling of molecules *via* light-induced inverse electron-demand hetero-Diels-Alder reaction between *o*NQMs and vinyl ethers. By means of genetic code expansion, UAAs bearing either *o*NQMPs (Labelling strategy I) or vinyl ethers (Labelling strategy II) are site-specifically introduced into the protein of interest (green sphere) in place of a premature amber stop codon in the mRNA. Irradiation of the *o*NQMP-containing protein affords *o*NQM which rapidly reacts with externally added vinyl-ether-containing labelling reagent (fluorophore or biotin tag represented by the red star) to give the desired cycloaddition product (Labelling strategy I). Conversely, proteins containing the vinyl-ether moiety react with externally added fluorophore- or biotin-containing *o*NQMPs after irradiation with UV light (Labelling strategy II).

From the start we expected the strategy of incorporating UAAs containing *o*NQMPs precursors (Labelling strategy I, Figure 2.1) to be more challenging; not only because it would involve multistep syntheses of different UAAs but it would likely require directed evolution for finding specific PylRS variants for the respective UAAs. Directed evolution is typically the bottleneck when designing novel UAAs for site-specific incorporation as it may or may not necessarily result with positive hits. However, we were encouraged by recent reports of successful incorporation of bulky aromatic UAAs by PylRS variants with enlarged amino acid binding pockets.<sup>[189]</sup> Furthermore, strategy I benefits from generating highly reactive species directly on a protein of interest, offering real spatial control as they cannot diffuse away. The addition of high excess of these intermediates that could, for example, undergo cross-reactions with endogenous nucleophiles or pose risk of cytotoxicity is thereby also avoided. In case the directed evolution approach would fail to deliver any PylRS variants specific for *o*NQMP-bearing UAAs, we aimed to simultaneously pursue the synthesis of vinyl ether UAAs (Labelling strategy II, Figure 2.1), knowing that the incorporation of structurally similar UAAs has been reported in the literature.<sup>[146]</sup> In terms of synthesis, these UAAs are also more easily obtainable. The following sections recount in detail how the two labelling strategies evolved with the rationale behind the performed experiments and give an overview of obtained results. Before focusing on these two strategies, however, it was of crucial

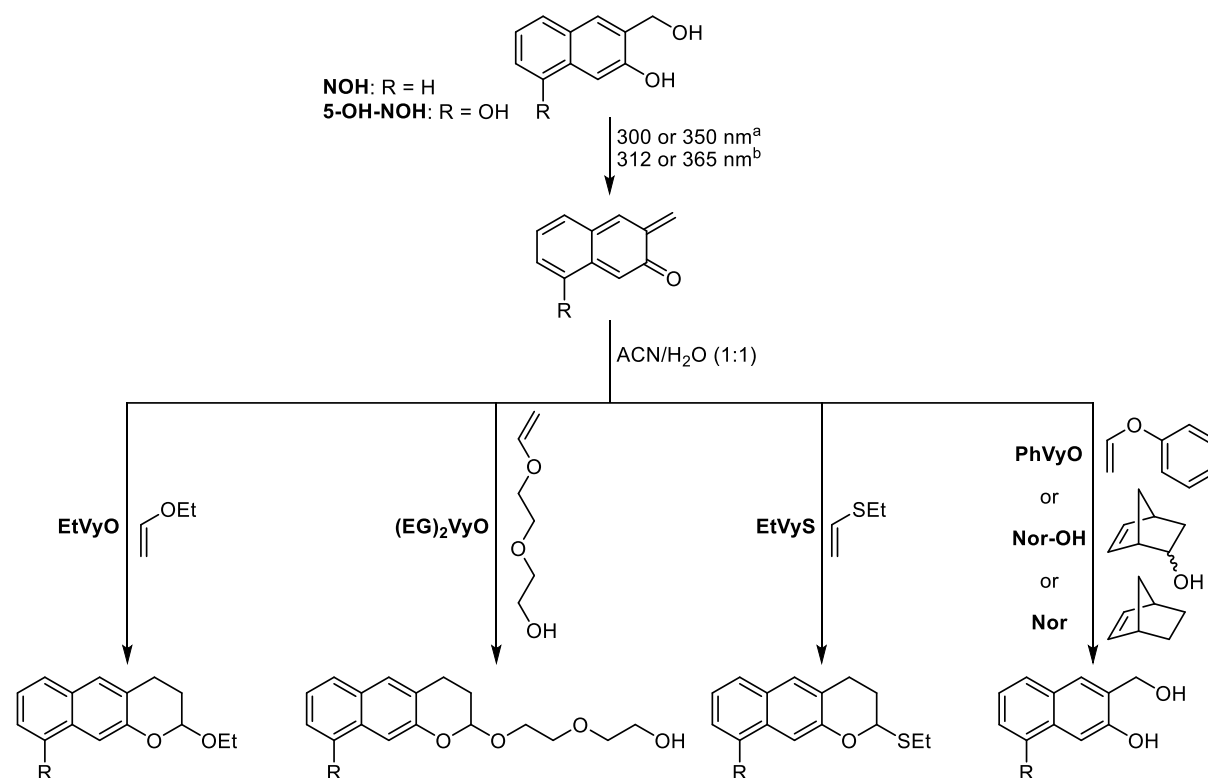
importance to establish and optimize a proper experimental procedure for photochemical generation of *o*NQMs.

### 2.3. Building the initial setup

In their original publications, the analytical photolysis experiments by Popik and co-workers involved irradiation of 0.1 mM solutions of 3-hydroxymethyl-2-naphthol (NOH) and 0.15 mM of ethyl vinyl ether (EtVyO) in 50 % ACN and 0.01 N aqueous biphosphate buffer (3 mL) respectively at 300 nm with  $8 \times 4$  W fluorescent tubes. Irradiation times of 3 minutes showed 15 % conversions, while more than 99 % conversion of the starting material was observed after 20 minutes.<sup>[137a]</sup> In order to establish the reaction in our lab, we first focused on analytical experiments between different *o*NQMPs and vinyl ethers, vinyl sulfides or even strained alkenes. These experiments were carried out in custom-made borosilicate tubes (Duran<sup>®</sup> glass) at 300 or 350 nm (Rayonet photochemical reactors containing a cylindrical array of 16 UV lamps with nominal power of 8 W, courtesy of the group of Prof. Thorsten Bach). Since this reaction setup is suitable for small molecule chemical transformations only and thus requires higher volumes of reaction mixtures, we also had to evaluate other light sources that would eventually be compatible with experiments performed in small volumes (< 1 mL) on purified proteins. Therefore, we chose broad-emitting high-power UV lamps ( $2 \times 15$  W) with  $\lambda_{\text{max}}$  of 312 or 365 nm that allow irradiation of reaction mixtures in cell culture plates from the top (see Chapter 4.3.1 for detailed information). The results of our initial analytical photolysis experiments, analysed by HPLC or LC-MS, are summarised in Scheme 2.4 and Table 2.1 while the respective chromatograms can be found in the Appendix (Supporting Figures S2.1-S2.14). The conversions of starting *o*NQMPs presented in Table 2.1 were approximately calculated from HPLC traces of analysed reaction mixtures. It should be noted that each of the reaction components had been irradiated on their own to exclude any possible side-reactions. Furthermore, the reactions performed without irradiation served as negative controls.

First photolysis experiments included irradiation of 1 mM solutions of NOH in 50 % aqueous ACN (3 mL) in the presence of different amounts of EtVyO (Entries 1-5, Table 2.1 and Figures S2.2-S2.6). Conversions higher than 80 % were achieved after irradiating the NOH solutions at 300 nm for 20 minutes with more than 10 equivalents of EtVyO added. Approximately 99 % conversions were achieved in the presence of 100 equivalents of EtVyO (Entry 5, Table 2.1 and Figure S2.6). As expected, the irradiation of NOH under the same conditions at 350 nm takes longer in order to achieve the same conversion as at 300 nm (Entry 6, Table 2.1 and Figure S2.7) as NOH absorbs only weakly above 350 nm.<sup>[137]</sup> All of these reactions were generally clean, affording the benzochromane as the major product, as determined by MS and NMR. Only minor secondary photoproducts were observed which are likely *o*NQM oligomers.<sup>[137b, 190]</sup> Same photoproducts were also observed when irradiating NOH on its own (Figure S2.1) or, more prominently, when the amounts of both NOH and EtVyO were increased 10 times (data not shown). When testing other olefinic targets for the hetero-DA reaction with *o*NQMs, we observed faster reactions with 10 mM ethyl vinyl

sulfide (EtVyS) with almost quantitative conversion after irradiation for 20 minutes at 300 nm (Entry 7, Table 2.1 and Figure S2.8). Contrarily, phenyl vinyl ether (PhVyO) did not give any benzochromane products (that is, only NOH was observed as a result of nucleophilic attack of water to the *o*NQM) and neither did strained alkenes such as norbornene (Nor) and 5-norbornen-2-ol (Nor-OH, *endo* and *exo* mixture; Entries 8-10, Table 2.1 and Figures S2.9-S2.11). Altogether, these observations confirm that *o*NQMs only form Diels-Alder cycloaddition adducts with electron rich polarized olefins.<sup>[137]</sup>



**Scheme 2.4.** Overview and outcome of analytical photolysis experiments between photochemically generated *o*NQMs and different olefins. In aqueous solutions, *o*NQMs undergo inverse electron-demand hetero-Diels-Alder cycloaddition reactions only with electron rich polarized olefins such as vinyl ethers and sulfides. Only hydration reactions of *o*NQMs are observed in case of strained alkenes or phenyl vinyl ether. <sup>a</sup> Rayonet photoreactor with a cylindrical array of UV lamps (16 × 8 W); <sup>b</sup> broad-emitting high-power UV lamp (2 × 15 W).

As mentioned earlier, the reaction setup which requires irradiation in a photochemical reactor is not suitable for labelling reactions on purified proteins due to huge differences in the required sample volumes. We, therefore, tested how the broad-emitting high-power UV lamps influence the reaction between NOH and vinyl ethers, when the samples inside cell culture plates are irradiated from the top (Entries 11-16, Table 2.1). Due to the low boiling point of EtVyO (33 °C), irradiation experiments had to be performed with di(ethylene glycol) vinyl ether ((EG)<sub>2</sub>VyO), which boils at 196 °C and is readily soluble in water. Furthermore, the change in the experimental setup allowed stirring of the solution during irradiation so that the light could equally penetrate through the entire mixture. Fortunately, the conversions achieved with the 300 nm lamps from the photoreactor and the 312 nm broad-emitting lamp after same irradiation times are comparable (above 90 %; Entries 5 and 11, Table 2.1 and

Figures S2.6 and S2.12), while irradiation with the broad-emitting 365 nm lamp unsurprisingly requires longer irradiation times (Entry 15, Table 2.1 and Figure S2.14a). Introduction of an alkoxy substituent into *o*NQMP has shown to result in ca. 10 nm bathochromic shift of its absorption bands which, consequently, enables a more efficient formation of *o*NQMs at 350 nm.<sup>[137a]</sup> We tried to compare the difference in reactivity between NOH and its 5-hydroxy derivative (5-OH-NOH) with (EG)<sub>2</sub>VyO depending on the light source, and observed that more of the benzochromane product is formed with NOH. Moderate amounts of by-products were also observed in reactions with 5-OH-NOH (Figures S2.13 and S2.14b). While small molecule analytical photolysis experiments on, for example, alkylated derivatives of 5-OH-NOH would possibly give cleaner reaction mixtures, these were, nevertheless, not pursued further in favour of preparing and characterising the reactivity of *o*NQMP-based fluorophore or biotin probes (described in further sections). These results should, furthermore, be taken with a grain of salt due to the broad-emission spectra of both high-power UV lamps ( $\lambda_{\text{max}}$  312 and 365 nm). Experiments with light sources with emissions focused to a certain wavelength (or at least with smaller spectral widths) would allow superior characterisation of the differences in reactivities between the two *o*NQMPs.

**Table 2.1.** Overview of analytical photolysis reactions between *o*NQMPs and different olefins. All reactions were performed in 50 % aqueous ACN. The conversions of starting *o*NQMPs were determined on the basis of HPLC traces of the analysed reaction mixtures.

Entry	<i>o</i> NQMP (A)	Olefin (B)	Ratio A:B	$\lambda_{\text{max}}$ / nm	Irradiation time / min	% Conversion of <i>o</i> NQMP
1	NOH	EtVyO	1:1.5	300 <sup>a</sup>	20	27
2	NOH	EtVyO	1:10	300 <sup>a</sup>	20	81
3	NOH	EtVyO	1:30	300 <sup>a</sup>	20	93
4	NOH	EtVyO	1:60	300 <sup>a</sup>	20	95
5	NOH	EtVyO	1:100	300 <sup>a</sup>	20	> 99
6	NOH	EtVyO	1:10	350 <sup>a</sup>	20	12
7	NOH	EtVyS	1:10	300 <sup>a</sup>	20	96
8	NOH	PhVyO	1:100	300 <sup>a</sup>	20	n.r. <sup>c</sup>
9	NOH	Nor	1:50	300 <sup>a</sup>	20	n.r. <sup>c</sup>
10	NOH	Nor-OH	1:50	300 <sup>a</sup>	20	n.r. <sup>c</sup>
11	NOH	(EG) <sub>2</sub> VyO	1:100	312 <sup>b</sup>	20	92
12	NOH	(EG) <sub>2</sub> VyO	1:100	312 <sup>b</sup>	40	> 99
13	5-OH-NOH	(EG) <sub>2</sub> VyO	1:100	312 <sup>b</sup>	20	53
14	5-OH-NOH	(EG) <sub>2</sub> VyO	1:100	312 <sup>b</sup>	40	74
15	NOH	(EG) <sub>2</sub> VyO	1:100	365 <sup>b</sup>	20	49
16	5-OH-NOH	(EG) <sub>2</sub> VyO	1:100	365 <sup>b</sup>	20	26

<sup>a</sup> Rayonet photoreactor with a cylindrical array of UV lamps (16 × 8 W)

<sup>b</sup> broad-emitting high-power UV lamp (2 × 15 W)

<sup>c</sup> “n.r.” denotes that only hydration reaction of *o*NQM was observed yielding the starting NOH

Although the above presented photolysis results are not in complete accordance with the published data, especially regarding the irradiation times needed for achieving comparable conversions and required amounts of reaction components,<sup>[137, 183b]</sup> we still managed to

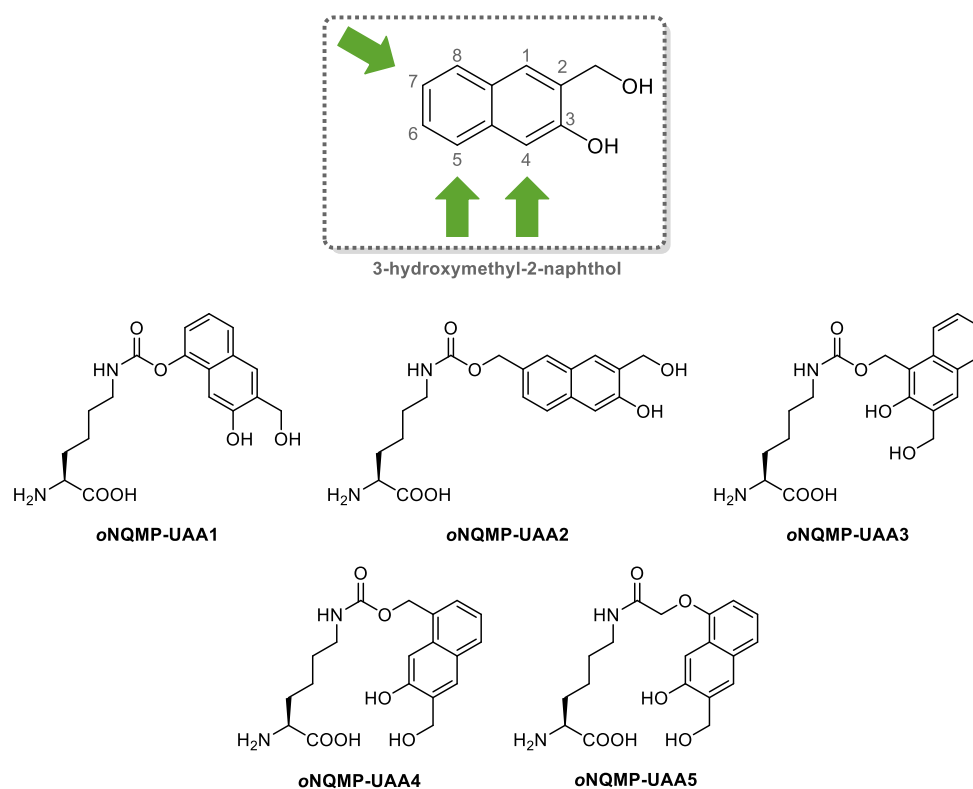


establish the reactivity of *o*NQMPs with vinyl ethers, mediated by light sources that were within our reach. Before moving to irradiation experiments on purified proteins, we needed to perform the small molecule reaction in the presence of nucleophiles such as cysteines, as they are known to undergo fast nucleophilic addition to *o*NQMs.<sup>[137]</sup> This reaction should, however, be reversible in the presence of vinyl ethers as recently shown on peptides and proteins.<sup>[183b]</sup> We, therefore, irradiated NOH in the presence of 100 equivalents of EtVyS and 10 equivalents of Boc-Cys-OMe, for 20 minutes at 300 nm (Figure S2.15). Apart from expected benzochromane product formation, only one by-product was observed – the thio-ene reaction product between excess of Boc-Cys-OMe and EtVyS. Similar results were obtained with EtVyO as well (Figure S2.16). With this data in hand, we were able to move on towards designing and performing labelling experiments as presented by the two labelling strategies in Figure 2.1.

#### 2.4. The rocky road towards obtaining *o*NQM-based UAAs

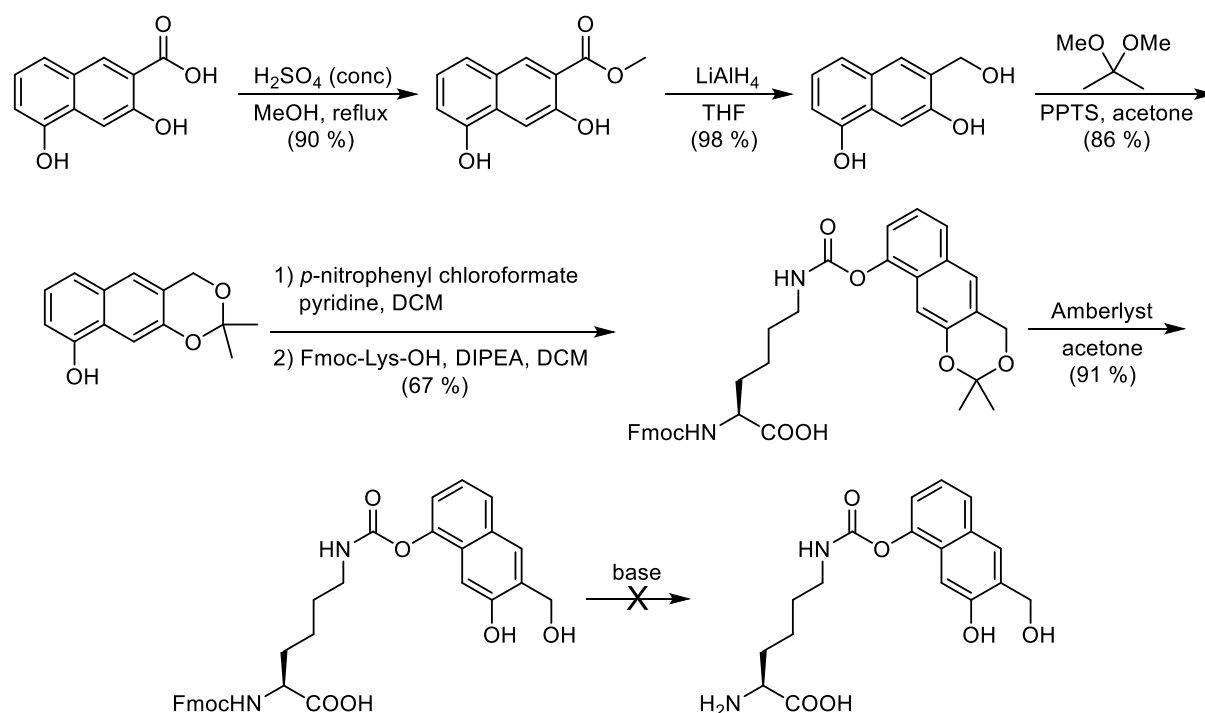
When approaching the design of *o*NQMP-containing UAAs, the most reasonable synthetic pathways involved introduction of necessary chemical modifications to the naphthalene core first, followed by coupling of the *o*NQMP-moiety to the  $\epsilon$ -amino group of L-lysine. Accordingly, it was of vital importance to only perform chemical reactions or introduce appropriate functionalities that would maintain the integrity of the 3-hydroxymethyl-2-naphthol moiety essential for *o*NQM formation. In order to avoid *de novo* synthesis of the naphthalene ring, we focused on positions in 3-hydroxymethyl-2-naphthol that could easily be derivatised (positions 4-, 5- and 7- as depicted in Scheme 2.5) and, correspondingly, designed five different UAAs (Scheme 2.5). It should be noted that we came up with different synthetic ideas and new amino acid designs as a response to stumbling upon various difficulties along the way. These developments are described in the following sections.

---

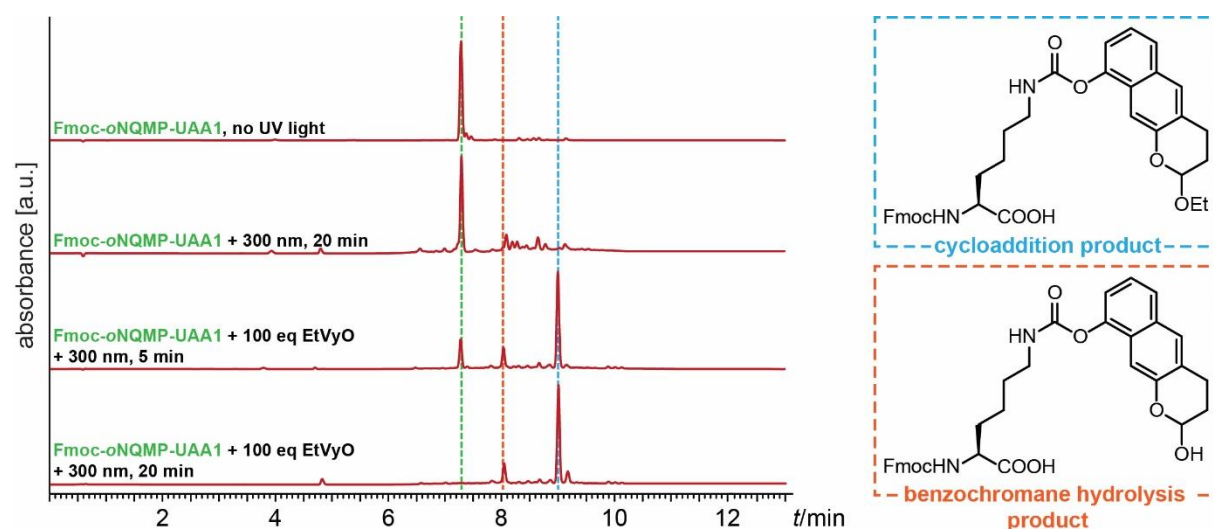


**Scheme 2.5.** Structures of devised UAAs containing *o*NQM precursors (*o*NQMPs).

The design of the first and synthetically easiest UAA, *o*NQMP-UAA1, allowed its preparation from commercially available 3,5-dihydroxy-2-naphthoic acid (Scheme 2.6). The carboxylic acid was first converted into the methyl ester *via* Fischer esterification, which was then reduced into the corresponding alcohol with  $\text{LiAlH}_4$ . The acid-labile acetonide group was subsequently introduced to protect the vicinal hydroxymethyl and naphthol moieties and allow selective derivatisation of the hydroxy group at position 5. After testing different alcohol-activating reagents such as DSC and CDI with poor success rates, we finally succeeded with *p*-nitrophenyl chloroformate in the presence of pyridine. The activated alcohol was directly coupled to *N* $\alpha$ -Fmoc-protected L-lysine without previous isolation to afford the protected *o*NQMP-UAA1. The acetonide protection group was readily removed in the presence of Amberlyst acidic resin while the reactivity of the respective UAA with vinyl ethers was verified upon irradiation at 300 nm. Upon irradiation with 100 equivalents of EtVyO, Fmoc-protected *o*NQMP-UAA1 predominantly afforded the desired benzochromane cycloaddition product (Figure 2.2). A minor peak (retention time around 8 minutes) corresponding to the benzochroman-2-ol was observed as well. This occurrence benzochromane hydrolysis is undesired as it would essentially cleave the fluorophore or biotin label from the UAA but the amount of the by-product did not increase with prolonged irradiation time. In spite of good reactivity with vinyl ethers, *o*NQMP-UAA1 was unfortunately unstable towards deprotection conditions of the *N* $\alpha$ -amine, as these resulted with cleavage of the carbamate moiety.



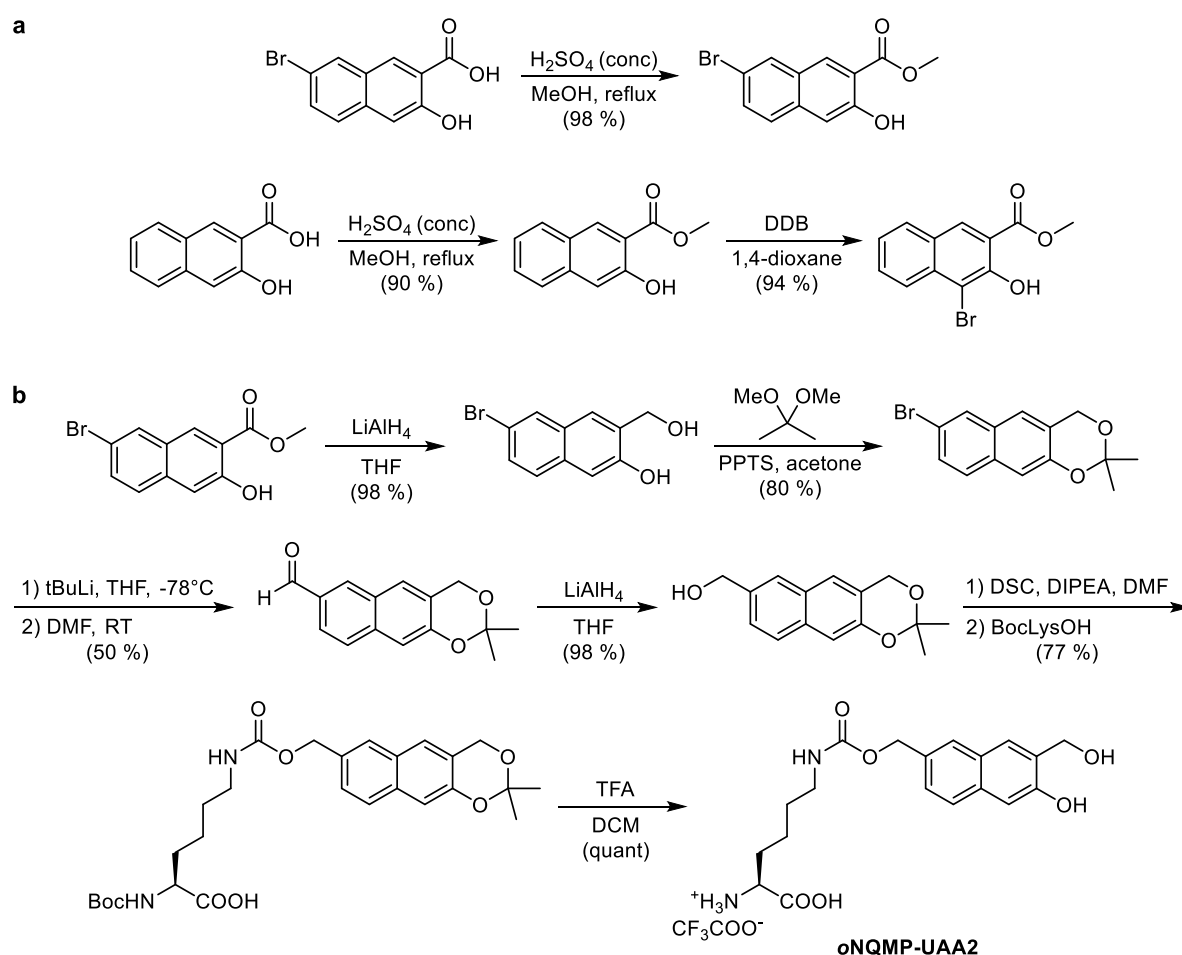
**Scheme 2.6.** Synthesis of *o*NQMP-UAA1. Due to undesired cleavage of the carbamate linkage during the final deprotection step, the synthetic pathway was never completed.



**Figure 2.2.** Reactivity of *N* $\alpha$ -Fmoc-protected *o*NQMP-UAA1 with ethyl vinyl ether (EtVyO). Irradiation of 1 mM of the UAA and 100 mM of EtVyO in 50 % aqueous ACN at 300 nm affords the desired benzochromane – product of the [4+2] cycloaddition (blue dashed line). Small amount of undesired hydrolysed benzochromane by-product, benzochroman-2-ol, was observed as well (orange dashed line). UV absorbance is measured at 280 nm. As a result, peaks corresponding EtVyO are not observable.

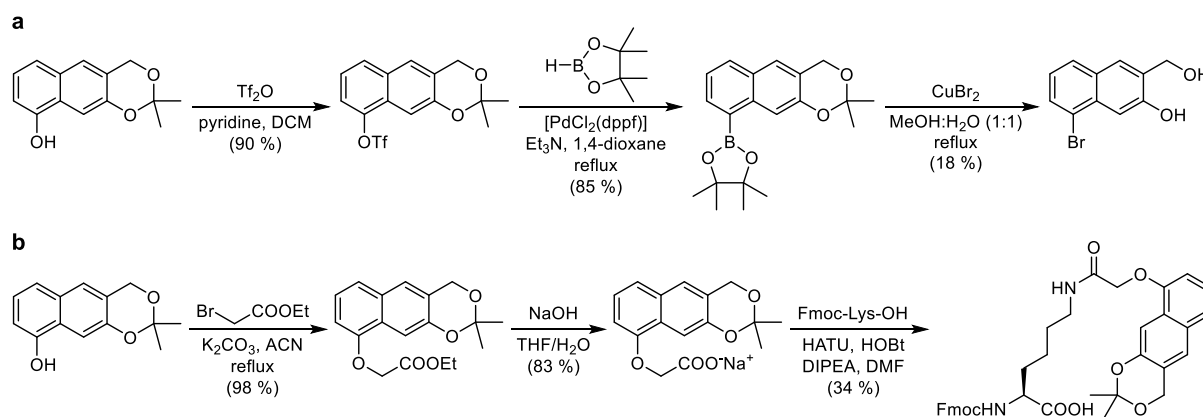
In order to make the UAAs more stable, we decided to couple L-lysine *via* carbamate linkages between the  $\epsilon$ -amine and hydroxymethyl groups introduced at different positions of the naphthalene core (*o*NQMP-UAAs 2-4). The most important synthetic steps towards *o*NQMP-UAA 2 and 3, presented in Scheme 2.7, included selective bromination of naphthalene precursors, halogen-lithium exchange followed by electrophilic quench with DMF in order to introduce a formyl group at respective positions and subsequent reduction of

the aldehyde into a free alcohol. Both compounds monobrominated either at position 4 or 7 can be prepared from 3-hydroxy-2-naphthoic acid (Scheme 2.7). While 1,4-dioxane dibromide (DDB) selectively brominates position 4 of the corresponding methyl ester,<sup>[191]</sup> 7-bromo-3-hydroxy-2-naphthoic acid, although commercially available, is typically prepared by treating the naphthoic acid with excess of bromine in glacial acetic acid, followed by selective debromination of the 4,7-dibromo hydroxyl acid with tin and HCl.<sup>[192]</sup> Monobrominated 3-hydroxy-2-naphthalene derivatives are converted into corresponding acetonides by a series of reactions already encountered during synthesis of *o*NQMP-UAA1: Fisher esterification (in case of the 7-bromo derivative), reduction with LiAlH<sub>4</sub> and ketal protection of the 1,3-diol in the presence of 2,2-dimethoxypropane (DMP). Subsequent rapid bromine-lithium exchange with *n*BuLi at -78 °C followed by quenching with DMF affords the aldehyde which is easily reduced to give the desired hydroxymethyl functionality. Finally, the free hydroxy group is activated with DSC and coupled to ε-amino group of *N*α-Boc-protected L-lysine to afford corresponding *o*NQMP-UAA2 and 3.



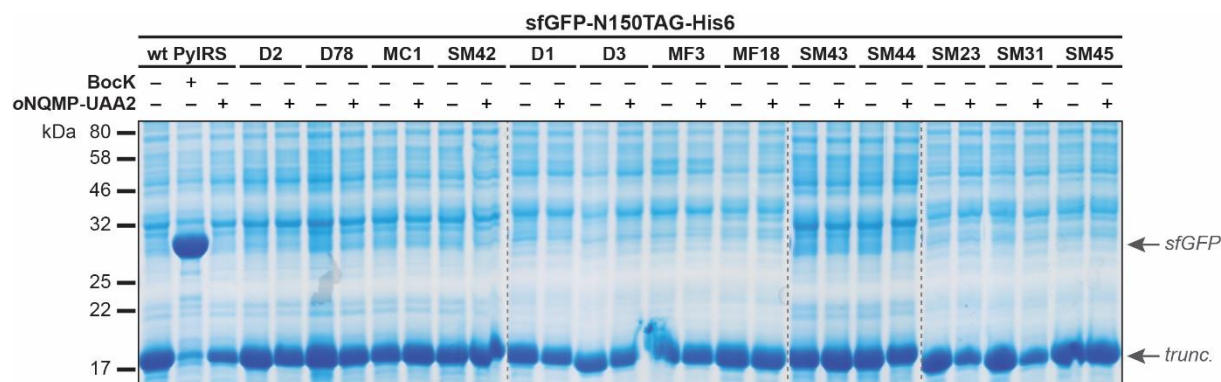
**Scheme 2.7.** Synthesis of *o*NQMP-UAA2 and 3. (a) Preparation of monobrominated methyl 3-hydroxymethyl-2-naphthoates – precursors for the desired UAAs. (b) Synthetic steps for obtaining *o*NQMP-UAA2 starting from the respective monobrominated precursor. The synthesis of *o*NQMP-UAA3 proceeds under analogous conditions from its respective precursor and is thus not shown. The yields of each step involved in the preparation *o*NQMP-UAA3 are as follows: ester reduction (99 %), acetonide protection (72 %), formylation (54 %), aldehyde reduction (92 %), coupling to Boc-Lys-OH and final deprotection (63 % over two steps).

Selective bromination of the position 5 had to be approached from a different synthetic perspective as electrophilic aromatic substitution reactions on 3-hydroxy-2-naphthoic acid are typically favoured at positions 4 and 7. We ultimately decided to target the 5-hydroxy group of acetonide-protected 3,5-dihydroxy-2-naphthol by converting it first into triflate ester followed by palladium catalysed Miyaura borylation with pinacolborane.<sup>[193]</sup> Preparation of the triflate ester was tried on methyl 3,5-dihydroxy-2-naphthoate as well, however, consumption of the starting material was not observed. We thereupon attempted to transform the boronate ester into the corresponding aryl bromide using copper(II) bromide,<sup>[194]</sup> however, the used conditions were not favourable for the acetonide protecting group and only minor amounts of the brominated 3-hydroxy-2-naphthol were obtained (Scheme 2.8a). The synthesis of *o*NQMP-UAA4 was accordingly not pursued any further. Instead, we opted to alkylate the acetonide-protected 3,5-dihydroxy-2-naphthol with ethyl bromoacetate, hydrolyse the ethyl ester with NaOH and couple the carboxylate thus obtained to *L*-lysine *via* standard amide bond coupling methods to afford protected *o*NQMP-UAA5, albeit in low yield (Scheme 2.8b).



**Scheme 2.8.** Synthetic steps devised for preparation of *o*NQMP-UAA4 and 5. (a) Attempted and ultimately abandoned synthetic pathway for preparation of the monobrominated precursor of *o*NQMP-UAA4. Synthesis of the UAA was not pursued any further due to poor outcomes during the transformation of the boronate ester into the corresponding aryl bromide. (b) Synthesis of *o*NQMP-UAA5.

Several different PylRS variants were tested in *E. coli* in order to find a suitable synthetase that would site-specifically incorporate *o*NQMP-UAA2 and 3 into sfGFP bearing an amber stop codon at position 150 (sfGFP-N150TAG-His6, Figure 2.3 and Table 4.2, Chapter 4.3.4). Both UAAs were dissolved as stock solutions in 1 M NaOH and added to the growing *E. coli* cell cultures so that the final concentration was 2 mM. Unfortunately, both UAAs precipitated out of the cell cultures at neutral pH. While this occurrence was only mild for *o*NQMP-UAA2, *o*NQMP-UAA3 precipitated completely out of the cell culture and was not taken into further consideration for incorporation experiments. Cell cultures expressing sfGFP and different PylRS variants in the presence of *o*NQMP-UAA2 were grown overnight and the lysed cells eventually analysed *via* SDS-PAGE. As visible in Figure 2.3, regrettably, none of the tested synthetase variants incorporated *o*NQMP-UAA2 into sfGFP-N150TAG-His6 even though the experiment included PylRS variants with enlarged amino acid binding pockets (see Table 4.2).

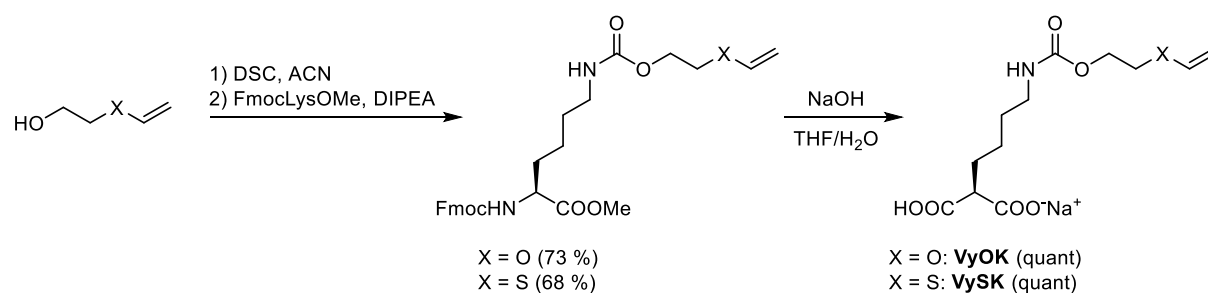


**Figure 2.3.** SDS-PAGE analysis of site-specific incorporation tests of *o*NQMP-UAA2 into sfGFP-N150TAG-His6 in *E. coli*, in the presence of different PylRS variants. BocK, incorporated by the wild-type (wt) PylRS, is included as the positive incorporation control. Details on active site mutations of tested PylRS variants can be found in Table 4.2 (Chapter 4.3.4).

Among the several factors that can attribute to the aforementioned negative incorporation outcome are: (i) the solubility problems of the tested UAA, (ii) the possibility that the UAA is not cell permeable and, naturally, (iii) that none of the tested variants recognizes the UAA as its substrate. While cellular uptake experiments have not been performed within the scope of this thesis, they should be carried out in the future if these UAAs are to be taken into consideration for site-specific incorporation into proteins. Directed evolution approach for finding novel synthetase variants is also likely needed. We, however, did not pursue these ideas any further as we had faster progress with the labelling strategy that includes vinyl ethers and sulfides as UAAs, which was simultaneously developed to *o*NQMP-based UAAs. These UAAs are easily synthesized and can, importantly, be site-specifically incorporated into proteins of interest.

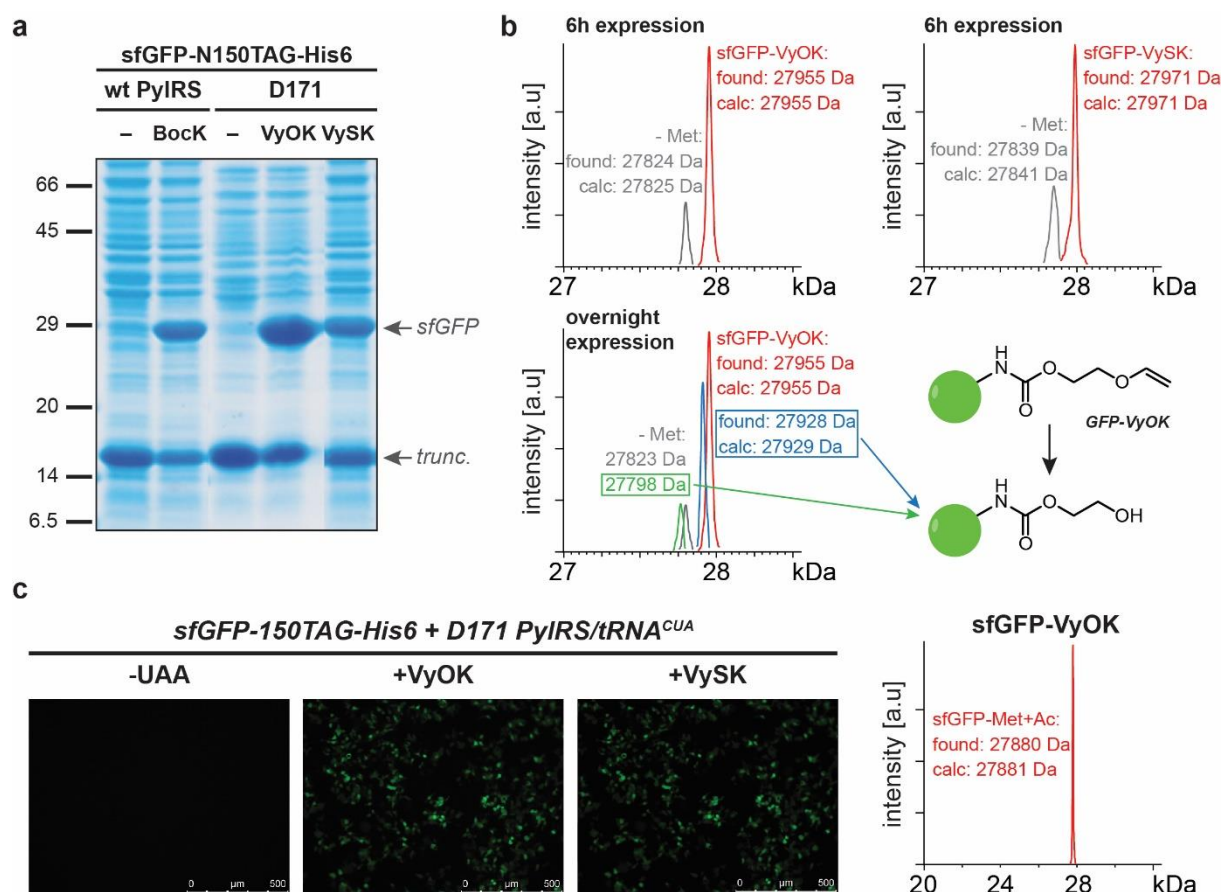
## 2.5. Incorporation of vinyl ether and vinyl sulfide UAAs and subsequent labelling experiments

As illustrated by Scheme 2.9, the straightforward synthesis of vinyl-containing UAAs was a far simpler challenge than the multistep routes of obtaining *o*NQMP-based UAAs. The hydroxy groups of either ethylene glycol vinyl ether or sulfide were activated by DSC and then coupled to the  $\epsilon$ -amino group of *N* $\alpha$ -Fmoc-protected methyl L-lysinate. Both protecting groups are labile under basic conditions (the vinyl group easily hydrolyses at acidic pH) and were simultaneously removed in the following step by treatment with NaOH to afford VyOK or VySK as sodium salts in approx. 70 % yield over two steps.



**Scheme 2.9.** Synthesis of vinyl-ether- and vinyl-sulfide-containing UAAs VyOK and VySK.

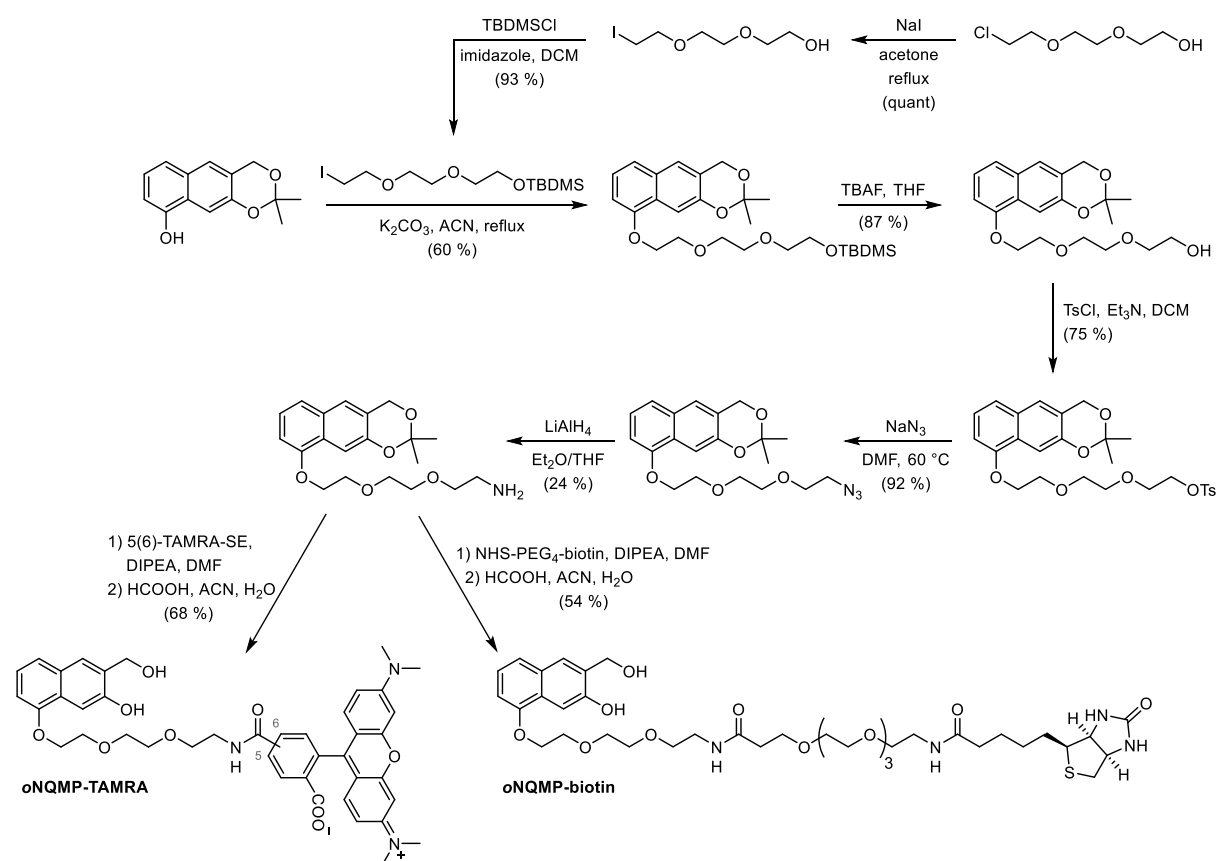
Both UAAs were, furthermore, auspiciously incorporated into sfGFP-N150TAG-His6 in *E. coli* by the same PylRS variant – D171 (containing mutations Y271M, L274A and C313A) (Figure 2.4a). Notwithstanding the outstanding incorporation efficiency, the purification of the modified sfGFP and subsequent characterisation by MS showed posttranslational hydrolysis of the vinyl group for both VyOK and VySK (Figure 2.4b, shown only for VyOK). We are unsure why this occurs, since both amino acids showed significant stability when stored for prolonged times as stock solutions (VyOK in water and VySK in NaOH) but, as a result of this observation, we limited the expression times in *E. coli* to 6 h where only masses of the desired full-length proteins were observed. It should be noted that even after shortened expression times, 25-30 mg of sfGFP variants per litre of cell culture were obtained, highlighting the striking efficiency of the D171 synthetase. Considering that the D171/tRNA<sup>CUA</sup> pair is orthogonal in eukaryotic cells as well, we transferred the respective mutations into a mammalian optimised *Mm*PylRS and successfully introduced both UAAs site-specifically into sfGFP in HEK293T cells (Figure 2.4c). Interestingly, the hydrolysis of the vinyl groups was not observed even after 48 h expression (as judged by MS). The ability to incorporate VyOK and VySK in mammalian cells, opens up an exciting opportunity for live-cell labelling applications of the *o*NQM-based cycloaddition reaction. In order to achieve this goal, the labelling reaction, naturally, has to be established on purified proteins *in vitro*.



**Figure 2.4.** Site-specific incorporation of VyOK and VySK into sfGFP. (a) Coomassie stained SDS-PAGE analysis of VyOK and VySK incorporation into sfGFP-N150TAG-His6 by the PylRS variant D171 in *E. coli*. Incorporation of BockK by the wild-type PylRS is used as a positive incorporation control. (b) MS analysis of VyOK- and VySK-bearing sfGFP variants after 6 h and overnight expression. Posttranslational hydrolysis of the vinyl group was observed after prolonged expression times, as seen on sfGFP-N150VyOK-His6. (c) Fluorescence microscopy of HEK293T cells expressing sfGFP-N150TAG-His6 in the absence or presence of 1 mM VyOK or VySK and MS analysis of purified sfGFP-N150VyOK-His6.

For labelling VyOK- and VySK-bearing proteins, we designed and successfully prepared TAMRA (rhodamine dye) and biotin precursors of *o*NQMs (Scheme 2.10). Acetonide-protected 3,5-dihydroxy-2-naphthol (turn to Scheme 2.6 for synthesis details) was first alkylated with the TBDMS-protected alkyl iodide, readily prepared from commercially available triethylene glycol monochlorohydrine in two steps (Finkelstein reaction and silyl protection of the free hydroxy group). Following the alkylation reaction, a four-step reaction sequence comprised of TBDMS deprotection, tosylation, azidation and reduction, afforded the *o*NQM precursor whose free amine group allowed derivatisation with activated succinimidyl ester of TAMRA (mixture of 5- and 6- isomers, 5(6)-TAMRA-SE) or NHS ester of PEGylated biotin (NHS-PEG<sub>4</sub>-biotin). Both conjugates were treated with formic acid in order to remove the acetonide group and purified *via* preparative HPLC to afford *o*NQMP-TAMRA and *o*NQMP-biotin ready for use.





**Scheme 2.10.** Synthesis of *o*NQMPs conjugated to a fluorophore (*o*NQMP-TAMRA) or biotin (*o*NQMP-biotin).

The labelling experiments were performed on purified sfGFP-N150VyOK-His6 by employing either the 312 nm or 365 nm lamp. The proteins were incubated in 96-well plates in the absence or presence of *o*NQMP-based labelling reagents while the experimental setup required the lamps to be positioned above the sample. After irradiation, the samples were analysed by SDS-PAGE and either *via* in-gel fluorescence measurements (for detection of TAMRA fluorophore) or western blotting with horseradish peroxidase (HRP)-streptavidin conjugate (for detection of biotinylated proteins). Importantly, the negative control included incubation of samples with or without labelling reagents added for a specific amount of time in the absence of light. It is also worth noting that pure sfGFP samples were exposed to different wavelengths of light in order to exclude the possibility of light-promoted side reactions and check the overall stability of sfGFP during irradiation. Furthermore, all irradiated samples were stirred during their exposure to light so that the light could penetrate equally through the entire reaction mixture.

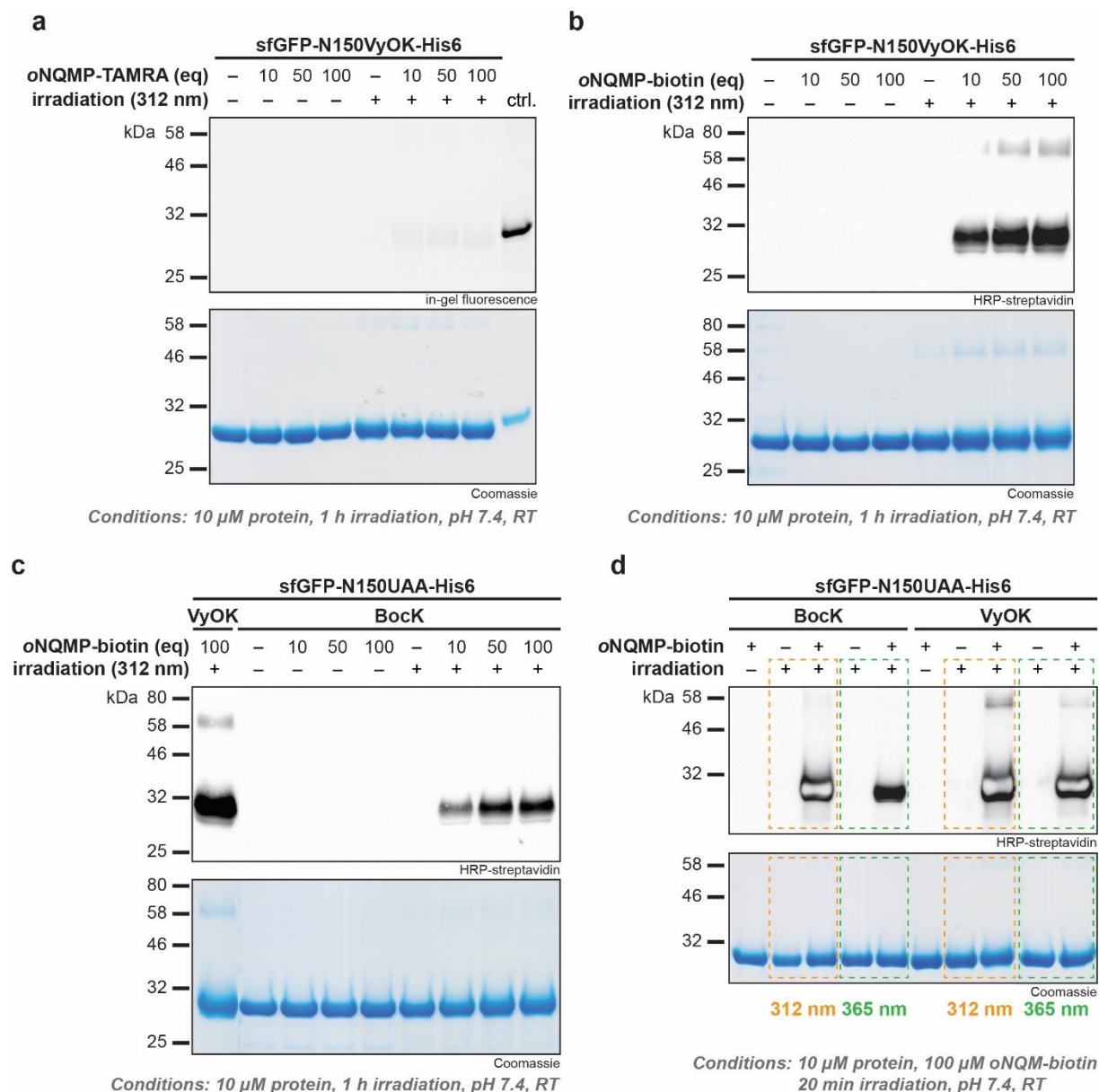
Initial labelling experiments with TAMRA fluorophore lamentably did not lead to the desired outcome (Figure 2.5a). While faintly fluorescent bands were observed for samples irradiated in the presence of different amounts of *o*NQM-TAMRA (10, 50 or 100 equivalents), these were barely noticeable next to the control sfGFP-N150BCNK-His6 sample pre-labelled with a tetrazine-TAMRA conjugate (lane labelled as “ctrl.” in the gel, used as a

positive control for in-gel fluorescence measurements, Figure 2.5a). In parallel, we tested the reactivity of *o*NQM-TAMRA on small molecule level with excess of vinyl ether but observed only minor amount of product formation (Figure S2.17). We can only assume that these results arise from the quenching effects of the rhodamine dye. Luckily, unambiguous bands of biotinylated proteins were observed through HRP-streptavidin western blot analysis of sfGFP samples irradiated in the presence of *o*NQM-biotin (Figure 2.5b). It must be emphasized that these bands were only observed for samples that were irradiated at 312 nm, thereby implying that the biotinylation reaction is triggered by light. The reactivity of *o*NQM-biotin was confirmed on small molecule level as well (Figure S2.18). The characterisation of the labelling reaction by MS was troublesome due to problems in deconvolution which afforded data with excessive signal-to-noise ratio. We did, however, observe masses that correspond to biotinylated sfGFP (Figure S2.19).

Even though the non-irradiated samples serve as nice negative controls that point out the indispensability of light, it was necessary for us to show the specificity of the labelling reaction on proteins that do not carry vinyl ethers. The high specificity of the labelling reaction is imperative if it was to be employed for site-specific labelling of biomolecules *in cellulo*. We therefore, repeated the aforementioned labelling experiments with *o*NQM-biotin on purified sfGFP-N150BocK-His6 (Figure 2.5c). To our dismay, we observed unambiguous bands of biotinylated proteins with sfGFP samples that were irradiated at 312 nm. Even though these protein bands were almost half the size of the respective bands seen with sfGFP-VyOK samples (compare lanes 1 and 9, Figure 2.5c), they indicate that our labelling reaction was not specific. Similar results were obtained in experiments with 365 nm lamps (Figure 2.5d).

It can be presumed that the undesired labelling of sfGFP-BocK arises from cross-reactivity of *o*NQM-biotin with nucleophilic amino acid residues such as cysteines, lysines or even the *N*-terminus of sfGFP. This could also explain why MS analysis of the irradiated reaction mixtures generally did not give any fruitful results as the deconvolution of multiple different protein species is challenging. Considering that the sulfhydryl groups of cysteines are by far the best nucleophiles among the canonical amino acid residues, they effortlessly outcompete reactions between *o*NQMs and other amino acids as well as water ( $k_2(\text{RSH}) \sim 2.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  versus  $k_2(\text{H}_2\text{O}) \sim 2.6 \text{ M}^{-1} \text{ s}^{-1}$ ).<sup>[137b]</sup> Nevertheless, the reaction between cysteines and *o*NQMs has been shown to be reversible in the presence of vinyl ethers,<sup>[183b]</sup> which we also observed on small molecule level photolysis experiments (Figures S2.15 and S2.16). When irradiated, *o*NQMP-labelled cysteines can readily be photo-released while the formed *o*NQMs trapped by vinyl ethers into photo-stable benzochromanes and are thus removed from the reaction mixture

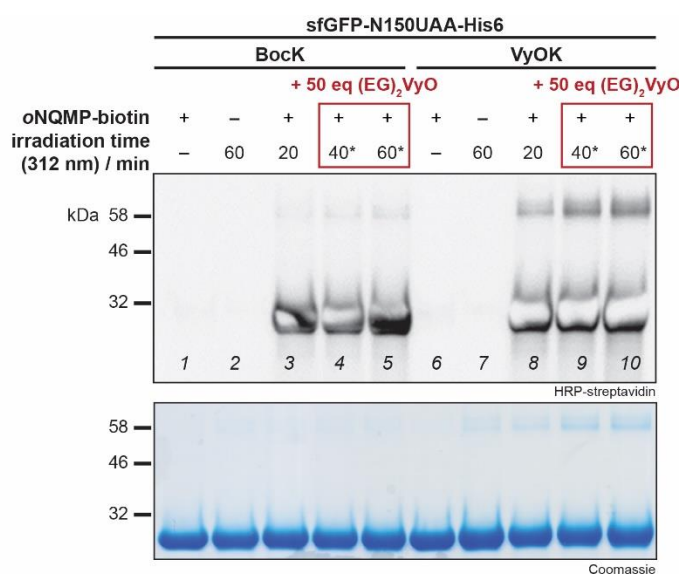
---



**Figure 2.5.** Light-induced *in vitro* labelling of purified sfGFP variants with different oNQMP labelling probes. (a) In-gel fluorescence analysis (upper panel) after SDS-PAGE (lower panel) of the labelling reaction between sfGFP-N150VyOK-His6 and different concentrations of oNQMP-TAMRA, in the absence and presence of UV light. sfGFP-N150BCNK-His6 pre-labelled with tetrazine-TAMRA conjugate is used as a fluorescence control (“ctrl.” lane). (b) HRP-streptavidin western blot (upper panel) and SDS-PAGE (lower panel) of the labelling reaction between sfGFP-N150VyOK-His6 and different concentrations of oNQMP-biotin, in the absence and presence of UV light. (c) Same as (b) for sfGFP-N150BocK-His6, displaying the non-selectivity of the light induced labelling reaction. Sample with sfGFP-N150VyOK-His6 is used for labelling comparison. (d) Effect of difference in irradiation wavelengths on the labelling of BocK- and VyOK-bearing sfGFP with oNQMP-biotin, analysed by HRP-streptavidin western blotting (upper panel) and SDS-PAGE (lower panel).

We, therefore, decided to perform photouncaging experiments on sfGFP bearing BocK and VyOK where we would first irradiate protein samples in the presence of oNQMP-biotin and then add excess of vinyl ether to the reaction mixture and continue irradiation for a specific amount of time. If proven successful, we would expect to see a change in the intensity of labelled protein bands in western blot as, after uncaging, only a single molecule of

*o*NQM should be bound to sfGFP-VyOK *via* the site-specific DA cycloaddition. More importantly, the corresponding bands in samples containing sfGFP-BocK should disappear completely in the presence of vinyl ether, as no photo-stable benzochromane products can be formed. Unfortunately, these experiments did not proceed as expected (Figure 2.6) as we observed bands corresponding to biotinylated proteins with both BocK- and VyOK-bearing sfGFP even after 40 minutes of irradiation at 312 nm in the presence of excess (50 equivalents) of (EG)<sub>2</sub>VyO. Furthermore, our results are in contradiction with previously reported data where photo-release of *o*NQMP-labelled cysteines was readily achieved within minutes in presence of equimolar amounts of vinyl ether.<sup>[183b]</sup>



Conditions: 10  $\mu$ M protein, 100  $\mu$ M oNQMP-biotin, pH 7.4, RT; after 20 min irradiation (samples in lanes 3 & 8) 500  $\mu$ M (EG)<sub>2</sub>VyO was added to the reaction mixtures and irradiation was continued for further 20 and 40 mins. \* denotes overall irradiation time

**Figure 2.6.** Light-mediated uncaging of sfGFP-*o*NQM adducts arisen from the undesired nucleophilic addition reaction in the presence of excess of vinyl ether, (EG)<sub>2</sub>VyO. BocK- and VyOK-bearing sfGFP samples were irradiated for 20 minutes with *o*NQMP-biotin to ensure unspecific labelling. Excess of (EG)<sub>2</sub>VyO was then added to these samples and irradiation continued for 20 and 40 minutes more. All of the protein samples were analysed by HRP-streptavidin western blot (upper panel) and SDS-PAGE (lower panel).

Even though the substantial discrepancy between the published and our data could possibly be attributed to the differences between the two photochemistry setups (with which we already encountered a fair share of difficulties on small molecule level) it seems odd that the uncaging reaction did not occur at all. For this reason, we tested the uncaging reaction on small molecule level in ACN:PBS (pH 7.4) mixture (1:1, v/v) and in presence of lysine and cysteine (Figure S2.20). Briefly, irradiation of a mixture containing 1 mM NOH and 10 mM of lysine or cysteine for 20 minutes at 312 nm was followed by addition of 20 mM (EG)<sub>2</sub>VyO and irradiation for further 20 minutes. Even though both reaction mixtures were not clean and formation of adducts was observed, not all of which were completely identifiable *via* MS (these were readily observed in the reaction with cysteine and to a much lesser extent with lysine), the formation of benzochromane product still occurred after addition of vinyl ether. Since the nucleophilicity of amino acid residues in proteins heavily depends on protein

microenvironment, cross-reactivity of *o*NQMs with amino acids other than cysteines cannot be excluded. Consequently, we determined the strategy of introducing VyOK into proteins *via* amber suppression and subsequent reaction with excess of NQMP-bearing probes not applicable for site-specific labelling of proteins.

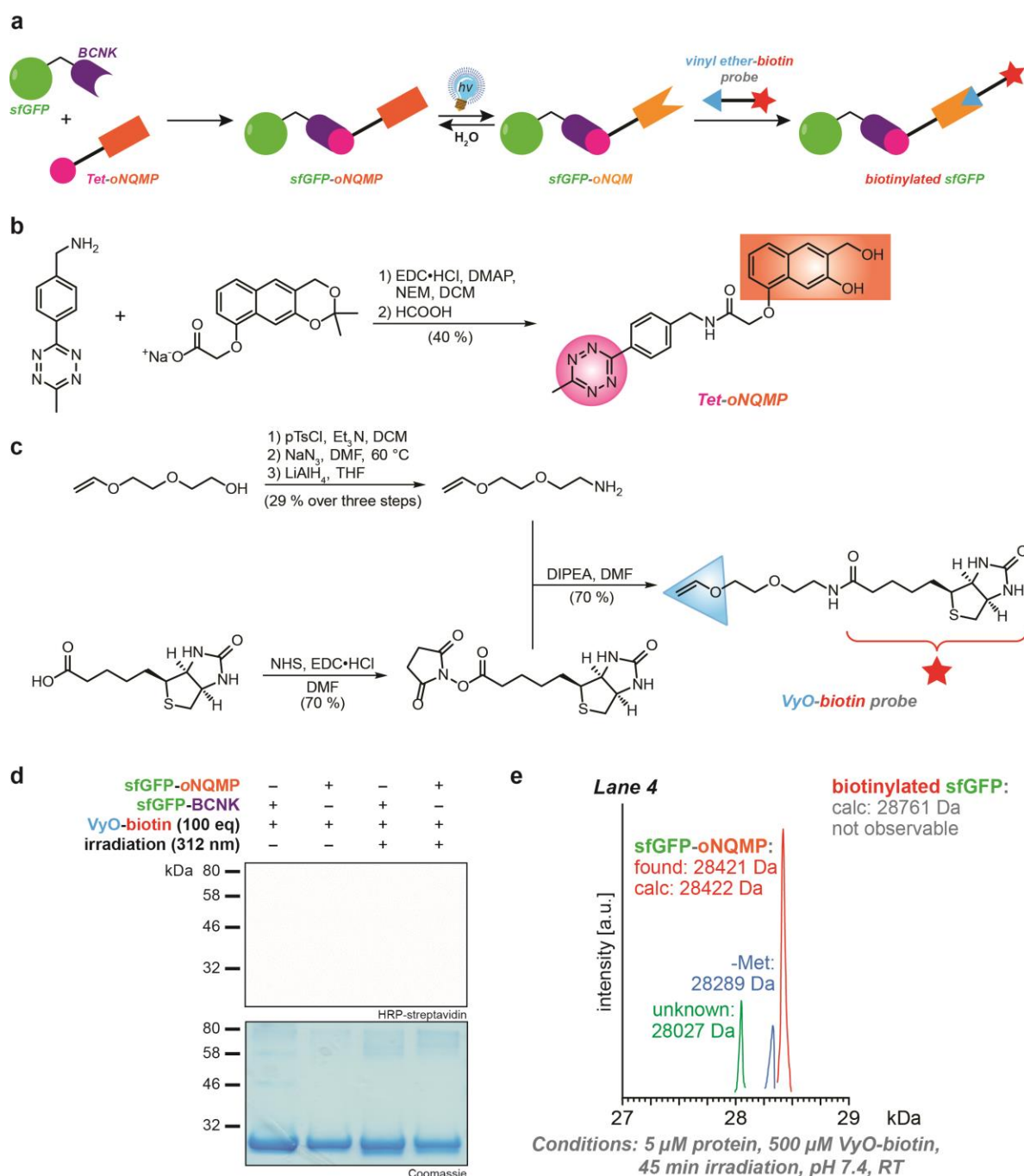
## 2.6. The final experiment

Slowly running out of options, especially after uncovering the non-specific labelling of proteins with *o*NQM-based probes, we turned our attention towards UAAs carrying *o*NQM precursors for one last time. To deem these UAAs worthy of undertaking the lengthy directed evolution approach for finding a novel synthetase, we decided to chemically introduce *o*NQMP into sfGFP and examine its reactivity with vinyl ethers. To this end we synthesized an *o*NQMP carrying a tetrazine moiety (Tet-*o*NQMP) which we could introduce site-specifically into sfGFP-N150BCNK-His6 *via* fast tetrazine ligation. With subsequent irradiation of the modified protein in the presence of excess of biotin-modified vinyl ether and analysis of the reaction mixture by western blotting, we sought out to determine the extent of the cycloaddition reaction (Figure 2.7a). The synthesis of both Tet-*o*NQMP and the VyO-biotin probe was fairly uncomplicated (Figure 2.7b,c). The acetonide-protected *o*NQMP bearing a free carboxylic group was coupled to (4-aminomethyl)phenyl methyl tetrazine using standard amide coupling methods, while the acetonide group was removed under acidic condition (40 % yield over two steps, Figure 2.7b). The biotinylated vinyl ether probe was prepared from commercially available di(ethylene glycol) vinyl ether. Briefly, the free hydroxyl group was first converted into an amine by a sequence of tosylation, azidation and reduction reactions and eventually coupled to the activated NHS ester of D-biotin (20 % yield over four steps, Figure 2.7c).

With both probes in hand, we incubated purified sfGFP-N150BCNK-His6 with excess of Tet-*o*NQMP (20 equivalents) for 2 hours at 37 °C in order to obtain the *o*NQMP-modified sfGFP. After washing the excess of label out of the modified sfGFP, we incubated the protein with VyO-biotin in the absence and presence of 312 nm light. sfGFP bearing BCNK was, furthermore, used as a negative control. The subsequent analysis of the reaction mixtures *via* western blotting with HRP-streptavidin unfortunately did not yield any positive results, that is, no bands of biotinylated proteins were observed (Figure 2.7d). Similarly, no masses that would correspond to biotinylated sfGFP were observed by MS (Figure 2.7e). Instead, the major observed species was the unreacted *o*NQMP-modified sfGFP. We repeated the analogous experiments on small molecule level by incubating BCNK with Tet-*o*NQMP, followed by labelling with excess of (EG)<sub>2</sub>VyO at 312 nm (Figure S2.21). The formation of the photocycloaddition product was observed by LC-MS, meaning that the introduction of the *o*NQM into a substrate (in this case, the free amino acid) *via* tetrazine ligation does not obstruct its reactivity. Nevertheless, the desired benzochromane proved to be only the minor product in this reaction since LC-MS revealed presence of another substance whose peak dramatically increased with irradiation time. This side product seemingly corresponds to the intramolecular nucleophilic addition of (possibly) the N-terminus of the UAA to the

---

methylene group of *o*NQM (Figure S2.21b). Taking the nucleophilic addition to *o*NQM aside, it remains puzzling why the photocycloaddition reaction was observed on BCNK but not on the BCNK-containing protein.



**Figure 2.7.** Establishing the *o*NQM-based labelling by chemically introducing the *o*NQMP moiety into sfGFP *via* tetrazine ligation. (a) Schematic representation of the devised *in vitro* labelling approach. Briefly, Tet-*o*NQMP is site-specifically introduced into purified sfGFP-N150BCNK-His6 by means of tetrazine ligation. Subsequent irradiation of the *o*NQMP-modified sfGFP affords *o*NQM which should readily react with the biotinylated vinyl ether probe. (b) Synthesis of Tet-*o*NQMP. (c) Synthesis of the biotinylated vinyl ether probe. (d) HRP-streptavidin western blot (upper panel) and Coomassie-stained SDS-PAGE (lower panel) analysis of the *in vitro* attempt to label the chemically modified *o*NQMP-sfGFP with excess of biotinylated vinyl ether. sfGFP-N150BCNK-His6 is used as the negative labelling control. (e) MS-analysis of the labelling reaction from (d).

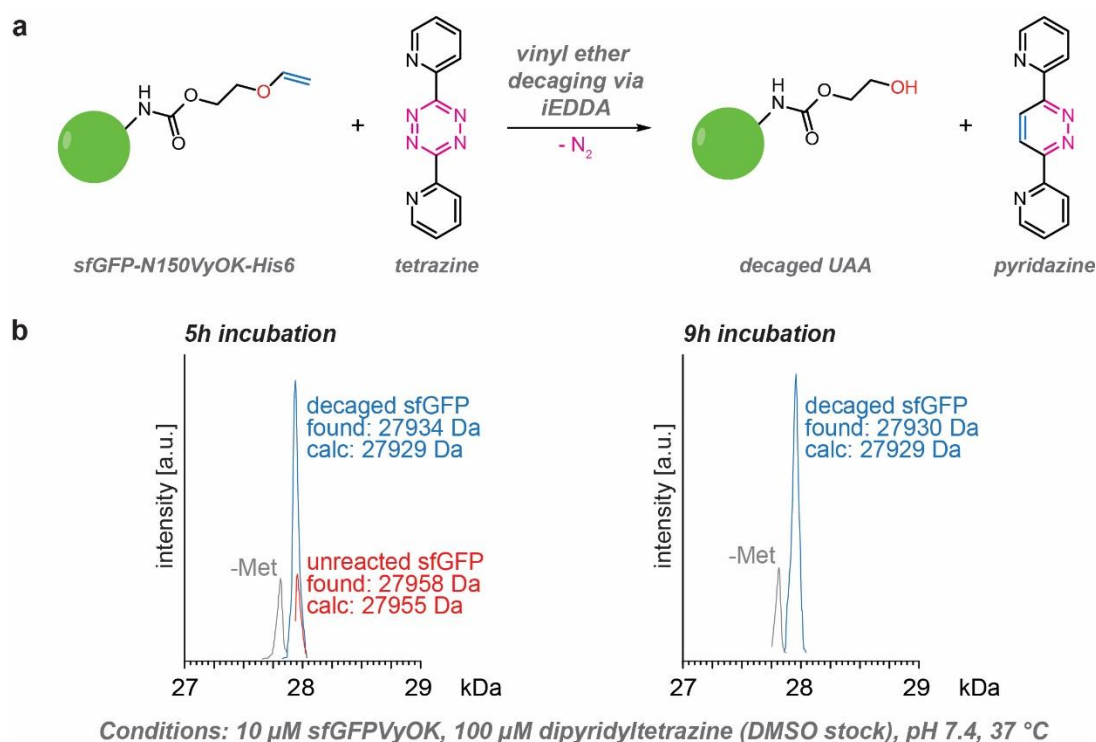
In view of these results, we opted against directed evolution of synthetases that would recognize prepared UAAs carrying *o*NQMPs as their substrates.

Considering all of the above experiments and results, it seems that the chemical nature and reactivity of photochemically-generated *o*NQM does not make them suitable for site-specific labelling of biomolecules. Although the overall data is not in accordance with the reports where the light-induced hetero-Diels-Alder cycloaddition between *o*NQMs and vinyl ethers was used for selective and spatiotemporally controlled modification of either functionalised proteins or surfaces, it is still consistently inconsistent. Joking aside, the problems involving the undesired non-selectivity of the *o*NQM-based labelling as well as the inability to circumvent them ultimately led up to the decision of closing the door on this photocycloaddition reaction.

## 2.7. Pursuing alternative bioorthogonal reactions for UAAs containing vinyl ether and vinyl sulfide moieties

The unfortunate demise of the *o*NQM project left us with two UAAs, VyOK and VySK that were efficiently incorporated in bacteria and mammalian cells by the PyIRS D171/tRNA<sup>CUA</sup> pair. Ergo, we sought out to find alternative bioorthogonal reactions that would make use of these UAAs as tools for a genetic code expansion approach. Vinyl ethers and sulfides are known to undergo Paternò-Büchi reactions with aliphatic and aromatic aldehydes and ketones, but these reactions are far from biocompatible due to the requirement of toxic UV light (280-350 nm).<sup>[195]</sup> Another option is the thiol-ene click reaction, yet it is limited to *in vitro* applications at best as it relies on usage of radical photoinitiators and produces radical intermediates, often at the expense of selectivity (e.g. cross-reactions with other thiols that are ubiquitously present in living cells). A further point against this reaction is the lack of novelty as UAAs that undergo thiol-ene click have been reported before (see Chapter I for more detail).<sup>[146-147]</sup> We also briefly considered the iEDDA reaction between vinyl ethers and tetrazines, however, this cycloaddition is immediately followed by the traceless release of alcohol (or, in other words, cleavage of the vinyl group) affording pyridazines as by-products (Figure 2.1).<sup>[196]</sup> Still, purified sfGFP-N150VyOK-His6 was incubated with 10 equivalents of dipyridyltetrazine (stock solution in DMSO) and samples were taken at different time points and analysed *via* MS. As visible in Figure 2.8, the predominant species after 5 h incubation is the released alcohol. As expected, the reaction is reasonably slower than the corresponding iEDDA with strained alkenes. The uncaging reaction between vinyl ethers and tetrazines was ultimately not pursued since the “free alcohol” form of VyOK is essentially still an unnatural amino acid. Masking of alcohol groups with vinyl ethers and subsequent uncaging with tetrazines would, for example, make more sense on natural amino acids that perform a catalytic or another important function in a protein of interest (e.g. tyrosine or serine), although the slow reaction rates could limit the applications of such approaches. Also, a couple of examples of aforesaid traceless unmasking of alcohols in cells have already been reported in recent literature.<sup>[197]</sup>

---



**Figure 2.8.** iEDDA reaction between vinyl ethers or sulfides and tetrazines. (a) Schematic representation of the respective reaction. (b) MS-analysis of the *in vitro* reaction between purified sfGFP-N150VyOK-His6 and dipyridyltetrazine.

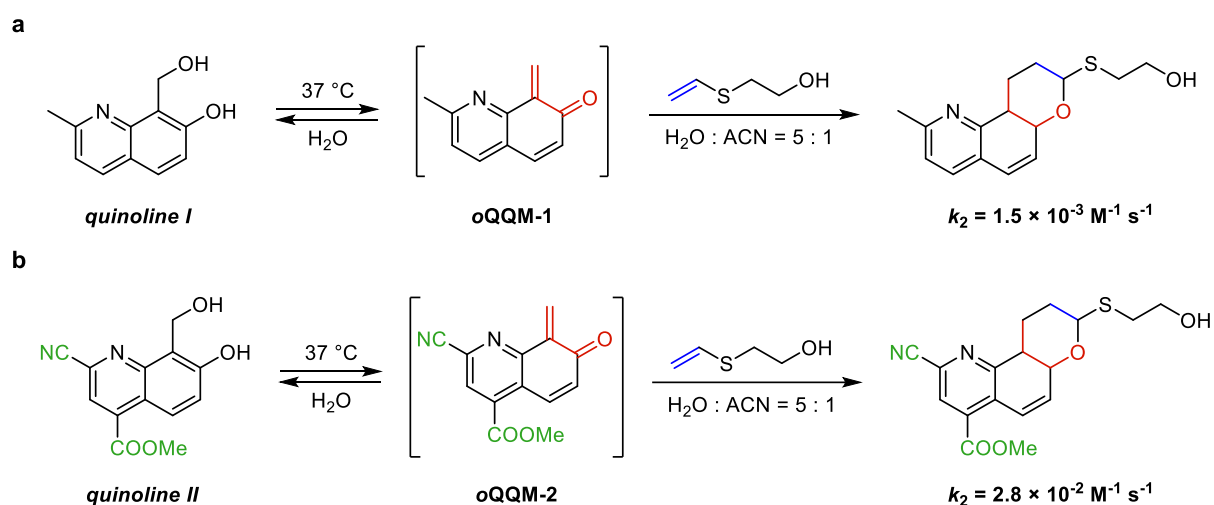
With the advances in recent literature, we finally opted to explore two alternative reactions: the hetero-Diels-Alder reaction between vinyl sulfides and thermally generated *ortho*-quinolinone quinone methides (*o*QQMs)<sup>[198]</sup> and the visible-light-induced [4+2] cycloaddition between vinyl ethers and 9,10-phenanthrenequinones (PQs).<sup>[138]</sup>

### 2.7.1. *ortho*-Quinoline quinone methides (*o*QQMs)

We became aware of the reactivity of *o*QQMs with vinyl sulfides with the publication of the so-called second generation TQ-ligation.<sup>[198b]</sup> The reaction was originally conceived as an improvement over other *o*NQMs through introduction of an electrophilic nitrogen atom into the *o*NQM core structure, which not only affects the reactivity of the molecule but raises its hydrophilicity as well (Scheme 2.11a).<sup>[198a]</sup> Consequently, the generation of *o*QQM-1 and its subsequent reaction with vinyl sulfides such as 2-(vinylthio)ethanol does not require employment of light as it takes place at 37 °C over 24 h in aqueous ACN (water:ACN, 5:1, v/v). The initial low reaction rate ( $k_2 = 1.5 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ )<sup>[198a]</sup> was, nevertheless, enhanced up to 18-fold with the design of *o*QQM-2 (Scheme 2.11b). Introduction of electron-withdrawing cyano and ester groups onto the quinoline core improved the reaction rate, with a reported second-order rate constant of  $2.8 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$  comparable to SPAAC.<sup>[198b]</sup> The utility of *o*QQM-1 was shown by labelling of proteins that were chemically pre-tagged with vinyl sulfides and imaging of a vinyl sulfide analogue of taxol inside living cells.<sup>[198a]</sup> *o*QQM-2 was employed for live cell imaging of a cellular membrane protein as well as membranes of organelles through derivatisation of organelle-specific peptides.<sup>[198b]</sup> Furthermore, vinyl

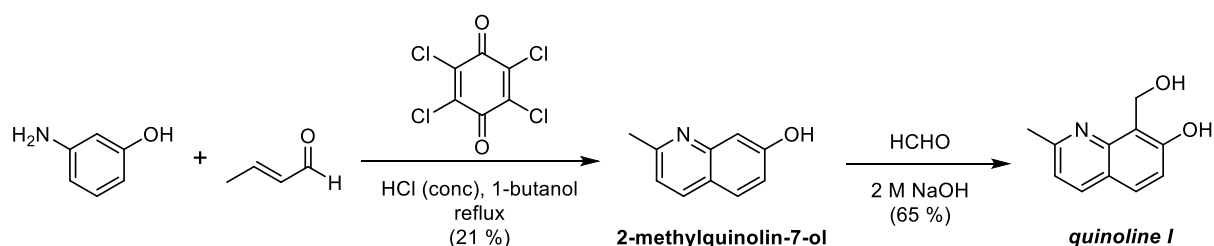


sulfide modified thymidine and a fluorescein derivative of *o*QQM-2 were recently used for bioorthogonal metabolic DNA labelling in mammalian cells.<sup>[199]</sup>



**Scheme 2.11.** Hetero-Diels-Alder cycloaddition between thermally generated *ortho*-quinolinone quinone methides (*o*QQMs) and vinyl sulfides. (a) First generation reaction.<sup>[198a]</sup> (b) Second generation reaction.<sup>[198b]</sup>

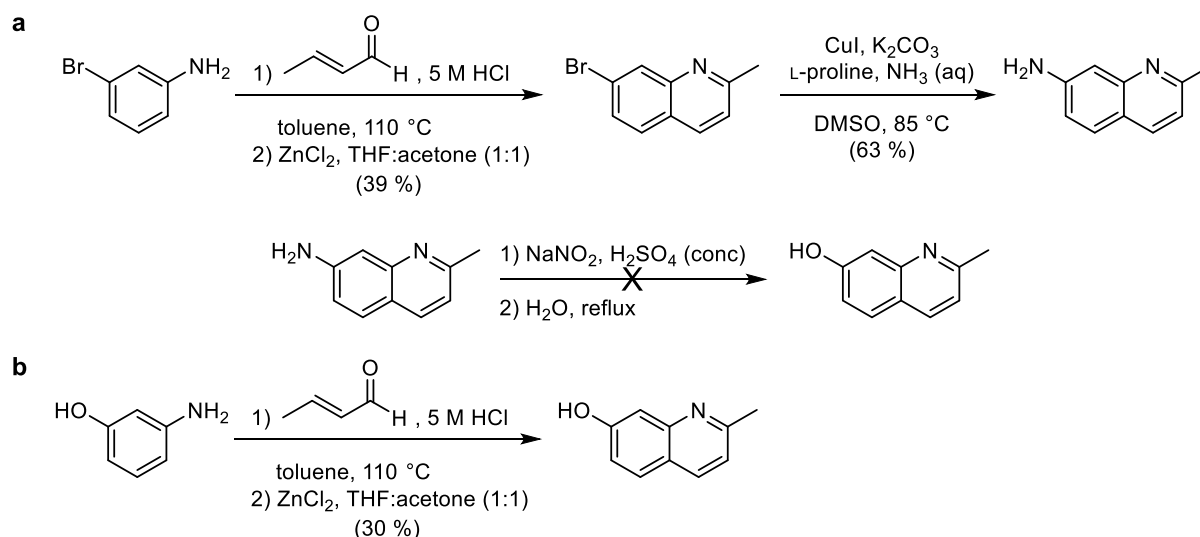
In spite of faster reaction kinetics, the preparation of the *o*QQM-2 precursors is synthetically more demanding when compared to *o*QQM-1. Hence, we decided to first focus on the reactivity of quinoline I with site-specifically incorporated VySK. Quinoline I can be prepared in two steps starting from 3-aminophenol and crotonaldehyde and a reaction known as Skraup-Döbner-von Miller quinoline synthesis.<sup>[200]</sup> 2-methylquinolin-7-ol thus obtained is then hydroxymethylated with formaldehyde in aqueous NaOH (Scheme 2.12).



**Scheme 2.12.** Skraup-Döbner-von Miller synthesis of 2-methylquinolin-7-ol from 3-aminophenol and crotonaldehyde.

Skraup-Döbner-von Miller synthesis of quinolines is carried out in refluxing hydrochloric acid and gives the desired product in poor yields which is generally attributed to the formation of many side-products and tedious purification of quinoline from the complex reaction mixture. Even though still a matter of debates, the mechanism of the reaction possibly involves a 1,4-nucleophilic addition of the aniline to the  $\alpha,\beta$ -unsaturated aldehyde, followed by a series of condensation and ring closure reactions to give 2-methyl-1,2-dihydroquinoline. The rearomatisation into quinoline is often mediated through oxidants such as *p*-chloranil.<sup>[200-201]</sup> When thinking about preparing various derivatives of *o*QQM-1 (e.g. fluorophores or biotin probes), we realised that the most reasonable synthetic pathways start

from 2-methylquinoline-7-ol.<sup>[198a]</sup> For that reason, we tried to optimize the reaction presented in Scheme 2.12 as it was low-yielding and turned out to be problematic during scale-up above 1 gram of starting 3-aminophenol. The longer and more complex route (Scheme 2.13a) involved the synthesis of 2-methyl-7-bromoquinoline from 3-bromoaniline and crotonaldehyde in the presence of hydrochloric acid, followed by the addition of Lewis acid (ZnCl<sub>2</sub>). Even though the yield of the reaction was not significantly improved (39%), the reaction was easily performed on a 10-gram scale of the starting aniline. Subsequent Cu(I)-catalysed *N*-arylation of 2-methyl-7-bromoquinoline with ammonia in the presence of L-proline afforded 2-methylquinoline-7-amine (so-called Ullmann-Goldberg-type C-N bond formation).<sup>[202]</sup> Nevertheless, the conversion of the amine into alcohol *via* diazonium salt did not afford any products, which is why we went back to the ZnCl<sub>2</sub>-mediated Skraup-Döbner-von Miller synthesis of quinolines and preformed the reaction with 3-aminophenol (Scheme 2.13b). Despite the low yield (30%) and difficulties during purification, the desired 2-methylquinoline-7-ol was ultimately attainable on a multi-gram scale.

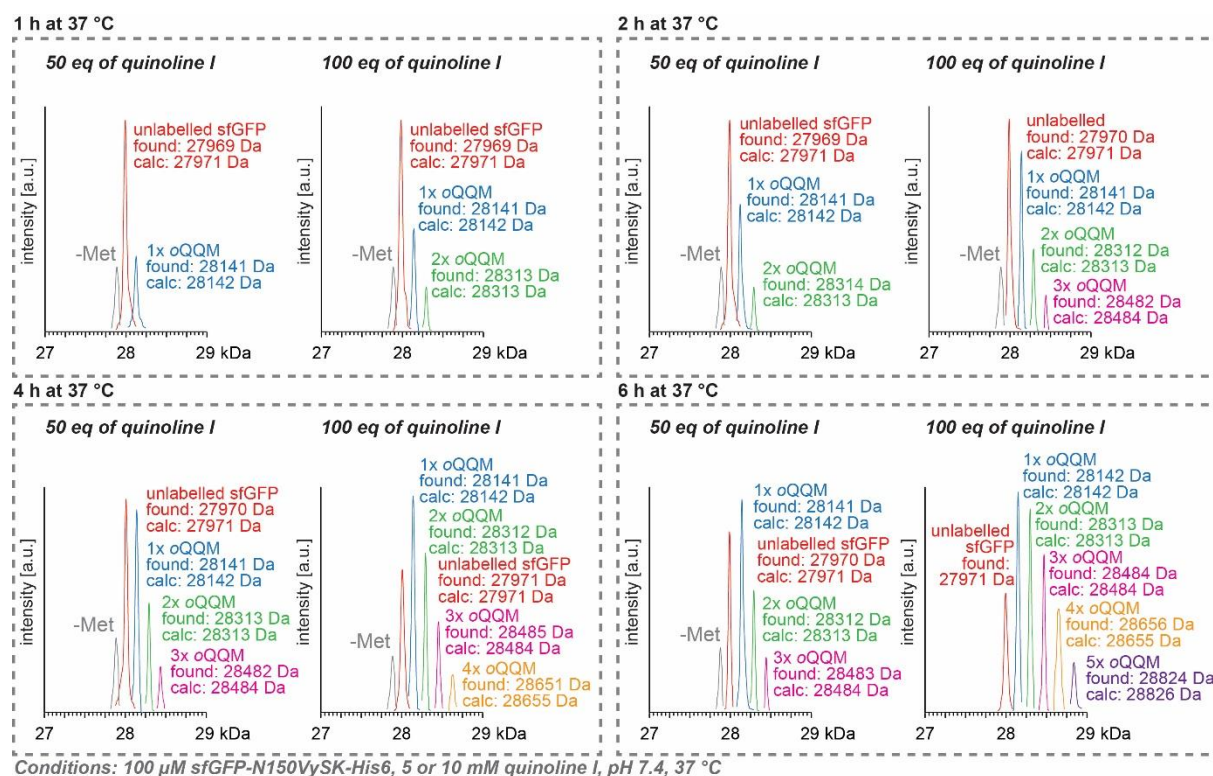


**Scheme 2.13.** Optimisation of the synthesis of 2-methylquinolin-7-ol. (a) Synthetic pathway from 3-bromoaniline. (b) ZnCl<sub>2</sub>-mediated Skraup-Döbner-von Miller synthesis of 2-methylquinolin-7-ol from 3-aminophenol.

The reactivity of quinoline I was first tested on small molecule level with ethyl vinyl sulfide (EtVyS). In agreement with the previous report,<sup>[198a]</sup> the reaction performed with 10 eq of EtVyS in water/ACN mixture (5:1, v/v) afforded the cycloaddition product after 24 h at 37 °C (Figure S2.22). Additionally, we tried to trigger the same reaction by light as well. We irradiated the reaction mixtures (same amount of reagents) for 20 minutes in Rayonet photoreactors containing lamps of different wavelengths (254, 300, 366 and 419 nm) and analysed them by HPLC (Figure S2.22). None of the irradiated reactions afforded the cycloaddition product, presumably due to relaxation of excited *o*QQM-1 through observed fluorescence.

We went on to perform first experiments on purified sfGFP-N150VySK-His6 by incubating the protein at 37 °C with excess of quinoline I (50 or 100 equivalents; stock solution in DMSO) and analysing the reaction at different time points *via* MS (Figure 2.9).

Startlingly, we observed masses corresponding to cycloaddition adducts (calculated mass difference of +171 Da) that gradually increased in number over incubation time. To put it differently, after 6 h of incubation we observed masses corresponding to species that had between one and five covalently bound *o*QQM-1 molecules (Figure 2.2). Puzzled by this lack of selectivity we repeated the same experiments on different proteins including Bock- and VyOK-bearing sfGFP and VyOK-bearing Myo and Ub, all giving the exact same results (Figure S2.23). Due to observed labelling non-selectivity on purified proteins, we did not pursue further applications of this cycloaddition reaction.

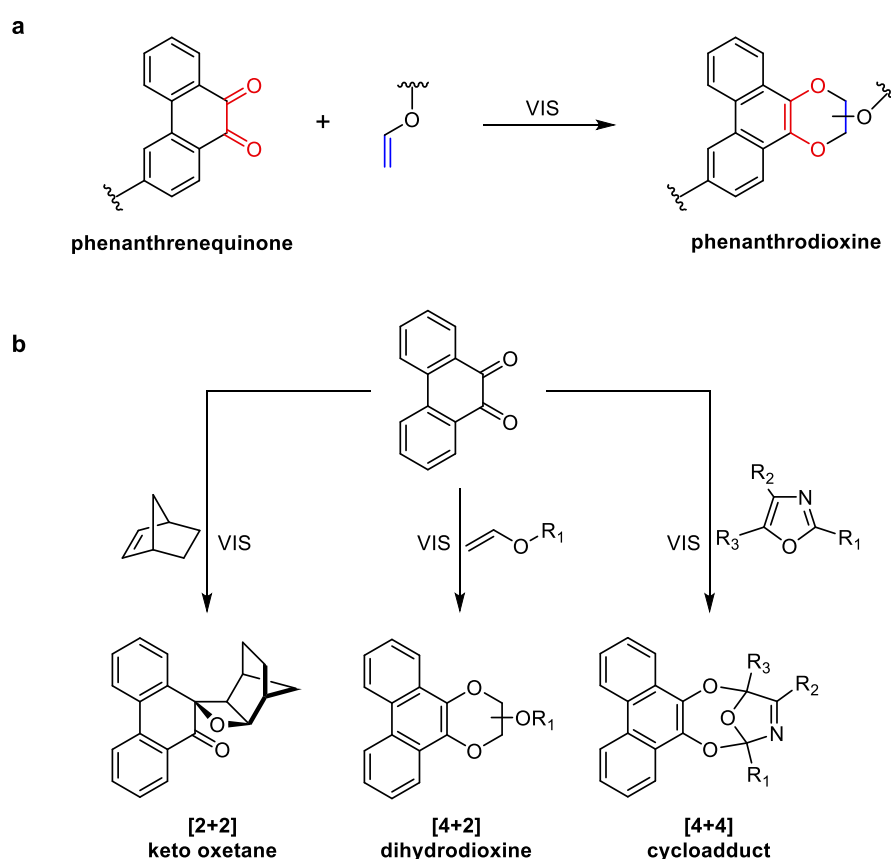


**Figure 2.9.** MS analysis of the *in vitro* labelling reaction between sfGFP-N150VySK-His6 and quinoline I. 100  $\mu$ M of protein was incubated with an excess of quinoline I (50 or 100 eq, stock solution in DMSO) at pH 7.4 and 37  $^{\circ}$ C while samples were taken at different time points (1, 2, 4 and 6 hours).

### 2.7.2. 9,10-phenanthrenequinones (PQs)

In late October 2018, another bioorthogonal photocycloaddition reaction saw the light of day.<sup>[138]</sup> Upon irradiation with visible light (> 420 nm), 9,10-phenanthrenequinones (PQs) readily undergo [4+2] cycloaddition reactions with vinyl ether to afford fluorogenic phenanthrodioxine products (Scheme 2.14a). The reaction is exceptionally fast with full conversions observed after minutes of irradiation with 10 W hand-held white light LED lamps, which makes it compatible with biological systems as well. Its utility was demonstrated through concurrent orthogonal labelling of two chemically pretagged proteins together with SPAAC or UV-assisted tetrazole click reaction. Moreover, the reaction was used for spatiotemporal live cell surface labelling of vinyl ether modified cetuximab.<sup>[138]</sup>

The visible-light mediated cycloaddition reactions of PQs with olefins have been, as a matter of fact, studied for decades and are known to afford two types of cycloaddition products: the [2+2] keto oxetanes and [4+2] dihydrodioxines (Scheme 2.14b).<sup>[203]</sup> While bicyclic alkenes such as norbornene exclusively give [2+2] keto oxetanes as cycloaddition products,<sup>[203a]</sup> electron-rich alkenes like vinyl ethers undergo [4+2] cycloaddition reactions with high chemoselectivity.<sup>[138, 204]</sup> Additionally, an unusual [4+4] photocycloaddition has so far only been reported between PQs and oxazoles (Scheme 2.14b).<sup>[205]</sup> The [4+2] cycloaddition reaction with vinyl ethers is initiated by the long-lived PQ triplet excited state ( $^3\text{PQ}^*$ ), followed by electron transfer between  $^3\text{PQ}^*$  and the electron-rich C=C bond to afford a 1,6-diradical intermediate. Fast intramolecular radical recombination ultimately gives the phenanthrodioxine products.<sup>[138, 206]</sup>



**Scheme 2.14.** Visible-light-induced cycloaddition reactions of 9,10-phenanthrenequinones (PQs). (a) Novel bioorthogonal [4+2] cycloaddition reaction between PQs and vinyl ethers. (b) Types of photocycloaddition products depending on the reaction pathway and olefin reaction partner.

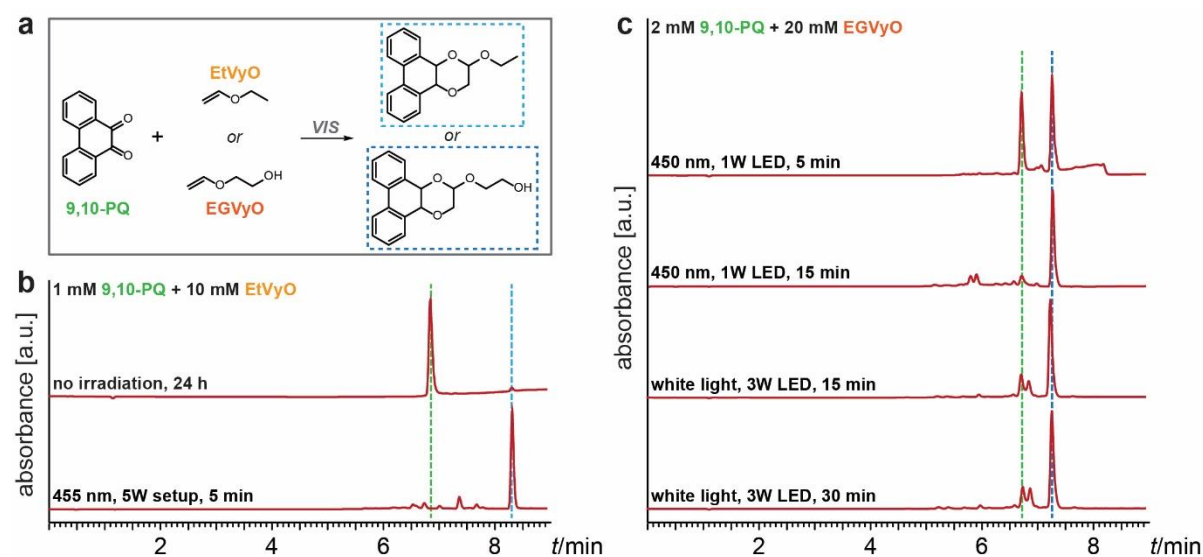
The facts that the photocycloaddition is rapid and inducible by visible light definitely sparked our interest; if proven successful with site-specifically incorporated VyOK, we anticipated a fast and specific spatiotemporal labelling reaction in our hands that circumvents the usage of toxic UV light. First of all, we decided to establish the reaction in our lab on a small molecule level, between the commercially available 9,10-phenanthrenequinone (9,10-PQ) and excess of ethyl vinyl ether (EtVyO) or ethylene glycol vinyl ether (EGVyO, Figure 2.10a). As with *o*NQM, we tried out several different light sources, taking into account the

absorption maximum of 9,10-PQ which is around 430 nm. These included the 455 nm 5 W setup where the reaction mixture is irradiated from the inside and 450 nm or white light single LEDs (either 1 or 3 W). In the 5 W setup, developed by the group of Prof. Thorsten Bach, the light is focused through a quartz rod which is immersed into the reaction mixture. The surface of the immersed part of the quartz rod is matted which scatters the light so that the entire reaction mixture is irradiated from the inside. Although, very pragmatic for small molecule reactions, this setup is not particularly useful for purified proteins as it typically requires 2 or more millilitres of reaction mixtures. We, therefore, wanted to test single LEDs as well, which would allow us to irradiate smaller volumes of reaction mixtures (< 1 mL) in cell culture plates. It should be pointed out that the 5 W setup is a closed system where the reaction mixture is in a Schlenk tube, while the LEDs were used to irradiate open cell culture plates from above. Furthermore, we considered different solvent systems for monitoring the reaction (with the goal to be as aqueous as possible). While it is readily soluble in polar aprotic DCE (up to 50 mM PQ solutions were used), 9,10-PQ was less soluble in ACN and especially in 50 % aqueous ACN. 9,10-PQ solutions of up to 10 mM were tested in the former case, and up to 2 mM in the latter. This is, by all means, not a major issue since the solubility of PQs in aqueous media can be increased upon derivatisation of its core structure (e.g. through introduction of PEG groups).

The reaction between 1 mM 9,10-PQ and 10 mM EtVyO at 455 nm in 50 % aqueous ACN achieved almost full conversion after 5 minutes of irradiation as judged by LC-MS (Figure 2.10b). Similar results were obtained for reactions in ACN (up to 10 mM 9,10-PQ) and DCE (up to 50 mM 9,10-PQ) while no reactivity was observed in the absence of light. The fluorescent [4+2] cycloaddition product was stable towards prolonged irradiation (up to 15 minutes) and was, furthermore, isolated and characterised by NMR (Appendix). Conversely, the reaction was very sluggish in DMSO with approx. 20 % conversion achieved after one-hour irradiation time at 455 nm (Figure S2.24). Due to the low boiling point of EtVyO (33 °C), irradiation experiments with single LEDs were performed with excess of EGVyO, which not only boils at 143 °C but is readily soluble in water as well. With the 1 W 450 nm single LED the reaction between 2 mM 9,10-PQ and 20 mM EGVyO shows 50 % conversion after 5 minutes and full conversion after 15 minutes. The same reaction with white light LEDs (both 1 W and 3 W) on the other hand required longer irradiation times with almost 90 % conversions achieved after 30 minutes (Figure 2.10c). The reaction was naturally slightly faster with the 3 W LED lamp. These results are not surprising if one compares the broad emission spectra of the white light LEDs to the emission of the LED focused to 450 nm and thus matching the absorption maximum of PQ (430 nm) very well. One can intuitively expect that the efficiency of a lamp rises with its power (e.g. our 1 W 450 nm LED lamp gives full conversion in 15 minutes, while the published experiments require only one minute irradiation with a 10 W hand-held white light lamp<sup>[138]</sup>), however, it is important to note that lamps with different powers can only be reliably compared through their irradiance (wavelength-dependent values which illustrate the flux of photons an irradiated surface

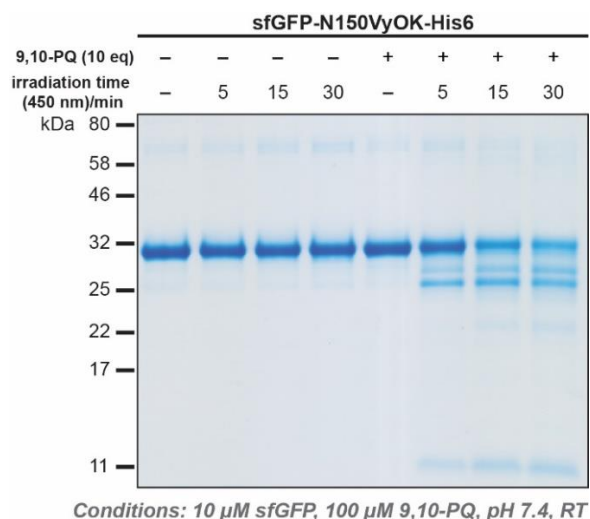
---

receives from the optical system). Both 3 W white light and 1 W 450 nm LEDs were eventually used in *in vitro* experiments on purified proteins.



**Figure 2.10.** Small molecule visible-light-induced cycloaddition reactions between 9,10-PQ and vinyl ethers. (a) Schematic representation of the respective reaction. The reactivity of 9,10-PQ with two different vinyl ethers (EtVyO and EGVyO) was tested, depending on the employed irradiation setup. (b) Almost quantitative cycloaddition product formation is observed after irradiating the 9,10-PQ and EtVyO mixture for 5 minutes at 455 nm. (c) The reaction between 9,10-PQ and EGVyO takes longer to achieve full conversion when 1 W 450 nm or 3 W white light LEDs are used. UV absorbance is measured at 254 nm. As a result, peaks corresponding to vinyl ethers are not visible.

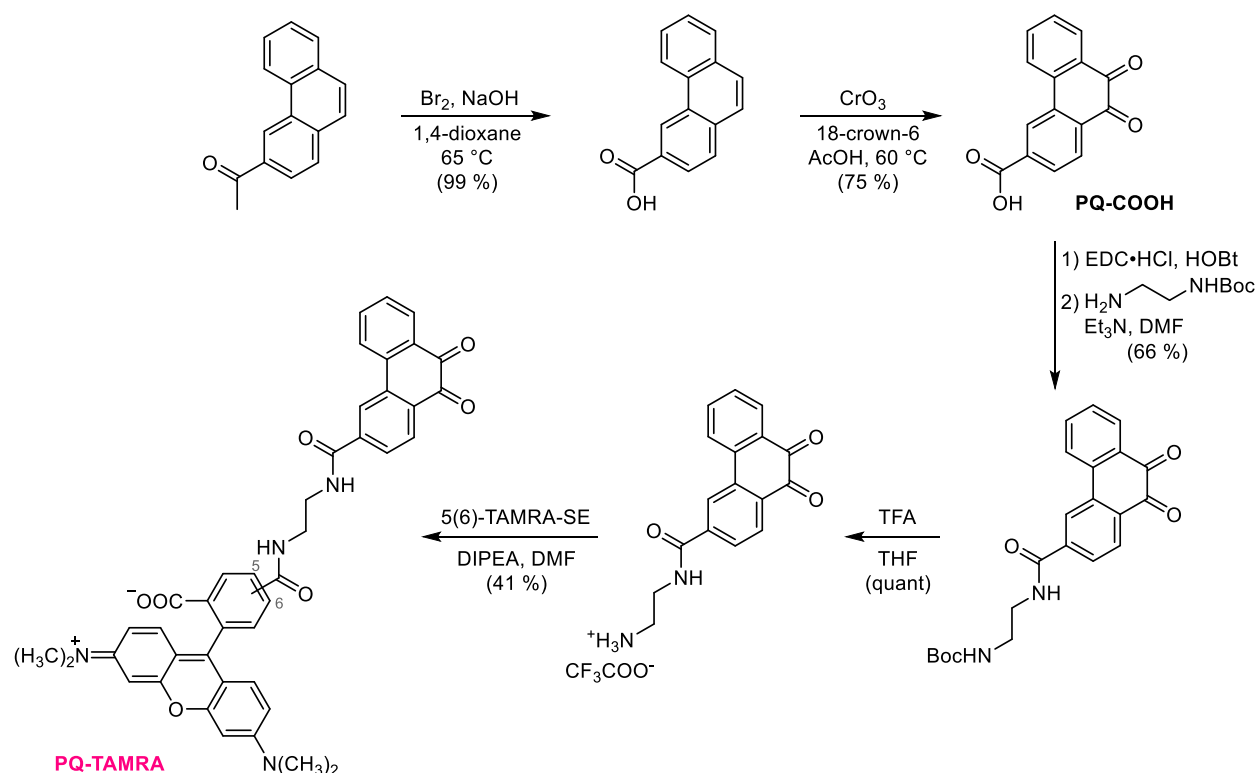
The first experiments on purified proteins were performed on our “golden standard” sfGFP. sfGFP-N150VyOK-His6 was initially irradiated with the 1 W 450 nm LED in the absence and presence of 9,10-PQ (10  $\mu$ M protein and 100  $\mu$ M 9,10-PQ; 9,10-PQ stock solution was dissolved in PBS (pH 7.4):ACN, 1:1 (v/v)). Surprisingly (but not in a good way), sfGFP showed significant degradation after prolonged irradiation (30 minutes) only with 9,10-PQ (Figure 2.11) making MS analysis difficult due to deconvolution problems of the degraded protein mixture (data not shown). Hoping to get better results with other proteins, we incorporated VyOK into myoglobin (Myo-S4VyOK-His6) and ubiquitin (Ub-K6VyOK-His6). While ubiquitin did not show any signs of degradation, myoglobin got degraded during irradiation in the absence and presence of 9,10-PQ (we can only assume that something had happened to the heme group). Analogous results were obtained with white light LEDs (data not shown). Due to these unexpected events, we decided to focus on labelling of purified ubiquitin with a PQ-TAMRA fluorophore.



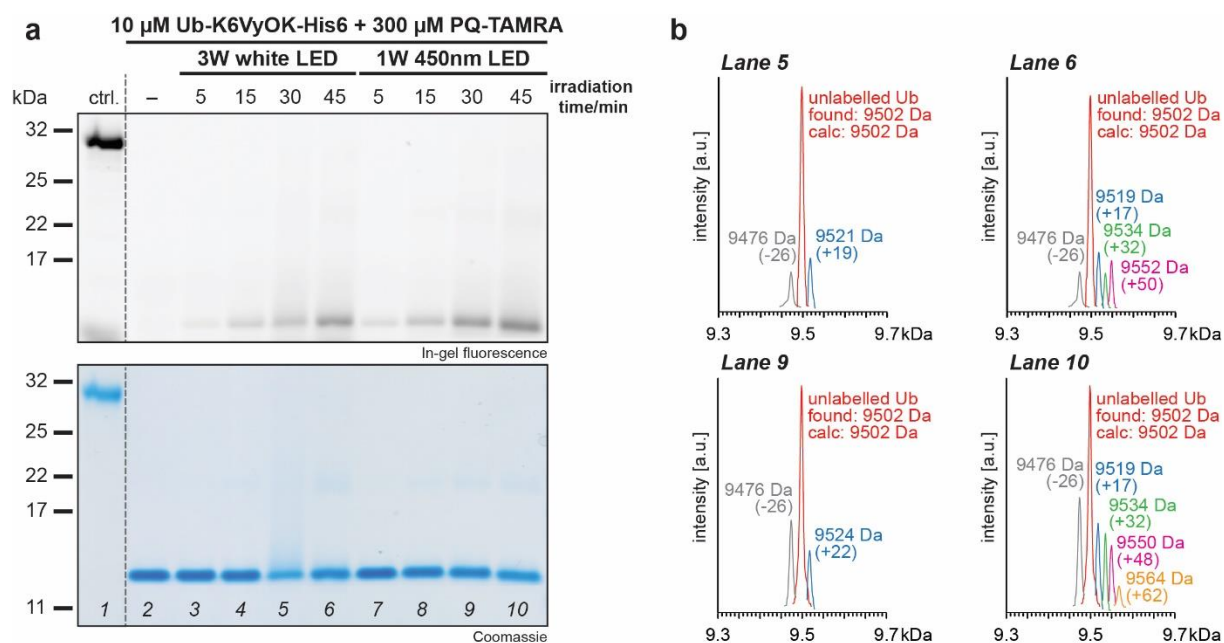
**Figure 2.11.** Degradation of sfGFP upon 450 nm irradiation only observed in the presence of 9,10-PQ, as analysed by SDS-PAGE.

PQ-TAMRA was synthesized in 20 % yield over five steps starting from the commercially available 3-acetylphenanthrene (Scheme 2.15). A haloform reaction was first performed with bromine under basic conditions to afford phenanthrene-3-carboxylic acid which was further oxidised into the corresponding PQ with chromium trioxide in acetic acid. Using standard amide coupling reagents, *N*-Boc-ethylenediamine was coupled to PQ-COOH, followed by removal of the Boc protection group with TFA. The ammonium salt thus obtained was reacted with the succinimidyl ester of TAMRA (mixture of 5- and 6- isomers, 5(6)-TAMRA-SE) in the presence of DIPEA to yield the desired PQ-fluorophore conjugate. PQ-TAMRA was purified by preparative HPLC and the stock solutions prepared in 25 % PBS (pH 7.4) in DMSO.

For subsequent experiments using PQ-TAMRA as the labelling reagent, 10  $\mu$ M Ub-6VyOK-His6 samples were first irradiated either with 3 W white light or 1 W 450 nm LEDs in the presence of 300  $\mu$ M PQ-TAMRA and analysed by in-gel fluorescence after SDS-PAGE (Figure 2.12a). There is no observable fluorescence in the sample incubated in the dark (line 2), showing that the reaction is indeed triggered by light. Although an increase of fluorescence is observed over prolonged irradiation times with both LEDs (up to 45 minutes, lanes 3-10), it is incredibly weak when compared to the standard sample in lane 1 (sfGFP-N150BCNK-His6 labelled with a TAMRA-tetrazine dye, denoted as “ctrl.” on the gel). One would also expect that the slight difference between the unlabelled (calc. mass 9502 Da) and the labelled ubiquitin (calc. mass 10208 Da) would be noticeable in the SDS-PAGE. Unfortunately, MS analysis of the reaction samples predominantly showed masses of unlabelled Ub, while no masses that would correspond to the labelled Ub were observed (Figure 2.12b). Instead, masses that possibly correspond to the hydrolysis of the vinyl group are present (9476 Da) as well as other adducts (mass differences of +17, +32, +48, etc.) that seem to indicate some sort of oxidation products of the amino acid side chains.



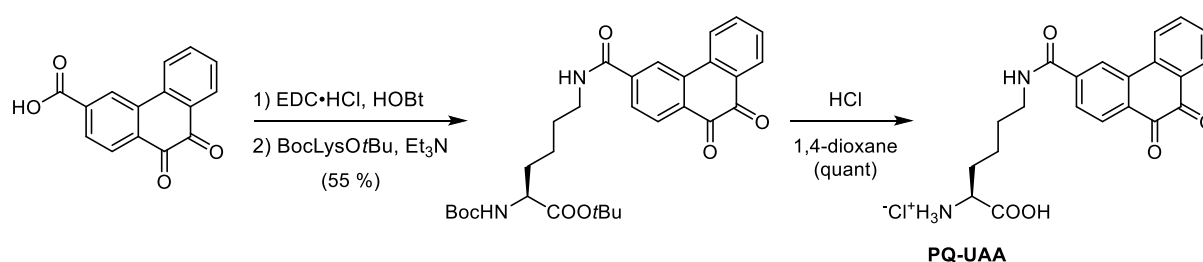
Scheme 2.15. Synthesis of rhodamine-based PQ-TAMRA.



**Figure 2.12.** *In vitro* visible-light-triggered labelling of Ub-K6VyOK-His6 with PQ-TAMRA. (a) In-gel fluorescence (upper panel) and SDS-PAGE (lower panel) of analysed probes. (b) MS characterisation of selected protein samples after photochemically triggered labelling.



In parallel to the aforementioned experiments, we prepared a PQ-bearing UAA as well by coupling PQ-COOH to the  $\epsilon$ -amino group of the Boc- and *tert*-butyl protected L-lysine, followed by deprotection with HCl (Scheme 2.16). Similar to the strategy of incorporating *ortho*NQM-based UAAs into proteins, PQ-UAA could (when incorporated into the protein of interest) allow site-specific labelling with vinyl ether probes, removing the need for excess of externally added PQs. A couple of synthetases were tested for their ability to incorporate PQ-UAA in *E. coli*, however, all cell cultures containing the UAA (2 mM per culture) had died during the process. The extent of the toxicity of PQs to living cells was not investigated within the scope of this thesis, nevertheless, past reports have shown that 9,10-PQ can elicit apoptotic cell death through production of reactive oxygen species (ROS).<sup>[207]</sup> Taken all of these results into consideration, we eventually dropped the reaction as a possible alternative to *ortho*NQMs.



**Scheme 2.16.** Synthesis of PQ-bearing UAA.

## 2.8. Conclusion and outlook

On a last note, this chapter provides a detailed description of our efforts to establish a light-induced hetero-Diels-Alder cycloaddition reaction as a tool for site-specific labelling of proteins *in vitro* and potentially *in vivo*. Our interest in light-induced reactions arises from the overall simplicity to control them as they proceed in space and time, just by applying light of suitable wavelengths and intensities and by regulating the overall irradiation time. We selected *ortho*-naphthoquinone methides (*ortho*NQMs) as reactive intermediates of interest due to several attractive reasons that spoke in their favour. First of all, *ortho*NQMs are readily generated under aqueous conditions from their corresponding precursors (*ortho*NQMPs) *via* photoelimination between 300 and 350 nm. Secondly, *ortho*NQMs rapidly react with vinyl ethers and sulfides in hetero-Diels-Alder cycloaddition reactions with inverse electron-demand, with reported second-order rate constants comparable to iEDDA reactions between tetrazines and strained alkenes/alkynes ( $k_2 \sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ). Employing fast and quantitative bioorthogonal reactions is of high importance in the community as these allow labelling on a very short time scale as well as labelling of low abundance proteins without the use of huge excess of labelling reagents. Owing to their polarised nature, *ortho*NQMs also undergo nucleophilic addition reactions. In regards to labelling of proteins, these biological nucleophiles are, naturally, nucleophilic residues of natural amino acids (Cys, Lys, Tyr, N-terminus, etc.) and water. Nevertheless, these reactions, especially the addition of thiols, have been shown to be reversible under prolonged irradiation in the presence of vinyl ethers.

Taking these features into account, we aspired to introduce *o*NQMP- or vinyl ether/sulfide-containing UAAs site-specifically into proteins of interest *via* amber suppression and label them with corresponding labelling reagents. While UAAs carrying *o*NQMPs proved to be troublesome from the synthetic and incorporation point of view, vinyl ether/sulfide UAAs VyOK and VySK were easily obtained and introduced into proteins of interest by a very efficient PylRS variant in both *E. coli* and mammalian cells. Regrettably, the labelling of proteins carrying these UAAs with *o*NQMPs-based reagents proved to be highly unspecific, presumably due to unwanted nucleophilic addition reactions with natural amino acid residues. Even though we expected such side-reactivity to be reversible during irradiation in the presence of vinyl ethers, we ultimately failed to circumvent this major issue. Consequently, we were obliged to terminate the idea of an *o*NQM-based labelling reaction. Instead, we attempted to find alternatives for labelling of proteins carrying VyOK or VySK. In this regard, we briefly considered the reactivity of thermally generated *ortho*-quinolinone quinone methides (*o*QQMs) with vinyl sulfides, nonetheless, we were confronted with the non-specificity of the labelling as seen with *o*NQMs. We then shifted our focus to the recently rediscovered visible-light-mediated [4+2] cycloaddition between vinyl ethers and 9,10-phenanthrenequinones (PQs). In spite of promising results on small molecule level, the labelling of proteins carrying VyOK proved to be very poor and on top of that PQs exhibited certain cellular toxicity.

Notwithstanding the attractive features related to employing light as an external trigger of bioorthogonal reactivity, these types of reactions are still uncommon and slightly overshadowed by other faster bioorthogonal reactions (e.g. tetrazine ligation). We hoped that quinone-methide-based reactions could arise as a new and powerful approach to decorate proteins with new functionalities by means of light, but our results ultimately highlight the difficulties in establishing highly specific reactions for use in complex biological environments. As for VyOK and VySK, we still hope to find a suitable bioorthogonal reaction that would make the site-specific incorporation of these UAAs worthwhile.

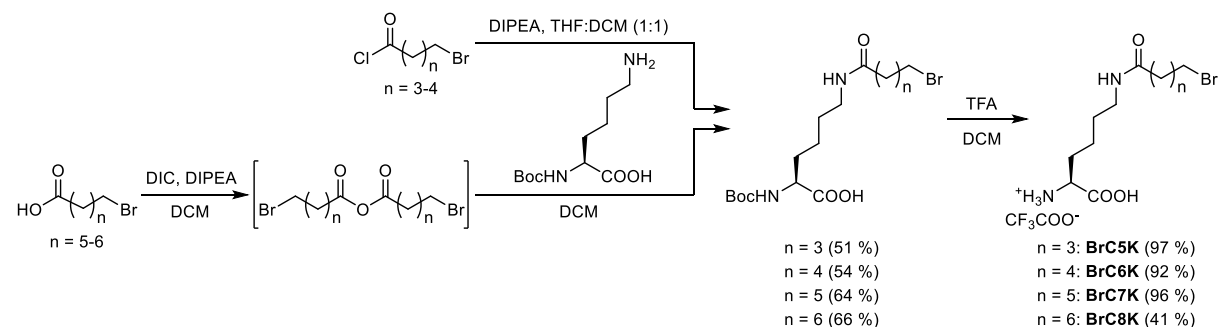
---

### § 3. PROXIMITY-TRIGGERED COVALENT STABILISATION OF LOW-AFFINITY PROTEIN COMPLEXES *IN VITRO* AND *IN VIVO*

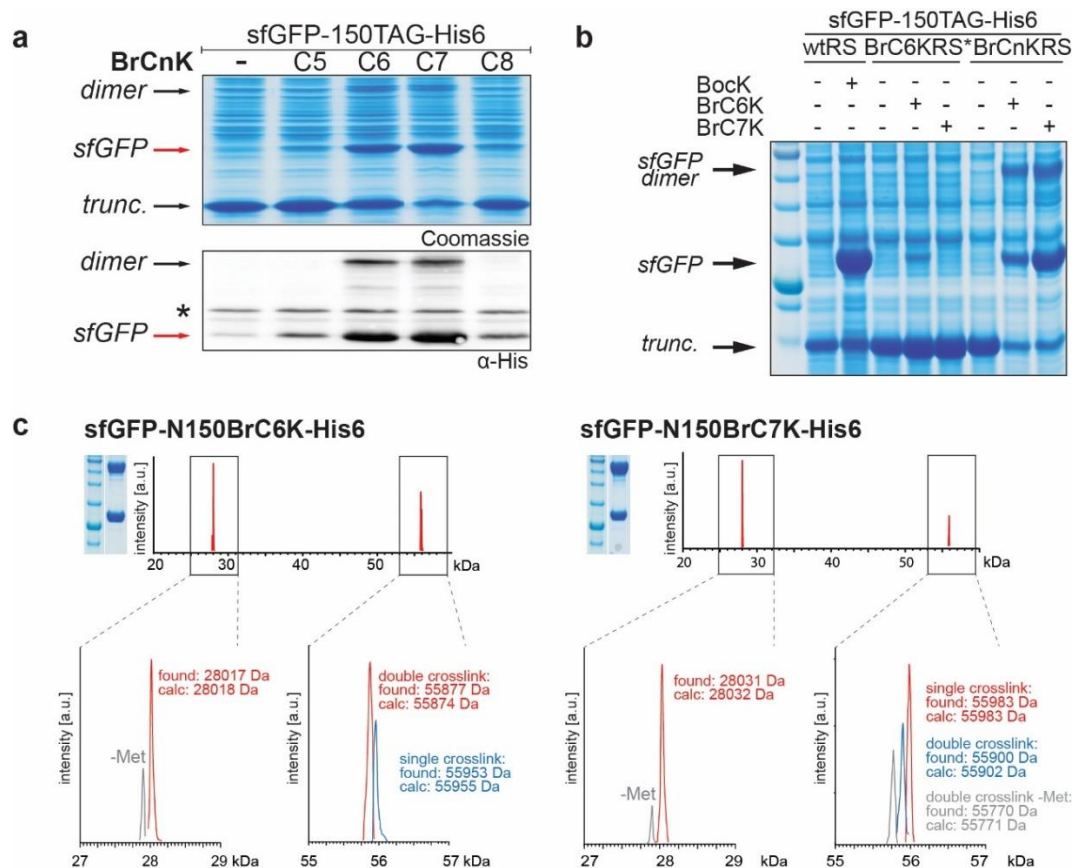
As already discussed in Chapter 1.5, the full potential of proximity-triggered crosslinking approaches has hitherto not been thoroughly exploited, with only few studies of biological processes reported in very recent literature.<sup>[53, 163f]</sup> We envisioned that this approach could be effectively used for the trapping of biologically relevant protein-protein interactions *in vivo*, in both bacteria and mammalian cells. We were especially interested in transient and low-affinity protein complexes as the attempts to study them have so far been challenging or elusive. In order to covalently trap such interactions, we needed a linker that would display enough rotational freedom when placed at the binding interface of the two protein partners without altering the nativity of their interaction and restraining the crosslinked complexes into artificial conformations. Importantly, the electrophilicity of the reactive group on the linker had to be fine-tuned to be inert under physiological conditions but react with natural nucleophilic amino acids when placed in close proximity (see Chapter 1.5). We, therefore, turned our focus towards genetically encoding bromoalkyl-bearing UAAs.<sup>[53, 163b]</sup>

#### 3.1. Synthesis and site-specific incorporation

We synthesized a series of lysine derivatives containing bromoalkyl chains with various lengths (BrC5K, BrC6K, BrC7K and BrC8K; Figure 1.19 and Scheme 3.1) in order to ensure highest possible conformational flexibility. Above all, these UAAs could span distances of up to 15 Å when reacting with proximal nucleophilic residues. The fairly simple syntheses involved the reactions of the free  $\epsilon$ -amino group of *N* $\alpha$ -Boc-protected L-lysine with activated derivatives (acyl chlorides or *in situ* generated anhydrides) of respective bromoalkyl carboxylic acids. Subsequent *N*-Boc deprotection with TFA afforded the desired UAAs as trifluoroacetate salts in moderate yields (27-60 %) over two steps (Scheme 3.1). All UAAs were used in biological experiments as 100 mM stocks in 1 M NaOH except for BrC8K (stock in 2 M NaOH) which displayed some solubility issues due to its longer aliphatic chain.



**Scheme 3.1.** Synthesis of bromoalkyl-containing lysine derivatives (BrCnK).

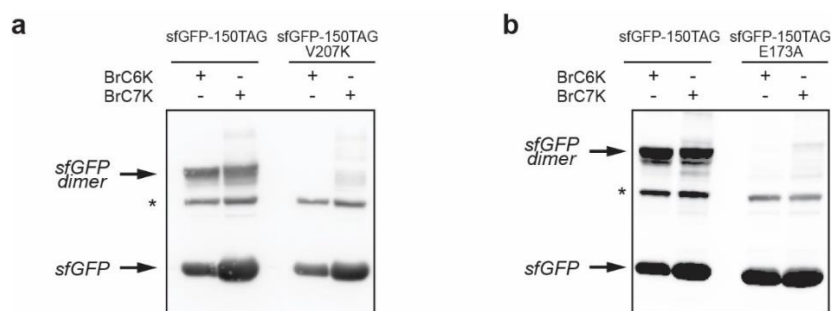


**Figure 3.1.** Site-specific incorporation of BrCnK into sfGFP-N150TAG-His6 in *E. coli*. (a) Coomassie stained SDS-PAGE (top) and  $\alpha$ -His western blot (bottom) analysis of the BrCnK incorporation by BrCnKRS. Asterisk (\*) denotes an unrelated  $\alpha$ -His binding protein. (b) Comparison of incorporation efficiencies between the novel *Mb*BrCnKRS and the previously published *Mm*BrC6KRS\*<sup>[163b]</sup>. *Mm*BrC6KRS\* only shows modest incorporation of BrC6K and does not incorporate BrC7K, while *Mb*BrCnKRS efficiently incorporates both UAAs. BockK incorporation by the wild-type *Mb*PylRS is used as a positive control. The dimerisation of sfGFP can only be observed when incorporating BrC6(7)K with our novel synthetase. (c) MS analysis of purified sfGFP-N150BrC6K-His6 (left) and sfGFP-N150BrC7K-His6 (right). Mass spectra clearly show the formation of the covalent crosslink within the sfGFP dimer during protein expression. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

In order to site-specifically incorporate the aforementioned UAAs, a panel of 25 different *Mb*PylRS variants was screened and a novel mutant (containing mutations Y271M, L274A and C313A, and termed BrCnKRS) was found to efficiently incorporate BrCnK into superfolder GFP (sfGFP) in *E. coli*, at position 150 (sfGFP-N150TAG-His6, Figure 3.1). The sfGFP construct additionally contained a C-terminal His6-tag for purification purposes while BockK, which is recognised by the wild-type PylRS, was used as a positive incorporation control. Although BrCnKRS is able to incorporate the entire BrCnK series, the incorporation efficiencies of BrC5K and BrC8K were almost 10-times lower when compared to BrC6(7)K (Figure 3.1a). It is likely that the linkers of these two UAAs were either too short or too long for optimal positioning into the active site of the synthetase. BrC5K and BrC8K were, therefore, discarded from further studies. We compared the efficiencies of our BrCnKRS variant with the previously published *Mm*PylRS mutant (containing a single Y384F mutation) evolved for BrC6K<sup>[163b]</sup> through site-specific incorporation of BrC6(7)K into sfGFP-

N150TAG-His6 (Figure 3.1b). The published *MmBrC6KRS\** shows very modest incorporation of BrC6K into sfGFP, however it does not recognize BrC7K as its substrate. BrCnKRS, on the other hand, shows remarkable incorporation efficiencies for both UAAs with 20-30 mg of proteins isolated per litre of cell culture after His6-tag affinity chromatography purification.

When expressing sfGFP over prolonged times (typically overnight, > 16 hours), a second band with conspicuous molecular weight of 60 kDa was observed in SDS-PAGE gels (Figures 3.1a and 3.1b). It is noteworthy that this only occurred when BrCnKRS and BrC6(7)K were used. Mass spectrometry (MS) of the purified sfGFP confirmed the band to be a mixture of single and double crosslinked homodimers of sfGFP; the single crosslink occurring when the UAA in one molecule reacts with a natural amino acid in the other, while the double crosslink results from the reaction of UAAs in both molecules with the corresponding residues in the other molecule (Figure 3.1c). Initially surprised by the dimer formation, we turned to MS/MS analysis for clarification and identified the crosslink as a stable ester linkage between BrC6K and a glutamate residue (E173 to be precise, Figure S3.1). To further support this finding, we introduced an E173A mutation into sfGFP which abolished the formation of the crosslink (Figure 3.2a). Moreover, we disrupted the dimerization tendency of sfGFP, which is of low-affinity ( $K_D \approx 100 \mu\text{M}$ ), through introduction of a V207K mutation into the hydrophobic dimerization interface,<sup>[2c, 208]</sup> confirming thereby that the formation of the crosslink was promoted through proximity (Figure 3.2b). Altogether, these findings demonstrated that our approach could trigger *in vivo* crosslink formation through proximity during protein expression, highlighting the need for a very efficient synthetase at the same time.

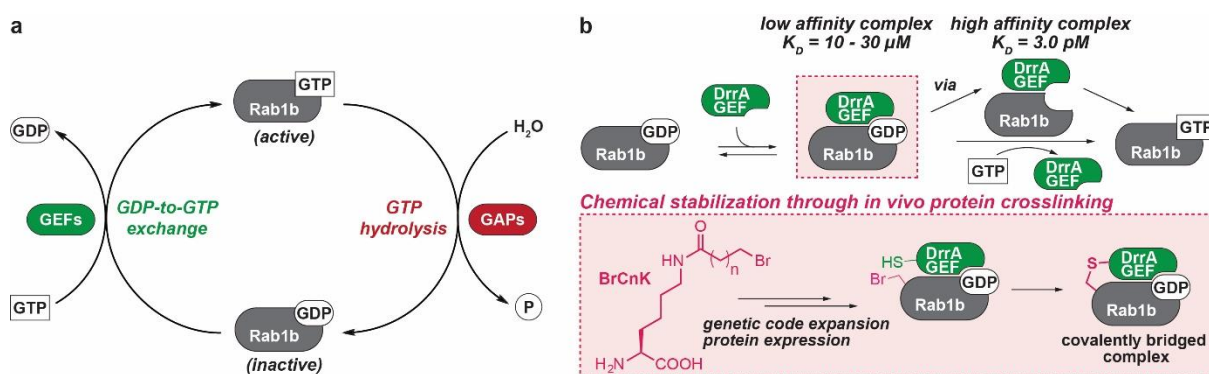


**Figure 3.2.** Introduction of a (a) V207K and (b) E173A mutation abolishes the covalent crosslink formation during expression of BrC6(7)K-bearing sfGFP mutants. Asterisk (\*) denotes an unrelated  $\alpha$ -His binding protein. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

### 3.2. Choosing the ideal proof-of-principle system

With encouraging initial results in hand, we aspired to establish our crosslinking approach on an appropriate low-affinity protein complex. In collaboration with the group of Prof. Aymelt Itzen, we chose the complex between the small human GTPase Rab1b and the guanine nucleotide exchange factor (GEF) domain of its effector protein DrrA from *Legionella pneumophila* as an ideal proof-of-principle target for several important reasons. Rab1b (or Ras-related protein Rab1b) is a member of a class of intracellular trafficking proteins and

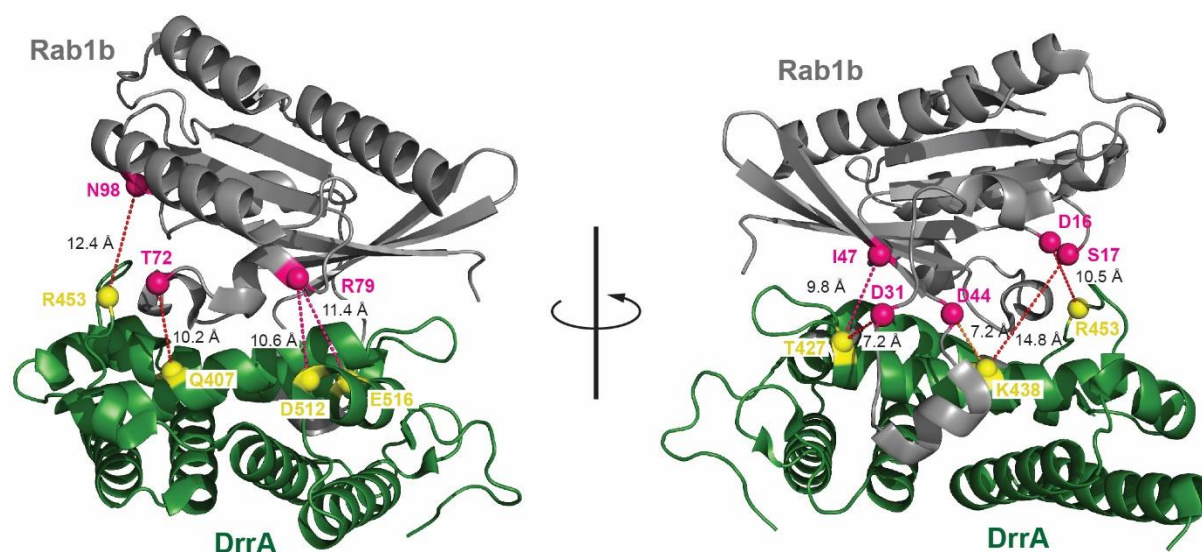
plays an important part in regulation of the vesicular transport from the endoplasmic reticulum to the Golgi apparatus.<sup>[209]</sup> Interestingly, Rab1b functions as a molecular switch; the protein is active in the GTP-bound state and inactive when bound to GDP. In its native environment, the inactivation of Rab1b through hydrolysis of GTP is stimulated by GTPase-activating proteins (GAPs) while the release of GDP and binding of GTP is modulated by GEFs (Figure 3.3a).<sup>[209b]</sup> The molecular mechanism of Rab1b switching involves five important structural features – loops G1-G5 – that bind GDP/GTP directly.<sup>[210]</sup> While the G1 loop (or the P-loop) binds the  $\beta$ -phosphate of GDP and GTP, G2 (switch I) and G3 (switch II) discriminate between GDP and GTP by recognizing the  $\gamma$ -phosphate. The guanine base is specifically recognised by loops G4 (NKxD) and G5 (SAK). The two switch loops, importantly, mediate the conformational changes of Rab1b upon nucleotide exchange.



**Figure 3.3.** (a) Schematic representation of the switching mechanism of Rab1b mediated by GAPs and GEFs. (b) *In vivo* protein crosslinking approach to stabilise the low-affinity ternary complex Rab1b:GDP:DrrA<sub>340-533</sub>. Figure in (b) was reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

As a response to the infection by bacterial pathogens *via* phagocytosis, eukaryotic cells employ different Rab proteins to eventually fuse the phagosome with the lysosome, killing the bacteria in the process. Nevertheless, pathogens like *L. pneumophila* have developed a smart way to circumvent the fusion with the lysosome by excreting over 250 virulence effector proteins including DrrA.<sup>[211]</sup> The central GEF domain of DrrA (amino acids 340-533, DrrA<sub>340-533</sub>) catalyses the GDP-to-GTP exchange in Rab1b, recruiting thereby the protein to become a part of the *Legionella*-containing vacuole during infection, ultimately preventing the lysis.<sup>[212]</sup> Depending on whether GDP is bound to Rab1b ( $K_D \approx 10-30 \mu M$ ) or not ( $K_D \approx 3 pM$ ), the binding affinity of the Rab1b:DrrA<sub>340-533</sub> complex differs by 6-7 orders of magnitude (Figure 3.3b).<sup>[213]</sup> Consequently, only the 3D structure of the artificial, high-affinity nucleotide-free Rab1b:DrrA<sub>340-533</sub> assembly has been reported (PDB: 3JZA)<sup>[213]</sup> while efforts to crystallize the low-affinity Rab1b:GDP:DrrA<sub>340-533</sub> complex have so far been elusive. Supported by the structural information from the Rab1b:DrrA<sub>340-533</sub> complex (Figure 3.4 and Table 3.1), we selected potential sites for Rab1b:DrrA<sub>340-533</sub> crosslinking in the presence of GDP. We opted for amino acid residues whose C $\alpha$  atoms are placed between 7 and 16 Å of the Rab1b/DrrA<sub>340-533</sub> binding interface (distances that our UAAs could span) and we substituted them pairwise with BrC6(7)K in Rab1b and cysteines in DrrA<sub>340-533</sub>. The

promising crosslinking positions included D16, S17, D31, Y37, D44, I47, T72, R79 and N98 in Rab1b and K377, Q407, T427, K438, R453, D512 and E516 in DrrA<sub>340-533</sub>.



**Figure 3.4.** Crystal structure of the high-affinity nucleotide-free Rab1b<sub>3-174</sub>:DrrA<sub>340-533</sub> complex (PDB: 3JZA).<sup>[213]</sup> Based on this structure, we selected several positions within the binding interface of Rab1b (grey) and DrrA (green) to be substituted pairwise with BrC6(7)K (pink spheres) and cysteines (yellow spheres), respectively. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

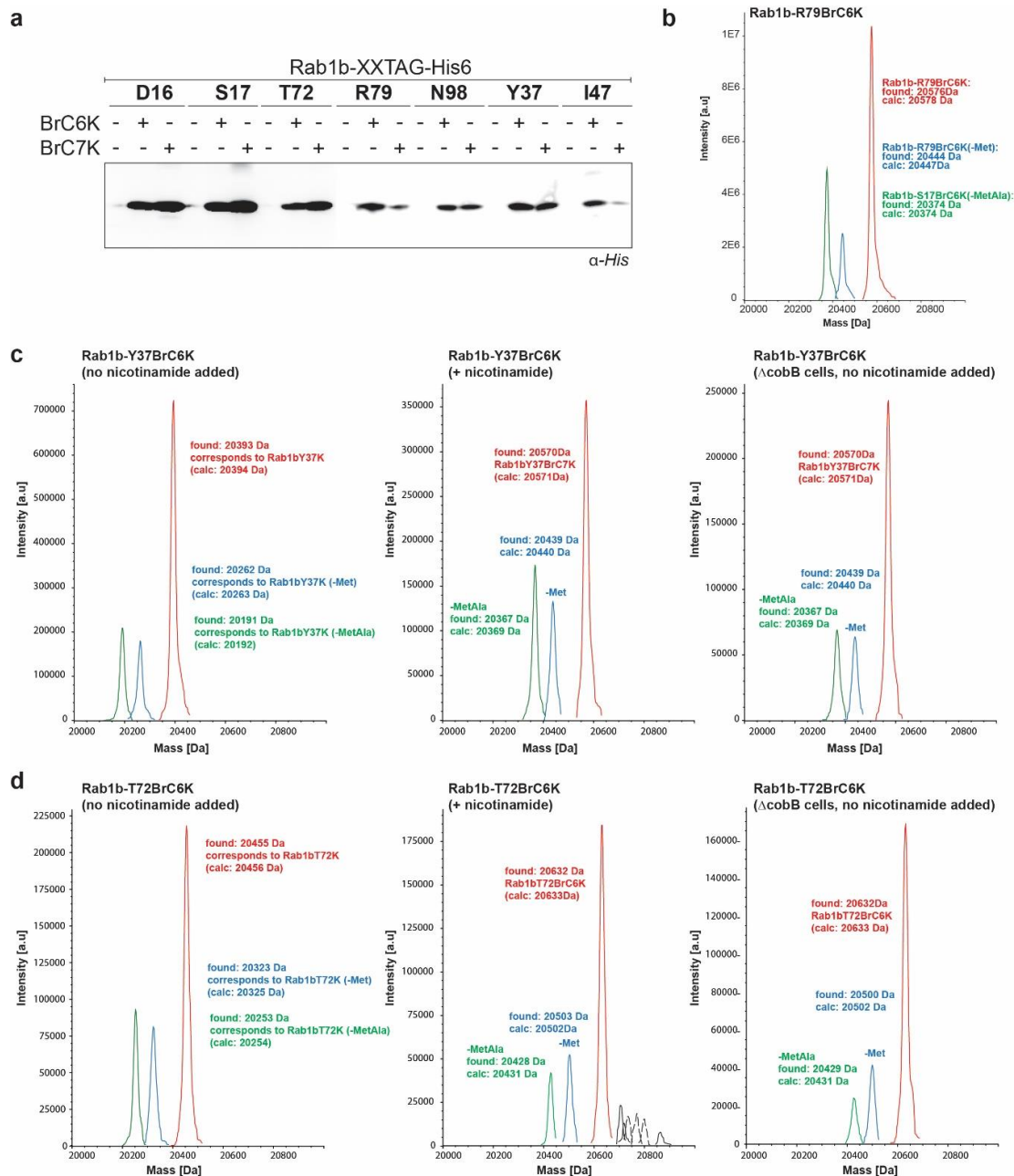
**Table 3.1.** Distances between the C $\alpha$  atoms in the nucleotide-free Rab1b<sub>3-174</sub>:DrrA<sub>340-533</sub> complex (PDB: 3JZA).<sup>[213]</sup>

Residue in Rab1b	Residue in DrrA <sub>340-533</sub>	Distances between the C $\alpha$ atoms in the Rab1b:DrrA <sub>340-533</sub> complex / Å
D16	R453	7.2
S17	K438	14.8
D31	T427	7.2
Y37	K377	10.2
D44	K438	7.2
I47	T427	9.8
T72	Q407	10.2
R79	D512	10.6
R79	E516	11.4
N98	R435	12.4
R79	R453	21.8 (negative control)

### 3.3. *In vitro* crosslinking experiments

As with sfGFP, the novel synthetase BrCnKRS efficiently incorporated BrC6(7)K into Rab1b-His6 in *E. coli* with all of the TAG mutants being expressed and purified in acceptable yields (2-6 mg per litre of cell culture) and characterized by MS (Figures 3.5a,b and Figure S3.2a in the Appendix). For two Rab1b variants (positions Y37 and T72), we observed

position-dependant hydrolysis of the amide bond of the UAAs. Similar observations have been reported for other UAAs containing  $N\epsilon$ -amide linkages and are in fact mediated by the nicotinamide adenine dinucleotide (NAD)-dependant bacterial sirtuin deacylase CobB, the only known deacylase in *E. coli*.<sup>[214]</sup> In order to solve this issue, we efficiently inhibited CobB by expressing all proteins in the presence of 10 mM nicotinamide (NAM) but we could also produce intact UAA-modified Rab1b mutants in a *cobB* knockout strain from *E. coli* DH10 $\beta$ , as judged by MS (Figures 3.5c,d and Figure S3.2b,c in the Appendix).

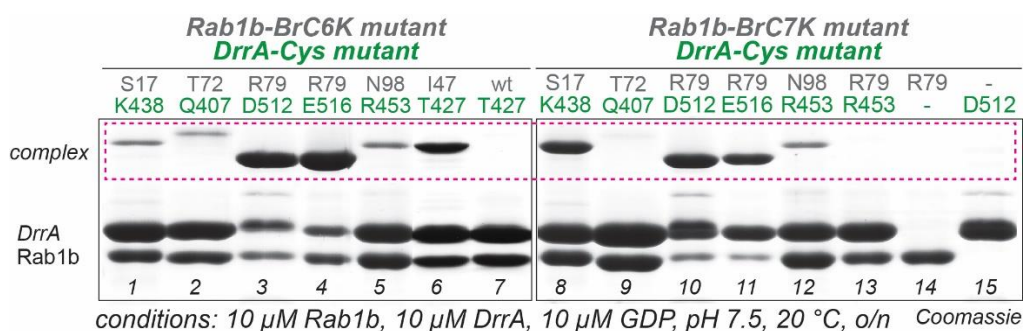


**Figure 3.5.** Expression and characterisation of purified Rab1b mutants bearing BrC6(7)K. (a)  $\alpha$ -His western blot analysis of site-specific incorporation of BrC6(7)K into Rab1b-His6 with TAG codons at different positions in *E. coli*. (b) Representative example of MS characterisation of purified Rab1b-R79BrC6K mutant. Red peak denotes the intact full-length protein mass. Blue peak denotes the observed mass where the N-terminal methionine was cleaved off, while the green peak shows the mass corresponding to cleavage of the N-terminal



Met-Ala sequence.<sup>[215]</sup> (c) MS characterisation of purified Rab1b-Y37BrC6K mutant. When expressed in *E. coli* DH10 $\beta$ , a peak corresponding to lysine incorporation (20455 Da) was observed (left) due to posttranslational degradation of BrC6K by the *E. coli* sirtuin CobB. The action of the deacylase could be inhibited by addition of nicotinamide to the cell culture during expression (middle) or by expressing the mutant in the DH10 $\beta$  *cobB* knockout strain (right). Peaks corresponding to the main observed species are denoted in red, while blue and green peaks denote N-terminal methionine and Met-Ala cleavage, respectively. (d) Same as (c) for Rab1b-T72BrC6K mutant. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

After expressing and purifying respective cysteine mutants of His6-DrrA<sub>340-533</sub>, we performed first *in vitro* crosslinking experiments in the presence of GDP (10  $\mu$ M of each Rab1b/DrrA<sub>340-533</sub> mutant and 10  $\mu$ M of GDP). As depicted in Figure 3.6, crosslinks (bands at approx. 44 kDa in SDS-PAGE) with modest to outstanding efficiencies were observed for most of the tested protein pairs. Rab1b-R79UAA and DrrA<sub>340-533</sub>-D512C or DrrA<sub>340-533</sub>-E516C, for instance, represent the most efficient crosslinking pairs. The significantly lower crosslinking of the Rab1b-N98/DrrA<sub>340-533</sub>-R453C pair can be explained by the intramolecular crosslink formation in Rab1b-N98 (presumably with the proximal K100), as judged by MS (Figure S3.2a). The Rab1b-Y37/DrrA<sub>340-533</sub>-K377 pair, on the other hand, was not crosslinked by BrC6(7)K at all (data not shown). Nevertheless, it has to be noted that there were no observed crosslinks where the distances between the respective C $\alpha$  atoms surpassed 20 Å (e.g. Rab1b-R79BrC7K and DrrA<sub>340-533</sub>-R453C), highlighting that only proximal sites can be spanned by the flexible thioether linker.



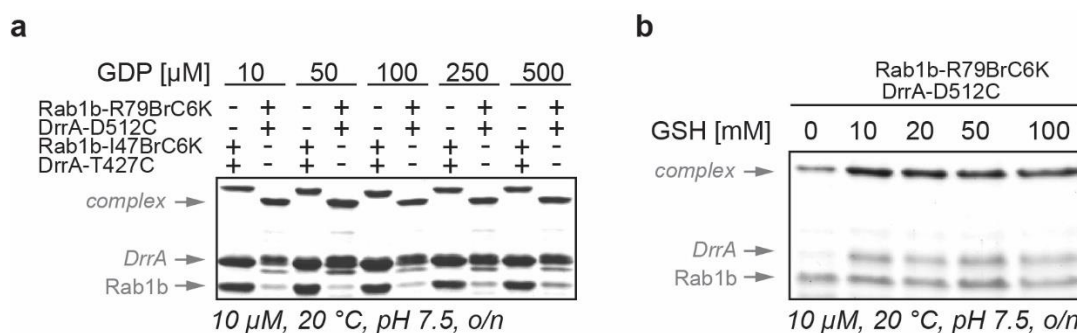
**Figure 3.6.** *In vitro* crosslinking between purified Rab1b-BrC6(7)K-His6 and His6-DrrA-cysteine mutants, analysed by SDS-PAGE. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Since our crosslinking approach is essentially an S<sub>N</sub>2 reaction, we accordingly screened various crosslinking conditions *in vitro* (varying incubation times, temperatures, pH and GDP concentrations) for two of the crosslinking pairs (Rab1b-R79BrC6K/DrrA<sub>340-533</sub>-D512C and Rab1b-I47BrC6K/DrrA<sub>340-533</sub>-T427C, Figure S3.3 in the Appendix and Figure 3.7a). Equimolar amounts (10  $\mu$ M) of Rab1b and DrrA mutants were used and the reactions stopped at different time points by denaturing the proteins in SDS loading buffer. The reaction mixtures were subsequently analysed by SDS-PAGE. Indeed, the crosslinking efficiencies improved with longer incubation times and higher pH and temperatures (as expected for an S<sub>N</sub>2 reaction) but it must be noted that the most efficient crosslink between Rab1b-R79UAA and DrrA<sub>340-533</sub>-D512C already formed within 10 minutes at pH 7.0 and 4 °C. The addition of 100-fold excess of GDP (1 mM) did not disrupt the complex formation

either (Figure 3.7a), pointing out that a low-affinity complex was undoubtedly crosslinked, as the non-covalently linked Rab1b: DrrA<sub>340-533</sub> complex is known to dissociate in the presence of excess GDP.<sup>[216]</sup> We also analysed the Rab1b-R79BrC6K/DrrA<sub>340-533</sub>-D512C and Rab1b-I47BrC6K/DrrA<sub>340-533</sub>-T427C reaction mixtures by MS and observed the masses corresponding to the crosslink formation (Figure S3.4a,b in the Appendix). Furthermore, we determined the  $K_D$  of the GDP-bound complex between Rab1b-R79BrC6K and DrrA<sub>340-533</sub>-D512C to be 13  $\mu$ M under pseudo-first-order conditions (Figure S3.5 in the Appendix).

### 3.4. *In vivo* crosslinking experiments in *E. coli* and mammalian cells

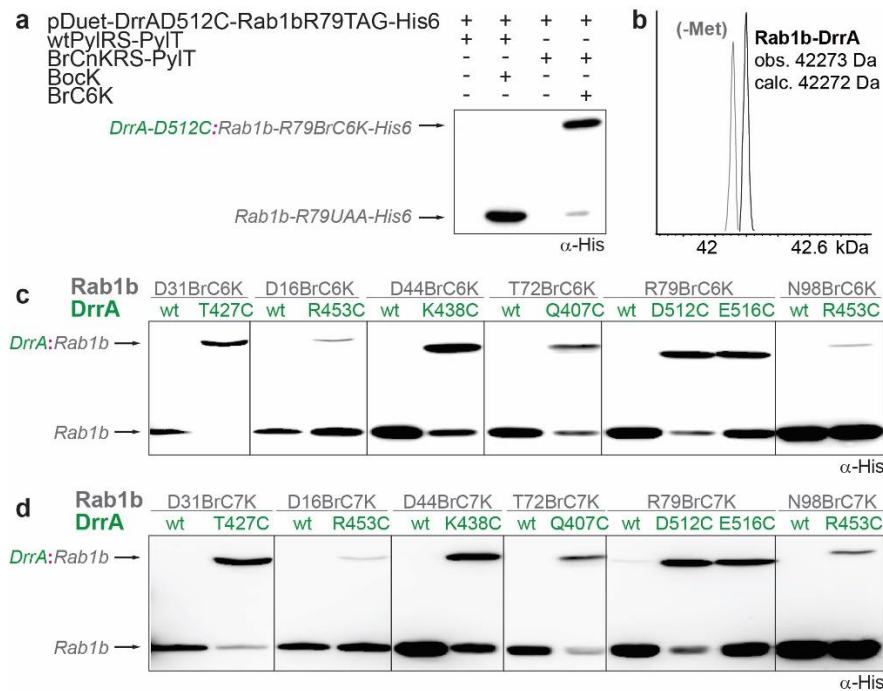
As a follow-up to the successful *in vitro* crosslinking experiments, we shifted our focus towards establishing the crosslinking approach under *in vivo* conditions. To this end we aspired to co-express the Rab1b and DrrA<sub>340-533</sub> mutant pairs containing BrC6(7)K and cysteines together with the orthogonal BrCnKRS/tRNA<sup>CUA</sup> pair and purify the crosslinked complexes *via* affinity tag chromatography. Preceding any crosslinking attempts *in vivo*, we had to test if the presence of increasing concentrations of glutathione (GSH) would affect our intramolecular crosslinking reaction (in other words, test if BrC6(7)K-bearing Rab1b variants would react with external nucleophiles such as GSH at physiological conditions, which would impede formation of crosslinks with corresponding DrrA variants). GSH is typically present in the cytosol of mammalian cells in the lower millimolar range (1-10 mM)<sup>[217]</sup> while the values in *E. coli* readily exceed 10 mM.<sup>[218]</sup> Gratifyingly, the presence of GSH concentrations as high as 100 mM (10,000-fold excess when compared to the concentrations of both protein partners added in equimolar amounts (10  $\mu$ M)) did not affect the crosslink formation efficiency (Figure 3.7b).



**Figure 3.7.** Effect of increasing (a) GDP and (b) glutathione (GSH) concentrations on *in vitro* crosslinking between Rab1b-R79BrC6K and DrrA<sub>340-533</sub>-D512C, analysed by SDS-PAGE. Equimolar amounts (10  $\mu$ M) of both proteins were used. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

We, therefore, constructed Duet vectors containing genes expressing different BrC6(7)K-bearing Rab1b (His6-tagged at the C-terminus) and cysteine-containing DrrA<sub>340-533</sub> mutant pairs (as established by successful crosslinking experiments *in vitro*). Duet vectors expressing wild-type DrrA<sub>340-533</sub> were used as negative controls. When co-expressed with the orthogonal BrCnKRS/tRNA<sup>CUA</sup> pair in *E. coli*, we observed *in vivo* crosslink formation only for protein pairs where the UAAs were placed at sites proximal to nucleophilic cysteine side chains.

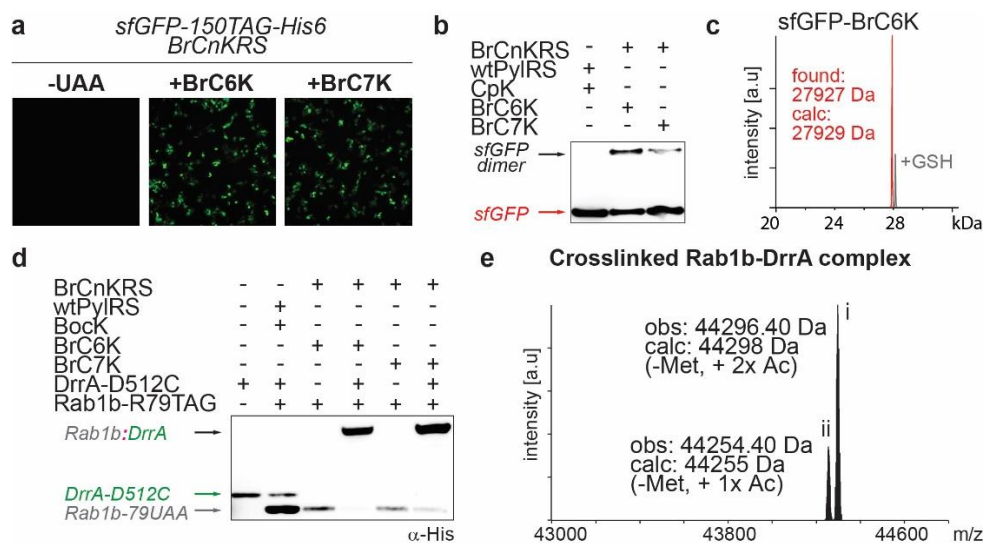
Analogous to the results obtained in *in vitro* experiments, the Rab1b-R79BrC6(7)K and DrrA<sub>340-533</sub>-D512C showed excellent crosslinking efficiencies and the obtained complex was isolated and characterized by MS (Figure 3.8 and Figure S3.4c).



**Figure 3.8.** *In vivo* crosslinking of the Rab1b:GDP:DrrA complex in *E. coli*. (a)  $\alpha$ -His western blot analysis of the *in vivo* crosslink formation between Rab1b-R79BrC6K and DrrA<sub>340-533</sub>\_D512C (both genes on the same Duet vector) in the presence of BrCnKRS/tRNA<sup>CUA</sup> pair. Incorporation of BocK into Rab1b-R79TAG by the wild-type PylRS is used as a negative crosslinking control. (b) MS characterisation of the purified *in vivo* formed crosslinked complex. (c) *In vivo* crosslink formation between different Rab1b-BrC6K and DrrA-cysteine mutant pairs, co-expressed with the BrCnKRS/tRNA<sup>CUA</sup> pair. Duet vectors co-expressing wild-type DrrA were used as negative crosslinking control. (d) Same as (c) for Rab1b-BrC7K mutants. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Considering that the BrCnKRS/tRNA<sup>CUA</sup> pair is orthogonal in mammalian cells as well, we transferred the mutations allowing incorporation of BrC6(7)K into a mammalian optimised *MmPylRS*<sup>[219]</sup> and we successfully incorporated BrC6(7)K into sfGFP-150TAG in HEK293T cells, as validated by fluorescence imaging of cells expressing full-length sfGFP and western blot analysis of cell lysates (Figure 3.9a,b). We purified the BrC6K variant of sfGFP as well and confirmed its identity by MS. It must be noted that a minor peak (approx. 30 %) of a GSH adduct was observed due to prolonged expression times (> 24 h), yet the formation of the sfGFP dimer still ensued, confirming the proximity-triggered nature of the crosslinking reaction (Figure 3.9c and Figure S3.6). Seeing that proximity-triggered crosslinking can be performed within the cytosol of mammalian cells, we went to demonstrate that the reaction occurs between the established Rab1b/DrrA<sub>340-533</sub> protein partners as well. To this end, we introduced the genes for Rab1b-R79TAG-His6 (plus a C-terminal His6-tag) and His6-DrrA<sub>340-533</sub>-D512C (plus an N-terminal His6-tag) into mammalian cell vectors, each containing four copies of the orthogonal tRNA PylT.<sup>[219]</sup> Next, we triple transfected HEK293T cells with these plasmids together with the plasmid encoding for either BrCnKRS

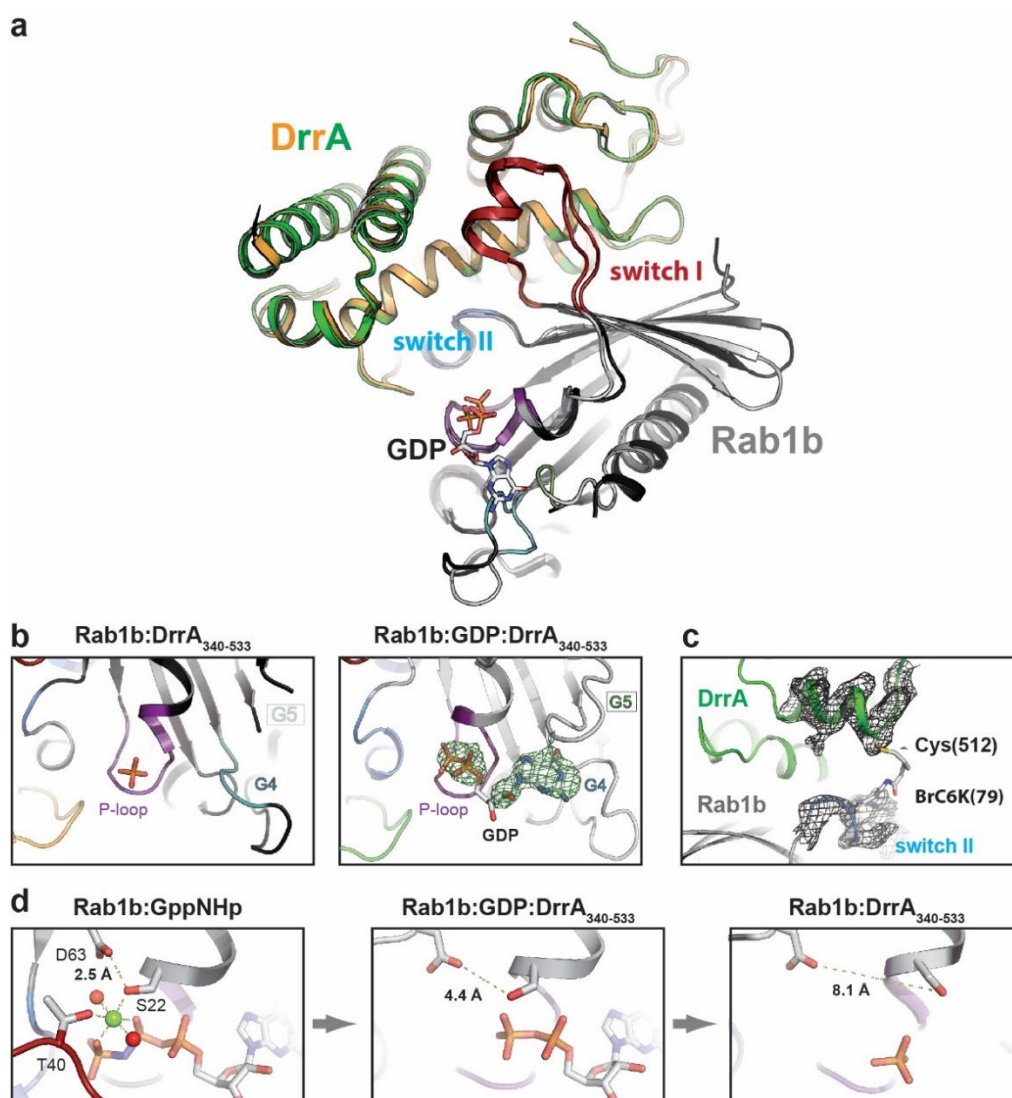
or the wild-type PylRS. The success of the crosslinking reaction was monitored by western blotting while the His-tag purified protein complex was analysed by MS. While the Rab1b variant containing BocK (incorporated by the wild-type PylRS) did not yield any crosslinked complex (negative control, lane 2, Figure 3.9d), BrC6(7)K-bearing Rab1b variants trapped DrrA<sub>340-533</sub>-D512C *via* flexible thioether bridge very efficiently so that the bands corresponding to His6-DrrA<sub>340-533</sub> are not visible in the western blot (lanes 4 and 6, Figure 3.9d). The ESI-MS analysis of the covalently crosslinked complex showed the monoisotopic masses of the twofold N-terminally acetylated complex (both on Rab1b or DrrA, 44296.40 Da) and the single N-terminally acetylated complex (either on Rab1b or DrrA, 44354.40 Da) missing an N-terminal methionine residue (Figure 3.9e).



**Figure 3.9.** Establishing *in vivo* crosslinking in mammalian cells. (a) Fluorescence microscopy of HEK293T cells expressing sfGFP-N150TAG-His6 in the absence or presence of 1 mM BrC6(7)K. (b)  $\alpha$ -His western blot analysis of sfGFP-N150TAG-His6 expression in the presence of BrC6(7)K shows formation of covalently crosslinked sfGFP homodimers. Incorporation of CpK by the wild-type PylRS is used as a positive incorporation control and a negative crosslinking control. (c) MS characterisation of purified sfGFP-N150BrC6K mutant after 24 h expression. The minor peak in grey corresponds to the mass of the GSH adduct. (d)  $\alpha$ -His western blot analysis of *in vivo* crosslink formation between Rab1b-R79BrC6(7)K and DrrA<sub>340-533</sub>-D512C in HEK293T cells. Expression of Rab1b-R79BocK in the presence of the wild-type PylRS is used as a positive incorporation and a negative crosslinking control. (e) MS characterisation of the purified covalently crosslinked Rab1b-R79BrC6K-His6 and DrrA<sub>340-533</sub>-D512C complex shows presence of single and double N-terminal acetylation. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

With the successful establishment of our crosslinking approach, mediated by site-specific incorporation of BrC6(7)K with the highly efficient orthogonal BrCnKRS/tRNA<sup>CUA</sup> pair, we applied it to determine the crystal structure of the low-affinity ternary Rab1b-R79BrC6K:GDP:DrrA<sub>340-533</sub>-D512C complex. The crosslinking reaction between Rab1b-R79BrC6K-His6 and DrrA<sub>340-533</sub>-D512C was performed on a preparative scale by incubating equimolar amounts (20  $\mu$ M) of the purified proteins overnight at pH 8.0 and 4 °C. The crosslinked complex was then purified by size-exclusion chromatography (SEC) in the presence of 500  $\mu$ M GDP so that any non-covalently linked Rab1b: DrrA<sub>340-533</sub> complexes would be removed. Collaborating with Matthias Müller (MPI Dortmund), the crystal structure

of the covalently linked ternary complex was determined at 2.5 Å resolution. The overall crystal structure is similar to the previously published GDP-free complex (PDB: 3JZA)<sup>[213]</sup> with Rab1b adopting the fold common for small GTPases (six stranded β-sheets surrounded by five α-helices and five nucleotide binding loops G1-G5, Figure 3.10a). The interaction with DrrA<sub>340-533</sub> occurs mainly through the P-loop and switch I/switch II regions of Rab1b. Importantly, unambiguous electron density for GDP is visible in the nucleotide-binding pocket. Distinct structural differences are, consequently, present in the G4 and G5 loops as they now adopt a nucleotide-binding-competent conformation (Figure 3.10b).



**Figure 3.10.** Structural characterisation of the covalently crosslinked Rab1b-R79BrC6K-His6:GDP:DrrA<sub>340-533</sub>-D512C complex. (a) Overlay of the previously published structure of the nucleotide-free Rab1b:DrrA<sub>340-533</sub> complex (PDB: 3JZA, DrrA in orange)<sup>[213]</sup> and the crosslinked GDP-bound ternary crosslink (DrrA in green). (b) Close up view comparison of the nucleotide-binding pockets (left: Rab1b:DrrA<sub>340-533</sub>; right: Rab1b:GDP:DrrA<sub>340-533</sub>). The  $F_o-F_c$  electron density map surrounding the nucleotide as green mesh ( $\sigma = 3.0$ ) shows the presence of GDP as well as the conformational reorganisation of the G4 and G5 loops. (c) Due to its high structural flexibility, no electron density is observed for the thioether linker. The surrounding  $2F_o-F_c$  electron density map is contoured at  $\sigma = 1.0$ . (d) A possible order of molecular changes during nucleotide-exchange reaction depicted from left (Rab1b:GppNHp, PDB: 3NKV)<sup>[220]</sup> to right (Rab1b:DrrA<sub>340-533</sub>, PDB: 3JZA)<sup>[213]</sup> Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Due to the high conformational flexibility of the thioether linker, its electron density is unsurprisingly not visible (Figure 3.10c). Above all, the artificially introduced covalent linker does not affect the general structural features of the protein complex. Furthermore, we postulate that the obtained structure of the Rab1b:GDP:DrrA<sub>340-533</sub> offers new insights into the molecular changes during GDP displacement as it likely represents the Rab1b structure just before the release of GDP. This idea is supported by the linear increase in the distances of amino acid residues, particularly S22 and D63, that coordinate the divalent magnesium ion needed for nucleotide binding in the active site. The distances between these two amino acids gradually increase from the structures of Rab1b:GppNHp<sup>3</sup> (2.5 Å, PDB: 3NKV)<sup>[220]</sup> to Rab1b:GDP:DrrA<sub>340-533</sub> (4.4 Å) to Rab1b:DrrA<sub>340-533</sub> (8.1 Å), suggesting that the displacement of Mg<sup>2+</sup> might happen prior to the GDP release (Figure 3.10d).

### 3.5. Expanding the crosslinking approach towards MS/MS-cleavable UAAs

Considering all of the results presented above, our crosslinking approach allowed successful entrapment of biologically relevant low-affinity protein-protein interactions in both *E. coli* and mammalian cells. We anticipate that this approach could become convenient for uncovering the structure and dynamics of protein-protein interactions and understanding their complex interaction networks. Although X-ray crystallography has majorly contributed to the structural elucidation of many proteins and stable protein complexes (with more than 135,000 crystal structures currently deposited in PDB<sup>[221]</sup>), it is not without significant limitations. Apart from difficult and sometimes even impossible crystallisation processes and reliance on high quality of the protein crystals, X-ray diffraction provides a static snapshot of a protein's conformation which can significantly differ from the one in solution. Consequently, the crystallographic data on its own does not fully reflect the protein dynamics. Access to high resolution structural information by solution-based methods such as NMR is similarly limited by protein size (e.g. relatively small proteins or protein domains). Still, NMR can provide information on the dynamics of various protein segments as a degree of probability that these segments can inhabit a certain conformation.

To circumvent the aforementioned limitations, recent years have seen an emergence of techniques based on MS such as hydrogen/deuterium exchange (HDX) MS<sup>[222]</sup>, high-resolution MS footprinting<sup>[223]</sup> and crosslinking MS (XL-MS)<sup>[224]</sup>. The advantages of these techniques lie in their sensitivity and high throughput, with the latter, XL-MS, capable of capturing native protein-protein interactions and providing detailed information about the interaction interface. XL-MS generally proceeds by reacting proteins (mostly pure proteins or protein complexes) with a bifunctional chemical crosslinker that covalently tethers amino acid residues within the distances defined by the linker length. The crosslinked proteins are then

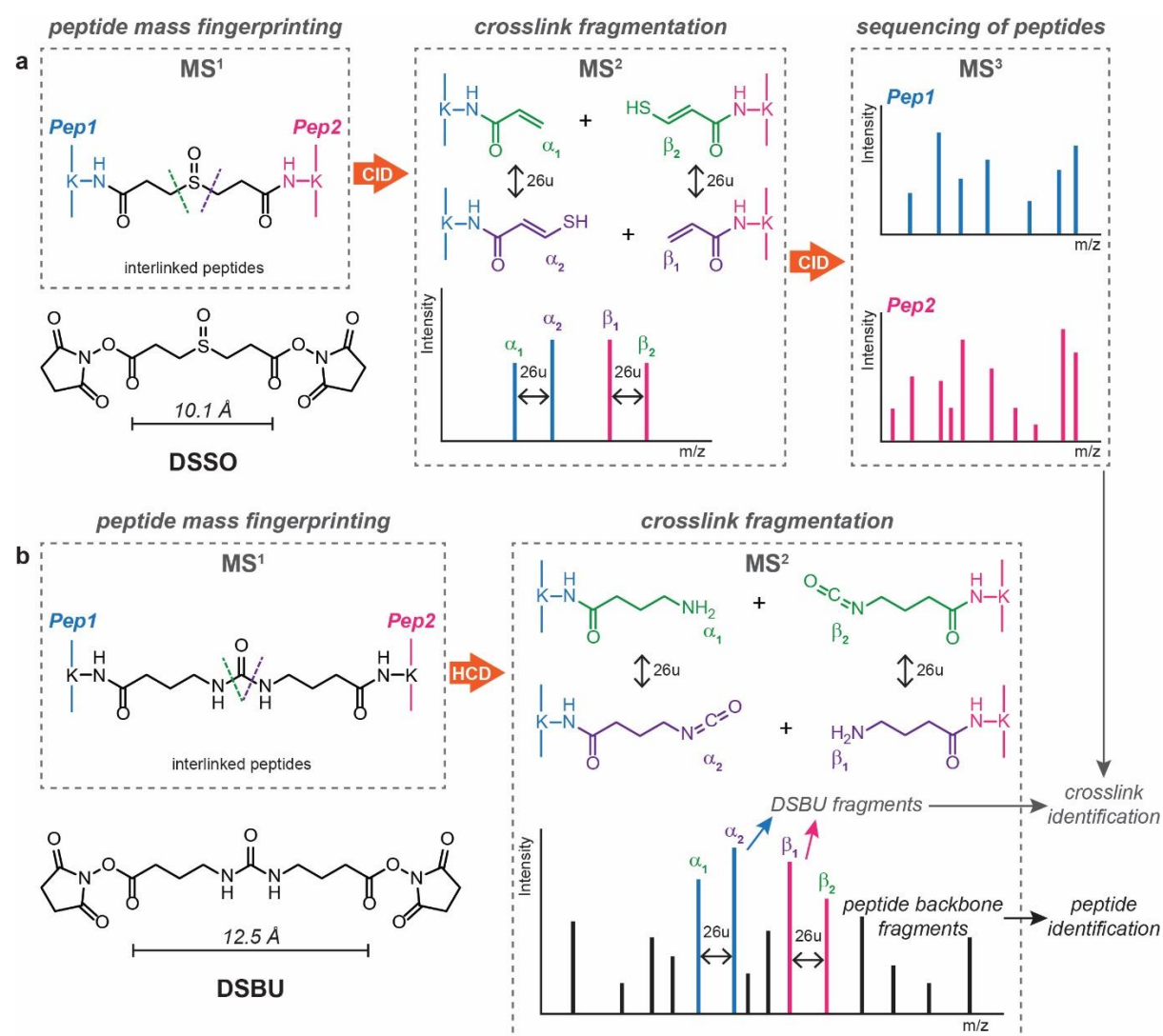
---

<sup>3</sup> 5'-guanylyl imidodiphosphate (GppNHp) is a non-hydrolysable analogue of GTP in which an amine group replaces one of the oxygen atoms of the triphosphate moiety. GppNHp binds tightly to G-proteins in the presence of Mg<sup>2+</sup> and forms stable Ras complexes.

---

digested enzymatically (e.g. trypsin digestion) while the peptide mixture is analysed by liquid chromatography-electrospray ionisation tandem MS (LC/ESI-MS/MS or MS<sup>2</sup>). The crosslinked peptides and the crosslinking sites are subsequently identified by bioinformatics analysis of the obtained MS data.<sup>[225]</sup> Although it sounds simple on paper, comprehensive analysis of the MS data is extremely challenging and requires powerful software tools due to low-abundance of crosslinked peptides in the peptide mixture, convoluted fragmentation patterns in MS/MS and non-specificity of the crosslinking reaction. Major efforts have, therefore, been undertaken not only to develop algorithms and software for meticulous characterisation of crosslinked peptides, but to design novel chemical crosslinkers which would facilitate the identification process as well.<sup>[226]</sup> These include cleavable crosslinkers<sup>[227]</sup>, videlicet reagents with labile covalent bonds within the spacer region whose specific cleavage is either chemically-,<sup>[228]</sup> photo-,<sup>[229]</sup> or MS-induced.<sup>[226, 230]</sup> While the photolabile and chemically-cleavable linkers allow the separation of the crosslinked peptides before MS analysis,<sup>[229]</sup> MS-cleavable linkers produce characteristic fragment ion patterns during MS<sup>2</sup>, enabling accurate identification of the crosslinked peptides. With the majority of the MS-cleavable crosslinkers the fragmentation method of choice is collision-induced dissociation (CID).<sup>[231]</sup> In short, the precursor ions, accelerated by an electric potential, enter a collision cell where they repeatedly collide with high pressure of chemically inert and energised gas (e.g. He, Ar, N<sub>2</sub>). The repeated collisions cause the kinetic energy of precursor ions to convert to internal energy and to build up until bond fragmentation occurs and the product ions can be analysed by MS<sup>2</sup>. Noteworthy examples of CID-cleavable bonds include the adjacent C-S bonds in sulfoxides<sup>[230a, 232]</sup> and the C-N bonds of ureas.<sup>[233]</sup> Apart from the assignment of the fragment ion patterns characteristic to the crosslink, additional backbone fragmentation of the connected peptides is necessary for definite characterisation of the crosslinking site. Therefore, sulfoxide-containing crosslinkers such as DSSO<sup>[230a]</sup>, whose C-S bonds are significantly more labile than peptide bonds, require multistage MS (MS<sup>n</sup>) analyses. After fingerprinting the digested peptide mixture (MS<sup>1</sup>), the crosslinker is cleaved at its labile C-S bonds by low-energy CID affording characteristic fragments that allow calculation of the masses of the crosslinked peptides (MS<sup>2</sup>). Consecutive MS<sup>3</sup> experiments on selected crosslinker ion fragments result with extensive fragmentation of the peptide backbone (peptide sequencing). Urea-based crosslinkers (e.g. DSBU<sup>[230b, 233a, 234]</sup>) on the other hand typically require higher dissociation energies, similar to those needed for peptide backbones, allowing concurrent analysis of fragment ions characteristic to the linker and the peptide backbone fragments at the MS<sup>2</sup> level (Figure 3.11). Nevertheless, one has to deal with a decreased signal-to-noise ratio, when compared to MS<sup>3</sup> experiments, due to the missing isolation step.

---



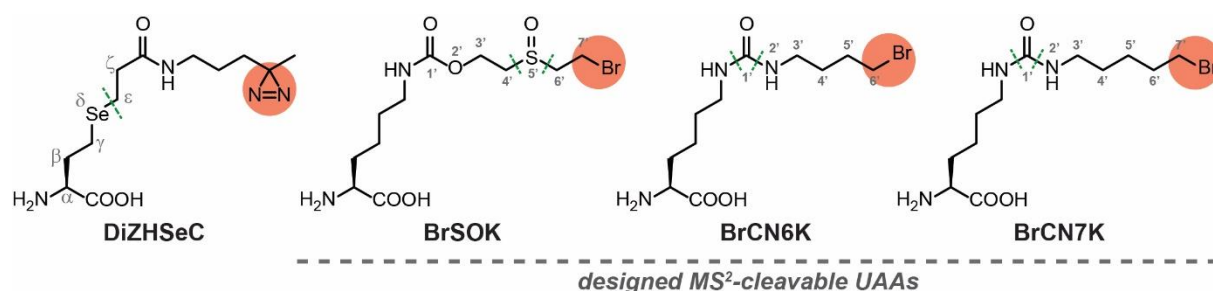
**Figure 3.11.** Commonly used MS<sup>2</sup>-cleavable chemical crosslinkers and MS analysis workflows. (a) Labile sulfoxide-bearing DSSO allows MS<sup>n</sup> analyses of the crosslinked peptides with CID fragmentation.<sup>[230a]</sup> Minimal peptide backbone cleavage at the MS<sup>2</sup> stage affords fragment ion patterns (blue and pink) characteristic to the crosslink and allows MS<sup>3</sup> fragment sequencing for unambiguous identification of crosslinked peptides. (b) MS<sup>2</sup> analysis workflow of peptides crosslinked with urea-containing DSBU during high-energy collision-induced dissociation (HCD or HCDI). Characteristic fragment ion patterns (blue and pink) allow identification of the interpeptide crosslinks. Due to parallel fragmentation of the peptide backbone (black) at higher energies, simultaneous identification of the crosslinking site is possible on the MS<sup>2</sup> level.

Despite the attractiveness of characteristic crosslinking fragmentation associated with MS-cleavable crosslinkers, the problem of the selectivity of the crosslinking reaction remains unsolved with three types of crosslinked products obtained: interlinked or type 2, intralinked or type 1 and “dead-end” or type 0 peptides. The interlinked peptides, naturally, give the most valuable structural information but the presence of all three types of peptides in an already complex mixture does obstruct the fragment analysis process. In order to further facilitate this and make the crosslinking reaction slightly more precise, we decided to expand our established crosslinking approach towards MS<sup>2</sup>-cleavable linkers by designing and synthesizing bromoalkyl-bearing UAAs containing either sulfoxide or urea moieties in the sidechain (Scheme 3.2). As with BrCnK, these UAAs would span distances up to 13 Å when



reacting with proximal nucleophilic residues. We anticipate a workflow that would encompass site-specific incorporation of these UAAs into a protein of interest, *in vivo* crosslinking upon protein expression and subsequent affinity-tag purification of the interlinked complex. Furthermore, after enzymatic digestion, MS<sup>2</sup> and computational analysis we envision that these UAAs could allow identification of new inter-protein interactions as well as exact crosslinking sites, mapping thereby the corresponding binding interfaces. It must be noted that hitherto only a single example of a crosslinking UAA carrying an MS-identifiable label has been published (DiZHSeC, Scheme 3.2).<sup>[235]</sup> As an expansion over a previous genetically encoded cleavable photocrosslinker<sup>[236]</sup>, DiZHSeC contains a diazirine moiety, a selenium atom in place of a  $\delta$ -carbon and a methylene group replacing the  $\zeta$ -amine. After crosslinking, treatment with H<sub>2</sub>O<sub>2</sub> causes oxidative cleavage of the Se $_{\delta}$ -C $_{\epsilon}$  bond to produce a stable acrylamide moiety that can be easily identified by MS. By incorporating DiZHSeC into a chaperone HdeA (bait protein), *in situ* cleavage and MS-label transfer after protein photocrosslinking (IMAPP) allowed identification of multiple new HdeA client proteins (prey proteins).

### 3.5.1. Design and site-specific incorporation of MS<sup>2</sup>-cleavable UAAs

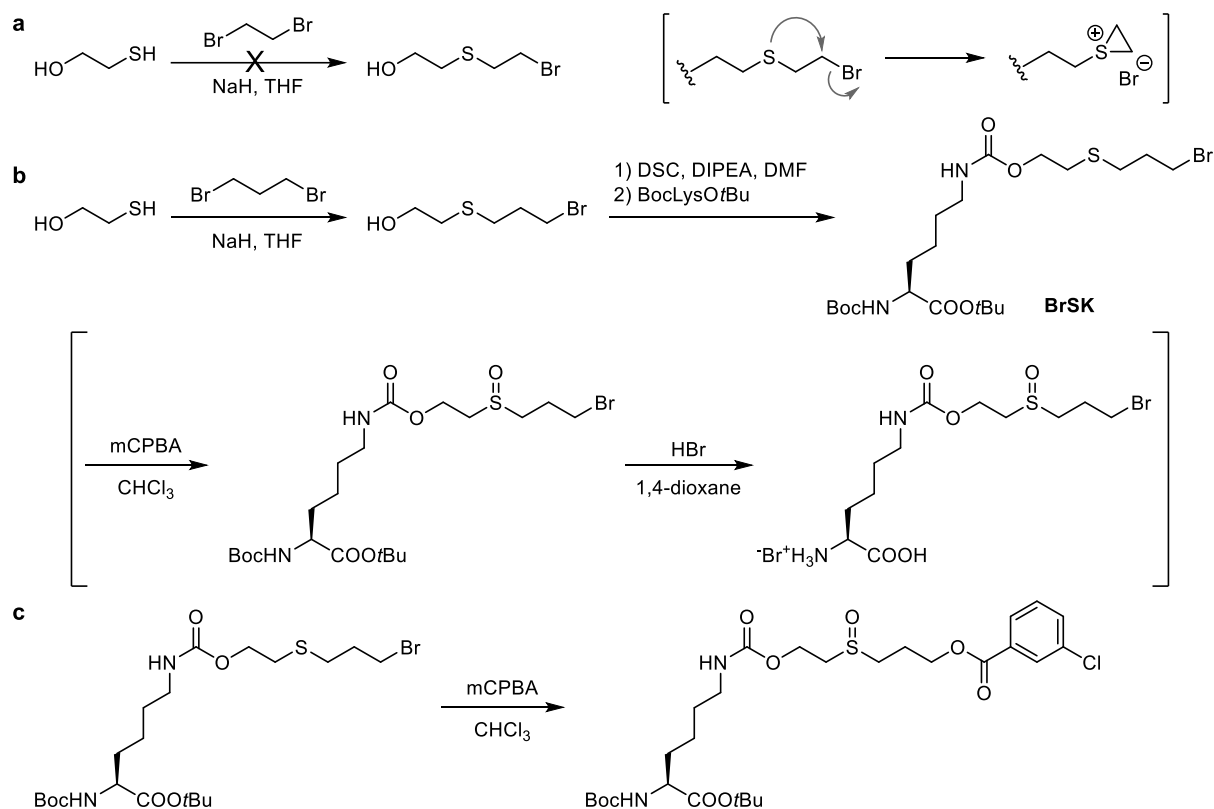


**Scheme 3.2.** Structures of the published UAA containing an MS-identifiable label (DiZHSeC)<sup>[235]</sup> and MS<sup>2</sup>-cleavable UAAs designed within the scope of this thesis. Red circles denote photo- or chemical crosslinking moieties. Green dashed lines represent cleavable bonds within the molecules.

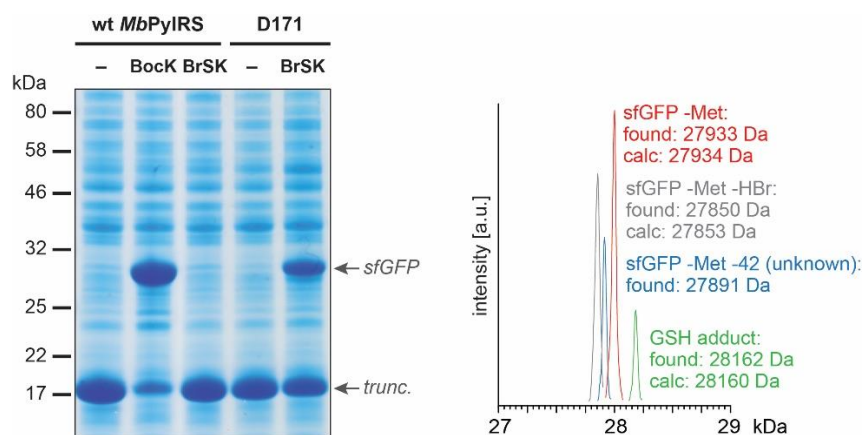
Our approach to designing MS<sup>2</sup>-cleavable UAAs was intuitive; since BrCnKRS efficiently incorporated BrC6(7)K while BrC5(8)K were incorporated poorly, we decided to restrain the sidechain length to that of BrC6(7)K so that our new UAAs are likely substrates for the same synthetase. While the easiest way to introduce a urea moiety into the BrC6(7)K scaffold was the replacement of the amide group, we opted to introduce the sulfoxide group as part of the linker so that we would not lose the similarity of BrSOK to Pyl. The amide group of BrC6(7)K was furthermore replaced with a carbamate, as we observed multiple times from past experiences that carbamate-bearing UAAs are generally better incorporated than the amide ones. Moreover, carbamates do not get cleaved by deacylases such as CobB.

Initial synthetic efforts for obtaining BrSOK are shown in Scheme 3.3 and involve selective S-alkylation of  $\beta$ -mercaptoethanol with 1,2-dibromoethane. The reaction was, however, unsuccessful presumably due to the elimination of the bromide ion by intramolecular nucleophilic substitution affording a highly-reactive three-membered cyclic sulfonium ion. The sulfonium ion gets readily trapped by nucleophiles – a reaction that is notoriously associated with compounds related to sulfur mustard agents and is considered to

be responsible for their toxicity (Scheme 3.3a).<sup>[237]</sup> Nevertheless, the S-alkylation reaction worked with 1,3-dibromopropane which would extend the length of the originally conceived linker for a methylene group (Scheme 3.3b). The bromoalcohol obtained this way was successfully coupled to a Boc- and *tert*-butyl protected L-lysine. Nevertheless, the following sulfide oxidation step ultimately led to the abandonment of this synthetic strategy. Aside from the sulfide oxidation, the *m*-chlorobenzoic acid by-product efficiently substituted the reactive bromide in the molecule (Scheme 3.3c). Interestingly, we were able to site-specifically incorporate the deprotected bromosulfide carrying UAA BrSK into sfGFP with BrCnKRS (denoted as D171), purify the protein and analyse it by MS (Figure 3.12). Other than the full-length protein mass without the N-terminal methionine (27933 Da), MS analysis showed presence of a glutathione adduct (28162 Da) and a species corresponding to the loss of HBr (27850 Da). The latter could arise either from elimination of the bromide ion and formation of an allyl thioether on the UAA itself or by intramolecular crosslinking reaction with a proximal nucleophilic amino acid residue. Sulfide oxidation attempts on purified protein with H<sub>2</sub>O<sub>2</sub> did not yield any desired products as well (data not shown).

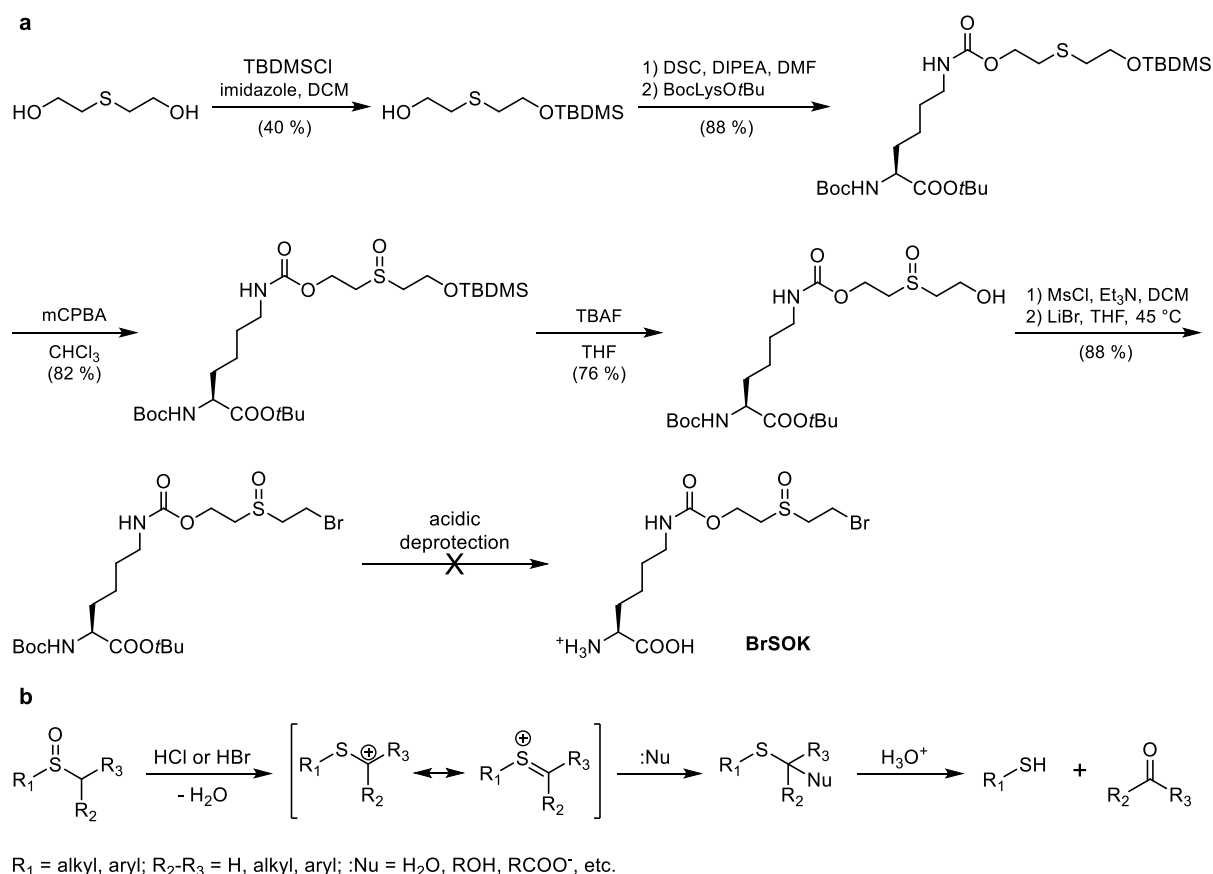


**Scheme 3.3.** (a) S-alkylation attempt of  $\beta$ -mercaptoethanol with 1,2-dibromoethane and the mechanism of the presumable elimination of the bromide ion through intramolecular nucleophilic substitution. (b) Proposed synthesis of a sulfoxide-bearing UAA via S-alkylation of  $\beta$ -mercaptoethanol with 1,3-dibromopropane. The steps in square brackets were unsuccessful. (c) Oxidation of BrSK with mCPBA affords an undesired *m*-chlorobenzoic acid adduct.



**Figure 3.12.** Site-specific incorporation of sulfide-bearing BrSK into sfGFP-N150TAG-His6 in *E. coli*. SDS-PAGE analysis of the BrSK incorporation by BrCnKRS (D171) (left) and MS characterisation of the purified sfGFP variant (right). The peak in blue with the mass of 27891 Da could not be assigned to any plausible species.

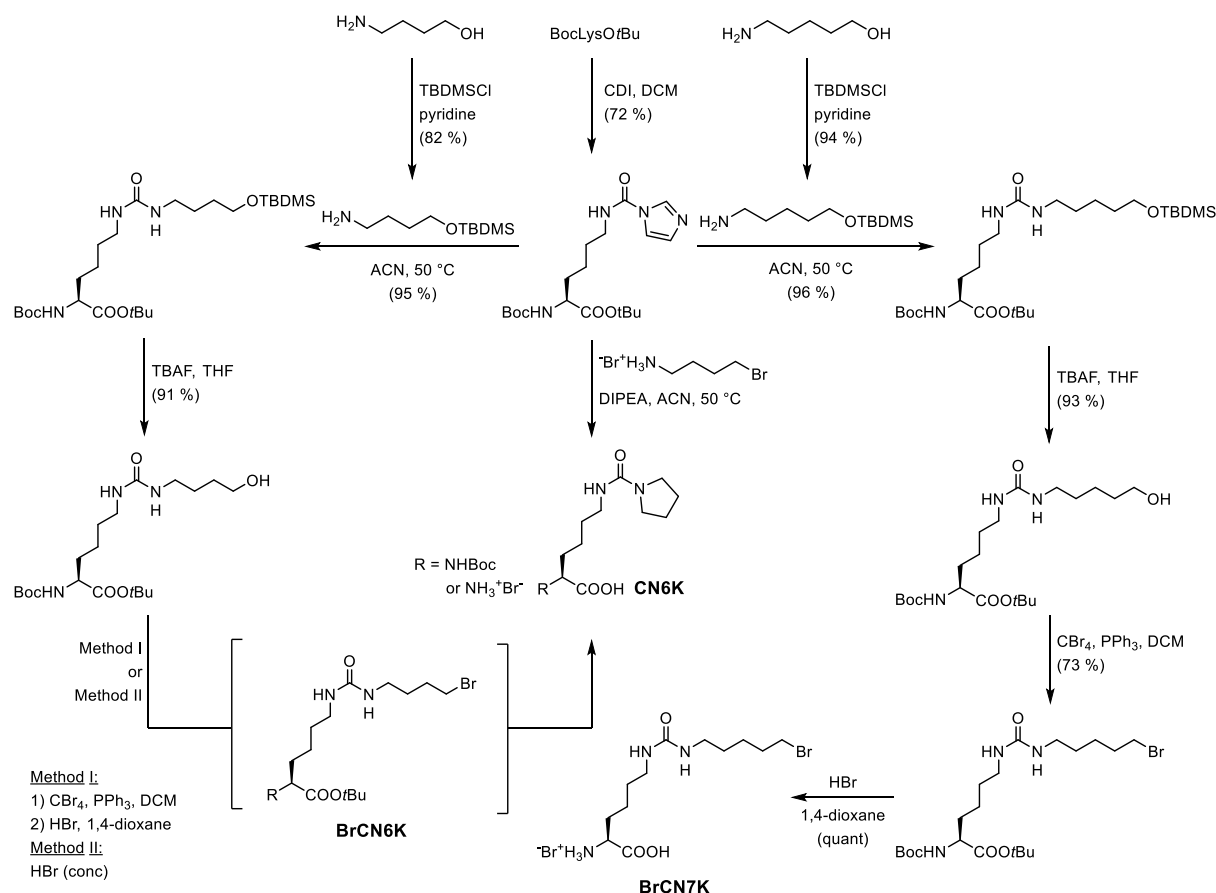
Considering the aforementioned difficulties, we were compelled to modify our synthetic strategy and decided to first introduce the sulfoxide moiety and only then the reactive bromide. Eventually, Boc- and *tert*-butyl-protected BrSOK was prepared in 19 % yield over five steps from commercially available 2,2'-thiodiethanol (Scheme 3.4a). A TBDMS protection group was first selectively introduced on one hydroxyl group of 2,2'-thiodiethanol. The product was then activated with DSC and coupled to Boc- and *tert*-butyl-protected L-lysine *via* a carbamate linkage. Oxidation of the sulfide was performed on the resulting amino acid with mild excess of mCPBA in order to prevent overoxidation into sulfone. After TBDMS deprotection with TBAF, the alkyl bromide was finally introduced by mesylation of the alcohol and subsequent bromination with LiBr. Unfortunately, the last deprotection step where both Boc- and the *tert*-butyl ester are simultaneously removed under acidic conditions did not yield the desired UAA. Deprotection with TFA afforded only lysine through expulsion of CO<sub>2</sub> and likely release of a vinylsulfinyl-containing sideproduct as major driving forces. Efforts to deprotect the UAA with HCl or HBr on the other hand caused side reaction within the sulfoxide-containing linker as judged by NMR. We assume that an acid catalysed Pummerer rearrangement had occurred followed by hydrolysis of the linker (Scheme 3.4b). Even though individual synthetic steps within the planned synthesis of BrSOK worked well, the final molecule appears to be extremely sensitive towards acidic conditions. A different combination of protecting groups might circumvent this problem, however we did not pursue them since, at the same time, the synthesis of urea-based UAAs showed more promising results.



**Scheme 3.4.** (a) Synthetic strategy for obtaining the sulfoxide-containing BrSOK. (b) Mechanism of Pummerer rearrangement. Mineral acids such as HCl and HBr cause the cleavage of the S-O bond to afford the thialcyl cation which easily captures a nucleophile. Upon acid hydrolysis, the  $\alpha$ -substituted sulfide affords a thiol and a carbonyl compound. Synthetic applications of Pummerer rearrangement widely employ acetic or trifluoroacetic anhydride as activating agents instead of mineral acids.

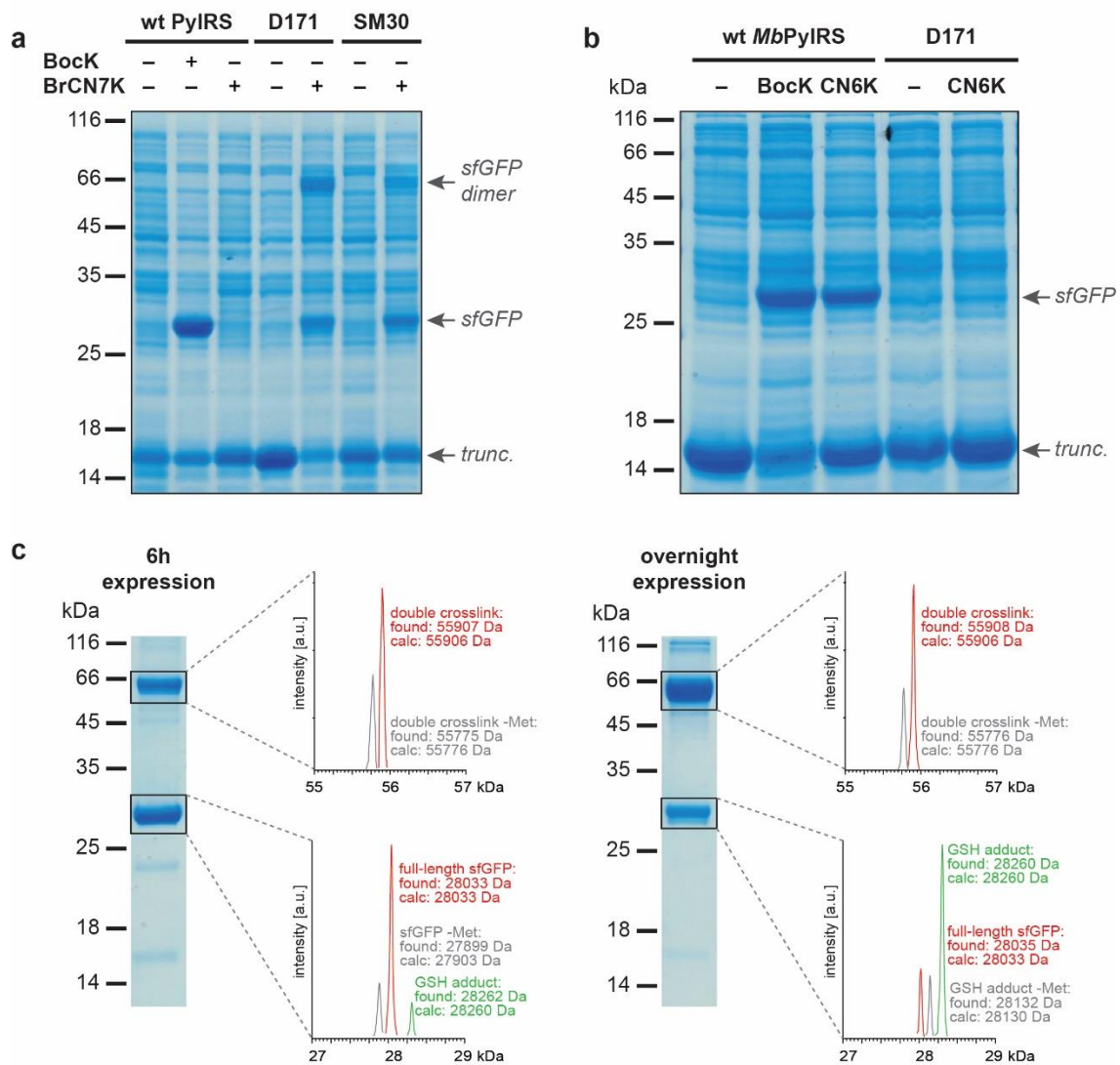
The urea linkage in BrCN6(7)K was efficiently introduced through a reaction of lysine carbonyldiimidazolide with TBDMS-protected butyl- and pentyl- amino alcohols (easily prepared from commercially available amino alcohols, Scheme 3.5). After TBDMS deprotection by TBAF, the free alcohol derivatives were converted into corresponding bromides *via* an Apple reaction. Finally, deprotection with HBr afforded UAAs as hydrobromide salts. BrCN7K was prepared by this method in satisfying 61 % yield over five steps and, more importantly, was efficiently incorporated by BrCnKS into sfGFP-N150TAG-His6 in *E. coli*. Another synthetase (SM30; contains mutations L274A, C313S and Y349F) published for a structurally similar diazirine-containing urea UAA<sup>[158d]</sup> incorporated BrCN7K as well, albeit with lower efficiency (Figure 3.12a). Although the deprotection with HBr afforded BrCN6K (as monitored by LC-MS), an intramolecular cyclisation reaction during work-up resulted in expulsion of the reactive bromide. The same result was obtained when the free alcohol was directly brominated with conc. HBr as well as during reaction of lysine carbonyldiimidazolide with 4-bromobutanamine hydrobromide. Addition of 1.2 equivalents of DIPEA in the latter step (which ensures the deprotonation of 4-bromobutanamine hydrobromide) was likely beneficial for the cyclisation reaction as was the increased reaction

temperature (50 °C). The resulting UAA is unfortunately not useful for our purposes anymore, however it did get moderately incorporated into sfGFP by the wild-type PylRS (Figure 3.12b).



**Scheme 3.5.** Overview of synthetic steps performed to obtain urea-based UAAs BrCN6K and BrCN7K. Due to intramolecular cyclisation, BrCN6K was unfortunately not synthesized.

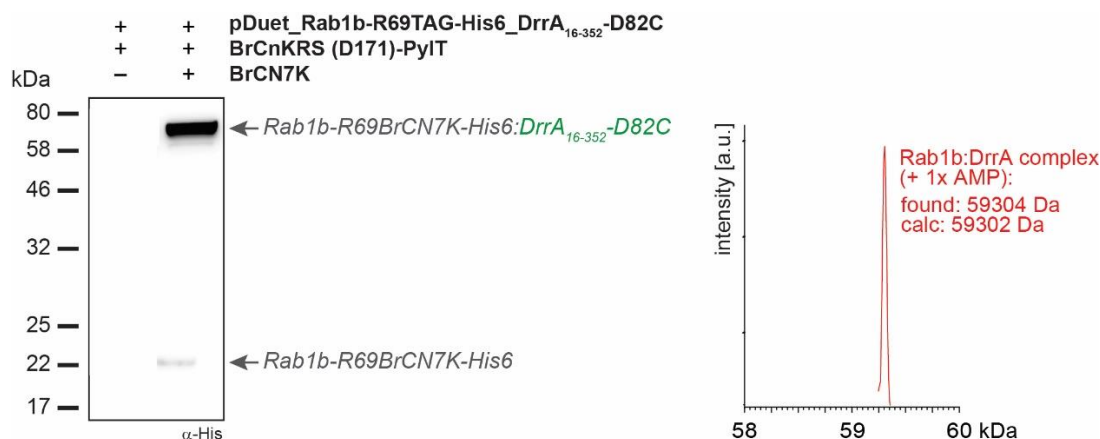
As it was the case with site-specific incorporation of BrC6(7)K, formation of the crosslinked sfGFP homodimer during protein expression was observed in the presence of BrCN7K as well. Both the monomeric sfGFP and the crosslinked homodimer were purified *via* His6-tag affinity chromatography and characterised by MS (Figure 3.13c). Over prolonged expression times (16-24 h), the amount of the crosslinked homodimer naturally increased with the MS data indicating that only the double crosslinked homodimer had formed (55908 Da). On the other hand, MS analysis of the uncrosslinked sfGFP-N150BrCN7K after overnight expression shows the GSH-adduct as the major species. It seems that the electrophilicity of BrCN7K somewhat increased with the introduction of the urea moiety into the amino acid scaffold, as the GSH-adduct formation was never observed during expression of proteins bearing BrC6(7)K in *E. coli*. Nevertheless, considering the great incorporation efficiency and the ability of the UAA to form covalent crosslinks *in vivo*, we were generally pleased with these initial results.



**Figure 3.13.** Incorporation of urea-based UAAs into sfGFP-N150TAG-His6 in *E. coli*. (a) BrCN7K is readily incorporated by the BrCnKRS (D171) and is also recognised by another PylRS (SM30), previously published for a structurally similar diazirine-containing urea UAA.<sup>[158d]</sup> Furthermore, sfGFP homodimer formation is observed on the SDS-PAGE, as it was the case for BrC6(7)K. (b) Although BrCN6K was not successfully isolated due to the intramolecular cyclisation reaction, the cyclic “CN6K” is recognised by the wild-type PylRS. (c) MS characterisation of His6-tag purified sfGFP monomer and homodimer mixtures after 6 h and overnight expression.

The crucial question that needs to be addressed and whose outcome would determine the fate of BrCN7K as an MS<sup>2</sup>-cleavable crosslinker is, undoubtedly, does the urea linkage of the UAA get cleaved during CID? We decided to explore this issue on the covalent complex between Rab1b and the AMPylation domain of DrrA (DrrA<sub>16-352</sub>); another complex that we and our collaborators at the group of Prof. Itzen have been studying for the last couple of years. In our previous efforts, we were able to identify a good crosslinking position between the two protein partners by covalently trapping Rab1b-R69TAG-His6 and DrrA<sub>16-352</sub>-D82C in the presence of BrC6(7)K (data not shown). Accordingly, we sought to replace BrC6(7)K with BrCN7K and co-express the two Rab1b/DrrA<sub>16-352</sub> variants (both genes on the same Duet vector) together with the BrCnKRS/tRNA<sup>CUA</sup> pair (Figure 3.14). An extremely efficient

crosslink formation was observed during western blot analysis of the lysed cell cultures (the expression was purposely limited to 6 h to avoid possible GSH-adducts observed during incorporation of BrCN7K into sfGFP). Furthermore, the crosslink was purified first by His6-tag affinity chromatography and a second time by SEC and validated by MS. The complex contained a single AMPylation (observed mass of 59304 Da) – a modification catalysed by DrrA at the position of Y77 of Rab1b.<sup>[220]</sup>



**Figure 3.14.** *In vivo* crosslinking of Rab1b-R69BrCN7K and DrrA<sub>16-352</sub>-D82C in *E. coli*.  $\alpha$ -His western blot analysis of cell lysates co-expressing Rab1b-R69TAG-His6 and DrrA<sub>16-352</sub>-D82C variants (both genes on the same Duet vector) together with the BrCnKRS/tRNA<sup>CUA</sup> pair in the absence and presence of BrCN7K (left) and MS characterisation of the purified Rab1b:DrrA complex (right). Mass corresponding to the complex AMPylated at position Y77 of Rab1b is observed (590304 Da).<sup>[220]</sup>

At the time of writing this thesis, MS<sup>2</sup> experiments are performed on the Rab1b-R69BrCN7K-His6:DrrA<sub>16-352</sub>-D82C complex with the help of our collaboration partners from the group of Prof. Stephan Sieber. If successful, BrCN7K has a possibility to become an important MS<sup>2</sup>-cleavable crosslinker for studying protein-protein interactions.

### 3.6. Conclusion and outlook

In summary, within this chapter we describe an approach that enables covalent stabilization of low-affinity protein complexes *in vitro* and *in vivo*. In order to trap such interactions, we relied on site-specific incorporation of UAAs bearing flexible bromoalkyl functionalities (BrCnK) mediated by a highly efficient PylRS/tRNA<sup>CUA</sup> pair in both *E. coli* and mammalian cells. While inert under physiological conditions, these UAAs react specifically with nucleophilic natural amino acids such as cysteines in a proximity-triggered manner. By substituting proximal positions within the interface of two interacting proteins pairwise with BrCnK and cysteines, we successfully illustrated the utility of our proximity-triggered crosslinking approach *in vitro* and *in vivo*. Furthermore, the approach allowed us to crystallise and determine the structure of a previously elusive low-affinity ternary complex between the GDP-bound small human GTPase Rab1b and the guanine nucleotide exchange factor domain of its effector DrrA from *L. pneumophila*.

With an approach that allows specific crosslinking of biologically relevant protein-protein interactions in hand, we decided to expand it towards UAAs that bear MS<sup>2</sup> cleavable

crosslinkers. In addition to the bromoalkyl functionality necessary for the proximity-triggered crosslinking, these UAAs contain an additional moiety that produces characteristic fragment ion patterns during MS<sup>2</sup> analysis, enabling thus accurate identification of the crosslinked peptides. By introducing MS<sup>2</sup>-cleavable UAAs into the protein of interest and subsequent *in vivo* proximity-triggered crosslinking with its protein partners during expression, we envision an approach that would, after affinity-tag purification of the complex, enzymatic digestion and MS<sup>2</sup> and computational analysis, facilitate identification and mapping of the respective binding interfaces. Among the designed UAAs, a urea-containing bromoalkyl derivative showed exceptional incorporation and crosslinking efficiencies during protein expression. Current studies that were not completed when writing up this thesis involve the establishment of collision-induced dissociation (CDI) experiments on purified crosslinked proteins that should elucidate if and to which extent the aforementioned UAA produces characteristic fragment ion patterns during MS<sup>2</sup> analysis. If these initial experiments prove to be successful, we would shift our attention towards establishing MS<sup>2</sup>-cleavable crosslinking UAAs as another important tool for covalently stabilizing and studying hitherto uncharted low-affinity protein-protein interactions.

---



## § 4. EXPERIMENTAL PART

### 4.1. General experimental details – Chemistry

All commercially available chemicals and dry solvents (*Sigma Aldrich*, *Acros*, *Carbolution*, *Alfa Aesar*, *TCI*) were used without further purification unless otherwise stated. All air or moisture sensitive reactions were carried out under argon atmosphere using standard Schlenk techniques in oven-dried glassware, which was further dried under vacuum using a heat gun before usage. Flash column chromatography used for product purification was performed on silica gel 60 (230-400 mesh). Thin-layer chromatography (TLC) was performed on *Merck* Millipore silica gel 60 F-254 plates. The developed silica plates were visualized by UV light (254 nm) and/or staining with potassium permanganate, ninhydrin or ceric ammonium molybdate (CAM) solutions, followed by heating.

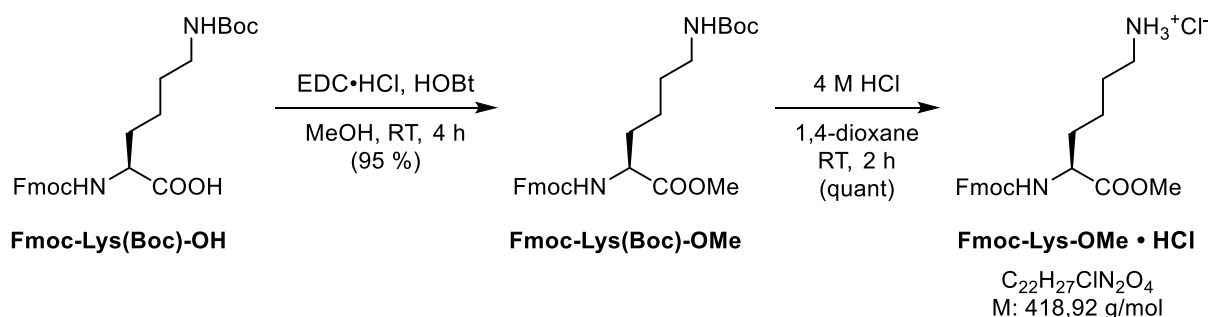
NMR spectra were recorded on *Bruker* AVHD300 (300 MHz for  $^1\text{H}$ -NMR, 75 MHz for  $^{13}\text{C}$ -NMR), *Bruker* AVHD400 (400 MHz for  $^1\text{H}$ -NMR, 101 MHz for  $^{13}\text{C}$ -NMR) or *Bruker* AVANCE 500 UltraShield<sup>TM</sup> spectrometer (500 MHz for  $^1\text{H}$ -NMR, 125 MHz for  $^{13}\text{C}$ -NMR). Chemical shifts ( $\delta$ ), reported in ppm, are referenced to the residual proton solvent signals (DMSO- $d_6$  – 2.50 ppm for  $^1\text{H}$ -NMR and 39.52 ppm for  $^{13}\text{C}$ -NMR;  $\text{CDCl}_3$  – 7.26 ppm for  $^1\text{H}$ -NMR and 77.16 ppm for  $^{13}\text{C}$ -NMR; MeOD – 3.31 ppm for  $^1\text{H}$ -NMR and 49.00 ppm for  $^{13}\text{C}$ -NMR;  $\text{D}_2\text{O}$  – 4.79 ppm for  $^1\text{H}$ -NMR spectra). Coupling constants ( $J$ ) are reported in Hertz (Hz) while peak multiplicities are described as follows: s (singlet), d (doublet), t (triplet), q (quartet), quint (quintet), m (multiplet), br (broad), obs (observed) or combinations thereof.

Small molecule LC-MS was carried out on an *Agilent Technologies* 1260 Infinity LC-MS system with a 6310 Quadrupole spectrometer. The solvent system consisted of 0.1 % formic acid in water as buffer A and 0.1 % formic acid in ACN as buffer B. Small molecule LC-MS was carried out on a *Phenomenex* Aeris<sup>TM</sup> Peptide XB-C18 column (100  $\times$  2.1 mm, 3.6  $\mu\text{m}$ ). The samples were analysed in both positive and negative mode and followed by UV absorbance at 193, 254 and/or 280 nm.

Reverse-phase HPLC purification was carried out with a *Shimadzu* LC-20AT Prominence system. A *Phenomenex* Luna C18, 5  $\mu\text{m}$  (4.6  $\times$  250 mm) column was used to separate compounds on an analytical scale, while a *Phenomenex* Luna C18, 5  $\mu\text{m}$  (10  $\times$  250 mm) and a *Phenomenex* Luna C18, 10  $\mu\text{m}$  (21.2  $\times$  250 mm) columns were used for preparative scale runs. Solvents A (MilliQ  $\text{H}_2\text{O}$  + 0.1 % formic acid) and B (ACN + 0.1 % formic acid) were used without filtration.

---

## 4.1.1. General procedure for preparation of protected amino acids

**Synthesis of methyl *N*<sup>2</sup>-(9-fluorenylmethoxycarbonyl)-L-lysinate hydrochloride (Fmoc-****Lys-OMe × HCl)**

Fmoc-Lys(Boc)-OH (1.00 g, 2.13 mmol) was dissolved in 10 mL of dry methanol, followed by addition of HOBT (0.39 mg, 2.56 mmol) and EDC × HCl (0.49 mmol, 2.56 mmol). The mixture was stirred at room temperature for 4 h, after which it was diluted with EtOAc and washed twice with 10 % citric acid and satur. NaHCO<sub>3</sub> solution. The obtained organic phases were combined, dried over MgSO<sub>4</sub> and removed under reduced pressure. Purification of the obtained oily residue by column chromatography on silica (packed in DCM, eluted with 2 % MeOH/DCM, v/v) afforded **Fmoc-Lys(Boc)-OMe** (0.98 mg, 95 %) as a white foam.

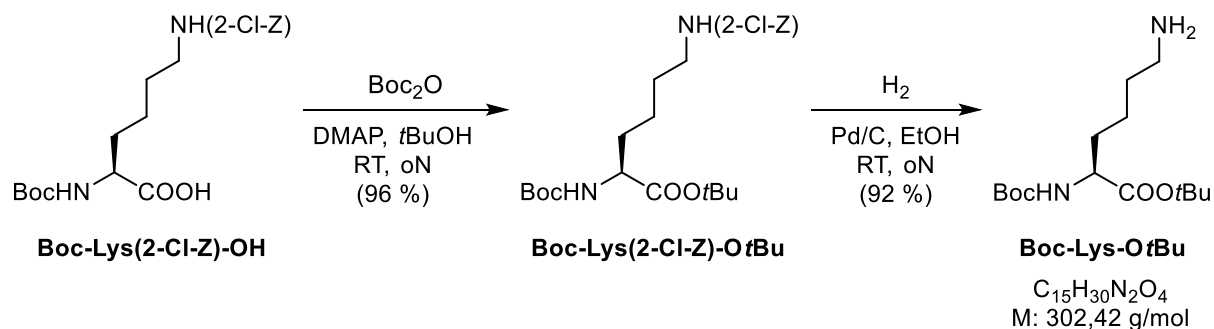
$R_f = 0.71$  (5 % MeOH/DCM, v/v) [UV, ninhydrin]

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ/ppm: 7.89 (d, *J* = 7.5 Hz, 2H), 7.75 (d, *J* = 7.8 Hz, 1H), 7.71 (dd, *J* = 7.5, 2.7 Hz, 2H), 7.43 (t, *J* = 7.4 Hz, 2H), 7.33 (td, *J* = 7.5, 1.1 Hz, 2H), 6.77 (t, *J* = 5.8 Hz, 1H), 4.32-4.27 (m, 2H), 4.25-4.20 (m, 1H), 4.02-3.95 (m, 1H), 3.62 (s, 3H), 2.92-2.85 (m, 2H), 1.71-1.54 (m, 2H), 1.36 (s, 9H), 1.35-1.21 (m, 4H).

Fmoc-Lys(Boc)-OMe (0.97 g, 2.01 mmol) was dissolved in 10 mL of dry 1,4-dioxane, followed by careful addition 4 M HCl solution in 1,4-dioxane (10 mL). The reaction mixture was stirred at room temperature for 2 h until completion, as judged by TLC (5 % MeOH/DCM). The solvent was then removed under reduced pressure, the residue dissolved in distilled water and the solution was extracted several times with cold ether. Lyophilisation of the aqueous phase afforded **Fmoc-Lys-OMe hydrochloride** (0.84 g, quant) as a pale yellow solid.

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ/ppm: 7.90 (d, *J* = 7.4 Hz, 2H), 7.86-7.75 (m, 3H), 7.75-7.68 (m, 2H), 7.46-7.39 (m, 2H), 7.34 (td, *J* = 7.4, 1.3 Hz, 2H), 4.39-4.28 (m, 2H), 4.28-4.19 (m, 1H), 4.01 (td, *J* = 8.7, 5.1 Hz, 1H), 3.57 (s, 3H), 2.82-2.69 (m, 2H), 1.75-1.46 (m, 4H), 1.42-1.26 (m, 2H).

**MS (ESI+):** m/z (%) 383.2 (100) [M+H]<sup>+</sup>

**Synthesis of *tert*-butyl *N*<sup>2</sup>-(*tert*butoxycarbonyl)-*L*-lysinate (**Boc-Lys-O*t*Bu**)**

*Note:* *Boc-Lys(Cbz)-OH* can be used as the starting material instead of *Boc-Lys(2-Cl-Z)-OH* as well. Under the same conditions, it affords *Boc-Lys-OtBu* in comparable yields.

*Boc-Lys(2-Cl-Z)-OH* (3.00 g, 7.23 mmol) and DMAP (0.27 g, 2.17 mmol) were dissolved in 20 mL of *tert*-butanol and  $\text{Boc}_2\text{O}$  was added in portions (2.21 g, 10.1 mmol) to the mixture (during which a significant amount of  $\text{CO}_2$  is produced). The resulting mixture was stirred at room temperature overnight. Afterwards, the solvent was removed under reduced pressure, the obtained residue dissolved in DCM and washed with 10 % citric acid solution and brine. Purification by column chromatography on silica (packed in DCM and washed with 2 → 5 % MeOH/DCM, v/v) afforded **Boc-Lys(2-Cl-Z)-O*t*Bu** (3.26 g, 96 %) as a yellow oil.

$R_f = 0.79$  (5 % MeOH/DCM, v/v) [ninhydrin]

<sup>1</sup>H NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 7.44-7.34 (m, 2H), 7.28-7.22 (m, 2H), 5.20 (s, 2H), 5.06 (d,  $J = 8.7$  Hz, 1H), 4.90 (br, 1H), 4.15 (q,  $J = 7.7$  Hz, 1H), 3.20 (q,  $J = 6.6$  Hz, 2H), 1.83-1.73 (m, 1H), 1.65-1.30 (m, 23H).

*Boc-Lys(2-Cl-Z)-O*t*Bu* (3.26 g, 6.93 mmol) was dissolved in 100 mL of absolute EtOH in a two-necked round-bottom Schlenk flask and palladium on carbon (105 mg, ~ 15 mg per mmol of the starting material) was carefully added to the mixture. The flask was evacuated and back-filled with argon three times, followed by flushing and finally filling the flask with hydrogen gas ( $\text{H}_2$  balloon). The reaction mixture was stirred at room temperature overnight. After completion (as judged by TLC in 5 % MeOH/DCM, v/v), the flask was evacuated and the  $\text{H}_2$  balloon removed. The reaction mixture was filtered off and the solvent removed under reduced pressure to afford **Boc-Lys-O*t*Bu** (1.93 g, 92 %) as a yellow oil, without the need of further purification.

<sup>1</sup>H NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 5.16-4.97 (m, 1H), 4.26-4.05 (m, 1H), 2.70 (t,  $J = 6.8$  Hz, 2H), 2.27 (br, 2H), 1.66-1.48 (m, 4H), 1.45 (s, 9H), 1.43 (s, 9H), 1.40-1.32 (m, 2H).

**MS (ESI+):**  $m/z$  (%) 303.2 (100)  $[\text{M}+\text{H}]^+$

## 4.2. General experimental details – Biology

Plasmid DNA was isolated from 8 mL of overnight *E. coli* cultures using the peqGOLD Plasmid MiniPrep kit (*Peqlab*) as indicated by the standard protocol with a minor modification: the DNA was eluted with MilliQ H<sub>2</sub>O instead of TE buffer. The antibiotic stock solutions were prepared as follows: ampicillin (100 mg/mL in ddH<sub>2</sub>O), kanamycin (50 mg/mL in ddH<sub>2</sub>O), tetracyclin (17 mg/mL in 70 % EtOH, v/v) and spectinomycin (50 mg/mL in ddH<sub>2</sub>O).

Agarose gels (1 % in 1x TAE buffer with addition of DNA stain Clear G (*Serva*, 4 µL per 100 mL of gel)) were cast in PerfectBlue™ gel systems (*Peqlab*) and run at 120 V for 20-25 minutes. The electrophoresis chamber contained 1x TAE buffer. Before loading them onto the gel, the DNA samples were mixed with 6x Purple Gel Loading Dye (*NEB*). GeneRuler 1 kb DNA Ladder (*Thermo Fisher Scientific*) was used as marker.

Analysis of protein samples was performed *via* SDS-PAGE on 15 % polyacrylamide gels. The electrophoresis chamber was filled with 1x SDS-MES buffer (for 20x buffer: 1 M MES, 1 M Tris pH 7.3, 2 % SDS, 20 mM EDTA). Cell culture pellets were resuspended in 1x SDS loading buffer (100 µL of buffer per OD<sub>600</sub> = 1), followed by heating for 10 minutes at 95 °C. After centrifugation at full speed for 10 minutes, 10 µL of samples were loaded onto gels. Purified proteins were mixed with 4x SDS loading buffer so that the final concentration of the dye was 1x. After cooking at 95 °C for 10 minutes, samples were centrifuged at full speed for a couple of seconds. Generally, 1-2 µg of proteins were loaded into the wells of the gels (4x SDS-loading buffer composition: 40 % glycerol (v/v), 240 mM Tris-HCl pH 6.8, 8 % SDS (w/v), 0.04 % Bromophenol blue (w/v), 5 % BME (v/v)). SDS-PAGE was carried out on either a Mini-PROTEAN® Tetra Cell (*Bio-Rad*) or a Bolt™ Mini Gel Tank (*Invitrogen*) system (110 V for 15 min, then 200 V for 45 min). Gels were stained with Quick Coomassie Stain (*Generon*) and subsequently destained with deionised water. Protein Marker III 6.5-200 kDa (*AppliChem*) or Color Prestained Protein Standard, Broad Range 11-245 kDa (*NEB*) were used as markers.

Western blotting was carried out on iBlot® 2 Dry Blotting System (*Life Technologies*) using Method P0 (20 V for 1 min, 23 V for 4 min, 25 V for 2 min). After blotting, the nitrocellulose membrane was blocked with 5 % skim milk powder solution in 1x TBST buffer (1 h at room temperature). The membrane was subsequently stained with 1:10000 Anti-His6-Peroxidase antibody (*Roche*) in 1 % skim milk powder solution in 1x TBST (1h at room temperature or overnight at 4 °C) for detection of His6-tagged proteins or 1:5000 HRP-streptavidin conjugate in 1 % skim milk powder solution in 1x TBST (overnight at 4 °C) for detection of biotinylated proteins. After washing several times with 1x TBST, the membrane was treated with ECL™ Prime Western Blotting Detection Reagent (*GE Healthcare*, RPN2236) and the proteins visualized Fusion Pulse 6 Imager (*Vilber Lourmat*).

DNA and protein concentrations were measured on NanoPhotometer® P330 (*Implen*) at 260 or 280 nm, respectively. Concentration of ubiquitin as well as of proteins used in labelling experiments was determined with a 3.1 mL Bradford assay, as described in the technical bulletin of the Bradford reagent (B6916, *Sigma Aldrich*).

LC-MS of full-length proteins was carried out on a *Phenomenex Aeris*<sup>TM</sup> Widepore C4 column (100 × 2.1 mm, 3.6 μm). The samples were analysed in positive mode, following protein UV absorbance at 280 nm. Total protein mass was calculated by deconvolution within the MS Chemstation software (*Agilent Technologies*). Theoretical masses of proteins were calculated using ProtParam (<http://web.expasy.org/protparam/>) and were manually corrected for the mass of the unnatural amino acid mutagenesis. Some full-length proteins were measured also on a LTQ-FT Ultra mass spectrometer coupled online to an Ultimate 3000 HPLC Instrument (both *Thermo Fisher Scientific*). Desalting was carried out with MassPREP On-Line Desalting Cartridges (*Waters*). Briefly, proteins were loaded in 0.1 % formic acid and eluted in a 3.5 min gradient (1.5 min from 6 to 95 %, 2 min 95 % acetonitrile, 0.1% formic acid) at a flow rate of 400 μL/min. Spectra were acquired in full scan mode with a resolution of 200.000 at m/z 400 and deconvoluted with the software Xtract or Promass (*Thermo Fisher Scientific*) using default deconvolution settings.

#### 4.2.1. Transformation of competent *E. coli* cells

An aliquot of chemically competent cells (50 μL) was thawed on ice for 5 minutes and added to the plasmid(s) of interest (typically 1 μL of single plasmid or 1 μL of each plasmid when co-transforming them). The sample was gently mixed with a fingertip and incubated on ice for 10 minutes. The cells were then heat shocked in a water bath at 42 °C for 45 seconds and incubated again on ice for 5 minutes. The cells were then recovered with 1 mL of SOC medium for 1 h at 37 °C and were cultured overnight either in 50 mL of LB medium or on LB agar plates, in the presence of appropriate antibiotic(s) (1:1000).

An aliquot of electrocompetent cells thawed on ice was mixed with the plasmid(s) of interest, as described above. The sample was gently mixed with a fingertip, transferred to an electroporation cuvette (0.2 cm gap) and electroporated at 2.0 kV. The cells were then recovered with 1 mL of SOC medium for 1 h at 37 °C and were cultured overnight either in 50 mL of LB medium or on LB agar plates, in the presence of appropriate antibiotic(s) (1:1000).

Either *E. coli* DH10β or BL21(DE3) competent cells (*NEB*) were used in this work. While the former were used both for cloning and protein expression, *E. coli* BL21(DE3) cells were used only for protein expression.

#### 4.2.2. QuikChange<sup>TM</sup> site-directed mutagenesis

Oligonucleotide primers for QuikChange<sup>TM</sup> site-directed mutagenesis were designed with PrimerX (<http://bioinformatics.org/primerx/>) and purchased from *Sigma*. Primers were

---

designed to be 25-45 bp in length with a  $T_m$  value  $\geq 78$  °C and a minimum GC content of 40 %. Furthermore, the desired mutation should be in the middle of the primer, while the primers should terminate in one or more C or G bases. PCR was carried out with a LifeECO Thermal Cycler (*Bioer*). Reactions (50  $\mu$ L total volume) were typically performed with PfuTurbo or PfuUltra II polymerases (*Agilent*) on 25 ng of template and in the presence of 5 % DMSO. The standardly used cycling parameters are summarised in Table 4.1. After the PCR, the parental plasmid was digested with DpnI (1  $\mu$ L; 4 h at 37 °C) and analysed by agarose gel electrophoresis while 7  $\mu$ L of the mixture were transformed into *E. coli* DH10 $\beta$  cells (50  $\mu$ L). After recovery with 1 mL of SOC medium for 1 h at 37 °C, the whole sample was plated onto LB-agar plates supplemented with appropriate antibiotics (1:1000) and incubated overnight at 37 °C. Obtained single colonies were picked and inoculated into 8 mL of LB medium containing appropriate antibiotics (1:1000). The cells were cultured overnight at 37 °C with shaking (200 rpm) and the plasmids isolated using peqGOLD Plasmid MiniPrep kit (*Peqlab*) and validated by sequencing (*GATC-Biotech*).

**Table 4.1.** Cycling parameters for site-directed mutagenesis with Pfu polymerases

Step	Temperature	Time	No. of cycles
Initial denaturation	95 °C	30 sec	1
Denaturation	95 °C	30 sec	20
Primer annealing	55 °C	1 min	
Elongation	68 °C	1 min per kb amplified + 2 min	
Final hold	10 °C	$\infty$	-

In case of a low yield of the PCR product, the DNA was precipitated by glycogen/ethanol before transformation into competent *E. coli* cells. Briefly, 2  $\mu$ L of 3 M sodium acetate, 22.5  $\mu$ L of isopropanol and 0.5  $\mu$ L of glycogen were added to 20  $\mu$ L of the digested PCR mixture and the mixture was incubated for 1 h at -20 °C. The mixture was afterwards centrifuged (30 min, 14,000 rpm, 4 °C), the supernatant carefully discarded and 50  $\mu$ L of 70 % EtOH (cooled at -20 °C) added to the pellet. The mixture was centrifuged again (20 min, 14,000 rpm, 4 °C) and the supernatant carefully discarded. After drying shortly on air, the pellet was re-suspended in 5  $\mu$ L of nuclease-free water and transformed into competent *E. coli* cells.

#### 4.2.3. Synthetase screening for site-specific incorporation of UAAs into sfGFP

*E. coli* DH10 $\beta$  cells (50  $\mu$ L) were co-transformed as described in section 4.2.1. with pPyIT\_sfGFP-N150TAG-His6 (which encodes a constitutively expressed tRNA<sub>CUA</sub> and a C-terminally His6-tagged sfGFP gene with an amber codon at position 150, under the control of araBAD promoter) and a plasmid encoding constitutively expressed *Mm* or *Mb* PylRS variant of choice (1  $\mu$ L of each plasmids). After recovery with 1 mL of SOC medium for 1 h at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing 1:1000 dilution of appropriate antibiotics (pPlyT\_sfGFP-N150TAG-His6 plasmid has a tetracycline resistance, while the *pylS*-containing plasmids are either kanamycin or ampicillin resistant, depending on

the used variant), at 37 °C with shaking (200 rpm). The overnight culture was then diluted to an OD<sub>600</sub> of 0.05 in 50 mL of fresh LB medium supplemented with half-strength of appropriate antibiotics (1:500) and was incubated at 37 °C until the OD<sub>600</sub> reached 0.3. Subsequently, 4 mL of the culture was transferred to cell culture tubes containing the UAA of interest (2 mM final concentration). The culture without the UAA of interest was used as a negative control, while typical positive incorporation control involved incorporation of BocK (100 mM stock solution in 100 mM NaOH) by the wild-type PylRS. The cultures with and without the UAA of interest were then grown at 37 °C until an OD<sub>600</sub> around 0.7 was reached. At this time point the expression was induced with the addition of 20 % arabinose solution (final concentration of 0.02 %). After 6 to 18 hours of expression at 37 °C with shaking, the cell pellet was harvested by centrifugation (5000 g, 5 min), suspended in 1x SDS loading buffer (depending on the OD<sub>600</sub> of the culture; 100 µL of buffer per OD<sub>600</sub> = 1) and analysed by SDS-PAGE. Observation of a green pellet, as a result of the full-length sfGFP expression, is also a visual indication that particular PylRS variant might incorporate the UAA of interest.

#### **4.2.4. General protocol for expression and purification of UAA-containing proteins**

*Note 1: While the following protocol refers only to the expression and purification of sfGFP-N150UAA variants, it was used for obtaining myoglobin (Myo-S4TAG) and ubiquitin (Ub-K6TAG) containing site-specifically incorporated UAAs of interest as well.*

As described in section 4.2.3., an overnight culture transformed with pPylT\_sfGFP-N150TAG-His6 and the plasmid encoding constitutively expressed PylRS variant of choice was diluted to an OD<sub>600</sub> of 0.05 in 200 mL of fresh LB medium supplemented with half-strength of appropriate antibiotics (1:500) and was incubated at 37 °C until the OD<sub>600</sub> reached 0.3. UAA of interest was then added to the culture (final concentration of 1.5 mM) and the cells were further incubated at 37 °C until an OD<sub>600</sub> of 0.6 was reached. At this time point the expression was induced with the addition of 20 % arabinose solution (final concentration of 0.02 %). After 6 to 18 hours of expression at 37 °C with shaking, the cell pellet was harvested by centrifugation (5000 g, 15 min, 4 °C), flash frozen in liquid nitrogen and stored at -80 °C until needed.

*Note 2: Only the expression of proteins containing site-specifically incorporated VyOK and VySK was limited to 6 h due to observed posttranslational degradation of the respective UAAs (see Chapter 2.5 for more detail). All other UAAs incorporated into sfGFP, Myo or Ub were expressed overnight.*

The obtained cell pellets were thawed on ice and resuspended in lysis buffer containing 20 mM Tris-HCl pH 8.0, 30 mM imidazole pH 8.0, 300 mM NaCl, 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and one cOmplete™ Mini EDTA-free protease inhibitor cocktail tablet (Roche). Generally, 100 mL of lysis buffer were used for 20 g of cell pellet. The cell suspension was incubated on ice for 30 min and sonicated with cooling in an ice-water bath.

---

The lysed cells were cleared by centrifugation (18,000 rpm, 45 min, 4 °C), Ni-NTA slurry (Jena Bioscience) was added (0.2 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 80 mL of wash buffer containing 20 mM Tris-HCl pH 8.0, 30 mM imidazole pH 8.0 and 300 mM NaCl. The protein was eluted in 1 mL fractions with the wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing the protein were pooled together, concentrated and rebuffed (20 mM Tris-HCl pH 8.0, 300 mM NaCl) with Amicon<sup>®</sup> centrifugal filter units with appropriate molecular weight cut-offs (*Merck*). Purified proteins were analyzed by SDS-PAGE and mass spectrometry.

### 4.3. Supporting information for Chapter 2

#### 4.3.1. Photochemistry equipment

##### Rayonet photoreactor

Small molecule photolysis experiments with *o*NQMs were carried out in custom-made borosilicate tubes, in Rayonet photochemical reactors in a positive geometry setup (cylindrical array of 16 UV lamps,  $\lambda_{\max}$  300 or 350 nm, with nominal power of 8 W), courtesy of the group of Prof. Thorsten Bach.

##### Specifications for the custom-made phototubes:



**Figure 4.1.** Custom-made phototubes used for photolysis reactions performed in Rayonet photoreactors.

Material:	Duran <sup>®</sup> glass	Total length:	250 mm
Glass joint:	NS14/23	Length from sphere to bottom:	200 mm
Volume:	10 mL	External diameter:	12 mm

##### Data for the Fluorescent Light Tube (300 nm), courtesy of the group of Prof. Thorsten Bach:

##### Basic information

Description:	Fluorescent light tube, Rayonet RPR-3000A
Manufacturer / supplier:	n/a / Southern New England Ultraviolet Company
Serial number:	FLT012
Type / size:	T5 tube, G5 socket
Mechanical specification:	16 mm diameter, 288 mm length
Electrical specification:	8 W

---



Wavelength (range, typ.): 300 nm  
Spectral width: ~ 40 nm

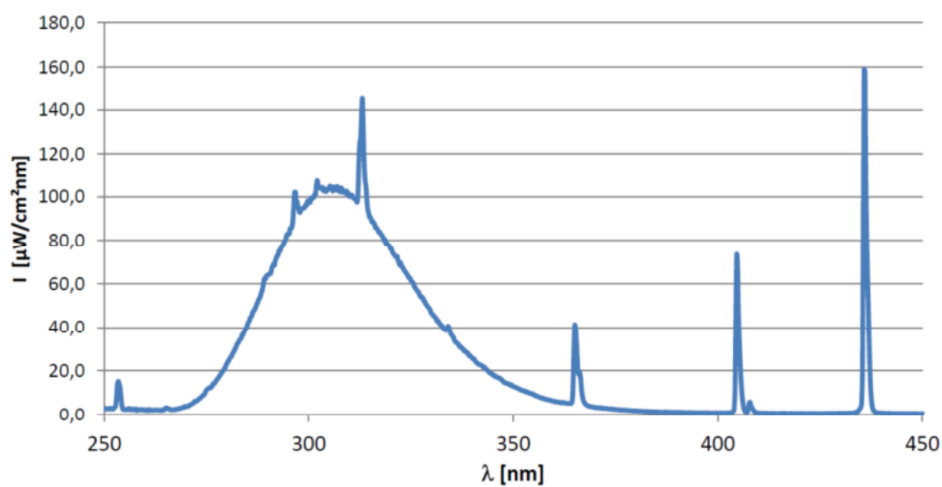
### **Characterisation**

Description of measurement: Measured with Ocean-optics USB4000 spectrometer using a calibrated setup (cosine corrector/fibre). The cosine corrector was placed at 20 mm distance from a single fluorescent tube at half height.

Measured dominant wavelength / Int.: 313 nm / 138  $\mu\text{W}/\text{mm}^2 \text{ nm}$

Measured spectral width: 40 nm

Integral reference intensity / range: 4275  $\mu\text{W}/\text{cm}^2$  / 260-380 nm



*Data for the Fluorescent Light Tube (350 nm), courtesy of the group of Prof. Thorsten Bach:*

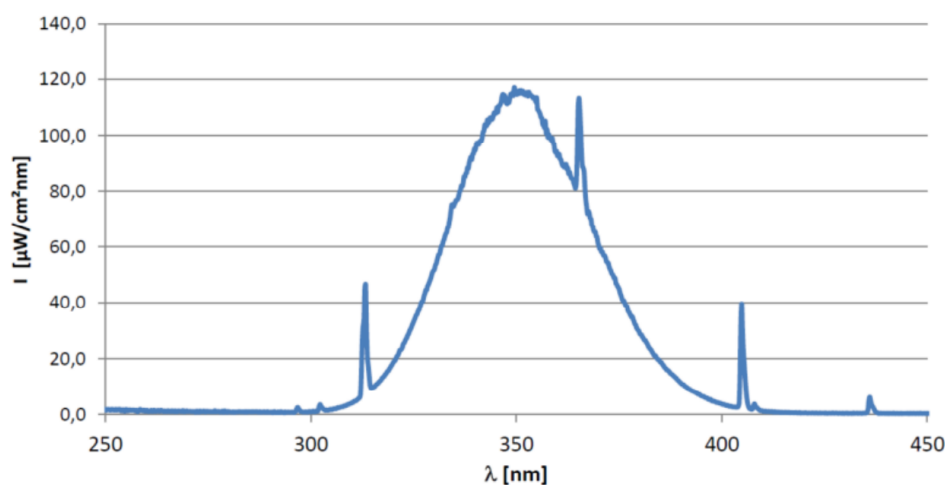
### **Basic information**

Description: Fluorescent light tube, Luzchem LZC-UVA  
Manufacturer / supplier: Hitachi / Luzchem  
Serial number: FLT021  
Type / size: T5 tube, G5 socket  
Mechanical specification: 16 mm diameter, 288 mm length  
Electrical specification: 8 W  
Wavelength (range, typ.): 300 – 400 nm, 350 nm, UV-A  
Spectral width: ~ 40 nm

### **Characterisation**

Description of measurement: Measured with Ocean-optics USB4000 spectrometer using a calibrated setup (cosine corrector/fibre). The cosine corrector was placed at 20 mm distance from a single fluorescent tube

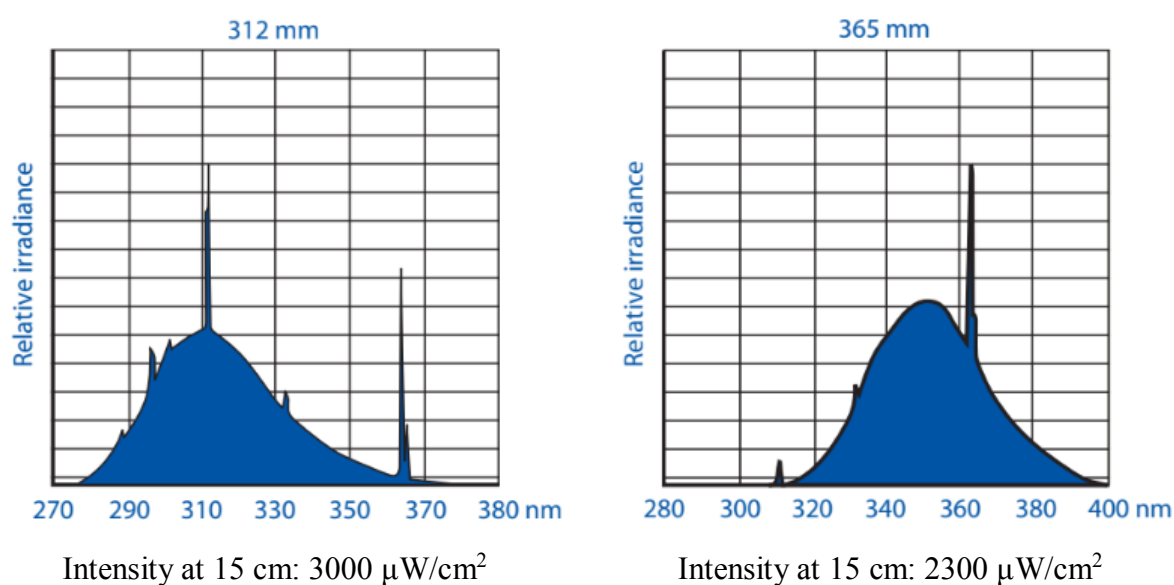
at half height.  
Measured dominant wavelength / Int.: 350 nm / 115  $\mu\text{W}/\text{mm}^2 \text{ nm}$   
Measured spectral width: 40 nm  
Integral reference intensity / range: 5017  $\mu\text{W}/\text{cm}^2$  / 300-425 nm



#### **Broad-emitting high-power UV lamp (2 × 15 W)**

Broad-emitting high-power filtered UV lamps (2 × 15 W) with  $\lambda_{\text{max}}$  of 312 nm (VL-215.M) or 365 nm (VL-215.L) were purchased from Vilber and used for photochemistry experiments on purified proteins with *o*NQMs (see Section 4.2.4). The samples stirred in cell culture plates were irradiated from the top, with the distance between the lamp and the sample not overcoming 10 mm. More details on the lamps can be found on [www.vilber.de](http://www.vilber.de).

#### Lamp emission spectra:



### **5 W LED (455 nm) setup**

Small molecule photochemistry experiments between 9,10-phenanthrenequinone and EtVyO were initially tested in a custom-made irradiation system developed by the group of Prof. Thorsten Bach (Figure 4.2). The light from a 455 nm 5 W high power LED (Avonec) is focused through a quartz rod which is immersed into the reaction mixture (min. 2 mL of total volume). The surface of the immersed part of the quartz rod is matted which scatters the light so that the entire reaction mixture is irradiated from the inside. Details on the LED lamp, including the emission spectrum, can be found on [www.avonec.de](http://www.avonec.de).

### **Single LEDs**

Single 1 W high power 450 nm LED (H2A1-H450, Roithner Lasertechnik) or 3 W high power LED 4000 K-4500 K (white) (Avonec) were used for labelling of VyOK-containing proteins with 9,10-phenanthrenequinones. Samples in cell culture well plates were irradiated by placing the single LED directly on the top of the well. Details on the 1 W 450 nm LED can be found on [www.roithner-laser.com](http://www.roithner-laser.com). Details on the 3 W white light LED lamp, including the emission spectrum, can be found on [www.avonec.de](http://www.avonec.de).



**Figure 4.2.** Custom-made apparatus and reaction setup for small molecule photochemistry experiments with 5 W high power LED (Avonec). The irradiation system was developed by the group of Prof. Thorsten Bach.

### ***4.3.2. Small molecule photochemistry experiments***

#### ***4.3.2.1. Small molecule photolysis experiments with oNQMPs***

A typical small molecule photolysis experiment carried out in custom-made phototubes (Figure 4.1) involved irradiation of 1 mM solution of 3-hydroxymethyl-2-naphthol (NOH) or its derivative in 50 % aqueous acetonitrile (3 mL) at 300 or 350 nm (Rayonet photoreactor), in the presence of excess (1.5, 10, 30, 60 or 100 eq) of vinyl ether/sulfide or another olefin (see Table 2.1 and Figure 2.4). Samples were taken at several time points and analysed by LC-MS or HPLC.

Similar setup was used for small molecule reactions that were irradiated with broad-emitting high-power filtered UV lamps ( $\lambda_{\max}$  312 or 365 nm,  $2 \times 15$  W) with the difference that the stirred sample in a cell culture plate was irradiated from top. Since this is essentially an open system, only vinyl ethers with high boiling points were used (e.g. EGVyO or (EG)<sub>2</sub>VyO). Samples were taken at several time points and analysed by LC-MS.

#### 4.3.3.2. Small molecule photochemistry experiments with PQs

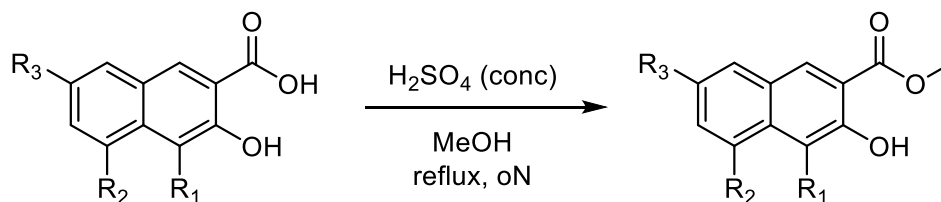
Small molecule [4+2] cycloaddition experiments were performed on 9,10-phenanthrenequinone (9,10-PQ) in the presence of excess of EtVyO or (EG)<sub>2</sub>VyO, depending on which light source was used (5 W LED system, Figure 4.2, or single 1 W 450 nm/3 W white light LEDs). The reactions were typically performed by adding 10 eq of vinyl ether to 1-2 mM solution of 9,10-PQ in 50 % aqueous ACN (min. 2 mL when using the 5 W LED system and up to 1 mL when using single LEDs), with irradiation times ranging between 5 and 30 minutes. Additionally, reactions were also performed on 10-50 mM solutions of 9,10-PQ dissolved in ACN or DCE. Samples were taken at several time points and analysed by LC-MS. A good indication that the starting material is being (or is completely) consumed is the discoloration of the reaction mixture (9,10-PQ solutions are golden yellow to orange, depending on the concentration) and especially the appearance of fluorescence of the cycloaddition product.

---

### 4.3.3. Chemical synthesis

#### 4.3.3.1. *ortho*-Naphthoquinone methide (oNQM) derivatives and related compounds

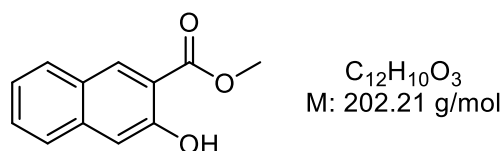
##### General procedure for esterification of 3-hydroxy-2-naphthoic acids



- 1a** : R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> = H  
**2a** : R<sub>1</sub>, R<sub>3</sub> = H; R<sub>2</sub> = OH  
**3a** : R<sub>1</sub>, R<sub>2</sub> = H; R<sub>3</sub> = Br

3-Hydroxy-2-naphthoic acid or its derivative (1.0 eq) was dissolved in MeOH, conc. H<sub>2</sub>SO<sub>4</sub> was carefully added and the mixture was stirred under reflux overnight. After cooling to ambient temperature, the solvent was removed under reduced pressure. The obtained residue was diluted with EtOAc and sequentially washed with 1 M Na<sub>2</sub>CO<sub>3</sub> solution, water and brine. The organic phase was dried over MgSO<sub>4</sub> and concentrated *in vacuo* to afford the desired ester without the need for further purification.

##### **Methyl 3-hydroxy-2-naphthoate (1a)**<sup>[238]</sup>

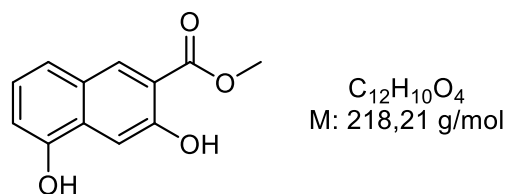


Title compound **1a** was obtained from commercially available 3-hydroxy-2-naphthoic acid (18.8 g, 99.9 mmol), conc. H<sub>2</sub>SO<sub>4</sub> (1.1 mL) and MeOH (100 mL) as a light yellow solid (18.2 g, 90 %).

**<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>) δ/ppm: 10.26 (s, 1H), 8.48 (s, 1H), 7.96 (dd, *J* = 8.6 Hz, 1H), 7.76 (dd, *J* = 8.0 Hz, 1H), 7.54 (ddd, *J* = 8.2, 6.8, 1.2 Hz, 1H), 7.40-7.31 (m, 2H), 3.95 (s, 3H).

**<sup>13</sup>C NMR** (101 MHz, DMSO-*d*<sub>6</sub>) δ/ppm: 168.6, 154.8, 136.9, 132.1, 129.1, 129.0, 126.6, 125.9, 123.9, 116.2, 111.0, 52.6.

**MS (ESI+)**: *m/z* (%) 203.1 (100), [M+H]<sup>+</sup>

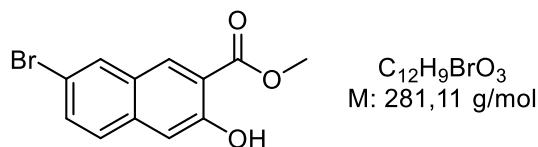
**Methyl 3,5-dihydroxy-2-naphthoate (2a)**

Title compound **2a** was obtained from commercially available 3,5-dihydroxy-2-naphthoic acid (10.0 g, 49.0 mmol), conc.  $H_2SO_4$  (3.3 mL) and MeOH (150 mL) as an ochre-coloured solid (9.59 g, 90 %).

$^1H$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 10.16 (s, 1H), 10.11 (s, 1H), 8.37 (s, 1H), 7.49 (d,  $J = 8.2$  Hz, 1H), 7.40 (t,  $J = 7.9$  Hz, 1H), 7.16 (dd,  $J = 7.5, 0.9$  Hz, 1H), 3.93 (s, 3H).

$^{13}C$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 168.6, 154.0, 151.6, 131.8, 128.6, 128.0, 124.3, 119.6, 116.5, 110.2, 105.9, 52.6.

MS (ESI+):  $m/z$  (%) 219.1 (100),  $[M+H]^+$

**Methyl 7-bromo-3-hydroxy-2-naphthoate (3a)**

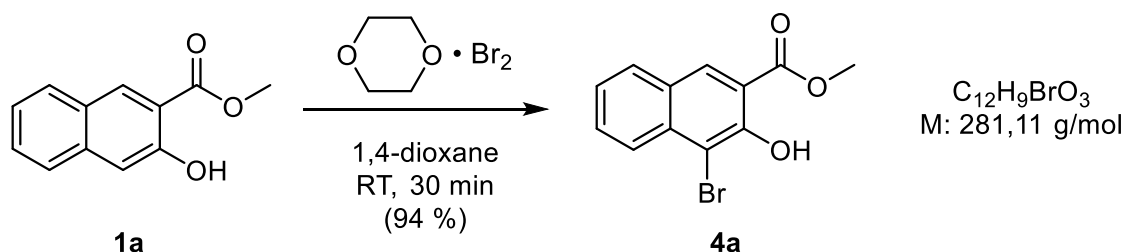
Title compound **3a** was obtained from commercially available 7-bromo-3-hydroxy-2-naphthoic acid (1.00 g, 3.74 mmol), conc.  $H_2SO_4$  (0.40 mL) and MeOH (50 mL) as a yellow solid (1.03 g, 98 %).

$^1H$  NMR (300 MHz, DMSO- $d_6$ )  $\delta$  ppm: 10.34 (s, 1H), 8.44 (s, 1H), 7.62 (dd,  $J = 8.9, 2.0$  Hz, 1H), 7.36 (s, 1H), 3.94 (s, 3H).

$^{13}C$  NMR (75 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 168.1, 155.0, 135.2, 131.7, 131.2, 130.7, 128.2, 127.7, 117.8, 116.6, 111.3, 52.7.

MS (ESI +):  $m/z$  (%) 281.0/283.0 (100),  $[M+H]^+$

---

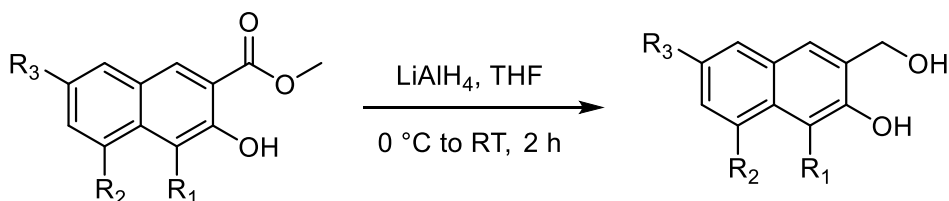
**Methyl 4-bromo-3-hydroxy-2-naphthoate (4a)**

A solution of 1,4-dioxane dibromide<sup>[239]</sup> (1.46 g, 5.88 mmol) in 10 mL of 1,4-dioxane was added to the solution of methyl 3-hydroxy-2-naphthoate **1a** (1.10 g, 5.44 mmol) in 10 mL of 1,4-dioxane and the reaction mixture was stirred at room temperature for 30 minutes. The reaction was quenched by addition of cold water until a precipitate had formed, which was then filtered off, washed with water and dried on air overnight. Methyl 4-bromo-3-hydroxy-2-naphthoate **4a** was obtained as a pale yellow solid (1.44 g, 94 %).

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 11.03 (br, 1H), 8.64 (s, 1H), 8.14-8.09 (m, 2H), 7.76 (ddd, *J* = 8.5, 6.8, 1.3 Hz, 1H), 7.50 (ddd, *J* = 8.0, 6.9, 1.0 Hz, 1H), 4.01 (s, 3H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  / ppm: 168.8, 152.1, 135.0, 132.1, 131.0, 130.2, 127.3, 124.9 (2C), 115.3, 106.1, 53.2.

**MS (ESI +):** *m/z* (%) 281.0/283.0 (100) [M+H]<sup>+</sup>

**General procedure for reduction of methyl 3-hydroxy-2-naphthoates**

**1a** : R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> = H

**2a** : R<sub>1</sub>, R<sub>3</sub> = H; R<sub>2</sub> = OH

**3a** : R<sub>1</sub>, R<sub>2</sub> = H; R<sub>3</sub> = Br

**4a** : R<sub>1</sub> = Br; R<sub>2</sub>, R<sub>3</sub> = H

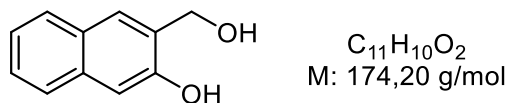
**1b** : R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> = H

**2b** : R<sub>1</sub>, R<sub>3</sub> = H; R<sub>2</sub> = OH

**3b** : R<sub>1</sub>, R<sub>2</sub> = H; R<sub>3</sub> = Br

**4b** : R<sub>1</sub> = Br; R<sub>2</sub>, R<sub>3</sub> = H

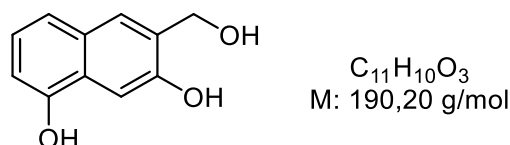
To a solution of the methyl ester (1.0 eq) in dry THF, 1 M LiAlH<sub>4</sub> solution in THF (2.0 eq) was added dropwisely over 30 min at 0 °C. The reaction mixture was warmed up to room temperature and stirring continued for 2 hours, after which the mixture was poured into cold Et<sub>2</sub>O and carefully quenched with 5 M HCl solution on ice under continuous shaking. The organic phase was separated, while the water phase was extracted 3 times with Et<sub>2</sub>O. Combined organic phases were washed with brine, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo* to afford the desired alcohol without the need for further purification.

**3-(hydroxymethyl)naphthalene-2-ol (1b)**

Reduction of **1a** (0.50 g, 2.47 mmol) with 1 M LiAlH<sub>4</sub> solution (5.04 mL, 5.04 mmol) in dry THF (10 mL) afforded the title compound **1b** as a pale yellow solid (0.38 g, 88 %).

**<sup>1</sup>H NMR** (400 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 9.80 (s, 1H), 7.80 (s, 1H), 7.76 (d,  $J = 8.0$  Hz, 1H), 7.74 (d,  $J = 8.1$  Hz, 1H), 7.33 (ddd,  $J = 8.2, 6.8, 1.3$  Hz, 1H), 7.24 (ddd,  $J = 8.2, 6.8, 1.2$  Hz, 1H), 7.09 (s, 1H), 5.15 (t,  $J = 5.5$  Hz, 1H), 4.63 (d,  $J = 4.5$  Hz, 2H).

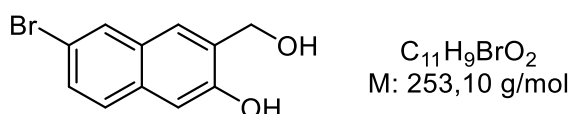
MS (ESI+):  $m/z$  (%) 175.1 (100) [M+H]<sup>+</sup>

**6-(hydroxymethyl)naphthalene-1,7-diol (2b)**

Reduction of **2a** (3.23 g, 14.8 mmol) with 1 M LiAlH<sub>4</sub> solution (17.8 mL, 17.8 mmol) in dry THF (30 mL) afforded the title compound **2b** as a yellow solid (2.76 g, 98 %).

**<sup>1</sup>H NMR** (500 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 9.72 (s, 1H), 9.64 (s, 1H), 7.70 (s, 1H), 7.36 (s, 1H), 7.19 (d,  $J = 8.1$  Hz, 1H), 7.02 (t,  $J = 7.8$  Hz, 1H), 6.71 (d,  $J = 7.4$  Hz, 1H), 5.11 (t,  $J = 5.6$  Hz, 1H), 4.65-4.57 (m, 2H).

MS (ESI+):  $m/z$  (%) 191.4 (100) [M+H]<sup>+</sup>

**6-bromo-3-(hydroxymethyl)naphthalene-2-ol (3b)**

Reduction of **3a** (3.30 g, 11.74 mmol) with 1 M LiAlH<sub>4</sub> solution (23.5 mL, 23.48 mmol) in dry THF (40 mL) afforded the title compound **3b** as a cream-coloured solid (2.91 g, 98 %).

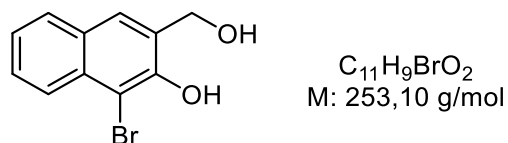
**<sup>1</sup>H NMR** (300 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 10.01 (s, 1H), 8.03 (d,  $J = 1.9$  Hz, 1H), 7.80 (s, 1H), 7.64 (d,  $J = 8.8$  Hz, 1H), 7.43 (dd,  $J = 8.7, 2.1$  Hz, 1H), 7.11 (s, 1H), 5.22 (br, 1H), 4.62 (s, 2H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 153.6, 133.3, 131.9, 129.1, 128.9, 128.2, 127.8, 124.5, 115.2, 108.0, 58.5.



MS (ESI +): m/z (%) 234.9/236.9 (100) [M+H]<sup>+</sup>

**1-bromo-3-(hydroxymethyl)naphthalene-2-ol (4b)**



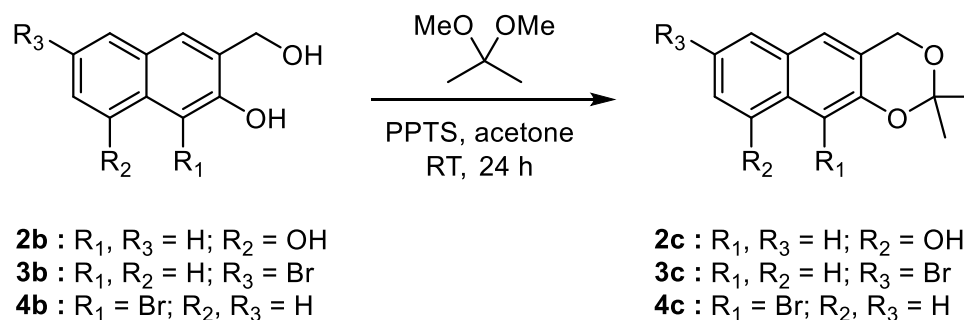
Reduction of **4a** (3.19 g, 11.35 mmol) with 1 M LiAlH<sub>4</sub> solution (22.7 mL, 22.69 mmol) in dry THF (40 mL) afforded the title compound **4b** as an ochre-coloured solid (2.86 g, 99 %).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ/ppm: 9.50 (s, 1H), 8.02-7.95 (m, 1H), 7.91-7.83 (m, 2H), 7.55 (ddd, *J* = 8.3, 6.9, 1.2 Hz, 1H), 7.39 (ddd, *J* = 8.1, 6.9, 1.1 Hz, 1H), 5.52 (br, 1H), 4.72 (s, 2H).

<sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>) δ/ppm: 150.0, 132.6, 131.3, 128.8, 128.0, 127.1, 125.3, 124.8, 123.9, 105.9, 59.5.

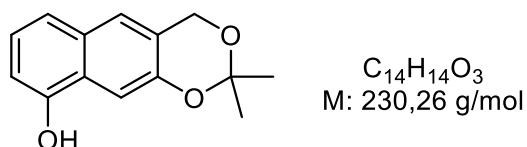
MS (ESI +): m/z (%) 234.9/236.9 (100) [M+H]<sup>+</sup>

**General procedure for acetonide protection of 1,3-diols**



To a solution of 1,3-diol (1.0 eq) in acetone were added 2,2-dimethoxypropane (3.5 eq) and a catalytic amount of pyridinium *p*-toluenesulfonate (PPTS, ~ 10 mol%). The mixture was stirred at room temperature for 24 h, after which the solvents were concentrated under reduced pressure. Purification of the crude product thus obtained by column chromatography on silica afforded the desired acetonide protected 1,3-diols.

**2,2-dimethyl-4H-naphtho[2,3-*d*][1,3]dioxine (2c)**



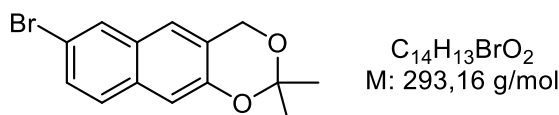
Reaction of **2b** (7.08 g, 37.2 mmol), 2,2-dimethoxypropane (16.1 mL, 130.2 mmol) and PPTS (~ 100 mg) in acetone (150 mL) followed by column chromatography purification on silica

(dry load; pentane:EtOAc, 1:1, v/v) afforded the title compound **2c** as a yellow solid (7.37 g, 86 %).

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 9.91 (s, 1H), 7.51 (br, 1H), 7.40 (br, 1H), 7.24-7.18 (m, 1H), 7.16-7.09 (m, 1H), 6.78 (dd, *J* = 7.3, 1.1 Hz, 1H), 5.01 (br, 2H), 1.51 (s, 6H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 152.0, 148.6, 129.4, 124.9, 124.1, 123.5, 121.4, 118.0, 107.5, 105.9, 99.5, 60.4, 24.8 (2C).

#### 7-bromo-2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxine (**3c**)



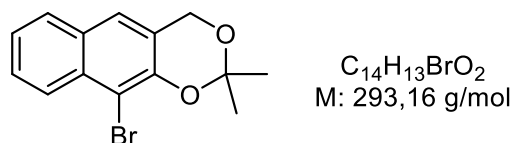
Reaction of **3b** (0.87 g, 3.45 mmol), 2,2-dimethoxypropane (1.48 mL, 12.1 mmol) and PPTS (~ 10 mg) in acetone (20 mL) followed by column chromatography purification on silica (DCM) afforded the title compound **3c** as a yellow solid (0.81 g, 80 %).

$R_f$  = 0.69 (DCM) [UV, CAM]

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 8.06 (d, *J* = 1.9 Hz, 1H), 7.71 (d, *J* = 9.0 Hz, 1H), 7.64 (s, 1H), 7.50 (dd, *J* = 8.8, 2.0 Hz, 1H), 7.28 (s, 1H), 5.09-5.02 (m, 2H), 1.52 (s, 6H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 149.9, 131.8, 129.1, 129.0, 128.8, 128.6, 123.1, 122.6, 116.5, 111.3, 99.8, 60.3, 24.8 (2C).

#### 10-bromo-2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxine (**4c**)

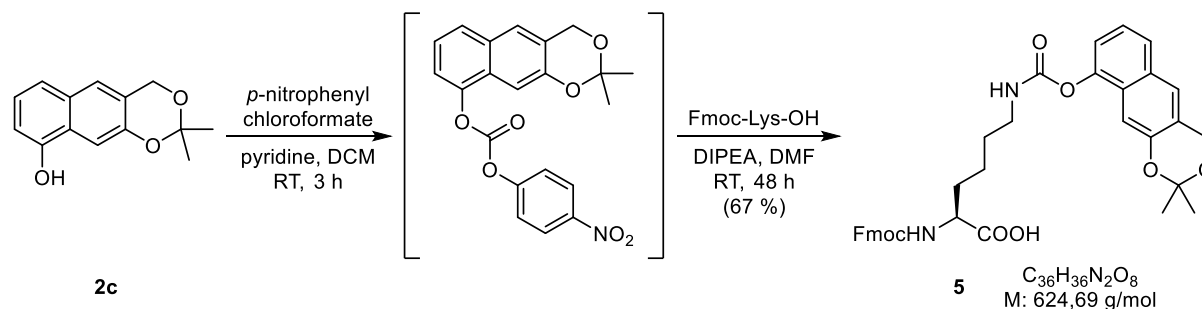


Reaction of **4b** (1.80 g, 7.39 mmol), 2,2-dimethoxypropane (3.18 mL, 25.9 mmol) and PPTS (~ 20 mg) in acetone (50 mL) followed by column chromatography purification on silica (pentane:DCM, 1:1, v/v) afforded the title compound **4c** as a pale yellow solid (1.56 g, 72 %).

$R_f$  = 0.62 (pentane:DCM, 1:1, v/v) [UV, CAM]

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 8.04 (d, *J* = 8.6 Hz, 1H), 7.86 (d, *J* = 8.1 Hz, 1H), 7.73 (s, 1H), 7.58 (ddd, *J* = 8.4, 6.9, 1.3 Hz, 1H), 7.44 (ddd, *J* = 8.1, 6.9, 1.2 Hz, 1H), 5.08 (d, *J* = 1.1 Hz, 2H), 1.57 (s, 6H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 146.6, 131.3, 128.6, 128.0, 127.5, 124.9, 124.6, 123.7, 122.0, 106.0, 101.2, 60.1, 24.9 (2C).

**Unsuccessful preparation of *o*NQMP-UAA1*****N*<sup>2</sup>-(9-fluorenylmethoxycarbonyl)-*N*<sup>6</sup>-(((2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxin-9-yl)oxy)carbonyl)-L-lysine (**5**)**

Acetonide-protected **2c** (0.59 g, 2.56 mmol) was dissolved in 10 mL of dry DCM, followed by addition of *p*-nitrophenyl chloroformate (1.03 g, 5.12 mmol) and pyridine (0.42  $\mu$ L, 5.12 mmol). The reaction mixture was stirred at room temperature for 3 h, after which the solvents were removed under reduced pressure. Purification of the crude residue was attempted by column chromatography on silica (DCM), but always afforded a mixture of the activated alcohol and *p*-nitrophenol. Nevertheless, the crude product (~ 1.27 g) was used directly in the following step.

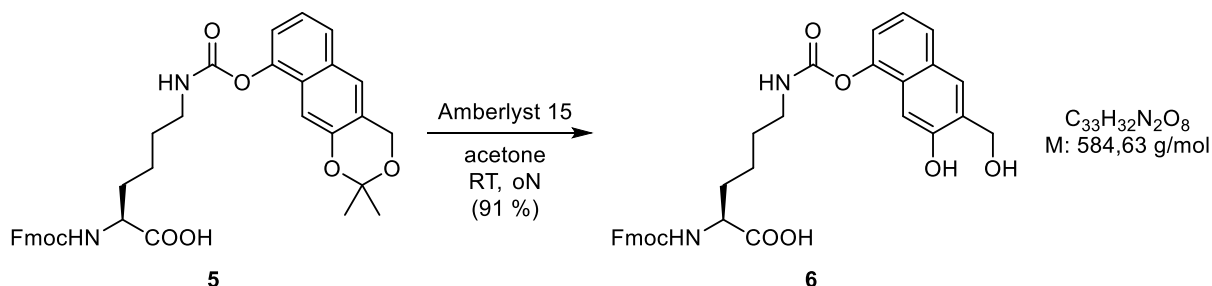
$R_f = 0.53$  (DCM) [UV, CAM]

The activated alcohol from the previous step was dissolved in 30 mL of dry DMF, followed by addition of Fmoc-Lys-OH (1.13 g, 3.07 mmol) and DIPEA (0.89  $\mu$ L, 5.12 mmol). The mixture was stirred at room temperature for 48 h, after which it was diluted with EtOAc. The organic phase was washed with 10 % citric acid solution, satur.  $NaHCO_3$  and brine, dried over  $MgSO_4$  and concentrated *in vacuo*. Purification of the crude residue by column chromatography on silica (packed in 5 % MeOH/DCM and eluted with 5  $\rightarrow$  10 % MeOH/DCM) afforded the desired acetonide- and Fmoc-protected *o*NQMP-UAA1 **5** as a yellow foam (1.07 g, 67 %).

**<sup>1</sup>H NMR** (300 MHz,  $DMSO-d_6$ )  $\delta$ /ppm: 7.95 (t,  $J = 5.6$  Hz, 1H), 7.87 (d,  $J = 7.4$  Hz, 2H), 7.74-7.60 (m, 4H), 7.44-7.35 (m, 2H), 7.35-7.12 (m, 6H), 5.04 (s, 2H), 4.34-4.17 (m, 3H), 3.87 (td,  $J = 8.2, 4.7$  Hz, 1H), 3.09 (q,  $J = 6.6$  Hz, 2H), 1.83-1.31 (m, 12H).

**MS (ESI+):**  $m/z$  (%) 647.4  $[M+Na]^+$

***N*<sup>2</sup>-(9-fluorenylmethoxycarbonyl)-*N*<sup>6</sup>-(((7-hydroxy-6-(hydroxymethyl)naphthalen-1-yl)-oxy)carbonyl)-L-lysine (**6**)**



Amberlyst<sup>®</sup> 15 resin (hydrogen form, 30 mg) was added to the solution of Fmoc- and acetonide-protected *o*NQMP-UAA1 **5** (35 mg, 0.056 mmol) in 5 mL of acetone and the mixture was stirred at room temperature overnight. The resin was removed by filtration through a cotton plug, while the filtrate was removed under reduced pressure to afford the Fmoc-protected *o*NQMP-UAA1 **6** as a yellow oil (30 mg, 91 %).

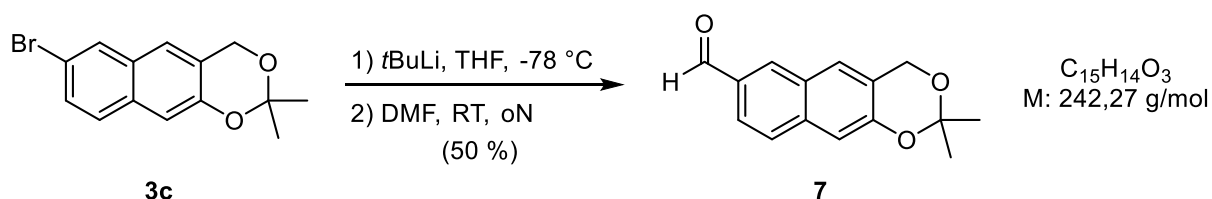
**<sup>1</sup>H NMR** (300 MHz, MeOD)  $\delta$ /ppm: 7.82 (s, 1H), 7.79 (d,  $J = 7.5$  Hz, 2H), 7.68-7.60 (m, 3H), 7.41-7.33 (m, 2H), 7.33-7.25 (m, 2H), 7.22 (t,  $J = 7.8$  Hz, 1H), 7.18 (s, 1H), 7.13-7.07 (m, 1H), 4.79 (d,  $J = 1.1$  Hz, 2H), 4.39-4.33 (m, 2H), 4.25-4.16 (m, 2H), 3.28-3.21 (m, 2H), 1.95-1.48 (m, 6H).

**MS (ESI+):**  $m/z$  (%) 607.2 [ $M+Na$ ]<sup>+</sup>

Treatment of **6** with NaOH resulted with the cleavage of the carbamate group on the side-chain of L-lysine.

**Synthesis of *o*NQMP-UAA2**

**2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxine-7-carbaldehyde (**7**)**



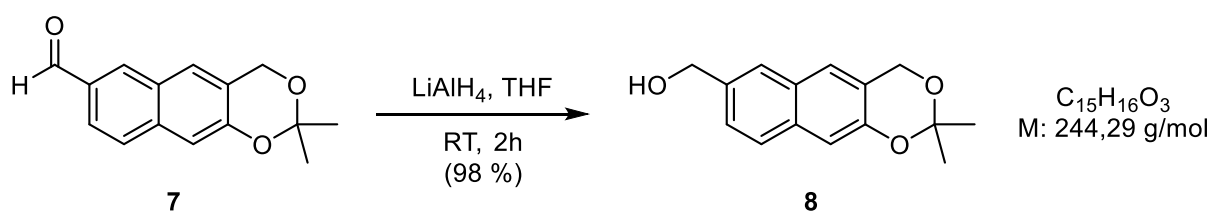
A two-neck round-bottom Schlenk flask was evacuated/Ar filled three times, charged with **4c** (1.00 g, 3.41 mmol) and then again evacuated/Ar filled three times. 30 mL of dry THF were added *via* canula and the solution cooled to -78 °C (dry ice/acetone cooling bath). *t*BuLi (2.14 mL, 3.75 mmol) was then carefully added to the cooled solution and the mixture was stirred at -78 °C for 1 h, followed by addition of dry DMF (0.53 mL, 6.82 mmol). The reaction mixture was warmed up to room temperature and stirring continued overnight. After quenching with water, most of the THF was removed under reduced pressure and the mixture extracted

several times with DCM. Combined organic phases were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under vacuum. Purification of the crude residue by column chromatography on silica (pentane:EtOAc, 5:1, v/v) afforded **7** (0.41 g, 50 %) as a yellow solid.

$R_f$  = 0.41 (pentane:EtOAc, 1:1, v/v) [UV, CAM]

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 10.07 (obs d,  $J$  = 0.5 Hz, 1H), 8.45 (br, 1H), 7.94-7.84 (m, 2H), 7.79 (dd,  $J$  = 8.6, 1.6 Hz, 1H), 7.37 (s, 1H), 5.10 (obs d,  $J$  = 0.9 Hz, 2H), 1.55 (s, 6H).

**(2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxin-7-yl)methanol (8)**



A 1 M LiAlH<sub>4</sub> solution in THF (1.11 mL, 1.11 mmol) was added to the solution of **7** (0.27 g, 1.11 mmol) in 10 mL of dry THF and the reaction mixture was stirred at room temperature for 2 h. After quenching the reaction with 5 % HCl solution (v/v), the mixture was extracted several times with Et<sub>2</sub>O. Combined organic phases were washed with satur. NaHCO<sub>3</sub> solution and brine, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under vacuum to afford **8** (0.26 g, 98 %) as a peach-coloured solid. No further purification was required.

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 7.72 (m, 2H), 7.60 (s, 1H), 7.38-7.33 (m, 1H), 7.21 (s, 1H), 5.23 (t,  $J$  = 5.7 Hz, 1H), 5.04 (obs d,  $J$  = 0.8 Hz, 2H), 4.60 (d,  $J$  = 5.7 Hz, 2H), 1.52 (s, 6H).

**N<sup>2</sup>-(tert-butoxycarbonyl)-N<sup>6</sup>-(((2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxin-7-yl)methoxy)carbonyl)-L-lysine (9)**



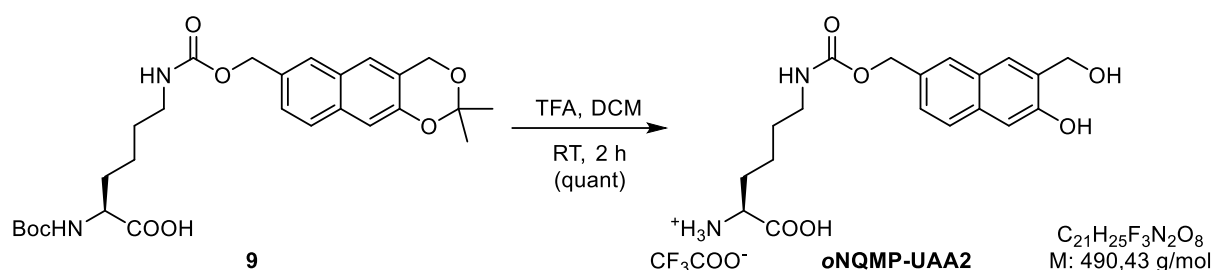
DSC (0.30 g, 1.18 mmol) and DIPEA (0.68 mL, 3.93 mmol) were added to the solution of **8** (0.24 g, 0.98 mmol) in 10 mL of dry DMF and the mixture was stirred at room temperature for 2 h, until TLC (DCM) showed that all of the starting material had been consumed. Boc-Lys-OH (0.29 g, 1.18 mmol) was then added to the mixture and stirring was continued overnight. After evaporation of the solvent, the residue was dissolved in DCM and washed

with 10 % citric acid, satur. NaHCO<sub>3</sub> solution and brine. The organic phase was dried over MgSO<sub>4</sub> and concentrated under reduced pressure. Purification by column chromatography on silica (5 % MeOH/DCM) afforded **9** (0.39 g, 77 %) as a pale orange foam.

$R_f$  = 0.31 (5 % MeOH/DCM) [UV, CAM]

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 7.76-7.71 (m, 2H), 7.63 (s, 1H), 7.37 (dd,  $J$  = 8.6, 1.6 Hz, 1H), 7.28-7.22 (m, 2H), 7.01 (d,  $J$  = 7.9 Hz, 1H), 5.10 (s, 2H), 5.05 (s, 2H), 3.85-3.78 (m, 1H), 2.97 (q,  $J$  = 5.9 Hz, 2H), 1.67-1.48 (m, 8H), 1.43-1.22 (13 H).

**N<sup>6</sup>-(((6-hydroxy-7-(hydroxymethyl)naphthalen-2-yl)methoxy)carbonyl)-L-lysine trifluoroacetate (*o*NQMP-UAA2)**



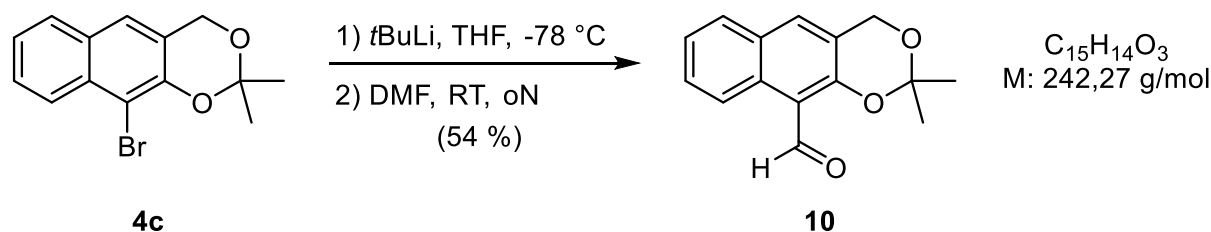
Protected amino acid **9** (50 mg, 0.097 mmol) was dissolved in 1.0 mL of DCM and 0.5 mL of TFA was added. The mixture was stirred at room temperature for 2 h, after which the solvent was removed under reduced pressure and the residue triturated with cold Et<sub>2</sub>O to afford ***o*NQMP-UAA2** as an orange solid (48 mg, quant). ***o*NQMP-UAA2** was stored at -20 °C and dissolved as a 100 mM stock solution in 1 M NaOH before usage.

<sup>1</sup>H NMR (300 MHz, MeOD)  $\delta$ /ppm: 7.77 (s, 1H), 7.69 (s, 1H), 7.61 (d,  $J$  = 8.5 Hz, 1H), 7.34 (dd,  $J$  = 8.4, 1.7 Hz, 1H), 7.09 (s, 1H), 4.80 (d,  $J$  = 1.0 Hz, 2H), 4.55 (s, 2H), 4.03-3.96 (m, 1H), 3.00-2.92 (m, 2H), 2.09-1.86 (m, 2H), 1.80-1.49 (m, 4H).

MS (ESI<sup>+</sup>):  $m/z$  (%) 399.8 (100) [M+Na]<sup>+</sup>

**Synthesis of *o*NQMP-UAA3**

**2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxine-10-carbaldehyde (**10**)**



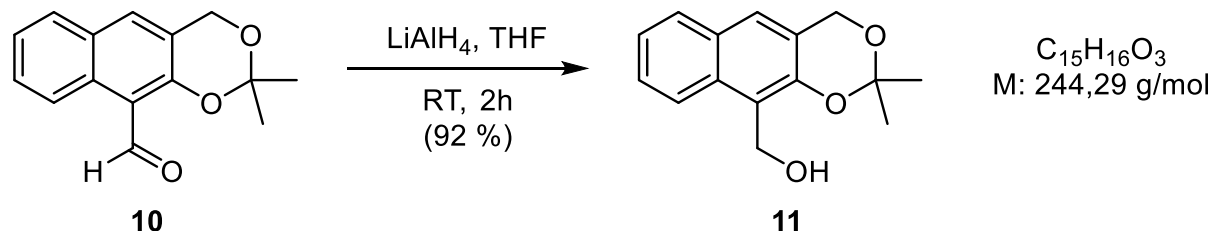
A two-neck round-bottom Schlenk flask was evacuated/Ar filled three times, charged with **4c** (1.00 g, 3.41 mmol) and then again evacuated/Ar filled three times. 30 mL of dry THF were added *via* canula and the solution cooled to -78 °C (dry ice/acetone cooling bath). *t*BuLi (2.14 mL, 3.75 mmol) was then carefully added to the cooled solution and the mixture was stirred at -78 °C for 1 h, followed by addition of dry DMF (0.53 mL, 6.82 mmol). The reaction mixture was warmed up to room temperature and stirring continued overnight. After quenching with water, most of the THF was removed under reduced pressure and the mixture extracted several times with DCM. Combined organic phases were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under vacuum. Purification of the crude residue by column chromatography on silica (pentane:EtOAc, 5:1, v/v) afforded **10** (0.45 g, 54 %) as a yellow solid.

$R_f = 0.57$  (pentane:EtOAc, 1:1, v/v) [UV, CAM]

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 10.71 (s, 1H), 9.08 (d,  $J = 8.6$  Hz, 1H), 8.01 (s, 1H), 7.88 (d,  $J = 8.0$  Hz, 1H), 7.61 (ddd,  $J = 8.5, 6.8, 1.4$  Hz, 1H), 7.46 (ddd,  $J = 8.1, 6.8, 1.2$  Hz, 1H), 5.12 (s, 2H), 1.63 (s, 6H).

<sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 190.6, 157.3, 132.2, 130.0, 129.3, 128.4, 127.3, 124.9, 123.8, 120.8, 114.7, 101.6, 60.1, 24.8 (2C).

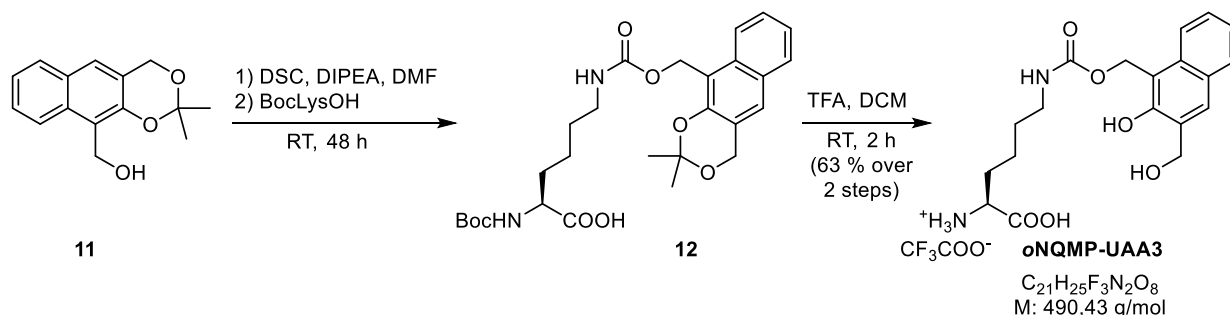
**(2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxin-10-yl)methanol (11)**



A 1 M LiAlH<sub>4</sub> solution in THF (1.03 mL, 1.03 mmol) was added the solution of **7** (0.25 g, 1.01 mmol) in 10 mL of dry THF and the reaction mixture was stirred at room temperature for 2 h. After quenching the reaction with 5 % HCl solution (v/v), the mixture was extracted several times with Et<sub>2</sub>O. Combined organic phases were washed with satur. NaHCO<sub>3</sub> solution and brine, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under vacuum to afford **11** (0.23 g, 92 %) as a white solid. No further purification was required.

<sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 8.12-8.07 (m, 1H), 7.80-7.75 (m, 1H), 7.59 (s, 1H), 7.45 (ddd,  $J = 8.4, 6.8, 1.4$  Hz, 1H), 7.34 (ddd,  $J = 8.0, 6.8, 1.2$  Hz, 1H), 5.04 (d,  $J = 1.0$  Hz, 2H), 4.89-4.85 (m, 3H), 1.54 (s, 6H).

***N*<sup>6</sup>-(((2-hydroxy-3-(hydroxymethyl)naphthalen-1-yl)methoxy)carbonyl)-L-lysine trifluoroacetate (*o*NQMP-UAA3)**

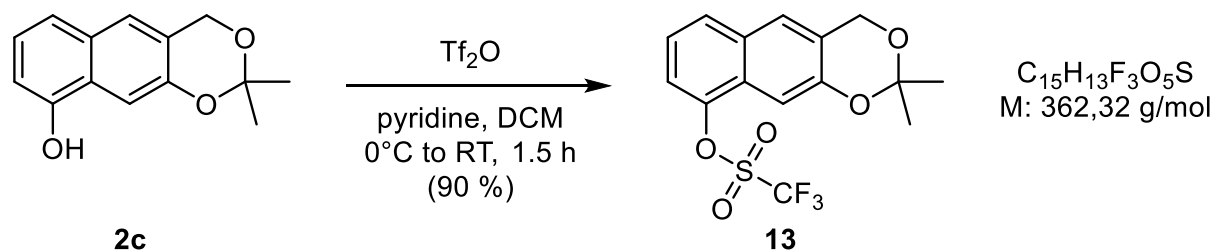


DSC (0.51 g, 1.99 mmol) and DIPEA (0.69 mL, 3.98 mmol) were added to the solution of **11** (0.33 g, 1.33 mmol) in 10 mL of dry DMF and the mixture was stirred at room temperature for 2.5 h, until TLC (DCM) showed that all of the starting material had been consumed. Boc-Lys-OH (0.39 g, 1.56 mmol) was then added to the mixture and stirring was continued for 48 h. After evaporation of the solvent, the residue was dissolved in DCM and washed with 10 % citric acid, satur. NaHCO<sub>3</sub> solution and brine. The organic phase was dried over MgSO<sub>4</sub> and concentrated under reduced pressure. Purification by column chromatography on silica (5 % MeOH/DCM) afforded **12** (~ 0.43 g) as a yellow foam which was directly dissolved in 10 mL of DCM, followed by addition of 5 mL TFA. The mixture was stirred at room temperature for 2 h, after which the solvent was removed under reduced pressure and the residue triturated with cold Et<sub>2</sub>O. **oNQMP-UAA3** (0.41 g, 63 %) was afforded as a yellow solid and was dissolved as a 100 mM stock solution in 2 M NaOH before usage.

**MS (ESI+):** m/z (%) 399.8 (100) [M+Na]<sup>+</sup>

**Unsuccessful preparation of *o*NQMP-UAA4**

**2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxin-9-yl trifluoromethanesulfonate (**13**)**



Pyridine (0.36 mL, 4.39 mmol) was added to a solution of acetone **2c** (0.51 g, 2.19 mmol) in 10 mL of dry DCM and the mixture was cooled to 0 °C. Triflic anhydride (0.44 mL, 2.63 mmol) was then added dropwise to the mixture and the reaction left stirring at room temperature for 1 h. The mixture was afterwards diluted with Et<sub>2</sub>O and quenched with 20 %



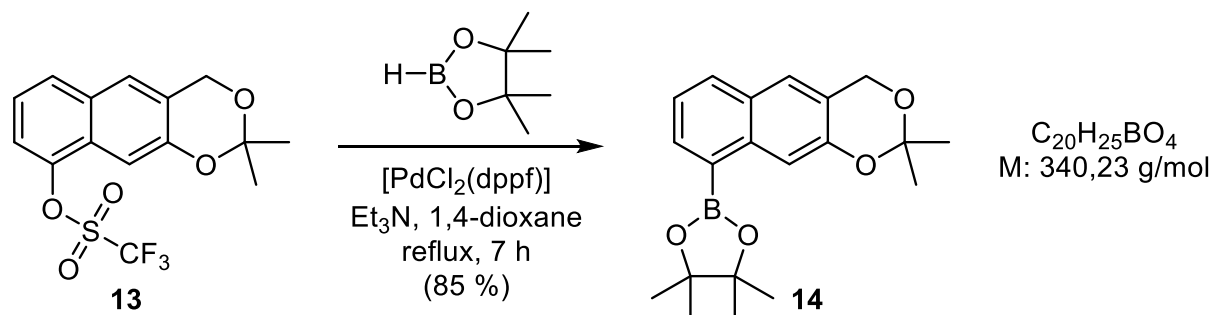
HCl solution (v/v). The organic phase was separated, washed sequentially with saturated NaHCO<sub>3</sub> solution and brine, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Purification by column chromatography on silica (packed in pentane and eluted with 50 % Et<sub>2</sub>O/pentane, v/v) afforded **13** (717 mg, 90 %) as a bright yellow solid.

$R_f$  = 0.72 (Et<sub>2</sub>O:pentane, 1:1, v/v) [UV, CAM]

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 7.97 (d,  $J$  = 8.3 Hz, 1H), 7.89 (s, 1H), 7.58 (d,  $J$  = 7.5 Hz, 1H), 7.45 (t,  $J$  = 8.0 Hz, 1H), 7.16 (s, 1H), 5.10 (s, 2H), 1.56 (s, 6H).

<sup>13</sup>C NMR (125 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 151.4, 143.8, 129.4, 128.7, 125.9, 125.1, 123.4 (d,  $J$  = 3.0 Hz), 119.5, 118.5, 116.9, 103.5, 100.6, 60.2, 24.8 (2C).

**2-(2,2-dimethyl-4H-naphtho[2,3-*d*][1,3]dioxin-9-yl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (14)**

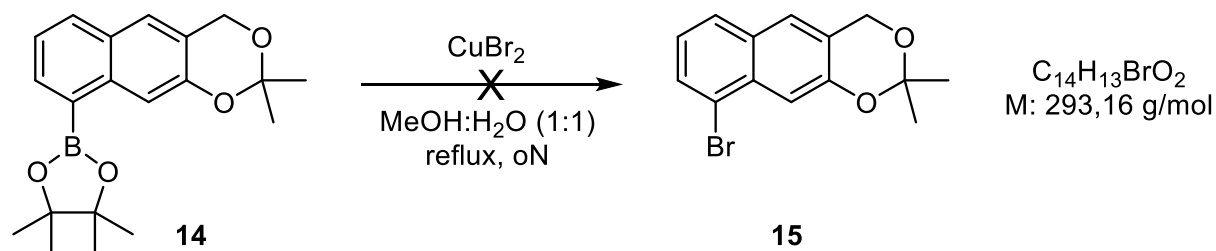


[PdCl<sub>2</sub>(dppf)] (0.34 g, 3 mol%) was suspended in 10 mL of dry 1,4-dioxane and triflate **13** (0.56 g, 1.53 mmol) and Et<sub>3</sub>N (0.64 mL, 4.60 mmol) were added to the mixture, followed by addition of pinacolborane (0.67 mL, 4.60 mmol). The mixture was refluxed for 7h, upon which the reaction reached complete conversion, as judged by TLC (50 % Et<sub>2</sub>O/pentane, v/v). After quenching the reaction with water (20 mL), the mixture was extracted several times with DCM. Combined organic phases were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. The obtained residue was re-dissolved in pentane and filtered off in order to remove the catalyst. Removal of pentane under reduced pressure yielded crude pinacolborane **14** (0.52 g, 85 %) as a yellow oil which was used in the next step without further purification.

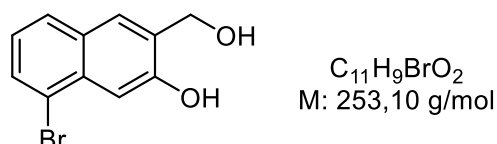
$R_f$  = 0.88 (Et<sub>2</sub>O:pentane, 1:1, v/v) [UV, CAM]

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 7.96 (s, 1H), 7.95-7.87 (m, 2H), 7.66 (s, 1H), 7.33 (dd,  $J$  = 8.1, 6.9 Hz, 1H), 5.05 (s, 2H), 1.54 (s, 6H), 1.36 (s, 12H).

<sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 149.8, 136.5, 135.8, 131.3, 127.9, 124.4, 122.9, 120.9, 111.7, 99.7, 83.5 (2C), 60.2, 24.8 (2C), 24.7 (4C).

**9-bromo-2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxine (15)**

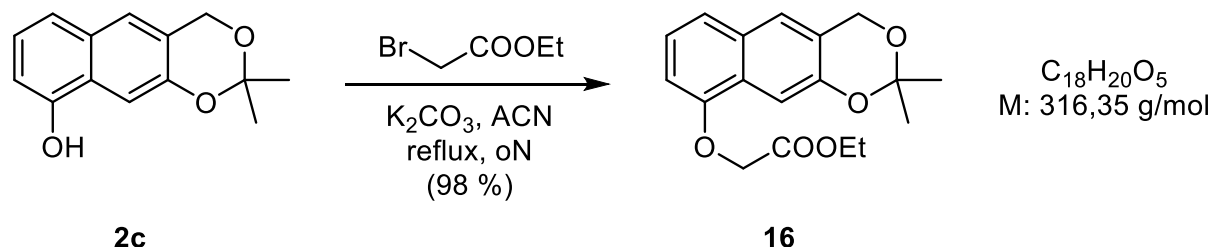
Pinacolborane **14** (1.53 mmol) was dissolved in 15 mL of MeOH and a solution of copper(II) bromide (1.03 g, 4.60 mmol) in 15 mL water was added. The mixture was refluxed overnight. After cooling to ambient temperature, the reaction mixture was extracted several times with Et<sub>2</sub>O. The combined organic phases were washed with brine, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Column chromatography of the crude residue on silica (Et<sub>2</sub>O:pentane, 1:1, v/v) did not afford the desired product acetone-protected **15**. Instead, **8-bromo-3-(hydroxymethyl)naphthalen-2-ol** was isolated as a yellow solid (0.19 mg, 43 %).



$R_f = 0.42$  (Et<sub>2</sub>O:pentane, 1:1, v/v) [UV, CAM]

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ/ppm: 10.27 (s, 1H, OH), 7.88 (s, 1H, H), 7.84 (d,  $J = 8.2$  Hz, 1H), 7.70 (dd,  $J = 7.5, 1.0$  Hz, 1H), 7.42 (s, 1H), 7.17 (dd,  $J = 8.0, 7.5$  Hz, 1H), 5.25 (t,  $J = 5.4$  Hz, 1H), 4.64 (d,  $J = 4.4$  Hz, 2H).

MS (ESI<sup>+</sup>): m/z (%) 234.9/236.9 (100) [M+H]<sup>+</sup>

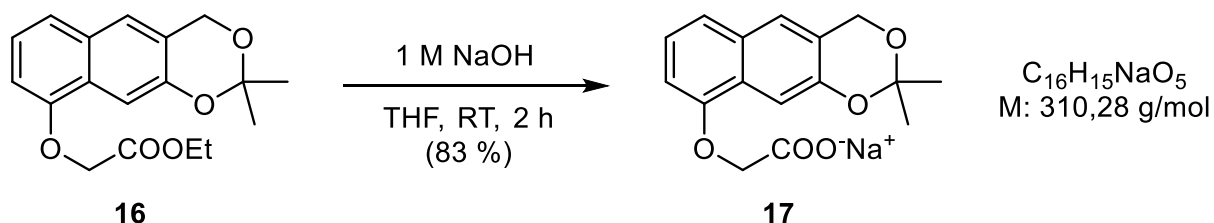
**Synthesis of oNQMP-UAA5****Ethyl 2-((2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxin-9-yl)oxy)acetate (16)**

Ethyl 2-bromoacetate (1.16 mL, 10.42 mmol) was added to the mixture of **2c** (1.20 g, 5.21 mmol) and ground K<sub>2</sub>CO<sub>3</sub> (1.44 g, 10.42 mmol) in 70 mL of dry ACN and the mixture

was stirred under reflux overnight. After completion, as judged by TLC (pentane:EtOAc, 4:1, v/v), the solvent was removed under reduced pressure. Purification by column chromatography on silica (pentane:EtOAc, 4:1, v/v) afforded **16** (1.62 g, 98 %) as a yellow oil.

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ/ppm: 7.62 (s, 1H), 7.46 (s, 1H), 7.39 (d, *J* = 8.3 Hz, 1H), 7.22 (t, *J* = 8.0 Hz, 1H), 6.80 (d, *J* = 7.6 Hz, 1H), 5.05 (s, 2H), 4.92 (s, 2H), 4.20 (q, *J* = 7.1 Hz, 2H), 1.53 (s, 6H), 1.23 (t, *J* = 7.1 Hz, 3H).

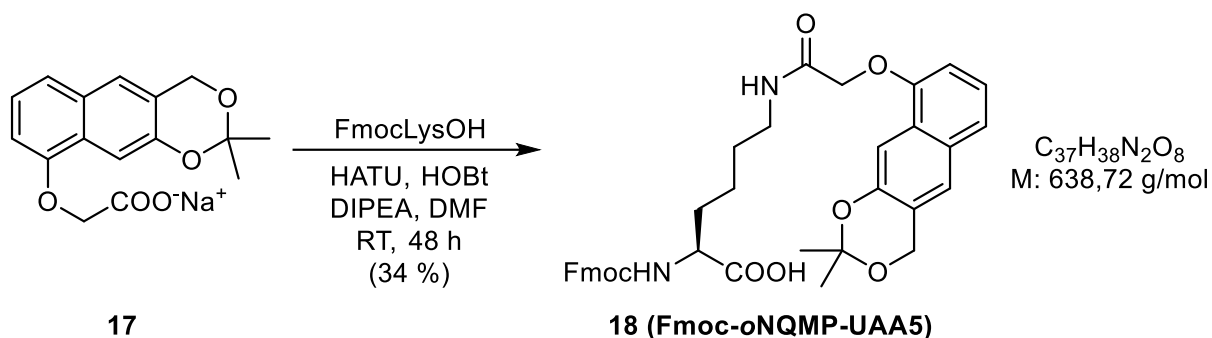
#### Sodium 2-((2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxin-9-yl)oxy)acetate (**17**)



In order to deprotect the ethyl ester, **16** (0.70 g, 2.21 mmol) was dissolved in 9 mL of THF and 1 M NaOH solution (4.42 mL, 4.42 mmol) was added. The mixture was stirred at room temperature for 2 h, after which the solvent was removed under reduced pressure. The residue was dissolved in EtOAc and washed with 10 % citric acid solution. The organic phase was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo* to afford **17** (0.57 g, 83 %) as a yellow solid.

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ/ppm: 7.62 (s, 1H), 7.46 (s, 1H), 7.38 (d, *J* = 8.3 Hz, 1H), 7.21 (t, *J* = 8.0 Hz, 1H), 6.78 (d, *J* = 7.6 Hz), 5.04 (s, 2H), 4.81 (s, 2H), 1.52 (s, 6H).

#### *N*<sup>2</sup>-(((9-fluorenylmethoxycarbonyl)-*N*<sup>6</sup>-(2-((2,2-dimethyl-4*H*-naphtho[2,3-*d*][1,3]dioxin-9-yl)oxy)acetyl)-*L*-lysine (**18**, Fmoc-*o*NQMP-UAA5)



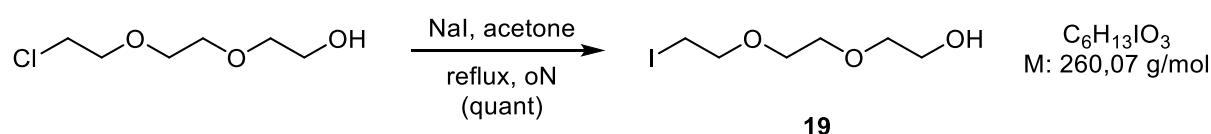
Carboxylic acid **17** (0.53 g, 1.84 mmol), DIPEA (0.64 mL, 3.67 mmol), HOBT (0.27 g, 2.02 mmol) and HATU (0.77 g, 2.02 mmol) were mixed together in 10 mL of DMF. After stirring for 15 minutes at room temperature, Fmoc-Lys-OH (0.81 g, 2.02 mmol) was added and the obtained suspension was left stirring over 48 h. The solvent was removed under reduced pressure, the obtained residue dissolved in EtOAc and washed with 10 % citric acid solution

and brine. The organic phase was dried over  $\text{Na}_2\text{SO}_4$  and concentrated under vacuum. Purification by column chromatography on silica (5 % MeOH/DCM) afforded **18** (0.40 g, 34 %) as an orange foam.

$^1\text{H NMR}$  (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$ /ppm: 8.23 (t,  $J = 5.9$  Hz, 1H), 7.88 (d,  $J = 7.5$  Hz, 2H), 7.75-7.70 (m, 3H), 7.65-7.59 (m, 2H), 7.48 (s, 1H), 7.43-7.35 (m, 3H), 7.34-7.29 (m, 2H), 7.24-7.19 (m, 1H), 5.04 (s, 2H), 4.81 (s, 2H), 4.32-4.18 (m, 3H), 3.97-3.88 (m, 1H), 3.22-3.14 (m, 2H), 1.78-1.58 (m, 2H), 1.53 (s, 6H), 1.51-1.29 (m, 4H).

### Synthesis of *o*NQMP-TAMRA and *o*NQMP-biotin

#### 2-(2-(2-iodoethoxy)ethoxy)ethan-1-ol (**19**)

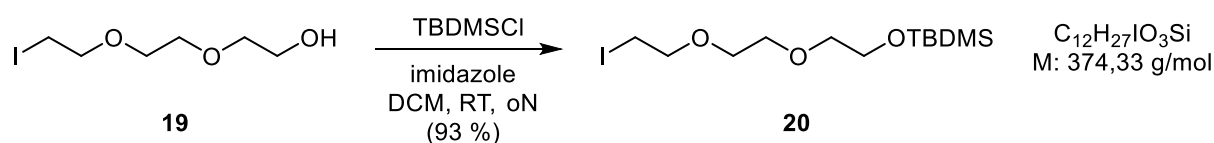


2-(2-(2-chloroethoxy)ethoxy)ethan-1-ol (0.84 g, 4.98 mmol) was dissolved in 8 mL of dry acetone and sodium iodide (1.50 g, 9.96 mmol) was added. The reaction mixture turned yellow and a white precipitate formed immediately. The mixture was refluxed overnight, upon which the reaction reached complete conversion, as judged by TLC (60 % EtOAc/pentane, v/v). The solvent was removed under reduced pressure and the obtained residue dissolved in DCM and washed several times with brine. The organics were dried over  $\text{MgSO}_4$  and concentrated *in vacuo* to yield crude **19** as a yellow oil (1.30 g, quant). The crude product was used in the next step without further purification.

$^1\text{H NMR}$  (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$ /ppm: 4.57 (t,  $J = 5,4$  Hz, 1H), 3.66 (t,  $J = 6.4$  Hz, 2H), 3.56 (ddd,  $J = 5.8, 3.6, 1.2$  Hz, 2H), 3.53 (ddd,  $J = 5.8, 3.6, 1.2$  Hz, 2H), 3.51-3.46 (m, 2H), 3.45-3.40 (m, 2H), 3.35-3.29 (m, 2H).

**MS (ESI+)**  $m/z$  (%): 154.9 (100)  $[\text{M-I}+\text{Na}]^+$ , 282.9 (40)  $[\text{M}+\text{Na}]^+$ , 260.9 (5)  $[\text{M}+\text{H}]^+$

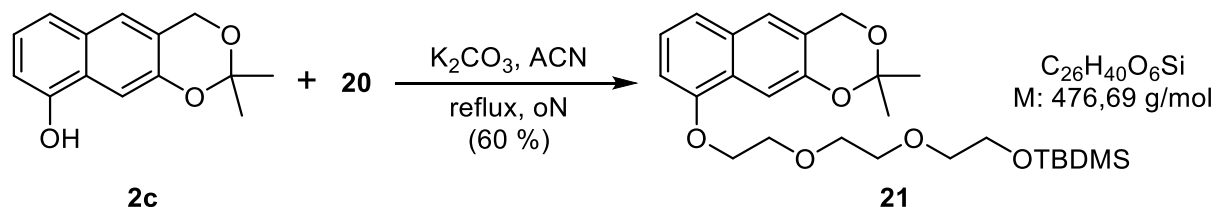
#### 12-iodo-2,2,3,3-tetramethyl-4,7,10-trioxa-3-siladodecane (**20**)



Imidazole (1.02 g, 15.0 mmol) and TBDMSCl (0.90 g, 6.0 mmol) were added to a solution of **19** (1.30 g, 5.0 mmol) in 50 mL of dry DCM and the obtained white suspension was stirred at room temperature overnight. The reaction mixture was then washed with water and brine, while the organic phase was separated, dried over  $\text{Na}_2\text{SO}_4$  and concentrated under reduced pressure to afford **20** (1.74 g, 93 %) as a pale yellow oil. The crude product was used in the next step without further purification.

$^1\text{H NMR}$  (500 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 3.72-3.64 (m, 5H), 3.59-3.52 (m, 4H), 3.48-3.44 (m, 2H), 3.33-3.29 (m, 1H), 0.86 (s, 9H), 0.04 (s, 6H).

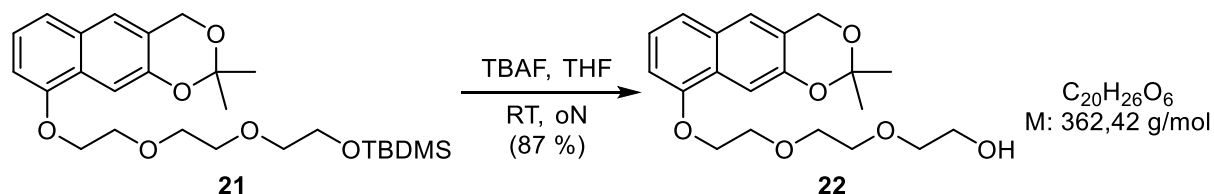
**12-((2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxin-9-yl)oxy)-2,2,3,3-tetramethyl-4,7,10-trioxa-3-siladodecane (21)**



Ground  $\text{K}_2\text{CO}_3$  (1.08 g, 7.82 mmol) was added to a solution of **2c** (0.75 g, 3.26 mmol) in 20 mL of dry ACN and the mixture was stirred for a couple of minutes before the addition of **20** (1.46 g, 9.38 mmol). The mixture was stirred at reflux for 48 h, after which the solvents were removed under reduced pressure. Purification of the crude residue by column chromatography on silica (EtOAc:pentane, 1:5, v/v) afforded **21** (0.93 g, 60 %) as a yellow oil.

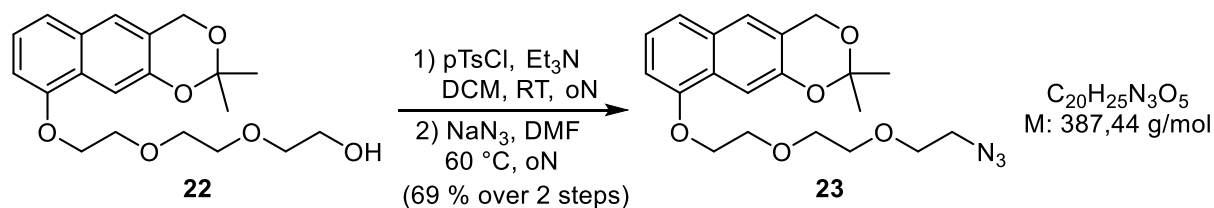
$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 7.66 (s, 1H), 7.41 (br, 1H), 7.32-7.27 (m, 1H), 7.23-7.16 (m, 1H), 6.72 (dd,  $J = 7.6, 1.0$  Hz, 1H), 5.06 (d,  $J = 1.2$  Hz, 2H), 4.31-4.23 (m, 2H), 4.00-3.95 (m, 2H), 3.81-3.73 (m, 4H), 3.73-3.68 (m, 2H), 3.62-3.55 (m, 2H), 1.60 (s, 6H), 0.89 (s, 9H), 0.06 (s, 6H).

**2-(2-(2-((2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxin-9-yl)oxy)ethoxy)ethoxy)ethan-1-ol (22)**



To a solution of **21** (1.83 g, 3.84 mmol) in 40 mL of dry THF, 1 M TBAF solution in THF (4.61 mL, 4.61 mmol) was added and the reaction mixture was stirred at room temperature overnight. After removing the solvent under reduced pressure, the residue was dissolved in EtOAc and washed twice with 10 % citric acid solution and once with brine. The organic phase was dried over  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. Purification by column chromatography on silica (pentane:EtOAc, 5:1  $\rightarrow$  1:1, v/v) afforded **22** as a pale yellow oil (1.21 g, 87 %).

$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 7.66 (s, 1H), 7.42 (br, 1H), 7.30 (d,  $J = 8.3$  Hz, 1H), 7.22-7.17 (m, 1H), 6.72 (dd,  $J = 7.5, 0.8$  Hz, 1H), 5.06 (d,  $J = 1.2$  Hz, 2H), 4.03-4.26 (m, 2H), 4.00-3.96 (m, 2H), 3.85-3.71 (m, 6H), 3.66-3.62 (m, 2H), 1.60 (s, 6H).

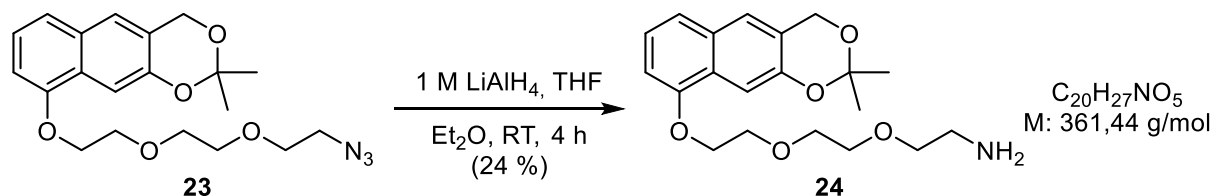
**9-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)-2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxine**
**(23)**<sup>[137a]</sup>


*p*-Toluenesulfonyl chloride (0.66 g, 3.48 mmol) and Et<sub>3</sub>N (0.81 mL, 5.80 mmol) were stepwisely added to the solution of **22** (1.05 g, 2.90 mmol) in 20 mL of dry DCM and the mixture was stirred at room temperature overnight. After completion (as judged by TLC in pentane:EtOAc, 4:1, v/v), the reaction mixture was washed with brine and the organic phase dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under vacuum. Purification by column chromatography on silica (pentane:EtOAc, 4:1 → 1:1, v/v) afforded the tosylated alcohol (1.12 g, 75 %) as a pale yellow oil.

The tosylated alcohol (1.12 g, 2.17 mmol) was then directly dissolved in 15 mL of dry DMF, NaN<sub>3</sub> (0.17 g, 2.60 mmol) was added and the reaction mixture stirred at 60 °C overnight. The solvent was removed under reduced pressure, the residue dissolved in Et<sub>2</sub>O and washed with brine. The organic phase was dried over Na<sub>2</sub>SO<sub>4</sub> and removed under reduced pressure. Purification by column chromatography on silica (pentane:EtOAc, 6:4, v/v) afforded the azide **23** (0.77 mg, 92 %) as a pale yellow oil.

$R_f = 0.63$  (pentane:EtOAc, 6:4, v/v) [UV, CAM]

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ/ppm: 7.60 (s, 1H), 7.43 (s, 1H), 7.35 (d, *J* = 8.2 Hz, 1H), 7.22 (t, *J* = 7.9 Hz, 1H), 6.86 (d, *J* = 7.5 Hz, 1H), 5.04 (s, 2H), 4.25-4.20 (m, 2H), 3.92-3.85 (m, 2H), 3.71-3.67 (m, 2H), 3.71-3.67 (m, 2H), 3.66-3.60 (m, 4H), 3.42-3.37 (m, 2H), 1.52 (s, 6H).

**2-(2-(2-((2,2-dimethyl-4H-naphtho[2,3-d][1,3]dioxin-9-yl)oxy)ethoxy)ethoxy)ethan-1-amine (24)**<sup>[137a]</sup>


Azide **23** (0.20 g, 0.52 mmol) was dissolved in 30 mL of dry Et<sub>2</sub>O and 1 M LiAlH<sub>4</sub> solution in THF (1.00 mL, 1.00 mmol) was dropwisely added at 0 °C. Stirring was continued at room temperature for 4 h. The reaction was then quenched with 10 % citric acid solution and the

aqueous phase washed a couple of times with Et<sub>2</sub>O. The combined organic phases were washed with brine, dried over MgSO<sub>4</sub> and concentrated under reduced pressure to afford **24** (45 mg, 24 %) as a yellow oil.

**MS (ESI+):** m/z (%) 362.2 (100) [M+H]<sup>+</sup>.

#### ***o*NQMP-TAMRA**

Succinimidyl ester of 5-(and-6)-carboxytetramethylrhodamine (mixed isomers, 5(6)-TAMRA-SE, *Thermo Fisher Scientific*; 10.0 mg, 0.019 mmol) was added to the mixture of **24** (13.7 mg, 0.038 mmol) and DIPEA (33.0 μL, 0.190 mmol) in 2 mL of dry DMF and the reaction was stirred at room temperature overnight. The solvent was removed under reduced pressure and the residue thus obtained dissolved in a mixture containing 5 mL of water, 1 mL ACN and 1 mL formic acid. The solution was stirred for 2 h at room temperature in order to remove the acetonide protecting group. Purification *via* RP-HPLC (C18, 250 × 10 mm, 4 mL/min, 20-80 % ACN in 15 min) and subsequent lyophilisation of the combined product-containing fractions afforded ***o*NQMP-TAMRA** (9.7 mg, 68 %) as a purple solid. A 10 mM stock solution of ***o*NQMP-TAMRA** used for labelling experiments was prepared in 50 % aqueous ACN, v/v.

**MS (ESI+):** m/z (%) 734.2 (100) [M+H]<sup>+</sup>; **(ESI-):** m/z (%) 732.2 (100).

#### ***o*NQMP-biotin**

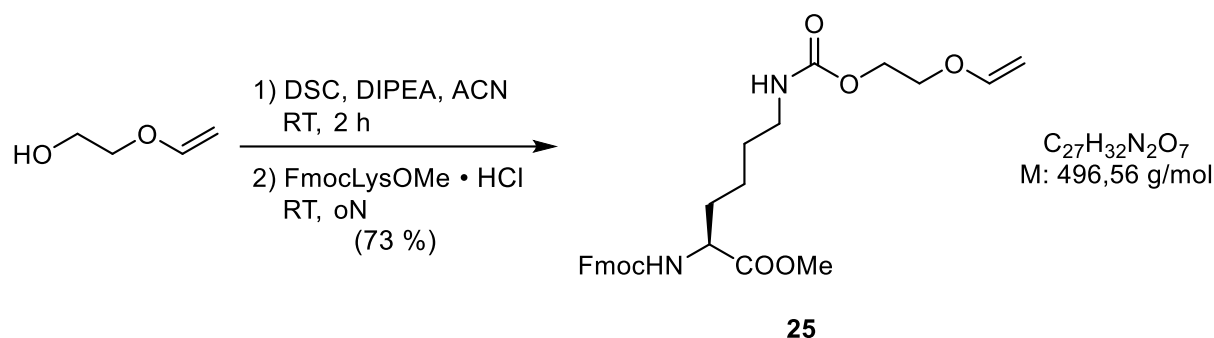
*N*-hydroxysuccinimidyl ester of PEGylated biotin (NHS-PEG<sub>4</sub>-biotin, *Thermo Fisher Scientific*; 10.0 mg, 0.017 mmol) was added to the mixture of **24** (12.3 mg, 0.034 mmol) and DIPEA (29.6 μL, 0.170 mmol) in 2 mL of dry DMF and the reaction was stirred at room temperature overnight. The solvent was removed under reduced pressure and the residue thus obtained dissolved in a mixture containing 5 mL of water, 1 mL ACN and 1 mL formic acid. The solution was stirred for 2 h at room temperature in order to remove the acetonide protecting group. Purification *via* RP-HPLC (C18, 250 × 10 mm, 4 mL/min, 20-80 % ACN in 15 min) and subsequent lyophilisation of the combined product-containing fractions afforded ***o*NQMP-biotin** (11 mg, 54 %) as a white solid. A 10 mM stock solution of ***o*NQMP-biotin** used for labelling experiments was prepared in 10 % ACN/water mixture, v/v.

**MS (ESI+):** m/z (%) 795.3 (100) [M+H]<sup>+</sup>, 817.3 (55) [M+Na]<sup>+</sup>; **(ESI-):** m/z (%) 793.2 (100) [M-H]<sup>-</sup>.

---

**Synthesis of *N*<sup>6</sup>-((2-(vinylloxy)ethoxy)carbonyl)-L-lysinate (VyOK)**

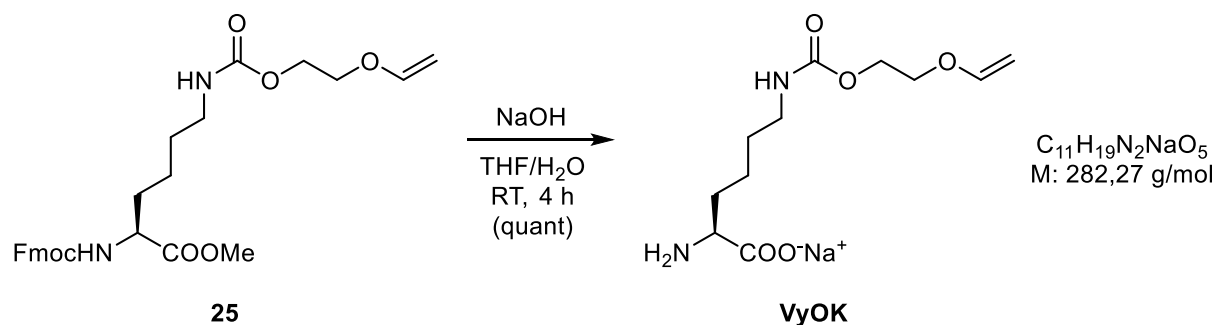
**methyl *N*<sup>2</sup>-(9-fluorenylmethoxycarbonyl)-*N*<sup>6</sup>-((2-(vinylloxy)ethoxy)carbonyl)-L-lysinate (25)**



DSC (4.36 g, 17.0 mmol) and DIPEA (5.93 mL, 34.0 mmol) were added to the solution of ethylene glycol vinyl ether (1.00 g, 11.3 mmol) in 100 mL of ACN and the mixture was stirred at room temperature for 2 h, until TLC (pentane:Et<sub>2</sub>O, 1:1, v/v) showed that all of the starting material had been consumed. Fmoc-Lys-OMe × HCl (5.19 g, 12.4 mmol) was then added to the mixture and stirring was continued overnight. After evaporation of the solvent, the oily residue was dissolved in DCM and washed with 10 % citric acid, satur. NaHCO<sub>3</sub> solution and brine. The organic phase was dried over MgSO<sub>4</sub> and concentrated under reduced pressure. Purification by column chromatography on silica (packed in DCM and washed with 2 % MeOH/DCM) afforded **25** (4.10 g, 73 %) as a white foam.

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ/ppm: 7.90 (d, *J* = 7.5 Hz, 2H), 7.75 (d, *J* = 7.8 Hz, 1H), 7.71 (dd, *J* = 7.4, 2.6 Hz, 2H), 7.42 (t, *J* = 7.4 Hz, 2H), 7.36-7.30 (m, 2H), 7.23 (t, *J* = 5.7 Hz, 1H), 6.50 (dd, *J* = 14.2, 6.7 Hz, 1H), 4.32-4.28 (m, 2H), 4.26-4.17 (m, 2H), 4.16-4.10 (m, 2H), 4.02-3.95 (m, 2H), 3.85-3.79 (m, 2H), 3.62 (s, 3H), 2.99-2.91 (m, 2H), 1.72-1.54 (m, 2H), 1.43-1.21 (m, 4H).

**sodium *N*<sup>6</sup>-((2-(vinylloxy)ethoxy)carbonyl)-L-lysinate (VyOK)**

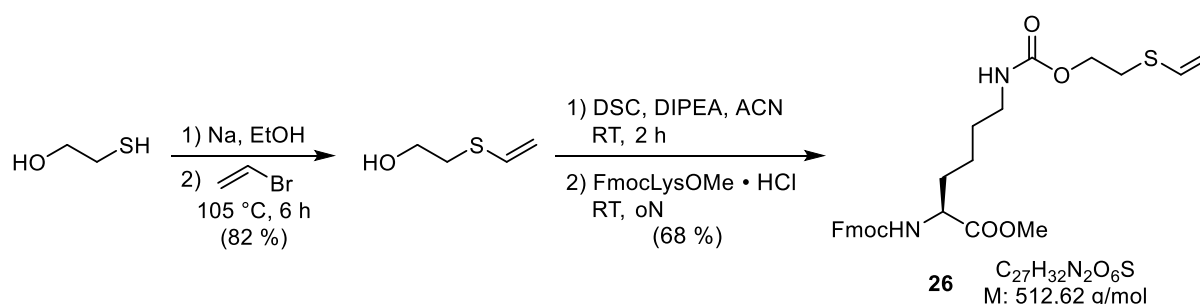




Protected amino acid **25** (0.22 g, 0.44 mmol) was dissolved in 5 mL of THF and 5 mL of 1 M NaOH solution was added. The mixture was stirred at room temperature for 4 h until it reached completion, as judged by TLC (5 % MeOH/DCM). After washing the mixture with Et<sub>2</sub>O, the aqueous phase was evaporated to afford **VyOK** (0.12 g, quant) as a white solid. VyOK was stored at -20 °C and dissolved as a 100 mM stock solution in distilled water before usage.

<sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ/ppm: 6.49 (dd, *J* = 14.2, 6.7 Hz, 1H), 4.36 (dd, *J* = 14.2, 2.2 Hz, 1H), 4.31-4.25 (m, 2H), 4.16 (dd, *J* = 6.7, 2.3 Hz, 1H), 4.01-3.95 (m, 2H), 3.21-3.19 (m, 1H), 3.12 (t, *J* = 6.8 Hz, 2H), 1.65-1.46 (m, 4H), 1.37-1.28 (m, 2H).

### Synthesis of *N*<sup>6</sup>-((2-(vinylthio)ethoxy)carbonyl)-L-lysine (**VySK**)



2-(vinylthio)ethan-1-ol was prepared according to a previously reported procedure. In an Ace pressure tube (heavy-wall borosilicate glass high pressure tube obtained from *Sigma Aldrich*) equipped with a magnetic stir bar, a solution of sodium ethanoate in EtOH was prepared by dissolving sodium (0.18 g, 8.00 mmol) in 4 mL of EtOH. BME (0.49 mL, 6.40 mmol) was subsequently added at room temperature and the mixture cooled to 0 °C, followed by addition of 1 M vinyl bromide solution in THF (10.2 mL, 10.2 mmol). The pressure tube was tightly closed with a teflon seal and heated to 105 °C for 6 h. The obtained brown suspension was concentrated under vacuum, followed by addition on water and extraction for several times with Et<sub>2</sub>O. Combined organic phases were dried under anhydrous MgSO<sub>4</sub> and the solvent was removed in vacuo. Crude 2-(vinylthio)ethan-1-ol (0.55 g, 82 %) was obtained as an orange-brown oil with satisfying purity as judged by NMR.

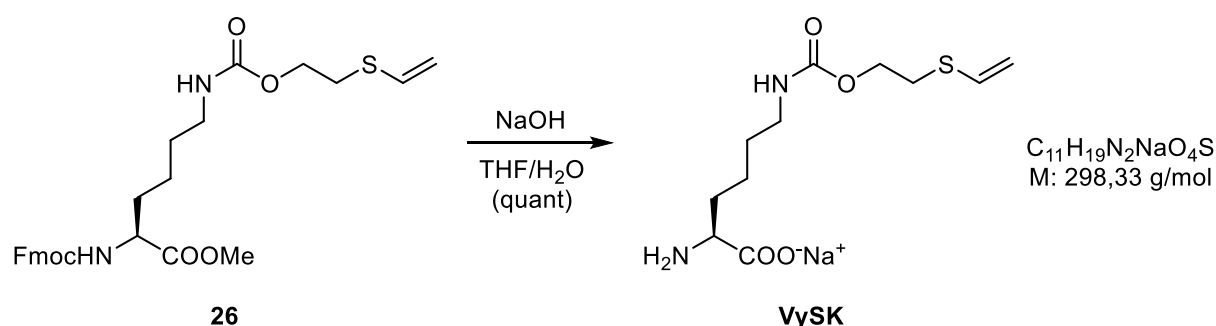
<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ/ppm: 6.32 (dd, *J* = 16.7, 10.0 Hz, 1H), 5.25 (d, *J* = 10.0 Hz, 1H), 5.23 (d, *J* = 16.7 Hz, 1H), 3.79 (t, *J* = 6.0 Hz, 2H), 2.91 (t, *J* = 6,0 Hz, 2H), 1.98 (br, 1H).

DSC (2.03 g, 7.92 mmol) and DIPEA (2.75 mL, 15.8 mmol) were added to the solution of 2-(vinylthio)ethan-1-ol (0.55 g, 5.28 mmol) in 40 mL of ACN and the mixture was stirred at room temperature for 2 h, until TLC (pentane:Et<sub>2</sub>O, 1:1, v/v) showed that all of the starting material had been consumed. Fmoc-Lys-OMe × HCl (2.43 g, 5.81 mmol) was then added to the mixture and stirring was continued overnight. After evaporation of the solvent, the oily residue was dissolved in DCM and washed with 10 % citric acid, satur. NaHCO<sub>3</sub> solution and brine. The organic phase was dried over MgSO<sub>4</sub> and concentrated under reduced pressure.

Purification by column chromatography on silica (packed in DCM and washed with 2 % MeOH/DCM) afforded **26** (1.84 g, 68 %) as a pale yellow foam.

**<sup>1</sup>H NMR** (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 7.89 (d, *J* = 7.5 Hz, 2H), 7.76 (d, *J* = 7.7 Hz, 1H), 7.74-7.69 (m, 2H), 7.42 (t, *J* = 7.4 Hz, 2H), 2.33 (t, *J* = 7.4 Hz, 2H), 7.20 (t, *J* = 5.6 Hz, 1H), 6.47 (dd, *J* = 16.7, 10.1 Hz, 1H), 5.20 (d, *J* = 10.1 Hz, 1H), 5.17 (d, *J* = 16.7 Hz, 1H), 4.34-4.27 (m, 2H), 4.25-4.19 (m, 1H), 4.07 (t, *J* = 6.7 Hz, 2H), 3.99 (dq, *J* = 8.9, 5.3 Hz, 1H), 3.62 (s, 3H), 2.99-2.89 (m, 4H), 1.72-1.55 (m, 2H), 1.43-1.25 (m, 4H).

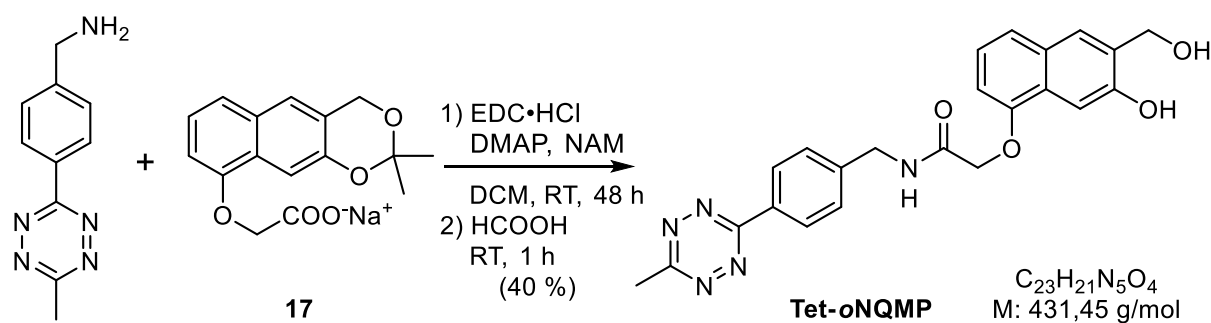
#### sodium *N*<sup>6</sup>-((2-(vinylthio)ethoxy)carbonyl)-L-lysinate (**VySK**)



Protected amino acid **26** (0.28 g, 0.55 mmol) was dissolved in 5 mL of THF and 5 mL of 1 M NaOH solution was added. The mixture was stirred at room temperature for 4 h until it reached completion, as judged by TLC (5 % MeOH/DCM). After washing the mixture with Et<sub>2</sub>O, the aqueous phase was evaporated to afford **VySK** (0.16 g, quant) as a pale yellow solid. **VySK** was stored at -20 °C and dissolved as a 100 mM stock solution in 1 M NaOH before usage.

**<sup>1</sup>H NMR** (500 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 6.43 (dd, *J* = 16.8, 10.2 Hz, 1H), 5.30 (d, *J* = 10.2 Hz, 1H), 5.38 (d, *J* = 16.8 Hz, 1H), 4.25 (t, *J* = 6.2 Hz, 2H), 3.23-3.19 (m, 1H), 3.11 (obs t, *J* = 6.8 Hz, 2H), 3.00 (obs t, *J* = 6.4 Hz, 2H), 1.65-1.45 (m, 4H), 1.36-1.28 (m, 2H).

#### Synthesis of Tet-*o*NQMP



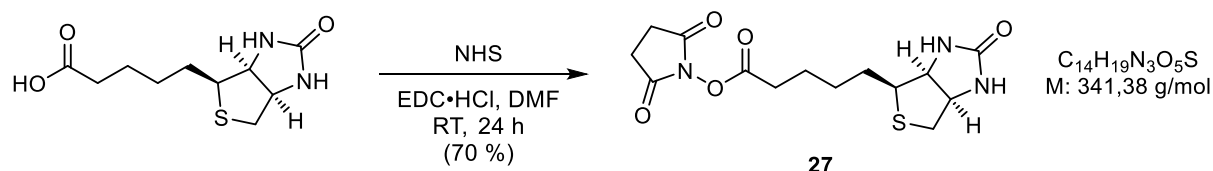
*o*NQMP **17** (76.0 mg, 0.025 mmol), DMAP (15.0 mg, 0.012 mmol) and EDC × HCl (70.5 mg, 0.037 mmol) were suspended in 5 mL of dry DCM, followed by addition of NEM (62.0

$\mu\text{L}$ , 0.049 mmol) and (4-(6-methyl-1,2,4,5-tetrazin-3-yl)phenyl)methanamine (64.1 mg, 0.027 mmol, courtesy of Susanne Mayer). After stirring the mixture for 48 h at room temperature, it was diluted with DCM and washed with satur.  $\text{NaHCO}_3$  solution and brine. The organic phase was dried over  $\text{Na}_2\text{SO}_4$  and concentrated under reduced pressure. Purification by column chromatography on silica (5 % MeOH/DCM, v/v) afforded a pink solid which was directly dissolved in 4 mL of THF. Formic acid (1 mL) was added to this solution, and the mixture was stirred at room temperature for 1 h. Removal of the solvent and subsequent trituration with  $\text{Et}_2\text{O}$  afforded **Tet-*o*NQMP** (43.1 mg, 40 %) as a pink solid. A 50 mM stock solution of Tet-*o*NQMP used for labelling experiments was prepared in DMSO.

**MS (ESI+):**  $m/z$  (%) 432.1 (100)  $[\text{M}+\text{H}]^+$

### Synthesis of biotinylated vinyl ether probe (VvO-biotin)

#### Succinimidyl ester of D-biotin (**27**)



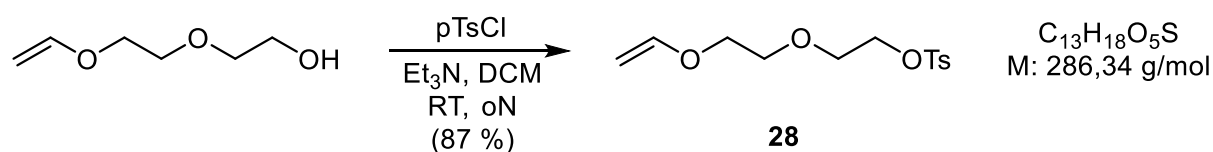
*N*-hydroxysuccinimide (0.35 g, 3.08 mmol) and  $\text{EDC} \times \text{HCl}$  (0.47 g, 2.46 mmol) were added to the solution of D-biotin (0.50 g, 2.05 mmol) in 30 mL of dry DMF and the reaction mixture was stirred at room temperature for 24 h. DMF was then removed under reduced pressure and the obtained white residue resuspended in EtOH:HOAc:H<sub>2</sub>O (95:1:4, v/v/v) mixture. The solid was filtered off, washed with the same solvent mixture (100 mL) and dried under vacuum. Succinimidyl ester **27** was afforded as a white solid (0.49 g, 70 %) and used in the following step without further purification.

**<sup>1</sup>H NMR** (300 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 6.41 (br, 1H), 6.35 (br, 1H), 4.35-4.26 (m, 1H), 4.15 (ddd,  $J = 7.5, 4.3, 1.8$  Hz, 1H), 3.15-3.06 (m, 1H), 2.87-2.84 (m, 1H), 2.81 (br, 4H), 2.74-2.55 (m, 3H), 1.72-1.34 (m, 6H).

**<sup>13</sup>C NMR** (75 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 170.2 (2C), 168.9, 162.7, 61.0, 59.2, 55.2, 39.9, 30.0, 27.8, 27.6, 25.4 (2C), 24.3.

**MS (ESI+):**  $m/z$  (%) 342.1 (100)  $[\text{M}+\text{H}]^+$

#### 2-(2-(vinylloxy)ethoxy)ethyl 4-methylbenzenesulfonate (**28**)

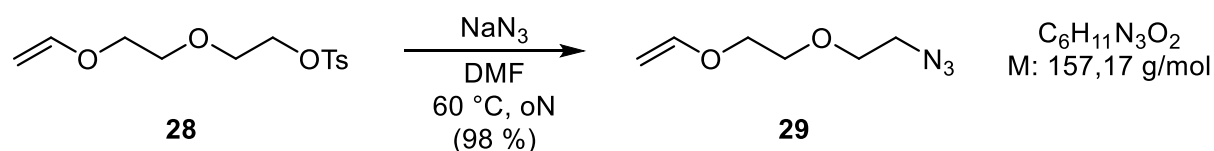


Et<sub>3</sub>N (4.22 mL, 30.2 mmol) and *p*-toluenesulfonyl chloride (3.46 g, 18.1 mmol) were carefully added to the solution of di(ethylene glycol) vinyl ether (2.00 g, 15.1 mmol) in 25 mL of dry DCM and the reaction mixture was stirred at room temperature overnight. The mixture was then washed with satur. NaHCO<sub>3</sub> solution and brine and DCM was removed under reduced pressure. Purification of the obtained residue by column chromatography on silica (DCM) afforded **28** (3.76 g, 87 %) as a pale yellow oil.

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ/ppm: 7.81-7.75 (m, 2H), 7.50-7.45 (m, 2H), 6.47 (dd, *J* = 14.3, 6.8 Hz, 1H), 4.16 (dd, *J* = 14.3, 1.8 Hz, 1H), 4.13-4.10 (m, 2H), 3.96 (dd, *J* = 6.7, 1.8 Hz, 1H), 3.72-3.67 (m, 2H), 3.60-3.57 (m, 2H), 3.55-3.52 (m, 2H), 2.41 (s, 3H).

MS (ESI+): *m/z* (%) 309.1 (100) [M+Na]<sup>+</sup>, 261.1 (38) [M+H]<sup>+</sup>

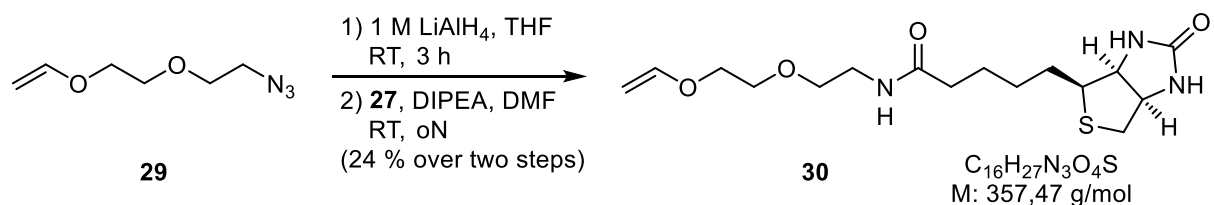
#### (2-(2-azidoethoxy)ethoxy)ethene (**29**)



Sodium azide (0.54 g, 8.38 mmol) was added to the solution of **28** (2.00 g, 6.98 mmol) in 15 mL of dry DMF and the mixture was stirred at 60 °C overnight. After cooling to ambient temperature, DMF was removed under reduced pressure and the obtained residue suspended in Et<sub>2</sub>O. The suspension was washed with brine, while the organic phase was separated, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Purification by column chromatography on silica (EtOAc:pentane, 4:6, v/v) afforded **29** (1.08 g, 98 %) as a white solid.

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ/ppm: 6.49 (dd, *J* = 14.3, 6.8 Hz, 1H), 4.20 (dd, *J* = 14.3, 2.2 Hz, 1H), 4.02 (dd, *J* = 6.8, 2.2 Hz, 1H), 3.89-3.83 (m, 2H), 3.77-3.72 (m, 2H), 3.72-3.66 (m, 2H), 3.44-3.36 (m, 2H).

#### VyO-biotin probe (**30**)



Azide **29** (0.56 g, 3.56 mmol) was dissolved in 3 mL of THF and 1 M LiAlH<sub>4</sub> solution in THF (6.77 mL, 6.77 mmol) was added at 0 °C. The mixture was allowed to warm up to room temperature and was stirred for 3 h. Afterwards, the mixture was quenched with 10 % NaOH (v/v) solution and the extracted several times with Et<sub>2</sub>O. The organic phase was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and removed under reduced pressure to afford the desired amine (0.16 g,

34 %) as a white solid. The amine was directly added to the mixture of 27 (0.21 g, 0.63 mmol) and DIPEA (0.44 mL, 2.51 mmol) in 5 mL of dry DMF. The reaction mixture was left stirring at room temperature overnight, after which the solvent was removed under reduced pressure and the residue purified *via* RP-HPLC (C18, 250 × 21.2 mm, 10 mL/min, 20-80 % ACN in 15 min). Lyophilisation of the combined product-containing fractions afforded the VyO-biotin probe **30** (0.45 g, 70 %) as a white solid.

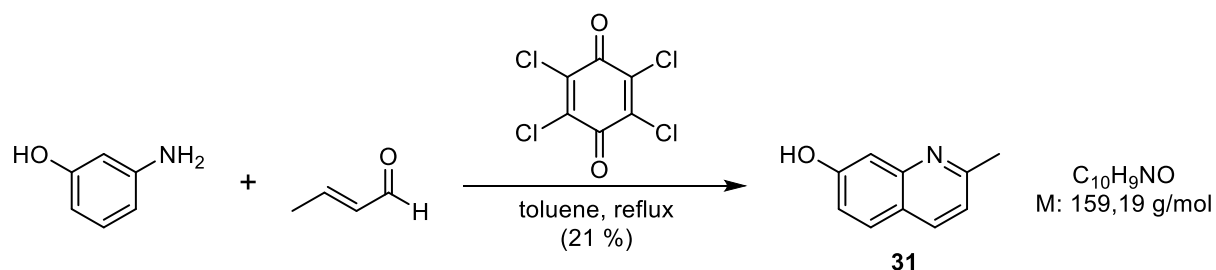
**MS (ESI+):** m/z (%) 358.2 [M+H]<sup>+</sup>.

#### **Synthesis of BCNK-*o*NQMP conjugate (31)**

BCNK (23.5 mg, 0.073 mmol) was dissolved in 2 mL of 50 % aqueous ACN (a little bit HCOOH was also added to the mixture until BCNK completely had dissolved) and 50 mM Tet-*o*NQMP stock in DMSO (14.5 mg, 0.034 mmol) was added. The mixture was incubated at 37 °C with shaking and the reaction progress was monitored *via* LC-MS, until the tetrazine peak was not observed anymore. The reaction mixture was concentrated under reduced pressure and the residue purified *via* RP-HPLC (C18, 250 × 21.2 mm, 10 mL/min, 20-80 % ACN in 15 min). Lyophilisation of the combined product-containing fractions afforded the BCNK-*o*NQMP **31** (19.3 mg, 79 %) as a pale yellow solid.

#### 4.3.3.2. *ortho*-Quinolinone quinone methide (*o*QQM) derivatives and related compounds

#### **Synthesis of 2-methylquinolin-7-ol (31) – Procedure A<sup>[200]</sup>**



3-aminophenol (1.00 g, 9.16 mmol) was suspended in conc. HCl (2.60 mL), followed by addition of *p*-chloranil (2.34 g, 9.16 mmol) and 3.20 mL of 1-butanol, and the reaction mixture was heated to reflux, during which time a colour change from bright yellow to dark brown was observed. Crotonaldehyde (predominantly *trans*, 1.00 mL, 11.91 mmol) dissolved in 0.40 mL of 1-butanol was then dropwisely added over 20 minutes and reflux was continued for further 30 minutes. After cooling to room temperature, the mixture was concentrated under reduced pressure and the obtained dark brown oily residue suspended in water. The suspension was first extracted a couple of times with Et<sub>2</sub>O, the aqueous phase was then neutralized with 1 M NaOH solution and extracted several times with EtOAc (until the organic phase was not yellow in colour anymore). The combined EtOAc extracts were dried over MgSO<sub>4</sub> and concentrated *in vacuo*. Purification by column chromatography on silica

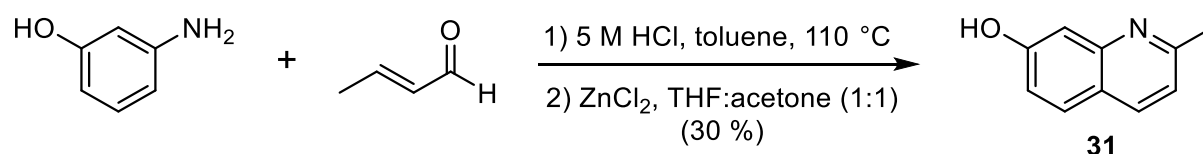
(EtOAc:pentane, 6:4, v/v) afforded 2-methylquinolin-7-ol **31** as a pale yellow solid (0.31 g, 21 %).

$R_f$  = 0.18 (EtOAc:pentane, 6:4) [UV, ninhydrin]

$^1\text{H NMR}$  (500 MHz, DMSO- $d_6$ )  $\delta$ /ppm: 10.03 (s, 1H), 8.06 (d,  $J$  = 8.3 Hz, 1H), 7.73 (d,  $J$  = 8.8 Hz, 1H), 7.15 (d,  $J$  = 8.3 Hz, 1H), 7.14 (d,  $J$  = 2.3 Hz, 1H), 7.06 (dd,  $J$  = 8.7, 2.4 Hz, 1H), 2.57 (s, 3H).

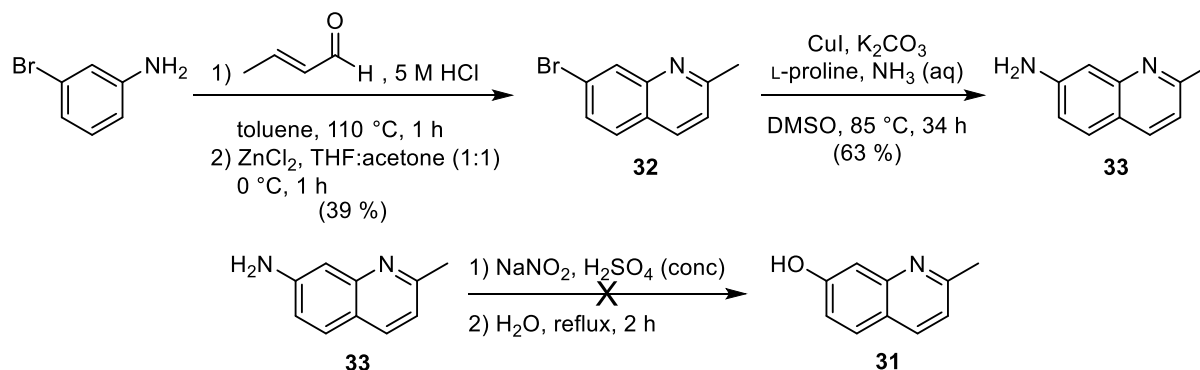
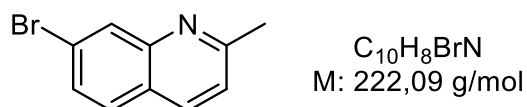
**MS (ESI +):**  $m/z$  (%) 160.0 (100)  $[\text{M}+\text{H}]^+$

#### **Synthesis of 2-methylquinolin-7-ol (31) – Procedure B**



3-aminophenol (9.39 g, 86 mmol) was dissolved in 60 mL of 5 M HCl solution, and the reaction mixture was heated to 110 °C. Crotonaldehyde (predominantly *trans*, 8.60 mL, 103 mmol) dissolved in 6 mL toluene was then slowly added over 45 minutes and the stirring continued at 110 °C for an hour, during which time a colour change from light yellow to dark orange was observed. The mixture was then cooled down to 60 °C, zinc chloride (11.7 g, 86 mmol) dissolved in 30 mL of THF/acetone (1:1, v/v) was slowly added and the reaction mixture was cooled on ice for an hour, during which time two phases had separated. Addition of concentrated NH<sub>3</sub> led to precipitation of a gummy solid which was further purified by column chromatography on silica (dry load; the column was packed in EtOAc/pentane, 7:3, v/v and eluted first with EtOAc/pentane, 8:2, v/v and then with EtOAc). Despite of smearing on the column, **31** was eventually obtained as a pale yellow solid (4.11 g, 30 %).

**MS (ESI +):**  $m/z$  (%) 160.0 (100)  $[\text{M}+\text{H}]^+$

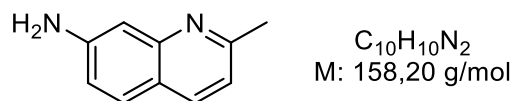
**Unsuccessful synthesis of 2-methylquinolin-7-ol (31) – Procedure C****2-methyl-7-bromoquinoline (32)<sup>[240]</sup>**

3-bromoaniline (9.40 mL, 86.0 mmol) was suspended in 60 mL of 5 M HCl solution and the suspension was heated at 110 °C until the brown solid had completely dissolved. Crotonaldehyde (predominantly *trans*, 8.60 mL, 103 mmol) dissolved in 6 mL of toluene was added dropwise to the mixture over a period of 1 hour, during which time a colour change from dark brown to orange was observed. The mixture was heated at 110 °C for a further hour before cooling down to 60 °C. Zinc chloride (11.7 g, 86.0 mmol) dissolved in 30 mL of THF:acetone (1:1, v/v) mixture was then slowly added and the reaction mixture cooled on ice for 1 hour. The precipitated solid was filtered off and sequentially washed with cold 5 M HCl solution (100 mL), THF (100 mL), isopropanol (100 mL) and Et<sub>2</sub>O (100 mL). Afterwards, the solid was dissolved in 25 % NH<sub>3</sub> solution (50 mL) and extracted with hexane (1 × 100 mL) and Et<sub>2</sub>O (3 × 50 mL). The combined organic phases were dried over MgSO<sub>4</sub> and concentrated under vacuum. Purification by column chromatography on silica (DCM to 2 % MeOH/DCM, v/v) afforded bromoquinoline **32** as a light brown solid (4.73 g, 39 %).

$R_f = 0.33$  (DCM) [UV, CAM]

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$ /ppm: 8.26 (d,  $J = 8.4$  Hz, 1H, H-4), 8.11 (d,  $J = 1.9$  Hz, 1H, H-8), 7.88 (d,  $J = 8.7$  Hz, 1H, H-5), 7.66 (dd,  $J = 8.7, 2.0$  Hz, 1H, H-6), 7.45 (d,  $J = 8.4$  Hz, 1H, H-3), 2.65 (s, 3H, CH<sub>3</sub>).

MS (ESI +):  $m/z$  (%) 223.9, 225.9 (100) [M+H]<sup>+</sup>

**2-methylquinolin-7-amine (33)**<sup>[240]</sup>

2-methyl-7-bromoquinoline (1.00 g, 4.50 mmol) was dissolved in 6 mL of DMSO, followed by addition of CuI (0.24 g, 1.26 mmol), L-proline (0.29 g, 2.53 mmol) and  $K_2CO_3$  (2.62 g, 18.96 mmol). The mixture was stirred for 10 minutes at room temperature, after which 25 %  $NH_3$  solution (0.72 mL, 10.58 mmol) was added and the reaction mixture stirred at 85 °C for 34 hours. Afterwards, the mixture was partitioned between DCM (50 mL) and sat.  $NH_4Cl$  (50 mL). The aqueous layer was extracted with DCM ( $2 \times 50$  mL) and the combined organic phases were washed with water, dried over  $MgSO_4$  and concentrated under vacuum. Purification by column chromatography on silica (gradient from pentane:EtOAc, 1:3, v/v to EtOAc (100 %) and EtOAc:MeOH: $NEt_3$ , 9:1:0.01, v/v/v to eluate the product) afforded **33** as a brown crystalline solid (0.45 g, 63 %).

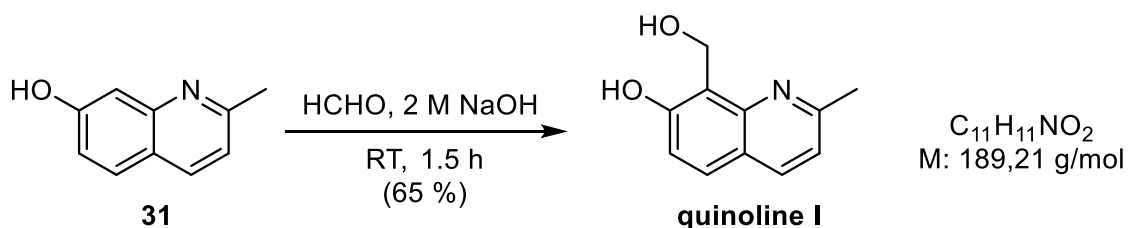
$R_f = 0.12$  (pentane:EtOAc, 1:3, v/v) [UV, ninhydrin]

$^1H$  NMR (500 MHz,  $CDCl_3$ )  $\delta$ /ppm: 7.92-7.84 (m, 1H), 7.56 (d,  $J = 8.63$  Hz, 1H), 7.20-7.16 (m, 1H), 7.03 (d,  $J = 8.24$  Hz, 1H), 6.94-6.89 (m, 1H), 4.04 (br, 2H), 2.69 (s, 3H).

MS (ESI +):  $m/z$  (%) 159.0 (100)  $[M+H]^+$

**2-methylquinolin-7-ol (31)**

2-methylquinolin-7-amine (450 mg, 2.84 mmol) was dissolved in 5 mL of conc.  $H_2SO_4$  and  $NaNO_2$  (235 mg, 3.41 mmol) was added at 0 °C in several portions. The mixture was stirred at room temperature for 1 hour and then poured on ice water (20 mL). The reaction mixture was then refluxed for 2 hours. After cooling to ambient temperature, the mixture was diluted with 1 M NaOH and the aqueous phase was extracted with DCM ( $3 \times 100$  mL). The combined organic phases were washed with brine, dried over  $MgSO_4$  and concentrated under reduced pressure. Unfortunately, no product was isolated in the reaction.

**8-(hydroxymethyl)-2-methylquinolin-7-ol (quinoline I)**<sup>[198a]</sup>

2-methylquinolin-7-ol (225 mg, 1.41 mmol) was dissolved in 1.8 mL of 2 M NaOH solution and formaldehyde (37 % in  $H_2O$ , 2.5 mL) was added. After stirring for 1.5 h at room



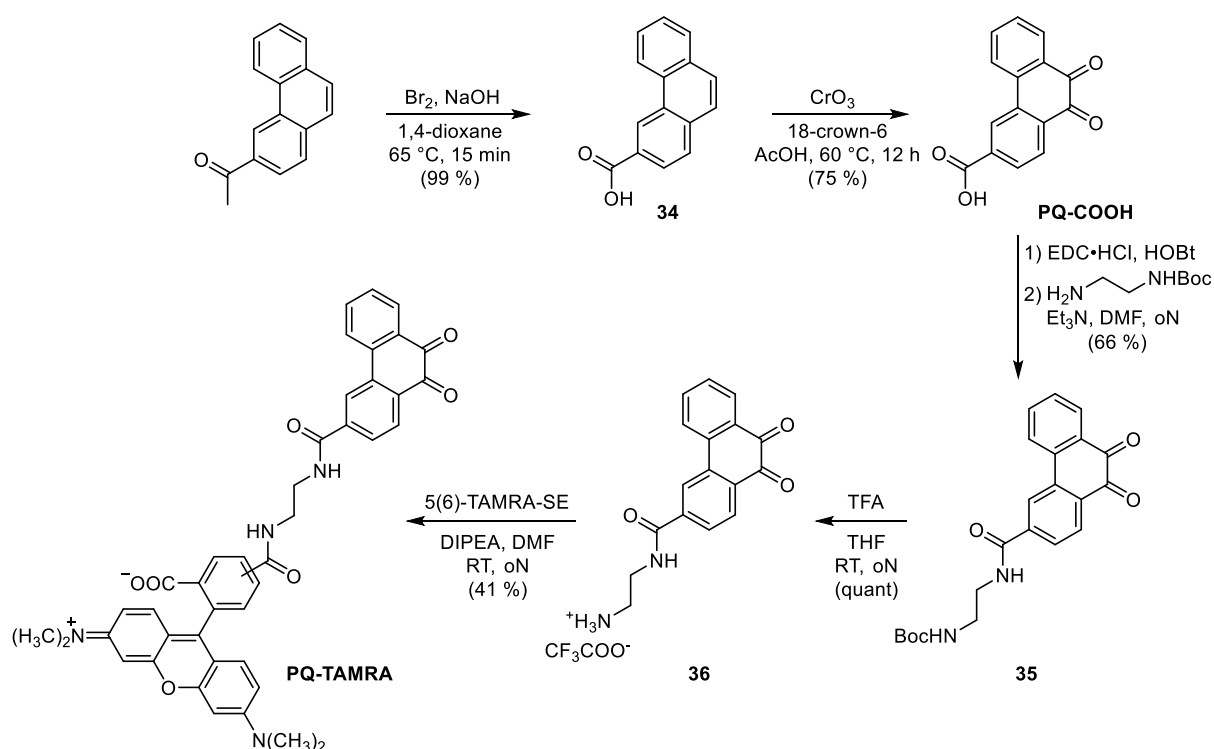
temperature, 1 mL of 5 M HCl solution was added and the stirring continued for further 3 minutes. The mixture was basified with sat. NaHCO<sub>3</sub> solution and extracted with *i*-PrOH/CHCl<sub>3</sub> (1:5, v/v) mixture (3 × 50 mL). The combined organic phases were washed with brine, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Purification by column chromatography on silica (packed in 25 % EtOAc/pentane and eluted with 25 to 50 % EtOAc/pentane) afforded quinoline I as a yellow solid (173 mg, 65 %).

<sup>1</sup>H NMR (500 MHz, MeOD) δ/ppm: 8.01 (d, *J* = 8.3 Hz, 1H), 7.62 (d, *J* = 8.8 Hz, 1H), 7.16 (d, *J* = 8.3 Hz, 1H), 7.10 (d, *J* = 8.8 Hz, 1H), 5.28 (s, 2H), 2.65 (s, 3H).

MS (ESI +): *m/z* (%) 190.1 (100) [M+H]<sup>+</sup>

#### 4.3.3.3. 9,10-phenanthrenequinones (PQs) derivatives and related compounds

##### Synthesis of TAMRA-PQ



##### 3-phenanthrenecarboxylic acid (34)

Bromine (4.70 mL, 91.0 mmol) was carefully added to an ice-cold solution of NaOH (12.7 g, 31.8 mmol) in 70 mL of water to form a basic solution of sodium hypobromite. The sodium hypobromite solution was then added dropwise over 15 minutes to a stirred solution of 3-acetylphenanthrene (4.32 g, 19.6 mmol) in 70 mL of 1,4-dioxane at 65 °C. Excess of sodium hypobromite was removed with addition of 20 mL of 10 % solution of sodium thiosulfate. The mixture was then diluted with 300 mL of water and approximately 100 mL of the solvent

were removed under reduced pressure. The reaction mixture was acidified with conc. HCl and the formed precipitate filtered off, washed with water and dried under vacuum to yield 3-phenanthrenecarboxylic acid **34** as an ochre-coloured solid (4.31 g, 99 %).

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 9.36 (s, 1H), 8.83 (d,  $J = 8.1$  Hz, 1H), 8.15 (dd,  $J = 8.3, 1.5$  Hz, 1H), 8.08 (d,  $J = 8.3$  Hz, 1H), 8.03 (dd,  $J = 7.7, 1.4$  Hz, 1H), 7.97 (d,  $J = 8.9$  Hz, 1H), 7.90 (d,  $J = 8.9$  Hz, 1H), 7.80-7.64 (m, 2H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 167.6, 134.3, 131.7, 129.8, 129.3, 129.1, 128.9, 128.84, 128.78, 127.6, 127.4, 126.6, 126.3, 124.5, 122.8.

#### **9,10-dioxo-9,10-dihydrophenanthrene-3-carboxylic acid (PQ-COOH)<sup>[241]</sup>**

A solution of 3-phenanthrenecarboxylic acid (1.00 g, 4.50 mmol) in 40 mL of acetic acid was added to the mixture of chromium trioxide (1.80 g, 18.0 mmol) and 18-crown-6 (0.16 g, 0.75 mmol) dissolved in 20 mL of acetic acid and 2 mL of water. After heating the reaction mixture at 60 °C for 12 h, the obtained green suspension was diluted with 50 mL of water and the solid filtered off, washed twice with acetic acid/water mixture (1:1, v/v) and twice with Et<sub>2</sub>O, and dried under vacuum. **PQ-COOH** was obtained as a bright orange solid (0.85 g, 75 %).

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 8.69-8.63 (m, 1H), 8.30 (d,  $J = 8.0$  Hz, 1H), 8.10 (d,  $J = 8.1$  Hz, 1H), 8.07-7.97 (m, 3H), 7.84-7.74 (m, 1H), 7.56 (t,  $J = 7.4$  Hz, 1H).

**<sup>13</sup>C NMR** (75 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 178.47, 178.40, 166.5, 136.3, 135.6, 135.4, 134.5, 134.0, 131.5, 129.7, 129.5, 129.4, 129.2, 124.8, 124.5.

#### ***tert*-butyl (2-(9,10-dioxo-9,10-dihydrophenanthrene-3-carboxamido)ethyl)carbamate (35)<sup>[138]</sup>**

PQ-COOH (100 mg, 0.39 mmol) was dissolved in 10 mL of dry DMF and EDC  $\times$  HCl (91.3 mg, 0.47 mmol), HOBt (64.4 mg, 0.47 mmol) and Et<sub>3</sub>N (166  $\mu$ L, 1.19 mmol) were added to the solution. After stirring at ambient temperature for 15 minutes, *tert*-butyl (2-aminoethyl)carbamate (75.4  $\mu$ L, 0.47 mmol) was added and the stirring was continued for 24 h. The reaction mixture was concentrated under reduced pressure and the crude residue was purified by column chromatography on silica (2 % MeOH/DCM) to afford the desired compound as a yellow solid (103 mg, 66 %).

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 8.86 (t,  $J = 5.4$  Hz, 1H), 8.64 (s, 1H), 8.38 (d,  $J = 7.9$  Hz, 1H), 8.12-8.04 (m, 2H), 7.97-7.91 (m, 1H), 7.84 (td,  $J = 7.8, 1.5$  Hz, 1H), 7.58 (td,  $J = 7.8, 0.8$  Hz, 1H), 6.98 (t,  $J = 5.5$  Hz, 1H), 3.41-3.33 (m, 2H), 3.16 (q,  $J = 6.2$  Hz, 2H), 1.38 (s, 9H).

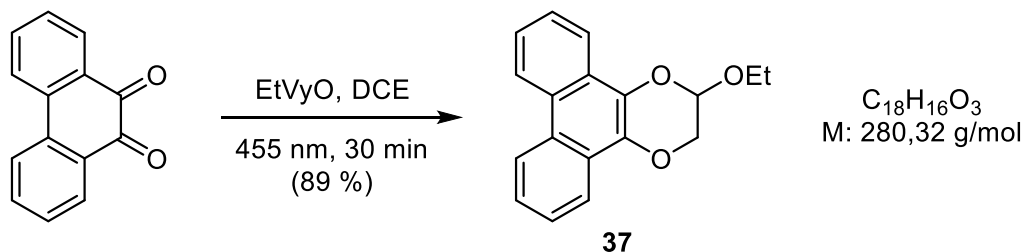
**TAMRA-PQ**

Boc-protected **35** (62 mg, 0.16 mmol) was dissolved in 4 mL of THF and 4 mL of TFA were added and the mixture stirred overnight. After the reaction was complete as judged by TLC (2 % MeOH/DCM) the reaction mixture was concentrated under reduced pressure to afford the crude **36** (~ 66 mg, quant) which was used in the next step without further purification.

**MS (ESI+):** m/z (%) 295.1 (100) [M+H]<sup>+</sup>, 278.1, 611.1 (23) [2M+Na]<sup>+</sup>; **(ESI-):** m/z (%) 293.1 (100) [M-H]<sup>-</sup>

The Boc-deprotected **36** (7.7 mg, 0.019 mmol) was dissolved in 2 mL of dry DMF, followed by addition of DIPEA (16.5  $\mu$ L, 0.095 mmol) and 5(6)-TAMRA-SE (5.0 mg, 0.009 mmol). The mixture was stirred for 24 h at room temperature, after which it was concentrated under reduced pressure. The crude product was purified *via* preparative RP-HPLC (C18, 250  $\times$  10 mm, 4 mL/min, 20-80 % ACN in 15 min). The fractions containing the desired product were combined and lyophilised to afford **TAMRA-PQ** (5.6 mg, 41 %) as a purple solid. A 5.6 mM stock solution of TAMRA-PQ used for labelling experiments was prepared in 25 % 1x PBS (pH 7.4) in DMSO.

**MS (ESI+):** m/z (%) 723.2 (100) [M+H]<sup>+</sup>; **(ESI-):** m/z (%) 721.2 (100) [M-H]<sup>-</sup>

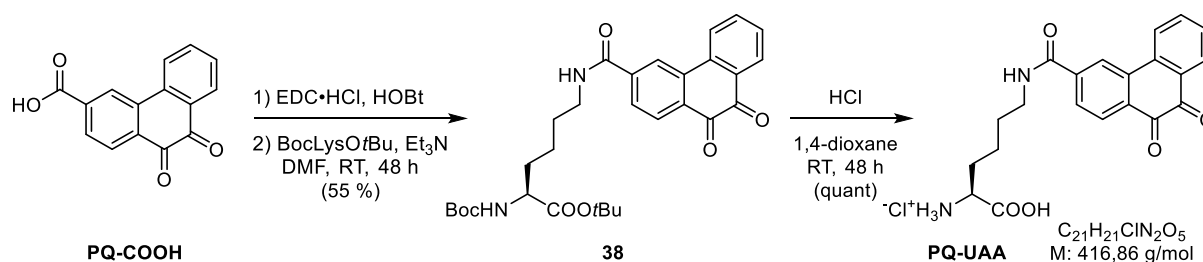
**2-ethoxy-2,3-dihydrophenanthro[9,10-*b*][1,4]dioxine (37)**

9,10-phenanthrene-9,10-dione (41.6 mg, 0.2 mmol) and ethyl vinyl ether (192  $\mu$ L, 2.0 mmol) were dissolved in 4 mL of dry dichloroethane and irradiated at 455 nm for 30 minutes. The solvent was evaporated and the obtained residue purified by silica gel column chromatography (pentane/EtOAc; 3:1, v/v) to afford **37** (49.9 mg, 89 %) as a pale yellow oil.

**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>)  $\delta$ /ppm: 8.67-8.56 (m, 2H), 8.26-8.16 (m, 2H), 7.66-7.53 (m, 4H), 5.52 (dd,  $J$  = 2.8, 1.8 Hz, 1H), 4.44 (dd,  $J$  = 11.0, 2.8 Hz, 1H), 4.35 (dd,  $J$  = 11.0, 1.8 Hz, 1H), 4.06 (dq,  $J$  = 9.7, 7.1 Hz, 1H), 3.84 (dq,  $J$  = 9.7, 7.1 Hz, 1H), 1.27 (t,  $J$  = 7.1 Hz, 3H).

**<sup>13</sup>C NMR** (75 MHz, CDCl<sub>3</sub>)  $\delta$ /ppm: 133.4, 130.7, 127.1, 126.9, 126.8, 126.5, 126.2, 125.2, 125.1, 122.7, 122.5, 121.2, 120.6, 94.4, 66.6, 64.8, 15.3.

**MS (ESI +):** m/z (%) 281.2 (100) [M+H]<sup>+</sup>

**Synthesis of PQ-bearing UAA (PQ-UAA)*****tert*-butyl *N*<sup>2</sup>-(*tert*-butoxycarbonyl)-*N*<sup>6</sup>-(9,10-dioxo-9,10-dihydrophenanthrene-3-carbonyl)-L-lysinate (38)**

PQ-COOH (0.51 g, 2.01 mmol) was dissolved in 20 mL DMF (with 2 mL of DMSO added due to its poor solubility) and EDC  $\times$  HCl (0.46 g, 2.42 mmol) and HOBT (0.31 g, 2.01 mmol) were added to the mixture. After stirring for 15 min at RT, Boc-Lys-OtBu (0.55 g, 2.22 mmol) and Et<sub>3</sub>N (0.56 mL, 4.03 mmol) were added and the stirring was continued for 48 h. The mixture was afterwards diluted with DCM and washed with 10 % LiCl, 10 % citric acid solution and brine. The organic phase was dried under anhydrous Na<sub>2</sub>SO<sub>4</sub> and the solvent removed under reduced pressure. Purification by column chromatography on silica (packed in DCM, eluted in 2 % MeOH/DCM, v/v) afforded **38** (0.59 mg, 55 %) as an ochre-coloured solid.

<sup>1</sup>H NMR (300 MHz, MeOD)  $\delta$ /ppm: 8.76 (t, *J* = 5.8 Hz, 1H), 8.53 (s, 1H), 8.22 (d, *J* = 8.1 Hz, 1H), 8.15-8.09 (m, 2H), 7.98 (s, 1H), 7.88-7.84 (m, 1H), 7.80-7.73 (m, 1H), 7.52 (t, *J* = 7.5 Hz, 1H), 4.02-3.95 (m, 1H), 3.50-3.40 (m, 2H), 1.89-1.35 (m, 24H).

MS (ESI<sup>+</sup>): *m/z* (%) 437.3 (100) [M-Boc+H]<sup>+</sup>, 537.3 (26) [M+H]<sup>+</sup>

***N*<sup>6</sup>-(9,10-dioxo-9,10-dihydrophenanthrene-3-carbonyl)-L-lysine hydrochloride (PQ-UAA)**

Protected amino acid **38** (0.56 g, 1.05 mmol) was dissolved in 20 mL of 1,4-dioxane and 20 mL of 4 M HCl solution in 1,4-dioxane were added. The mixture was stirred at room temperature for 48 h hours (the Boc group got readily deprotected, but not the *tert*-butyl ester). After completion, the reaction mixture was concentrated under reduced pressure and the residue triturated with cold Et<sub>2</sub>O to afford PQ-UAA (0.44 g, quant) as an ochre-coloured solid. PQ-UAA was stored at -20 °C and was dissolved as a 100 mM stock solution in DMSO before usage.

MS (ESI<sup>+</sup>): *m/z* (%) 381.1 (100) [M+H]<sup>+</sup>

#### 4.3.4. Biological methods and labelling experiments

##### **Synthetase screening for site-specific incorporation of *o*NQMP-UAA2 into sfGFP**

The synthetase screening experiment was performed as described in section 4.2.3. The active site mutations of the tested PylRS variants are listed in Table 4.2.

**Table 4.2.** Active site mutations of PylRS variants tested for site-specific incorporation of *o*NQMP-UAA2 into sfGFP.

<b>PylRS variant</b>	<b>Active site mutations</b>
D1	Y271M, L274G, C313A
D2	Y271V, L274M
D3	M241F, A267S, Y271C, L274M
D78	Y271A, L274M, C313A
MC1	Y271A, L274M
MF3	Y271G, C313V
MF18	L274A, N311Q, C313S
SM23	Y271A, Y349F
SM31	Y271A, Y349F, I378L
SM42	M241A, Y271A, L274V, C313V, M315Y, V370R
SM43	Y271A, L274V, C313V, M315Y, V370R
SM44	Y271A, L274V, C313V, M315Y, Y349F, V370R
SM45	M241A, Y271A, L274V, C313V, M315Y, Y349F, V370R

##### **Expression and purification of VyOK- and VySK-containing proteins**

Both VyOK and VySK were site-specifically incorporated into proteins of interest (sfGFP-N150TAG, Myo-S4TAG and Ub-K6TAG) by the same PylRS variant, which contains mutations: Y271M, L274A and C313A (referred to as D171). Detailed expression and purification protocol is described in Section 4.2.4.

##### **Light-induced *in vitro* labelling of purified sfGFP-VyOK with *o*NQMP-TAMRA**

A series of 200  $\mu$ L reaction mixtures containing purified sfGFP-N150VyOK-His6 (10  $\mu$ M final concentration) and 10, 50 or 100 eq of *o*NQMP-TAMRA (10 mM stock in 50 % aqueous ACN) in 1x PBS buffer pH 7.4 were incubated either for 1 h in the dark or were irradiated for 1 h in a 96-well plate with a broad-emitting 312 nm lamp from the top (Section 4.3.1). The samples were slowly stirred with a magnetic stir bar during irradiation. A further protein sample irradiated in the absence of *o*NQMP-TAMRA was used as a negative labelling control. All samples were mixed with 4x SDS loading buffer, cooked at 95  $^{\circ}$ C and analysed by SDS-PAGE and TAMRA in-gel fluorescence excited at 520 nm and captured using ImageQuant LAS 4000 with a 575DF20 Cy3 filter.

### **Light-induced *in vitro* labelling of purified sfGFP-VyOK with *o*NQMP-biotin**

A series of 200  $\mu\text{L}$  reaction mixtures containing purified sfGFP-N150VyOK-His6 or sfGFP-N150BocK-His6 (10  $\mu\text{M}$  final concentration) and 10, 50 or 100 eq of *o*NQMP-biotin (10 mM stock in 10 % ACN/water, v/v) in 1x PBS buffer pH 7.4 were incubated either for 1 h in the dark or were irradiated for 1 h in a 96-well plate with a broad-emitting 312 nm lamp from the top (Section 4.3.1). The samples were slowly stirred with a magnetic stir bar during irradiation. Furthermore, protein samples irradiated in the absence of *o*NQMP-biotin were used as negative labelling controls. All samples were mixed with 4x SDS loading buffer, cooked at 95  $^{\circ}\text{C}$  and analysed by SDS-PAGE and western blotting with HRP-streptavidin.

A similar experiment with 10  $\mu\text{M}$  of either sfGFP-N150VyOK-His6 or sfGFP-N150BocK-His6 and 100  $\mu\text{M}$  of *o*NQMP-biotin in PBS buffer pH 7.4 was performed with the broad-emitting 365 nm lamp as well, in order to compare both light sources (Section 4.3.1).

### **Light-mediated uncaging of unspecific *o*NQM-sfGFP adducts with excess of externally added vinyl ether**

Mixtures of 10  $\mu\text{M}$  of either sfGFP-N150VyOK-His6 or sfGFP-N150BocK-His6 and 100  $\mu\text{M}$  of *o*NQMP-biotin in 1x PBS buffer pH 7.4 were irradiated in a 96-well plate for 20 minutes at 312 nm, under slow stirring (Section 4.3.1). Samples for SDS-PAGE analysis were taken after 20 minutes, followed by addition of (EG)<sub>2</sub>VyO (500  $\mu\text{M}$  final concentration). Irradiation was continued for further 40 minutes, with samples taken after 20 minutes and at the end. Furthermore, protein samples irradiated in the absence of *o*NQMP-biotin were used as negative labelling controls. All samples were mixed with 4x SDS loading buffer, cooked at 95  $^{\circ}\text{C}$  and analysed by SDS-PAGE and western blotting with HRP-streptavidin.

### **Preparation of *o*NQMP-GFP conjugate and subsequent labelling with biotinylated vinyl ether derivates (VyO-biotin)**

Purified sfGFP-N150BCNK-His6 was diluted to the concentration of 20  $\mu\text{M}$  in 1x PBS buffer pH 7.4 (total volume of 250  $\mu\text{L}$ ) and 50 mM stock solution of Tet-*o*NQMP in DMSO (final concentration 400  $\mu\text{M}$ ) was added. The mixture was incubated at 37  $^{\circ}\text{C}$  with mild shaking for 2.5 h, after which the tetrazine ligation reaction was judged complete by LC-MS. Excess of Tet-*o*NQMP was removed by re-buffering the sfGFP-*o*NQMP conjugate into 1x PBS buffer with Amicon<sup>®</sup> Ultra 0.5 mL Centrifugal Filter units (10 kDa cut-off, Merck).

Subsequent labelling was performed by diluting the sfGFP-*o*NQMP conjugate or sfGFP-N150BCNK-His6 (negative control) to the concentration of 5  $\mu\text{M}$  in 1x PBS buffer pH 7.4 (total volume of 100  $\mu\text{L}$ ), followed by addition of 10 mM stock solution of VyO-biotin in 1x PBS buffer pH 7.4 (final concentration 500  $\mu\text{M}$ ). The reaction mixtures were either incubated in the dark for 45 minutes or irradiated at 312 nm for the same amount of time, with slow stirring. Afterwards, the samples were mixed with 4x SDS loading buffer, cooked at 95  $^{\circ}\text{C}$  and analysed by SDS-PAGE and western blotting with HRP-streptavidin. The reaction progress was furthermore monitored by LC-MS.

---

**Labelling of purified sfGFP-VyOK with dipyriddyltetrazine**

Purified sfGFP-N150VyOK-His6 was diluted to the concentration of 10  $\mu$ M in 1x PBS buffer pH 7.4 and 2 mM stock solution of dipyriddyltetrazine in DMSO (final concentration 100  $\mu$ M) was added. The samples were incubated at 37 °C and the reaction progress monitored by LC-MS.

**Labelling of purified VySK-containing proteins with thermally generated oQOM-1**

Purified sfGFP-N150VySK-His6 was diluted to the concentration of 0.1 mM in 1x PBS buffer pH 7.4 and 20 mM stock solution of quinoline I in DMSO (final concentration 5 mM or 10 mM) was added. The samples were incubated at 37 °C and the reaction progress monitored by LC-MS.

The labelling was performed on purified sfGFP-N150VyOK-His6, sfGFP-N150BocK-His6, Myo-S4VyOK-His6 and Ub-K6VyOK-His6 as well, under same conditions as described above.

**Visible-light-induced labelling of purified Ub-VyOK with PQ-TAMRA**

Purified Ub-K6VyOK-His6 was diluted to the concentration of 10  $\mu$ M in 1x PBS buffer pH 7.4 and 5.6 mM stock solution of TAMRA-PQ in 25 % 1x PBS (pH 7.4)/DMSO (final concentration 300  $\mu$ M) was added. The reaction mixtures were then irradiated with 1 W high power 450 nm and 3 W high power white light LEDs (Section 4.3.1). Samples were taken at different time points and analysed by SDS-PAGE, in-gel fluorescence and LC-MS. TAMRA in-gel fluorescence was excited at 520 nm and captured using ImageQuant LAS 4000 with a 575DF20 Cy3 filter.

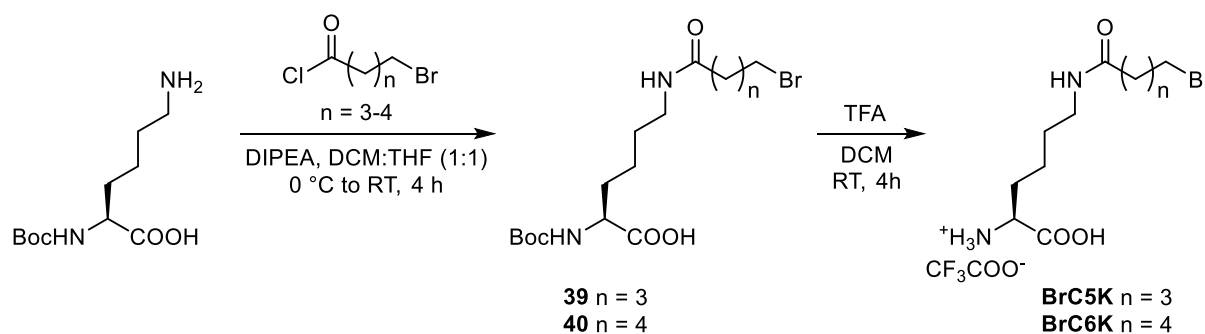
---

## 4.4. Supporting information for Chapter 3

### 4.4.1. Chemical synthesis

#### 4.4.1.1. Bromoalkyl-bearing UAAs

#### Synthesis of *N*<sup>6</sup>-(5-bromopentanoyl)-L-lysine (BrC5K) and *N*<sup>6</sup>-(6-bromohexanoyl)-L-lysine (BrC6K)



#### *N*<sup>6</sup>-(5-bromopentanoyl)-L-lysine trifluoroacetate (BrC5K)

Boc-Lys-OH (2.50 g, 10.2 mmol) was dissolved in 50 mL of dry THF:DCM (1:1, v/v) mixture and DIPEA (2.12 mL, 12.2 mmol) was added under inert atmosphere (Ar balloon). The mixture was cooled to 0 °C (ice/water bath) and 5-bromovaleryl chloride (1.77 mL, 13.2 mmol) was added dropwise. The reaction mixture was stirred in the melting ice bath for 4 h, after which it was diluted with ethyl acetate. The organics were washed four times with 10 % citric acid and twice with brine, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The obtained oily residue was purified by column chromatography on silica (1 → 5 % MeOH, 0.25 % AcOH in DCM) to yield **39** (2.10 g, 51 %) as a yellow oil which was used in the next step.

<sup>1</sup>H-NMR (500 MHz, DMSO-d<sub>6</sub>) δ/ppm 7.78 (t, *J* = 5.6 Hz, 1H), 7.01 (d, *J* = 8.0 Hz, 1H), 3.87-3.79 (m, 1H), 3.52 (t, *J* = 6.6 Hz, 2H), 3.04-2.96 (m, 2H), 2.08 (t, *J* = 7.4 Hz, 2H), 1.80-1.72 (m, 2H), 1.65-1.48 (m, 4H), 1.38 (s, 9H), 1.37-1.23 (m, 4H).

**MS (ESI<sup>+</sup>):** m/z (%) 309.1 (100) [M-Boc+H]<sup>+</sup>; **(ESI<sup>-</sup>):** m/z (%) 407.1 (100) [M-H]<sup>-</sup>

TFA (10 mL) was added dropwise to a solution of **39** (2.10 g, mmol) in 20 mL of dry DCM at 0 °C (ice/water bath) and the mixture was stirred until TLC showed complete conversion (after 2 hours the deprotection was complete, TLC in 7 % MeOH/DCM) and the mixture was concentrated under reduced pressure. The obtained residue was dissolved in filtered water and extracted four times with Et<sub>2</sub>O. The aqueous phase was evaporated under reduced pressure to give **BrC5K** trifluoroacetate (2.10 g, 97 %) as a yellow oil. BrC5K was stored at -20 °C and dissolved as a 100 mM stock solution in 1 M NaOH before usage.



**<sup>1</sup>H NMR** (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm 8.27 (br, 3H), 7.82 (t, *J* = 5.6 Hz, 1H), 3.91-3.81 (m, 1H), 3.52 (t, *J* = 6.7 Hz, 2H), 3.06-2.97 (m, 2H), 2.07 (t, *J* = 7.3 Hz, 2H), 1.83-1.51 (m, 6H), 1.48-1.21 (m, 4H).

**MS (ESI+):** *m/z* (%) 309.1 (100) [M + H]<sup>+</sup>

***N*<sup>6</sup>-(6-bromohexanoyl)-L-lysine trifluoroacetate (BrC6K)**

BrC6K was prepared according to a previously reported procedure.<sup>[163b]</sup>

Boc-Lys-OH (5.00 g, 20.3 mmol) was dissolved in 200 mL of dry THF:DCM (1:1, v/v) mixture (200 mL) and DIPEA (4.24 mL, 24.4 mmol) was added under inert atmosphere (Ar balloon). The mixture was cooled to 0 °C (ice/water bath) and 6-bromohexanoyl chloride (4.04 mL, 26.4 mmol) was added dropwise. The reaction mixture was stirred in the melting ice bath for 4 h, after which it was diluted with EtOAc. The organics were washed four times with 10 % citric acid and twice with brine, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The obtained oily residue was purified by column chromatography on silica (1 → 5 % MeOH, 0.25 % AcOH in DCM) to yield **40** (4.67 g, 54 %) as an orange oil which was used in the next step.

**<sup>1</sup>H NMR** (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm 7.76 (t, *J* = 5.7 Hz, 1H), 7.00 (d, *J* = 8.0 Hz, 1H), 3.81 (ddd, *J* = 9.5, 7.8, 4.7 Hz, 1H), 3.51 (t, *J* = 6.7 Hz, 2H), 3.04-2.95 (m, 2H), 2.04 (t, *J* = 7.4 Hz, 2H), 1.78 (dt, *J* = 14.3, 6.8 Hz, 2H), 1.69-1.45 (m, 4H), 1.37 (s, 9H), 1.36-1.21 (m, 6H).

**MS (ESI+):** *m/z* (%) 322.8 (100) [M-Boc+H]<sup>+</sup>; **(ESI-):** *m/z* (%) 421.1 (100) [M-H]<sup>-</sup>

TFA (25 mL) was added dropwise to a solution of **40** (4.67 g, 11.0 mmol) in 50 mL of DCM at 0 °C (ice/water bath) and the mixture was stirred until TLC showed complete conversion (after 2 hours the deprotection was complete, monitored by TLC in 7 % MeOH/DCM) and the mixture was concentrated under reduced pressure. The obtained residue was dissolved in filtered water and extracted four times with Et<sub>2</sub>O. The aqueous phase was evaporated under reduced pressure to give **BrC6K** trifluoroacetate (4.46 g, 92 %) as a yellow oil. BrC6K was stored at -20 °C and dissolved as a 100 mM stock solution in 1 M NaOH before usage.

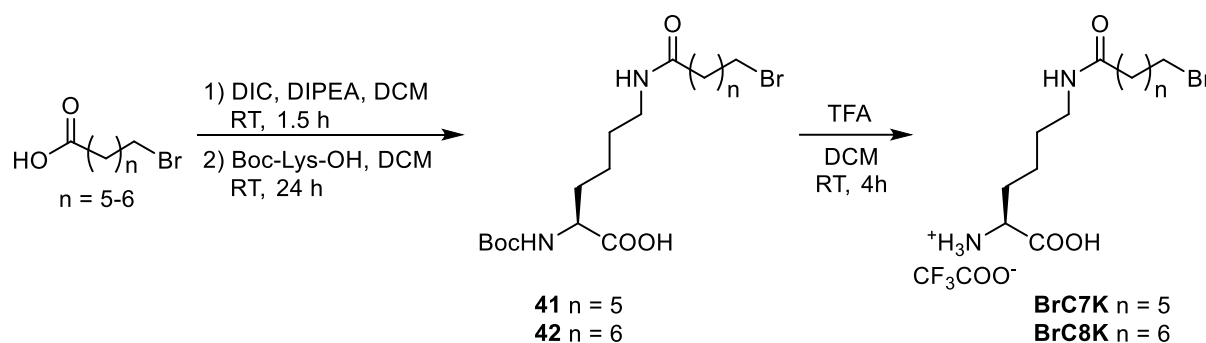
**<sup>1</sup>H NMR** (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm 8.21 (br, 3H), 7.77 (t, *J* = 5.6 Hz, 1H), 3.91-3.84 (m, 1H), 3.51 (t, *J* = 6.7 Hz, 2H), 3.06-2.97 (m, 2H), 2.04 (t, *J* = 7.4 Hz, 2H), 1.83-1.67 (m, 4H), 1.50 (quint, *J* = 7.5 Hz, 2H), 1.43-1.21 (m, 6H).

**<sup>13</sup>C NMR** (125 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm 171.9, 171.2, 51.9, 38.1, 35.3, 35.2, 32.1, 29.8, 28.8, 27.3, 24.5, 21.8

**MS (ESI+):** *m/z* (%) 323.1 (100) [M + H]<sup>+</sup>

---

### Synthesis of *N*<sup>6</sup>-(7-bromoheptanoyl)-L-lysine (BrC7K) and *N*<sup>6</sup>-(8-bromooctanoyl)-L-lysine (BrC8K)



#### *N*<sup>6</sup>-(7-bromoheptanoyl)-L-lysine trifluoroacetate (BrC7K)

DIPEA (2.83 mL, 16.0 mmol) and DIC (1.26 mL, 8.0 mmol) were added to a solution of 7-bromoheptanoic acid (3.40 g, 16.0 mmol) in 200 mL of dry DCM and the reaction mixture was allowed to stir for 1.5 h at room temperature to form the corresponding anhydride. Meanwhile, Boc-Lys-OH (2.0 g, 8.0 mmol) was suspended in 200 mL of dry DCM and the anhydride mixture was added dropwise to the prepared suspension. The reaction mixture was stirred at room temperature for 24 hours, after which it was washed four times with 10 % citric acid solution and twice with brine. The organic phase was dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated under reduced pressure. The obtained crude product was purified by column chromatography on silica (0 → 7 % MeOH, 0.25 % AcOH in DCM) to yield **41** (2.2 g, 64 %) as a peach-colored oil which was used in the next step.

**<sup>1</sup>H NMR** (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$ /ppm 7.73 (t,  $J = 5.6$  Hz, 1H), 7.00 (d,  $J = 8.0$  Hz, 1H), 3.85-3.77 (m, 1H), 3.51 (t,  $J = 6.7$  Hz, 2H), 3.04-2.95 (m, 2H), 2.03 (t,  $J = 7.4$  Hz, 2H), 1.78 (q,  $J = 6.8$  Hz, 2H), 1.68-1.52 (m, 2H), 1.48 (quint,  $J = 7.4$  Hz, 2H), 1.37 (s, 9H), 1.36-1.19 (m, 8H).

**MS (ESI+):**  $m/z$  (%) 336.8 (100)  $[\text{M}-\text{Boc}+\text{H}]^+$ ; **(ESI-):**  $m/z$  (%) 435.1 (100)  $[\text{M}-\text{H}]^-$

TFA (10 mL) was added dropwise to the solution of **41** (2.2 g, 5.0 mmol) in anhydrous DCM (20 mL) at 0 °C (ice/water bath) and the mixture was stirred until completion. After 2 hours, the deprotection was complete (monitored by TLC in 7 % MeOH/DCM) and the mixture was concentrated under reduced pressure. The obtained residue was washed several times with cold  $\text{Et}_2\text{O}$ , dissolved in filtrated water and lyophilized in order to give **BrC7K** trifluoroacetate (2.18 g, 96 %) as a white powder. BrC7K was stored at -20 °C and dissolved as a 100 mM stock solution in 1 M NaOH before usage.

**<sup>1</sup>H NMR** (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$ /ppm 8.29-8.17 (m, 3H), 7.76 (t,  $J = 5.7$  Hz, 1H), 3.91-3.83 (m, 1H), 3.51 (t,  $J = 6.7$  Hz, 2H), 3.05-2.97 (m, 2H), 2.03 (t,  $J = 7.5$  Hz, 2H), 1.84-1.68 (m, 4H), 1.48 (quint,  $J = 7.5$  Hz, 2H), 1.43-1.20 (m, 8H).

$^{13}\text{C}$  NMR (125 MHz, DMSO- $d_6$ )  $\delta$ /ppm 172.0, 171.2, 52.0, 38.1, 35.4, 35.2, 32.2, 29.8, 28.8, 27.8, 27.4, 25.2, 21.8.

**MS (ESI+):**  $m/z$  (%) 337.1 (100)  $[\text{M} + \text{H}]^+$

**$N^6$ -(8-bromooctanoyl)-L-lysine trifluoroacetate (BrC8K)**

DIPEA (0.781 mL, 4.5 mmol) and DIC (0.333 mL, 1.1 mmol) were added to a solution of 8-bromooctanoic acid (1.0 g, 4.5 mmol) in 75 mL of dry DCM and the reaction mixture was allowed to stir for 1.5 h at room temperature to form the corresponding anhydride. Meanwhile, Boc-Lys-OH (0.552 g, 2.2 mmol) was suspended in 75 mL of dry DCM and the anhydride mixture was added dropwise to the prepared suspension. The reaction mixture was stirred at room temperature for 24 hours, after which it was washed four times with 10 % citric acid and twice with brine. The organic phase was dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered and concentrated under reduced pressure. The obtained crude product was purified by column chromatography on silica (0  $\rightarrow$  7 % MeOH, 0.25 % AcOH in DCM) to yield **42** (0.71 g, 66 %) as an off-white oil which was used in the next step.

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$ /ppm 7.72 (t,  $J = 5.7$  Hz, 1H), 7.00 (d,  $J = 8.0$  Hz, 1H), 3.84-3.79 (m, 1H), 3.52 (t,  $J = 6.7$  Hz, 2H), 3.04-2.95 (m, 2H), 2.03 (t,  $J = 7.4$  Hz, 2H), 1.78 (quint,  $J = 7.1$  Hz, 2H), 1.67-1.51 (m, 2H), 1.47 (quint,  $J = 7.4$  Hz, 2H), 1.37 (s, 9H), 1.34-1.17 (m, 10H).

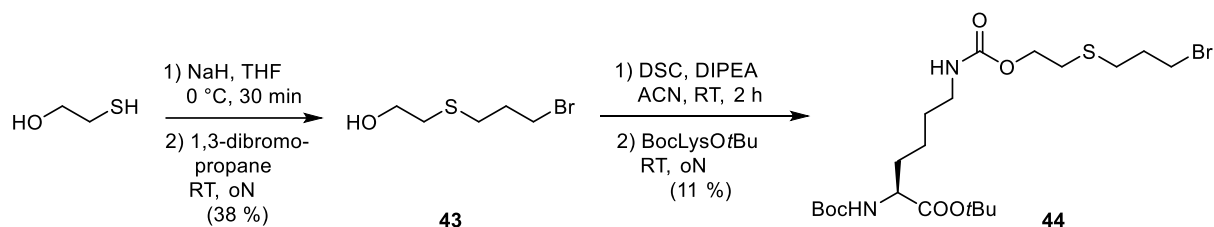
TFA (10 mL) was added dropwise to a solution of **42** (0.71 g, 1.6 mmol) in 20 mL of dry DCM at 0 °C (ice/water bath) and the mixture was stirred until completion. After 2 hours, the deprotection was complete (monitored by TLC in 7 % MeOH/DCM) and the mixture was concentrated under reduced pressure. The obtained residue was washed several times with cold  $\text{Et}_2\text{O}$  and dried under vacuum in order to give **BrC8K** trifluoroacetate (0.43 g, 41 %) as an off-white powder. BrC8K was stored at -20 °C and dissolved as a 90 mM stock solution in 2 M NaOH before usage.

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$ /ppm 8.24-8.10 (m, 3H), 7.76 (t,  $J = 5.7$  Hz, 1H), 3.91-3.83 (m, 1H), 3.50 (t,  $J = 6.7$  Hz, 2H), 3.07-2.96 (m, 2H), 2.03 (t,  $J = 7.5$  Hz, 2H), 1.82-1.66 (m, 4H), 1.47 (quint,  $J = 7.4$  Hz, 2H), 1.44-1.15 (m, 10H).

**MS (ESI+):**  $m/z$  (%) 351.1 (100)  $[\text{M} + \text{H}]^+$ .

---

## 4.4.1.2. Sulfoxide-containing UAAs

**Synthesis of *tert*-butyl *N*<sup>6</sup>-((2-((3-bromopropyl)thio)ethoxy)carbonyl)-*N*<sup>2</sup>-(*tert*-butoxy-carbonyl)-*L*-lysinate trifluoroacetate (BrSK × TFA)**

NaH (60 % suspension in mineral oil, 1.50 g, 64.0 mmol) was added to the solution of BME (4.46 mL, 64.0 mmol) in 250 mL THF at 0 °C (ice bath). After stirring at 0 °C for 30 minutes, 1,3-dibromopropane (19.5 mL, 192.0 mmol) was added and the mixture stirred at room temperature overnight. Afterwards, the solvent was removed under reduced pressure and the residue dissolved in DCM and washed with 10 % citric acid solution. The organic phase was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Purification of the obtained residue by column chromatography on silica (packed in DCM and eluted with 1 % MeOH/DCM, v/v) afforded **43** (4.79 g, 38 %) as a yellow oil.

The alcohol **43** (1.00 g, 5.02 mmol), DSC (1.54 g, 6.03 mmol) and DIPEA (0.87 mL, 5.02 mmol) were dissolved in 20 mL of ACN and the mixture was stirred at room temperature until all of the starting **43** was consumed, as judged by TLC (DCM). Boc-Lys-OH (1.48 g, 6.03 mmol) as well as more DIPEA (1.28 mL, 7.35 mmol) and 6 mL of dry DMF were then added and the mixture was stirred at room temperature overnight. The reaction mixture was washed several times with 10 % citric acid solution, 10 % LiCl solution and brine while the organic phase was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. Purification by column chromatography on silica (packed in DCM and eluted with 5 % MeOH/DCM, v/v) afforded protected BrSK **44** (0.25 g, 11 %) as a pale yellow foam.

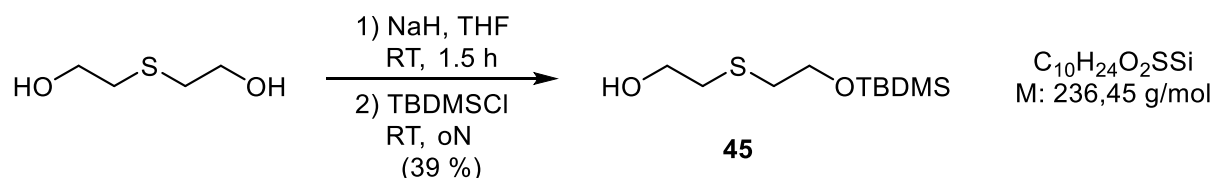
Treatment of **44** with mCPBA not only resulted with oxidation of the sulfide into sulfoxide, but with the nucleophilic substitution of the bromide with *m*-chlorobenzoic acid. Another approach of obtaining sulfoxide-containing alkyl bromide UAAs was therefore needed.

Deprotection of the UAA was performed by treating a solution of **44** (0.25 g, 0.53 mmol) in 4 mL of DCM with 2.5 mL of TFA. After stirring at room temperature for 4 h, the solvents were evaporated while the residue dissolved in water and washed a couple of times with cold Et<sub>2</sub>O. Lyophilisation of the aqueous phase afforded **BrSK** (0.26 g, quant) as an orange foam. A 100 mM stock solution in water was prepared for site-specific incorporation tests in *E. coli*.

**MS (ESI+):** m/z (%) 371.1/373.1 (100) [M+H]<sup>+</sup>

### Unsuccessful preparation of *N*<sup>6</sup>-((2-((2-bromoethyl)sulfinyl)ethoxy)carbonyl)-L-lysine (BrSOK)

#### 2-((2-((*tert*-butyldimethylsilyl)oxy)ethyl)thio)ethan-1-ol (**45**)

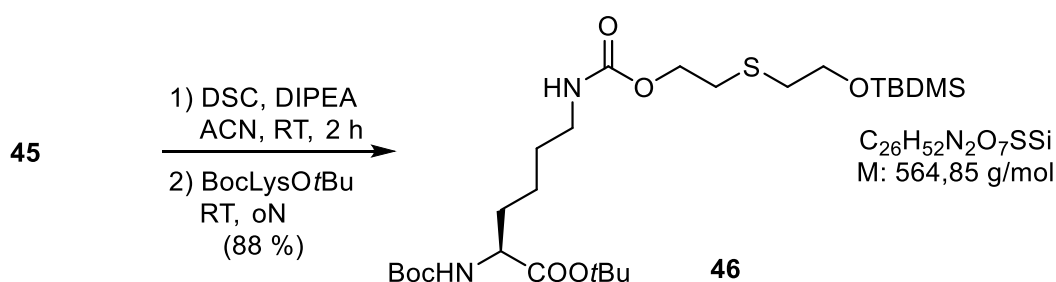


2,2'-thioethanol (5.80 mL, 58.0 mmol) was added to the suspension of NaH (60 % in mineral oil, 3.20 g) in 50 mL of dry THF and the mixture was stirred at room temperature for 1.5 h. After cooling to 0 °C (ice bath), TBDMSCl (8.60 g, 57.1 mmol) was slowly added to the reaction mixture in several portions and the stirring of the mixture continued overnight at room temperature. The reaction mixture was then quenched with 20 mL of 10 % Na<sub>2</sub>CO<sub>3</sub> solution and THF was removed under reduced pressure. The residue thus obtained was dissolved in Et<sub>2</sub>O and washed with 10 % citric acid and brine. The ether phase was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under vacuum. Purification by column chromatography on silica (pentane:EtOAc, 8:2, v/v) afforded **45** (5.28 g, 39 %) as a golden-yellow liquid.

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ/ppm: 3.78 (t, *J* = 6.6 Hz, 2H), 3.73 (t, *J* = 5.8 Hz, 2H), 2.76 (t, *J* = 5.8 Hz, 2H), 2.67 (t, *J* = 6.6 Hz, 2H), 2.56 (br, 1H), 0.90 (s, 9H), 0.07 (s, 6H).

#### *tert*-butyl

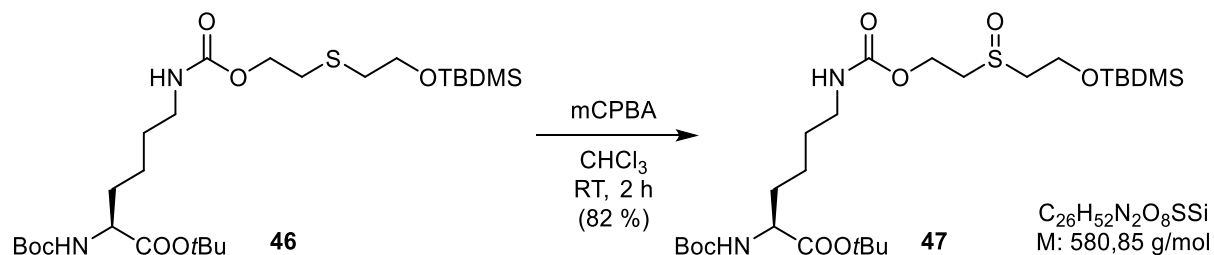
#### *N*<sup>2</sup>-((*tert*-butoxycarbonyl)-*N*<sup>6</sup>-((2-((*tert*-butyldimethylsilyl)oxy)ethyl)thio)ethoxy)carbonyl)-L-lysinate (**46**)



DSC (0.65 g, 2.53 mmol) and DIPEA (0.92 mL, 5.28 mmol) were added to the solution of **45** (0.50 g, 2.11 mmol) in 20 mL of ACN and the mixture was stirred at room temperature for 2 h, until TLC (DCM) showed that all of the starting material had been consumed. Boc-Lys-O*t*Bu (0.77 g, 2.53 mmol) was then added to the mixture and stirring was continued overnight. After evaporation of the solvent, the oily residue was dissolved in DCM and washed with 10 % citric acid, satur. NaHCO<sub>3</sub> solution and brine. The organic phase was dried over MgSO<sub>4</sub> and concentrated under reduced pressure. Purification by column chromatography on silica (packed in DCM and washed with 2 % MeOH/DCM) afforded **46** (1.05 g, 88 %) as a pale yellow oil.

**MS (ESI+):** m/z (%) 465.3 (100) [M-Boc+H]<sup>+</sup>

***tert*-butyl *N*<sup>2</sup>-(*tert*-butoxycarbonyl)-*N*<sup>6</sup>-((2-((2-((*tert*-butyldimethylsilyl)oxy)ethyl)sulfinyl)ethoxy)carbonyl)-L-lysinate (47)**



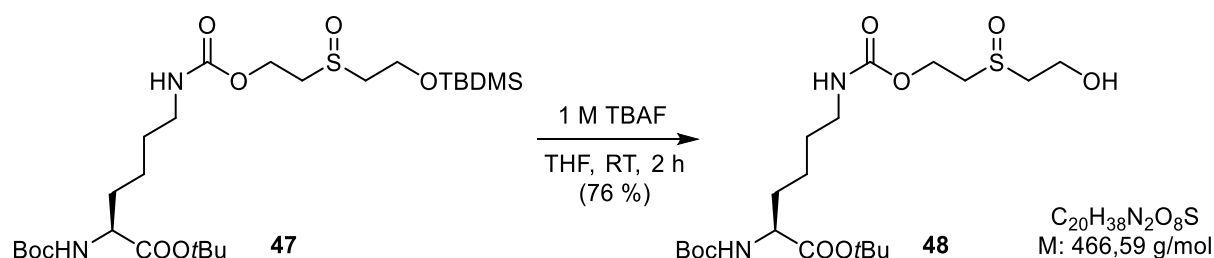
mCPBA (77 %, 0.26 g, 1.52 mmol) was added to the solution of **46** (0.86 g, 1.52 mmol) in 50 mL CHCl<sub>3</sub> at 0 °C and the mixture was allowed to warm up to room temperature under stirring for 2 h. After washing the mixture with satur. NaHCO<sub>3</sub> solution and brine, the organic phase was removed under reduced pressure. Purification of the obtained residue by column chromatography on silica (packed in DCM and washed with 1 → 5 % MeOH/DCM) afforded **47** (0.72 g, 82 %) as a pale yellow oil.

*R<sub>f</sub>* = 0.44 (2 % MeOH/DCM) [ninhydrin]

**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>) δ/ppm: 5.06 (d, *J* = 8.2 Hz, 1H), 4.96 (br, 1H), 4.60-4.49 (m, 1H), 4.48-4.36 (m, 1H), 4.19-4.09 (m, 1H), 4.09-4.00 (m, 2H), 3.23-2.82 (m, 6H), 1.84-1.48 (m, 4H), 1.45 (s, 9H), 1.43 (s, 9H), 1.40-1.28 (m, 2H), 0.88 (s, 9H), 0.08 (d, *J* = 1.6 Hz, 6H).

**MS (ESI+):** m/z (%) 481.3 (100) [M-Boc+H]<sup>+</sup>, 583.3 (55) [M+H]<sup>+</sup>, 425.2 (15) [M-Boc-OtBu+H]<sup>+</sup>; **(ESI-):** m/z (%) 625.3 (100) [M+HCOOH-H]<sup>-</sup>

***tert*-butyl *N*<sup>2</sup>-(*tert*-butoxycarbonyl)-*N*<sup>6</sup>-((2-((2-hydroxyethyl)sulfinyl)ethoxy)carbonyl)-L-lysinate (48)**



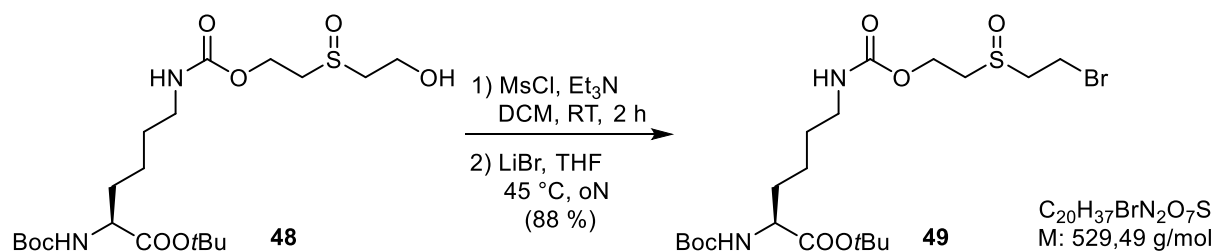
TBDMS-protected **47** (0.82 g, 1.42 mmol) was dissolved in 15 mL of dry THF and 1 M solution of TBAF in THF (1.64 mL, 1.64 mmol) was added. The mixture was stirred at room temperature for 4 h, after which the solvent was removed under reduced pressure. The obtained residue was dissolved in DCM and washed with 10 % citric acid solution. After drying over Na<sub>2</sub>SO<sub>4</sub>, the organic phase was concentrated *in vacuo*. Purification by column

chromatography on silica (packed in DCM and washed with 2 → 7 % MeOH/DCM) afforded **48** (0.50 g, 76 %) as a pale yellow oil.

$R_f = 0.25$  (5 % MeOH/DCM) [ninhydrin]

**MS (ESI+):**  $m/z$  (%) 367.2 (100) [M-Boc+H]<sup>+</sup>, 311.1 (43) [M-Boc-OtBu+H]<sup>+</sup>, 467.2 (40) [M+H]<sup>+</sup>; **(ESI-):**  $m/z$  (%) 511.2 (100) [M+HCOOH-H]<sup>-</sup>

**tert-butyl N<sup>6</sup>-((2-((2-bromoethyl)sulfinyl)ethoxy)carbonyl)-N<sup>2</sup>-(tert-butoxycarbonyl)-L-lysinate (**49**)**



Et<sub>3</sub>N (1.21 mL, 8.65 mmol) and MsCl (0.63 mL, 8.07 mmol) were sequentially added to the solution of **48** (2.69 g, 5.76 mmol) in 40 mL of dry DCM and the mixture was stirred at RT for 2 h, when the TLC (DCM) showed the full consumption of the product. The mixture was washed with 10 % citric acid solution and brine and the solvent was removed under reduced pressure. The obtained residue was then dissolved in 40 mL of THF, LiBr (5.00 g, 57.6 mmol) was added and the suspension was stirred at 45 °C overnight. The solvent was removed *in vacuo* and the oily residue purified by column chromatography on silica (packed in DCM and eluted with 2 % MeOH/DCM) to afford **49** (2.68 g, 88 %) as a yellow oil.

$R_f = 0.33$  (5 % MeOH/DCM) [ninhydrin]

**<sup>1</sup>H NMR** (300 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ /ppm: 7.25 (t,  $J = 5.5$  Hz, 1H), 7.04 (d,  $J = 7.6$  Hz, 1H), 4.37 (dt,  $J = 11.6, 5.0$  Hz, 1H), 4.22 (ddd,  $J = 12.1, 9.0, 4.2$  Hz, 1H), 3.92-3.81 (m, 1H), 3.79-3.67 (m, 2H), 3.36-3.28 (m, 2H), 3.23-3.09 (m, 1H), 3.08-2.88 (m, 3H), 1.65-1.19 (m, 24H).

**MS (ESI+):**  $m/z$  (%) 430.1 (100) [M-Boc+H]<sup>+</sup>

Deprotection of **49** under acidic conditions (TFA, HCl or HBr) unfortunately led to decomposition of the UAA. An undesired Pummerer rearrangement had likely occurred.

#### 4.4.1.3. Bromoalkyl-bearing urea derivatives

#### General procedure for TBDMS protection of amino-alcohols

TBDMSCl (1.2 eq) was added to a solution of amino-alcohol (1.0 eq) in dry pyridine and the mixture was stirred overnight at room temperature. The mixture was then concentrated under reduced pressure (water bath at 40 °C) and the obtained residue was dissolved in DCM. The DCM phase was washed with satur. NaHCO<sub>3</sub> solution, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Purification of the crude product by column chromatography on silica (packed in 10 % MeOH/DCM and eluted with 2 % Et<sub>3</sub>N/10 % MeOH/DCM, v/v) afforded the desired product.

#### 4-((*tert*-butyldimethylsilyl)oxy)butan-1-amine (**50**)



4-amino-1-butanol (2.00 g, 22.4 mmol) and TBDMSCl (4.10 g, 27.2 mmol) in 4 mL of dry pyridine afforded **50** as a yellow liquid (3.72 g, 82 %).

$R_f$  = 0.35 (2 % Et<sub>3</sub>N/10 % MeOH/DCM) [ninhydrin]

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ/ppm: 3.62 (t,  $J$  = 6.2 Hz, 2H), 2.72 (t,  $J$  = 6.8 Hz, 2H), 1.86 (s, 2H), 1.61-1.43 (m, 4H), 0.88 (s, 9H), 0.04 (s, 6H).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ/ppm: 63.2, 42.1, 30.3, 30.1, 26.1, 18.5, -5.2.

MS (ESI +):  $m/z$  (%) 204.2 (100) [M+H]<sup>+</sup>

#### 5-((*tert*-butyldimethylsilyl)oxy)pentan-1-amine (**51**)



5-amino-1-pentanol (3.00 g, 29.1 mmol) and TBDMSCl (5.26 g, 34.9 mmol) in 6 mL of dry pyridine afforded **51** as a yellow liquid (5.91 g, 94 %).

$R_f$  = 0.38 (2 % Et<sub>3</sub>N/10 % MeOH/DCM) [ninhydrin]

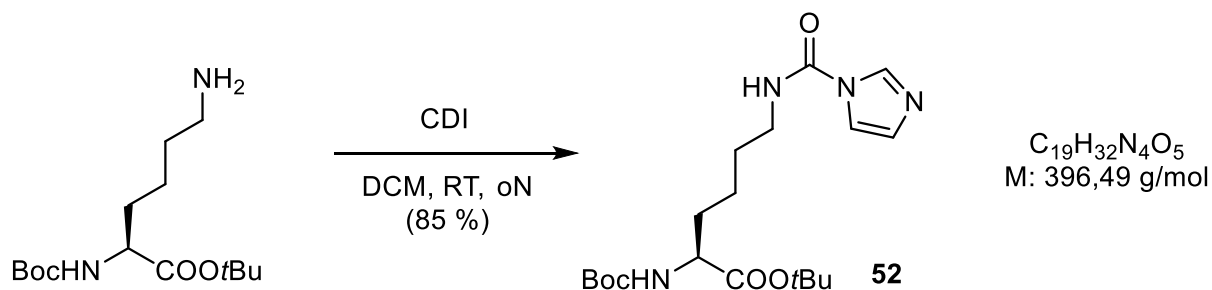
<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ/ppm: 3.58 (t,  $J$  = 6.4 Hz, 2H), 2.75-2.63 (m, 2H), 2.31 (s, 2H), 1.58-1.26 (m, 6H), 0.86 (s, 9H), 0.01 (s, 6H).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ/ppm: 63.2, 42.0, 33.1, 32.7, 26.1, 23.2, 18.5, -5.2.

MS (ESI +):  $m/z$  (%) 218.2 (100) [M+H]<sup>+</sup>

---



**tert-butyl N<sup>2</sup>-(tert-butoxycarbonyl)-N<sup>6</sup>-(1H-pyrrole-1-carbonyl)-L-lysinate**

Boc-Lys-OtBu (2.73 g, 9.04 mmol) dissolved in 60 mL of dry DCM was added over the course of 1h to the solution of CDI (2.93 g, 18.1 mmol) in 20 mL of dry DCM at 0 °C. During this time, the initially white suspension gets clear. The stirring was continued overnight at room temperature, after which the mixture was washed with brine and the separated aqueous phase extracted twice with DCM. The combined organic phases were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. Purification by column chromatography on silica (packed in DCM and eluted with 2 → 4 % MeOH/DCM, v/v) yielded **52** as a white foam (3.06 g, 85 %).

$R_f = 0.24$  (4 % MeOH/DCM) [ninhydrin]

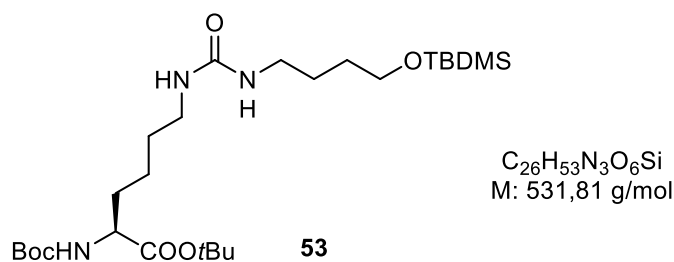
**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>)  $\delta$ /ppm: 8.21 (obs t,  $J = 1.1$  Hz, 1H), 7.51 (obs t,  $J = 1.5$  Hz), 7.27 (br, 1H), 7.03 (obs dd,  $J = 1.6, 0.9$  Hz, 1H), 5.18 (d,  $J = 8.4$  Hz, 1H), 4.11 (q,  $J = 8.1, 7.6$  Hz, 1H), 3.40 (q,  $J = 6.5$  Hz, 2H), 1.86-1.53 (m, 4H), 1.47-1.36 (m, 2H), 1.44 (s, 9H), 1.40 (s, 9H).

**<sup>13</sup>C NMR** (75 MHz, CDCl<sub>3</sub>)  $\delta$ /ppm: 171.9, 155.9, 149.3, 136.1, 130.1, 116.5, 82.3, 80.1, 53.8, 40.8, 33.0, 28.6, 28.4, 28.1, 22.7.

**MS (ESI+):**  $m/z$  (%): 397.2 (100) [M+H]<sup>+</sup>, 419.2 (10) [M+Na]<sup>+</sup>; **(ESI-):**  $m/z$  (%): 509.1 (100) [M+TFA-H]<sup>-</sup>, 196.0 (70) [M-2H]<sup>2-</sup>, 253.1 (68), 431.1 (55) [M+Cl]<sup>-</sup>, 395.2 (18) [M-H]<sup>-</sup>, 327.1 (10).

**General procedure for urea bond formation**

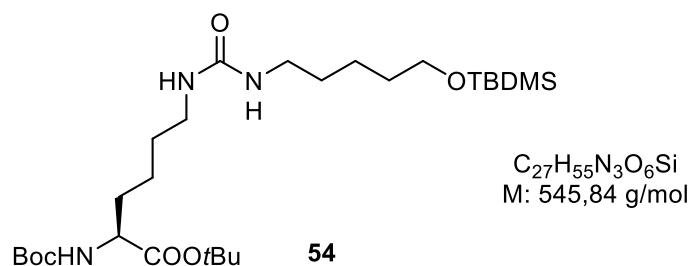
TBDMS-protected amino-alcohol (1.1 eq) and **52** (1.0 eq) were dissolved in dry ACN and the mixture was stirred at 50-60 °C for 24 h. After completion, as judged by TLC (4 % MeOH/DCM), the reaction mixture was diluted with DCM and subsequently washed with 1 M HCl and satur. NaHCO<sub>3</sub> solution. The organics were dried over Na<sub>2</sub>CO<sub>3</sub> and removed under reduced pressure to afford the crude residue which was purified by column chromatography on silica (packed in 2 % MeOH/DCM and eluted with 2 → 4 % MeOH/DCM, v/v).

***tert*-butyl *N*<sup>2</sup>-(*tert*-butoxycarbonyl)-*N*<sup>6</sup>-((4-((*tert*-butyldimethylsilyl)oxy)butyl)carbamoyl)-L-lysinate (**53**)**

Amine **50** (0.64 g, 3.17 mmol) and **52** (1.14 g, 2.88 mmol) in 20 mL of dry ACN afforded **53** as a pale yellow oil (1.45 g, 95 %).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ/ppm: 5.11 (d, *J* = 8.2 Hz, 1H), 4.53 (t, *J* = 5.3 Hz, 2H), 4.13 (q, *J* = 7.8 Hz, 1H), 3.62 (dq, *J* = 5.8, 3.2 Hz, 2H), 3.25-3.08 (m, 4H), 1.80-1.51 (m, 8H), 1.45 (s, 9H), 1.44 (s, 9H), 1.41-1.31 (m, 2H), 0.88 (s, 9H), 0.04 (s, 6H).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ/ppm: 171.9, 158.3, 155.6, 81.9, 79.7, 62.8, 53.7, 40.4, 40.2, 32.8, 30.1, 29.6, 28.4, 28.0, 26.8, 26.0, 22.5, 18.3, -5.3.

***tert*-butyl *N*<sup>2</sup>-(*tert*-butoxycarbonyl)-*N*<sup>6</sup>-((5-((*tert*-butyldimethylsilyl)oxy)pentyl)carbamoyl)-L-lysinate (**54**)**

Amine **51** (0.55 g, 2.56 mmol) and **52** (0.92 g, 2.32 mmol) in 15 mL of dry ACN afforded **54** as a pale yellow oil (1.22 g, 96 %).

*R<sub>f</sub>* = 0.50 (5 % MeOH/DCM) [ninhydrin]

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ/ppm: 5.12 (d, *J* = 7.6 Hz, 1H), 4.58 (t, *J* = 5.6 Hz, 1H), 4.50 (t, *J* = 5.6 Hz, 1H), 4.18-4.06 (m, 1H), 3.58 (t, *J* = 6.4 Hz, 2H), 3.25-3.05 (m, 4H), 1.88-1.45 (m, 8H), 1.45 (s, 9H), 1.43 (s, 9H), 1.41-1.30 (m, 4H), 0.87 (s, 9H), 0.03 (s, 6H).

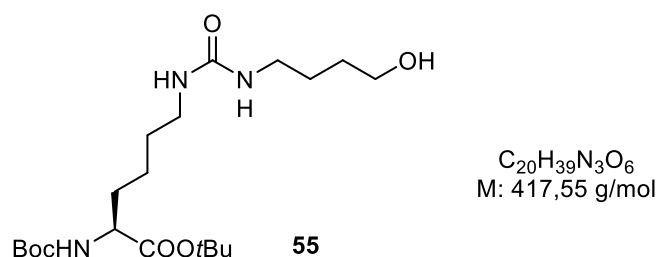
<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ/ppm: 172.0, 158.5, 155.7, 82.0, 79.8, 63.2, 53.8, 40.7, 40.3, 32.9, 32.7, 30.2, 29.7, 28.5, 28.2, 26.1, 23.3, 22.7, 18.5, -5.1.

**General procedure for TBDMS deprotection**

TBDMS-protected urea-containing L-lysine derivative (1.00 eq) was dissolved in dry THF and 1 M TBAF solution in THF (1.16 eq) was added to the solution. After stirring at room temperature for 4-5 h, the reaction was judged complete by TLC (5 % MeOH/DCM) and the

solvent was evaporated under reduced pressure. The obtained residue was dissolved in DCM and washed with 10 % citric acid and satur.  $\text{NaHCO}_3$  solution. The organics were dried over  $\text{Na}_2\text{CO}_3$  and removed under reduced pressure to afford the crude residue which was purified by column chromatography on silica (packed in 2 % MeOH/DCM and eluted with 4 → 8 % MeOH/DCM, v/v).

***tert*-butyl  $N^2$ -(*tert*-butoxycarbonyl)- $N^6$ -(4-hydroxybutyl)carbamoyl)-L-lysinate (**55**)**

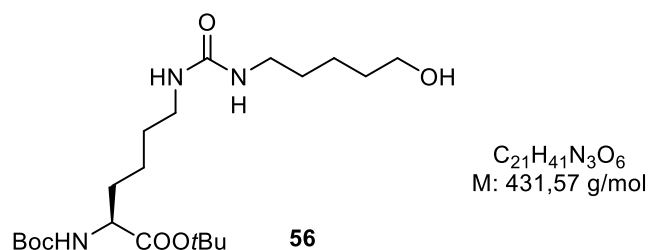


Amino acid **53** (1.43 g, 2.69 mmol) and TBAF (3.12 mL, 3.12 mmol) in 20 mL of dry ACN afforded **55** as a pale yellow oil (1.02 g, 91 %).

$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 5.20 (d,  $J = 8.0$  Hz, 1H), 5.06-4.90 (m, 2H), 4.16-4.03 (m, 1H), 3.71-3.60 (m, 2H), 3.29-3.04 (m, 4H), 2.77 (br, 1H), 1.83-1.47 (m, 8H), 1.45 (s, 9H), 1.43 (s, 9H), 1.34-1.29 (m, 2H).

$^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 172.1, 159.0, 155.8, 82.1, 79.9, 62.5, 54.0, 40.2 (2C), 32.8, 30.0, 28.5, 28.1, 27.1, 22.7.

***tert*-butyl  $N^2$ -(*tert*-butoxycarbonyl)- $N^6$ -(5-hydroxypentyl)carbamoyl)-L-lysinate (**56**)**

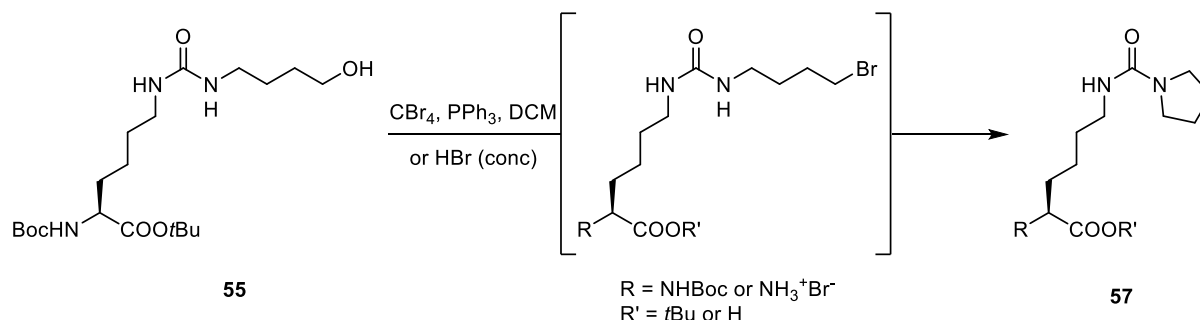


Amino acid **54** (1.20 g, 2.20 mmol) and TBAF (2.55 mL, 2.55 mmol) in 20 mL of dry ACN afforded **56** as a pale yellow oil (0.89 g, 93 %).

$R_f = 0.32$  (5 % MeOH/DCM) [ninhydrin]

$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 5.19 (d,  $J = 7.8$  Hz, 1H), 4.95-4.80 (m, 2H), 4.10 (q,  $J = 7.5$  Hz, 1H), 3.62 (t,  $J = 6.3$  Hz, 2H), 3.24-3.04 (m, 4H), 1.83-1.47 (m, 8H), 1.45 (s, 9H), 1.43 (s, 9H), 1.41-1.30 (m, 4H).

$^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ )  $\delta$ /ppm: 172.1, 158.9, 155.8, 82.1, 79.9, 62.5, 53.9, 40.3, 40.2, 32.8, 32.3, 30.1, 29.8, 28.5, 28.1, 23.1, 22.7.

**Unsuccessful preparation of *N*<sup>6</sup>-((4-bromobutyl)carbamoyl)-L-lysine (BrCN6K)**

Urea-containing UAA **55** (1.02 g, 2.44 mmol) and  $\text{CBr}_4$  (0.99 g, 2.99 mmol) were dissolved in 10 mL of dry DCM, the mixture was cooled to 0 °C and triphenylphosphine (1.03 g, 3.93 mmol) was added in several portions during 15 minutes. The mixture was stirred at room temperature overnight and subsequently diluted with more DCM, washed with water and brine and concentrated under vacuum. The crude residue thus obtained was purified by column chromatography on silica (packed in DCM and washed with 1 → 3 % MeOH/DCM). Instead of the desired monobrominated UAA, *tert*-butyl *N*<sup>2</sup>-(*tert*-butoxycarbonyl)-*N*<sup>6</sup>-(pyrrolidine-1-carbonyl)-L-lysinate (**57**), the product of intramolecular cyclisation reaction, was isolated as a white foam (0.71 g, 73 %).

Similarly, stirring of imidazolide **52** (0.21 g, 0.52 mmol) in the presence of 4-bromobutan-1-amine hydrobromide (0.14 g, 0.62 mmol) and DIPEA (0.11 mL, 0.62 mmol) in 5 mL of dry ACN at 50-60 °C, followed by column chromatography on silica (2 → 4 % MeOH/DCM), only afforded the undesired **57** as a white foam (0.20 g, 95 %).

$R_f = 0.16$  (2 % MeOH/DCM) [ninhydrin]

**MS (ESI+):**  $m/z$  (%) 400.1  $[\text{M}+\text{H}]^+$ , 300.1  $[\text{M}-\text{Boc}+\text{H}]^+$ , 422  $[\text{M}+\text{Na}]^+$

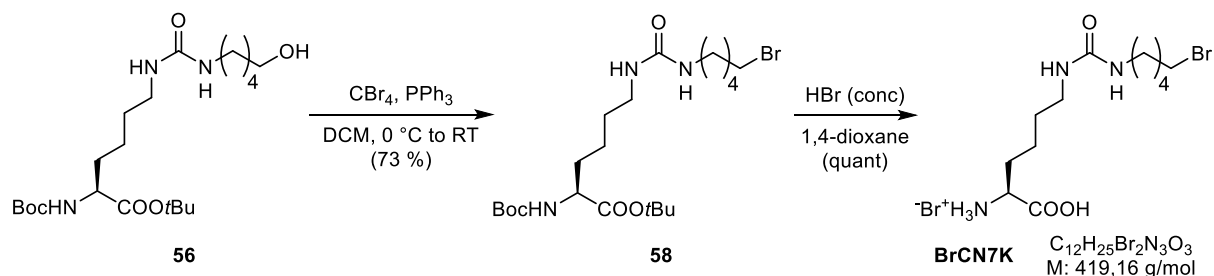
Alternatively, the UAA **55** (1.49 g, 3.57 mmol) was treated with 20 mL of conc. HBr (48 % in  $\text{H}_2\text{O}$ ) at 60 °C. The reaction was monitored by LC-MS which showed full conversion after 48 h. After concentrating the mixture in vacuo, the residue was dissolved in water, washed with cold  $\text{Et}_2\text{O}$  and lyophilised. An orange oil was obtained (1.44 g) which only showed the presence of the undesired fully deprotected cyclisation product. Incorporation test with this UAA was still performed as described in the general incorporation protocol. **57** was therefore dissolved as a 100 mM stock solution in water.

*Before workup:*

**MS (ESI+):**  $m/z$  (%) 324.2/326.2 (100)  $[\text{M}+\text{H}]^+$

*After workup:*

**MS (ESI+):**  $m/z$  (%) 244.1  $[\text{M}+\text{H}]^+$

**Preparation of  $N^6$ -((5-bromopentyl)carbamoyl)-L-lysine (BrCN7K)**

Urea-containing UAA **56** (0.87 g, 2.01 mmol) and  $\text{CBr}_4$  (0.80 g, 2.42 mmol) were dissolved in 40 mL of dry DCM, the mixture was cooled to 0 °C and triphenylphosphine (0.63 g, 2.42 mmol) was added in several portions during 15 minutes. The mixture was stirred at room temperature overnight and concentrated under reduced pressure. The crude residue thus obtained was purified by column chromatography on silica (packed in DCM and washed with 1 → 3 % MeOH/DCM) to afford the desired product as a pale yellow foam (0.72 g, 73 %) which was directly used in the following step.

$R_f$  = 0.53 (5 % MeOH/DCM) [ninhydrin]

MS (ESI+): m/z (%) 394.1 (100) [M-Boc+H]<sup>+</sup>, 494.1 (60) [M+H]<sup>+</sup>

The UAA **58** (0.28 g, 0.57 mmol) was dissolved in 5 mL of 1,4-dioxane and 5 mL of conc. HBr (48 % in water) were added to the solution. The reaction mixture was stirred at room temperature for 4 h, after which it was concentrated under reduced pressure. The obtained residue was dissolved in water, washed with cold ether and lyophilised to afford **BrCN7K** as a pale orange solid (0.24 g, quant). BrCN7K was stored at -20 °C and dissolved as a 100 mM stock solution in water before usage.

MS (ESI+): m/z (%) 338.1/340.1 (100) [M+H]<sup>+</sup>; (ESI-): m/z 336.0/338.0 (100) [M-H]<sup>-</sup>, 673.1/675.1/677.1 (65) [2M-H]<sup>-</sup>

**4.4.2. Biological methods**

Details on cloning of constructs for bacterial expression, generation of the *E. coli* DH10βΔ*cobB* strain, MS/MS analysis of crosslinked sfGFP homodimer, cloning of constructs for mammalian cell applications, site-specific incorporation of BrCnK in mammalian cells as well as crystallisation experiments and crystal structure determination can be found under the reference 53.

Crystallisation and structure determination was performed by our collaboration partner Matthias Müller (MPI Dortmund), while MS/MS experiments have been performed with the help of Nina Bach and Matthias Stahl from the group of Prof. Stephan Sieber (TUM). *E. coli* DH10βΔ*cobB* strain was generated by Marie-Kristin von Wrisberg. Mammalian cell experiments were performed by Daniel Horn-Ghetko. Cloning of constructs for bacterial

expression was performed by Thorsten Müller and myself, as were *in vitro* crosslinking assays. Maximilian Fottner significantly contributed to the latter experiments as well.

#### **Cloning of constructs for co-expression of Rab1b-BrCnK and DrrA<sub>340-533</sub> cysteine mutants**

For co-expression of Rab1b-His6 bearing a TAG codon and cysteine variants of DrrA<sub>340-533</sub> we constructed the dual expression vector pTM-Duet based on pETDuet1 (Novagen) and pNHD Ub (a tetracycline resistant plasmid containing LacI and Ub(S65TAG)-His6, a gift from Jason Chin)<sup>[242]</sup> as detailed under the ref. 53. These experiments were performed by Thorsten Müller.

Other pTM Duet vectors containing different Rab1b-DrrA mutant pairs were prepared through QuikChange<sup>TM</sup> site-directed mutagenesis (Section 4.2.2), using pTM Duet\_Rab1b-R79TAG-His6\_DrrA-D512C as template. pTM Duet vectors containing DrrA<sub>340-533</sub> wt gene were used as controls in *in vivo* crosslinking assays. Sequences of used primers are displayed in Table 4.3. pKW1-BrCnK-RS was constructed by recloning the sequence of BrCnK-RS from pBK MbPylRS BrCnKRS into pKW1 (a streptomycin resistant plasmid encoding PylRS wt and PylT under constitutive promoters, a gift from Jason Chin)<sup>[242]</sup> using *Bam*HI and *Pst*I restriction sites.

#### **sfGFP expression and purification**

Expression and purification of sfGFP variants containing BrCnK and BrCN7K was performed as described in Section 4.2.4. In case of BrCnK, a slight modification involved the addition of 1 M nicotinamide solution in water (final concentration of 10 mM) at the same time as the addition of the UAA (OD<sub>600</sub> around 0.3), as a way of inhibiting the *E. coli* sirtuin deacylase CobB.

**Table 4.3.** Primer sequences used for QuickChange™ site-directed mutagenesis.

QuickChange Primer	Sequence
<b>Rab1b-H6 D16TAG</b>	Fw 5' CTGCTGATCGGCTAGTCTGGTGTGGG 3' Rv 5' CCCACACCAGACTAGCCGATCAGCAG 3'
<b>Rab1b-H6 D31TAG</b>	Fw 5' CTGCTGCTGCGTTTTGCTGATTAGACGTATACCGAATCCTATATC 3' Rv 5' GATATAGGATTCGGTATACGTCTAATCAGCAAACGCAGCAGCAG 3'
<b>Rab1b-H6 D44TAG</b>	Fw 5' CAGTACCATTGGCGTCTAGTTCAAAATCCGTACG 3' Rv 5' CGTACGGATTTTGAAGTAGACGCCAATGGTACTG 3'
<b>Rab1b-H6 T72TAG</b>	Fw 5' CTGGTCAAGAACGTTTCCGTTAGATTACCAGCTCCTATTATC 3' Rv 5' GATAATAGGAGCTGGTAATCTAACGGAAACGTTCTTGACCAG 3'
<b>Rab1b-H6 R79TAG</b>	Fw 5' CATTACCAGCTCCTATTATTAGGGTGCTCATGGCATCATTG 3' Rv 5' CAATGATGCCATGAGCACCTAATAATAGGAGCTGGTAATG 3'
<b>Rab1b-H6 N98TAG</b>	Fw 5' GACCAAGAATCTTATGCCTAGGTGAAACAGTGGCTGCAG 3' Rv 5' CTGCAGCCACTGTTTACCTAGGCATAAGATTCTTGGTC 3'
<b>DrrA<sub>340-533</sub> Q407C</b>	Fw 5' CAAAAGGATATGCTATTTTATGCTCTTTATGGGGAGCAGCTTCG 3' Rv 5' CGAAGCTGCTCCCATAAAGAGCATAAAATAGCATATCCTTTTG 3'
<b>DrrA<sub>340-533</sub> T427C</b>	Fw 5' GCTGCAACTCTCACAGAATCATGCGTAGAGCCTGGTTTAGTATC 3' Rv 5' GATACTAAACCAGGCTCTACGCATGATTCTGTGAGAGTTGCAGC 3'
<b>DrrA<sub>340-533</sub> K438C</b>	Fw 5' CCTGGTTTAGTATCAGCTGTTAACTGCATGTCAGCATTCTTTATGGATTG 3' Rv 5' CAATCCATAAAGAATGCTGACATGCAGTTAACAGCTGATACTAAACCAGG 3'
<b>DrrA<sub>340-533</sub> R453C</b>	Fw 5' GATTGTAAATTAAGTCCAAATGAGTGC GCAACCCCTGATCCAGATTTTAAG 3' Rv 5' CTAAAATCTGGATCAGGGGTTGCGCACTCATTTGGACTTAATTTACAATC 3'
<b>DrrA<sub>340-533</sub> D512C</b>	Fw 5' GAGAGAAGCAATAATGTTAATTGCGAAACATTAGAGAGTGTGCTG 3' Rv 5' CAGCACACTCTAATGTTTCGCAATTAACATTATTGCTTCTCTC 3'
<b>DrrA<sub>340-533</sub> E516C</b>	Fw 5' GAGAAGCAATAATGTTAATGATGAAACATTATGCAGTGTGCTGAGCTCTAAAG GTGAAAATTTATC 3' Rv 5' GATAAATTTTACCTTTAGAGCTCAGCACACTGCATAATGTTTCATCATTAAACAT TATTGCTTCTC 3'

### **Rab1b expression and purification**

Rab1b-His6 proteins containing the UAAs BrC6K and BrC7K at different positions were prepared following the same procedure as described in Section 4.2.4., by using pPylT\_Rab1b-*xxTAG*-His6 (which encodes *MbtRNA*<sub>CUA</sub> and a C-terminally His6-tagged Rab1b gene with a TAG codon at different positions: D16TAG, S17TAG, Y37TAG, I47TAG, T72TAG, R79TAG and N98TAG) and pBK *MbPylS* BrCnRS (which encodes *MbBrCnRS*). Expressions were carried out at 37°C in the presence of 1.5 mM BrCnK for 18 hours. After cell lysis (lysis buffer: 50 mM HEPES pH 8.0, 20 mM imidazole pH 8.0, 500 mM LiCl, 1 mM MgCl<sub>2</sub>, 2 mM BME, 10 μM GDP, 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and one cOmplete™ Mini EDTA-free protease inhibitor cocktail tablet) and centrifugation, the cleared cell lysate was incubated with Ni-NTA slurry at 4°C for 1 h. The Ni-NTA beads with bound protein were washed with wash buffer containing 50 mM HEPES pH 8.0, 20 mM imidazole

pH 8.0, 500 mM LiCl, 1 mM MgCl<sub>2</sub>, 2 mM BME and 10 μM GDP and the protein was eluted with wash buffer supplemented with 500 mM imidazole at pH 8.0. The fractions containing the protein were pooled together, concentrated and rebuffered (50 mM HEPES pH 8.0, 500 mM LiCl, 1 mM MgCl<sub>2</sub>, 2 mM BME and 10 μM GDP) with Amicon® Ultra-4 10K NMWL centrifugal filter units (*Merck*). Purified proteins were analyzed by SDS-PAGE and mass spectrometry.

Rab1b expression in DH10BΔ*cobB* strain was performed as described above.

#### **Expression and purification of DrrA<sub>340-533</sub> cysteine mutants**

*E. coli* BL21(DE3) cells were transformed with pET19mod\_His6-TEV-DrrA<sub>340-533</sub>xyC (which encodes an N-terminally His6-tagged DrrA<sub>340-533</sub> gene with cysteine mutants at different positions: Q407C, K438C, T427, K377, Q373, R453C, D512C and E516C, and a TEV cleavage site). After recovery with 1 mL of SOC medium for 1 h at 37 °C, the cells were cultured in 50 mL of LB medium containing ampicillin (100 μg/mL) overnight at 37 °C with shaking (200 rpm). The overnight culture was diluted in 2.7 L of fresh TB medium containing ampicillin (50 μg/mL) to an OD<sub>600</sub> of 0.05 and was incubated at 20 °C until the OD<sub>600</sub> reached 0.6, when the expression was induced with 1 M IPTG solution. The proteins were expressed overnight at 20 °C with shaking (200 rpm), after which the cells were harvested by centrifugation (5000 g, 15 min, 4 °C), flash frozen in liquid nitrogen and stored at -80 °C.

The obtained cell pellets were thoroughly resuspended in 100 mL of lysis buffer containing 50 mM HEPES pH 8.0, 20 mM imidazole pH 8.0, 500 mM LiCl, 2 mM BME, 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and two cComplete™ Mini EDTA-free protease inhibitor cocktail tablets (*Roche*). The cell suspension was incubated on ice for 30 min and sonicated with cooling in an ice-water bath. The lysed cells were centrifuged (18,000 rpm, 1 h, 4 °C), the cleared lysate added to 2.4 mL of Ni-NTA slurry (*Jena Bioscience*) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the lysate was poured into an empty plastic column and washed with 200 mL of wash buffer containing 50 mM HEPES pH 8.0, 20 mM imidazole pH 8.0, 500 mM LiCl and 2 mM BME. The proteins were eluted in ten 1 mL fractions with wash buffer supplemented with 500 mM imidazole pH 8.0. The fractions containing the protein were pooled together, concentrated and rebuffered (50 mM HEPES pH 8.0, 20 mM imidazole pH 8.0, 500 mM LiCl, 2 mM BME) with Amicon® Ultra-4 10K NMWL centrifugal filter units (*Merck*). Purified proteins were analyzed by SDS-PAGE and mass spectrometry.

In order to obtain DrrA<sub>340-533</sub>D512C without His6-tag, which was used for preparative crosslinking assay, the protein containing a TEV cleavage site was expressed and purified as described above. The fractions containing the protein were pooled together and 500 μL of TEV protease (1.8 mg/μL) was added. The mixture was transferred to SERVAPOR® dialysis tubing (*Serva*) and the dialysis bag was immersed in 5 L of cold dialysis buffer containing 50



mM HEPES pH 8.0, 100 mM NaCl and 2 mM BME, and stirred at 4 °C overnight. The protein mixture was recovered from the dialysis tubing and centrifuged (full speed, 4 °C) in order to precipitate the TEV protease. 2 mL of Ni-NTA slurry (*Jena Bioscience*) were added to the supernatant and the mixture was incubated with agitation for 1 h at 4 °C. The mixture was then poured into an empty plastic column and the flow-through was collected. The Ni-NTA beads were washed twice with 15 mL of wash buffer containing 50 mM HEPES pH 8.0, 20 mM imidazole pH 8.0, 500 mM LiCl and 2 mM BME and the bound protein eluted in 1.5 mL fractions with wash buffer supplemented with 500 mM imidazole pH 8.0. Using 15 % SDS-PAGE, we determined that the flow-through and fractions collected after washing the column contained the protein of interest. These were combined and concentrated with Amicon® Ultra-4 10K NMWL centrifugal filter units (*Merck*). Purified proteins were analyzed by mass spectrometry and stored at -80 °C.

### **MS/MS analysis**

MS analysis of digested samples was performed on an Orbitrap Fusion instrument coupled to an Ultimate3000 Nano-HPLC via an electrospray easy source (all *Thermo Fisher Scientific*). Samples were loaded on a 2 cm PepMap RSLC C18 trap column (particles 2 µm, 100A, inner diameter 75 µm, *Thermo Fisher Scientific*) with 0.1% TFA and separated on a 50 cm PepMap RSLC C18 column (particles 2 µm, 100A, inner diameter 75 µm, *Thermo Fisher Scientific*) constantly heated at 50 °C. The gradient was run from 5-35% acetonitrile, 0.1% formic acid during a 62 min method (7 min 5 %, 30 min to 28 %, 5 min to 35 %, 10 min wash at 90 %, 10 min equilibration at 5%) at a flow rate of 300 nl/min. Survey scans (m/z 300-1500) were acquired in the orbitrap with a resolution of 120,000 at m/z 200 and the maximum injection time set to 50 ms (target value 2e5). Most intense ions were selected for fragmentation with high-energy collisional dissociation at collision energy of 30 %. The instrument was operated in top speed mode and spectra acquired in the orbitrap with the maximum injection time set to 120 ms (target value 5e4). The option to inject ions for all available parallelizable time was enabled. Dynamic exclusion of sequenced peptides was set to 60 s. Data were acquired using Xcalibur software version 3.0sp2 (*Thermo Fisher Scientific*).

Raw mass data files were converted into mzML files using MSConvert from ProteoWizard<sup>[243]</sup> Cross link searches were performed with the software Kojak, version 1.5.5 (<http://www.kojak-ms.org>)<sup>[244]</sup> against a database consisting of sfGFP-N150BrC6K and contaminant proteins downloaded from the Andromeda configuration in the MaxQuant software package.<sup>[245]</sup> The search parameters followed the sample configuration file provided with the Kojak software with the following exceptions: MS1 and MS2 resolution were set to 120000 and 30000 at m/z 400, respectively. Crosslinks were allowed between BrC6K and K, C, H, D, E. Variable modifications were oxidation on methionine and carbamidomethylation. Enzyme specifications were set to trypsin or chymotrypsin depending on the samples. Ppm tolerance for the prescan was changed to 10. Kojak results were exported to Percolator<sup>[246]</sup> to produce a statistically validated set of cross-linked peptide identifications. Spectra were visualized with the Kojak Spectrum viewer.<sup>[247]</sup>

### **In vitro crosslinking assay**

Purified proteins were mixed at a concentration of 10  $\mu\text{M}$  each in buffer containing 50 mM HEPES pH 7.5, 500 mM LiCl, 1 mM  $\text{MgCl}_2$ , 2 mM BME and 10  $\mu\text{M}$  GDP, and were incubated at 20 °C overnight. The samples were afterwards denatured by addition of SDS loading buffer and heated to 95 °C for 10 minutes. The obtained samples were subsequently analyzed by SDS-PAGE.

The effects of pH, temperature and time as well as GDP and glutathione (GSH) concentrations on crosslink formation were further analyzed in respective screening experiments by modifying the conditions described above. Screenings for the optimal pH for the crosslinking were performed on purified Rab1b-R79BrC6K-His6/His6-DrrA-D512C and Rab1b-I47BrC6K-His6/His6-DrrA-T427C pairs with pH values varying from pH 6.5 to 9.0 at different time points (10 min, 60 min, 3 h and 18 h) and at 20 °C. The screenings for optimal temperature were performed on the same protein pairs as for pH screening experiments at 4 °C, 20 °C and 37 °C and different time points (10 min, 60 min, 3 h and 18 h). The pH was held constant at 7.5. To prove that the GSH concentration does not interfere with proximity-enabled crosslinking reaction, the two protein pairs were incubated overnight at 20 °C and pH 7.5 with GSH concentrations ranging from 10 mM to 100 mM. Similarly, dependence of crosslinking reaction on GDP concentrations was monitored overnight at 20 °C and pH 7.5 with GDP concentrations ranging from 10  $\mu\text{M}$  to 500  $\mu\text{M}$ . These experiments were performed by Maximilian Fottner.

### **Preparative crosslinking of Rab1b-R79BrC6K-His6 and DrrA-D512C and purification of covalently linked complex**

Both purified proteins Rab1b-R79BrC6K-His6 and DrrA-D512C (whose His6-tag was removed by TEV protease cleavage) were mixed at a concentration of 20  $\mu\text{M}$  in buffer containing 20 mM HEPES pH 8.5, 50 mM NaCl, 5 mM  $\text{MgCl}_2$ , 2 mM BME and 10  $\mu\text{M}$  GDP (total volume of 15 mL). The mixture was incubated with agitation first for 2 h at room temperature, then overnight at 4 °C and again for 2 h at room temperature before it was concentrated to around 2 mL with Vivaspin® 2 30K MWCO centrifugal filter units (*Sartorius*) and then centrifuged (3234 g, 10 min) to remove the precipitates. The crosslinked complex was purified by size-exclusion chromatography on HiLoad™ 16/60 Superdex™ 200 pg column (*GE Life Sciences*) at a flow rate of 0.75 mL/min (buffer: 20 mM HEPES pH 8.0, 50 mM NaCl, 1 mM  $\text{MgCl}_2$ , 2 mM dithioerithriol, 0.5 mM GDP). Fractions containing the complex were pooled together, rebuffered into buffer containing 20 mM HEPES pH 8.0, 50 mM NaCl and 2 mM dithioerithriol, and concentrated with Vivaspin® 2 30K MWCO centrifugal filter units (*Sartorius*) to a concentration of approx. 20 mg/ml.

### **$K_D$ determination of Rab1b:GDP:DrrA complex**

Purified Rab1b-R79BrC6K-His6 at a concentration of 2  $\mu\text{M}$  was incubated with increasing amounts of DrrA-D512C (5  $\mu\text{M}$ , 10  $\mu\text{M}$ , 30  $\mu\text{M}$ , 80  $\mu\text{M}$  and 150  $\mu\text{M}$ ) in the presence of 1 mM GDP in crosslinking buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM  $\text{MgCl}_2$  and 2 mM BME). A fixed concentration of sfGFP-His6 (0.2  $\mu\text{M}$ ) was added as internal standard for subsequent quantification. The mixture was incubated at 25  $^\circ\text{C}$  with 300 rpm shaking. 15  $\mu\text{L}$  samples were taken at different time points (0 min, 10 min, 30 min, 2 h and 6 h), quenched with 5  $\mu\text{L}$  of hot 4x SDS loading buffer and flash frozen in liquid nitrogen. The samples were afterwards denatured at 95  $^\circ\text{C}$  for 15 minutes and analyzed by SDS and  $\alpha$ -His6 western blotting. The amount of formed crosslinked complex was quantified using ImageJ<sup>[248]</sup> and normalized to the GFP signal, plotted against time and fitted with an exponential equation to determine the rate constant  $k_{\text{obs}}$ .  $k_{\text{obs}}$  values were plotted against DrrA concentration and the  $K_D$  was determined from a hyperbolic fit. A  $K_D$  of  $13 \pm 2.2$   $\mu\text{M}$  was determined. All data processing was performed using Kaleidagraph software (Synergy Software, Reading, UK). The experiments were performed in duplicates.

### **In vivo crosslinking**

Electrocompetent *E. coli* BL21(DE3) cells were co-transformed with both pTM Duet\_Rab1bxyTAG-His6\_DrrAxyC (which encodes C-terminally His6-tagged Rab1b gene with an amber codon at different positions and DrrA<sub>340-533</sub> gene with cysteine mutants at different positions) and pKW BrCnKRS (which encodes BrCnKRS and MbtRNA<sub>CUA</sub>) plasmids and the cells were cultured overnight in 50 mL of LB medium containing tetracycline (12.5  $\mu\text{g}/\text{mL}$ ) and spectinomycin (50  $\mu\text{g}/\text{mL}$ ), at 37  $^\circ\text{C}$  with shaking (200 rpm). The overnight culture was diluted to an OD<sub>600</sub> of 0.05 in 50 mL of fresh TB medium supplemented with tetracycline (6.25  $\mu\text{g}/\text{mL}$ ) and spectinomycin (25  $\mu\text{g}/\text{mL}$ ) and was incubated at 37  $^\circ\text{C}$  until an OD<sub>600</sub> of 0.3 was reached. Both 1 M nicotinamide solution in water (final concentration of 10 mM) and 100 mM BrC6K or BrC7K solution in 1M NaOH (final concentration of 2 mM) were added to 4 mL of cell culture, which was further incubated at 37  $^\circ\text{C}$  until OD<sub>600</sub> of 0.6, when the expression was induced with the addition of 1 M IPTG (1:1000). The proteins were expressed up to 12 h, after which 1 mL of the culture was taken, centrifuged (5000 rpm, 5 min), the supernatant discarded and the obtained cell pellets resuspended in 200  $\mu\text{L}$  of 1x SDS loading buffer. The cell lysates were kept at 95  $^\circ\text{C}$  for 10 minutes, centrifuged (20.238 rpm, 10 min) and 10  $\mu\text{L}$  of each sample was loaded on 15 % SDS gels. After SDS-PAGE, the gels were blotted onto nitrocellulose membrane and stained with  $\alpha$ -His6-Peroxidase antibody (Roche) as described in General Methods. For preparative expression and purification of *in vivo* covalently formed complex, electrocompetent *E. coli* BL21(DE3) cells were co-transformed with both pTM Duet\_Rab1bR79TAG-His6\_DrrAD512C and pKW BrCnKRS plasmids and the complex was expressed as described above in a 100 ml LB culture containing 1.5 mM BrC6K for 12 hours. The cells were pelleted and the complex was purified *via* His6-tag affinity column as

described above. The covalently linked complex was analyzed by full-length mass spectrometry.

---

## § 5. REFERENCES

- [1] O. Shimomura, F. H. Johnson, Y. Saiga, *Journal of Cellular and Comparative Physiology* **1962**, *59*, 223-239.
- [2] (a) O. Shimomura, *FEBS Letters* **1979**, *104*, 220-222; (b) M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, S. J. Remington, *Science* **1996**, *273*, 1392; (c) F. Yang, L. G. Moss, G. N. Phillips, *Nature Biotechnology* **1996**, *14*, 1246-1251.
- [3] (a) D. P. Barondeau, C. D. Putnam, C. J. Kassmann, J. A. Tainer, E. D. Getzoff, *Proc Natl Acad Sci U S A* **2003**, *100*, 12111-12116; (b) L. J. Pouwels, L. Zhang, N. H. Chan, P. C. Dorrestein, R. M. Wachter, *Biochemistry* **2008**, *47*, 10111-10122.
- [4] R. Y. Tsien, *Annual Review of Biochemistry* **1998**, *67*, 509-544.
- [5] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier, *Gene* **1992**, *111*, 229-233.
- [6] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, D. C. Prasher, *Science* **1994**, *263*, 802.
- [7] D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, *Physiological Reviews* **2010**, *90*, 1103-1163.
- [8] G. U. Nienhaus, *Angewandte Chemie International Edition* **2008**, *47*, 8992-8994.
- [9] (a) A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nature Biotechnology* **2003**, *21*, 86-89; (b) A. Gautier, A. Juillerat, C. Heinis, I. R. Corrêa, M. Kindermann, F. Beaufils, K. Johnsson, *Chemistry & Biology* **2008**, *15*, 128-136; (c) G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klauert, R. F. Bulleit, K. V. Wood, *ACS Chemical Biology* **2008**, *3*, 373-382.
- [10] (a) X. Sun, A. Zhang, B. Baker, L. Sun, A. Howard, J. Buswell, D. Maurel, A. Masharina, K. Johnsson, C. J. Noren, M.-Q. Xu, I. R. Corrêa Jr, *Chembiochem* **2011**, *12*, 2217-2226; (b) T. Komatsu, K. Johnsson, H. Okuno, H. Bito, T. Inoue, T. Nagano, Y. Urano, *Journal of the American Chemical Society* **2011**, *133*, 6745-6751; (c) E. I. Kovalenko, S. Ranjbar, L. D. Jasenosky, A. E. Goldfeld, I. A. Vorobjev, N. S. Barteneva, *BMC Research Notes* **2011**, *4*, 340.
- [11] B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science* **1998**, *281*, 269.
- [12] S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis, R. Y. Tsien, *Journal of the American Chemical Society* **2002**, *124*, 6063-6076.
- [13] J. Zhang, R. E. Campbell, A. Y. Ting, R. Y. Tsien, *Nature Reviews Molecular Cell Biology* **2002**, *3*, 906-918.
- [14] S. Kottegoda, P. C. Aoto, C. E. Sims, N. L. Allbritton, *Anal Chem* **2008**, *80*, 5358-5366.
- [15] (a) M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi, A. Y. Ting, *Nature Biotechnology* **2007**, *25*, 1483; (b) C. Uttamapinant, K. A. White, H. Baruah, S. Thompson, M. Fernández-Suárez, S. Puthenveetil, A. Y. Ting, *Proceedings of the National Academy of Sciences* **2010**, *107*, 10914; (c) J. D. Cohen, P. Zou, A. Y. Ting, *Chembiochem* **2012**, *13*, 888-894; (d) D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox, A. Y. Ting, *Journal of the American Chemical Society* **2012**, *134*, 792-795; (e) H. Baruah, S. Puthenveetil,

- Y.-A. Choi, S. Shah, A. Y. Ting, *Angewandte Chemie International Edition* **2008**, *47*, 7018-7021.
- [16] (a) L. Davis, J. W. Chin, *Nature Reviews Molecular Cell Biology* **2012**, *13*, 168; (b) K. Lang, J. W. Chin, *Chemical Reviews* **2014**, *114*, 4764-4806.
- [17] A. R. Fersht, M. M. Kaethner, *Biochemistry* **1976**, *15*, 818-823.
- [18] J. J. Perona, I. Gruic-Sovulj, in *Aminoacyl-tRNA Synthetases in Biology and Medicine* (Ed.: S. Kim), Springer Netherlands, Dordrecht, **2014**, pp. 1-41.
- [19] R. B. Loftfield, D. Vanderjagt, *Biochemical Journal* **1972**, *128*, 1353.
- [20] J. Ling, D. Söll, *Proceedings of the National Academy of Sciences* **2010**, *107*, 4028.
- [21] J. M. Berg, J. L. Tymoczko, L. Stryer, G. J. Gatto, *Biochemistry*, Palgrave Macmillan, Basingstoke, **2012**.
- [22] (a) R. Young, H. Bremer, *Biochemical Journal* **1976**, *160*, 185; (b) P. P. Dennis, H. Bremer, *Journal of Molecular Biology* **1974**, *84*, 407-422; (c) P. P. Dennis, M. Nomura, *Proceedings of the National Academy of Sciences of the United States of America* **1974**, *71*, 3819-3823; dH. Bremer, P. Dennis, *EcoSal Plus* **2008**.
- [23] R. Reyes-Lamothe, C. Possoz, O. Danilova, D. J. Sherratt, *Cell* **2008**, *133*, 90-102.
- [24] T. Kimmerlin, D. Seebach, *The Journal of Peptide Research* **2005**, *65*, 229-260.
- [25] (a) C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182; (b) N. Lee, Y. Bessho, K. Wei, J. W. Szostak, H. Suga, *Nature Structural Biology* **2000**, *7*, 28-33.
- [26] P. E. Dawson, S. B. H. Kent, *Annual Review of Biochemistry* **2000**, *69*, 923-960.
- [27] (a) J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, *Current Opinion in Chemical Biology* **2010**, *14*, 774-780; (b) J. T. Ngo, D. A. Tirrell, *Accounts of Chemical Research* **2011**, *44*, 677-685; (c) C. C. Liu, P. G. Schultz, *Annual Review of Biochemistry* **2010**, *79*, 413-444.
- [28] (a) J. C. M. van Hest, D. A. Tirrell, *FEBS Letters* **1998**, *428*, 68-70; (b) K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Letters* **2001**, *502*, 25-30; (c) K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc Natl Acad Sci U S A* **2002**, *99*, 19-24.
- [29] H. Ovaa, k. wals, *Frontiers in Chemistry* **2014**, *2*.
- [30] (a) Y. Nakamura, T. Gojobori, T. Ikemura, *Nucleic Acids Research* **2000**, *28*, 292-292; (b) J. Xie, P. G. Schultz, *Methods* **2005**, *36*, 227-238.
- [31] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
- [32] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, *301*, 964.
- [33] (a) T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochemical and Biophysical Research Communications* **2008**, *371*, 818-822; (b) A. Gautier, D. P. Nguyen, H. Lusic, W. An, A. Deiters, J. W. Chin, *Journal of the American Chemical Society* **2010**, *132*, 4086-4088; (c) K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, J. W. Chin, *Nature Chemistry* **2012**, *4*, 298.
- [34] F. Li, H. Zhang, Y. Sun, Y. Pan, J. Zhou, J. Wang, *Angewandte Chemie International Edition* **2013**, *52*, 9700-9704.
- [35] (a) S. Greiss, J. W. Chin, *Journal of the American Chemical Society* **2011**, *133*, 14196-14199; (b) A. Bianco, F. M. Townsley, S. Greiss, K. Lang, J. W. Chin, *Nature Chemical Biology* **2012**, *8*, 748.
- [36] W. Wan, J. M. Tharp, W. R. Liu, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014**, *1844*, 1059-1070.
- [37] V. Beránek, J. C. W. Willis, J. W. Chin, *Biochemistry* **2019**, *58*, 387-390.
-

- [38] (a) R. Jiang, J. A. Krzycki, *J Biol Chem* **2012**, *287*, 32738-32746; (b) T. Yanagisawa, R. Ishii, R. Fukunaga, O. Nureki, S. Yokoyama, *Acta Crystallographica Section F* **2006**, *62*, 1031-1033.
- [39] (a) T. Yanagisawa, T. Sumida, R. Ishii, S. Yokoyama, *Acta Crystallographica Section D* **2013**, *69*, 5-15; (b) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Journal of Molecular Biology* **2008**, *378*, 634-652; (c) J. M. Kavran, S. Gundllapalli, P. Donoghue, M. Englert, D. Söll, T. A. Steitz, *Proceedings of the National Academy of Sciences* **2007**, *104*, 11268; dK. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Arais, R. Ishitani, T. Umehara, D. Söll, O. Nureki, *Nature* **2008**, *457*, 1163.
- [40] A. Ambrogelly, S. Gundllapalli, S. Herring, C. Polycarpo, C. Frauer, D. Söll, *Proceedings of the National Academy of Sciences* **2007**, *104*, 3141.
- [41] (a) S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nature Biotechnology* **2002**, *20*, 1044-1048; (b) D. R. Liu, P. G. Schultz, *Proceedings of the National Academy of Sciences* **1999**, *96*, 4780; (c) L. Wang, P. G. Schultz, *Chemistry & Biology* **2001**, *8*, 883-890.
- [42] (a) M. R. Fleissner, E. M. Brustad, T. Kálai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz, W. L. Hubbell, *Proceedings of the National Academy of Sciences* **2009**, *106*, 21637; (b) T. Braun, M. Drescher, D. Summerer, *International Journal of Molecular Sciences* **2019**, *20*.
- [43] (a) L. Wang, A. Brock, P. G. Schultz, *Journal of the American Chemical Society* **2002**, *124*, 1836-1837; (b) H. S. Lee, J. Guo, E. A. Lemke, R. D. Dimla, P. G. Schultz, *Journal of the American Chemical Society* **2009**, *131*, 12921-12923; (c) J. Wang, J. Xie, P. G. Schultz, *Journal of the American Chemical Society* **2006**, *128*, 8738-8739; (d) D. Summerer, S. Chen, N. Wu, A. Deiters, J. W. Chin, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2006**, *103*, 9785; (e) G. Charbon, J. Wang, E. Brustad, P. G. Schultz, A. L. Horwich, C. Jacobs-Wagner, E. Chapman, *Bioorganic & Medicinal Chemistry Letters* **2011**, *21*, 6067-6070.
- [44] (a) A. Deiters, B. H. Geierstanger, P. G. Schultz, *ChemBiochem* **2005**, *6*, 55-58; (b) J. T. Hammill, S. Miyake-Stoner, J. L. Hazen, J. C. Jackson, R. A. Mehl, *Nature Protocols* **2007**, *2*, 2601; (c) J. C. Jackson, J. T. Hammill, R. A. Mehl, *Journal of the American Chemical Society* **2007**, *129*, 1160-1166; (d) S. E. Cellitti, D. H. Jones, L. Lagpacan, X. Hao, Q. Zhang, H. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. Ryu, T. Uno, P. G. Schultz, B. H. Geierstanger, *Journal of the American Chemical Society* **2008**, *130*, 9268-9281; (e) J. N. Lampe, R. Brandman, S. Sivaramakrishnan, P. R. O. de Montellano, *Journal of Biological Chemistry* **2010**, *285*, 9594-9603; (f) C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl, G. J. Pielak, *Journal of the American Chemical Society* **2010**, *132*, 321-327.
- [45] (a) K. C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera, P. G. Schultz, *Journal of the American Chemical Society* **2006**, *128*, 13984-13985; (b) S. Ye, T. Huber, R. Vogel, T. P. Sakmar, *Nature Chemical Biology* **2009**, *5*, 397; (c) S. Ye, E. Zaitseva, G. Caltabiano, G. F. X. Schertler, T. P. Sakmar, X. Deupi, R. Vogel, *Nature* **2010**, *464*, 1386.
- [46] (a) D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *Journal of the American Chemical Society* **2009**, *131*, 14194-14195; (b) D. P. Nguyen, M. M. G. Alai, S. Virdee, J. W. Chin, *Chemistry & Biology* **2010**, *17*, 1072-1076.
-

- [47] (a) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nature Chemical Biology* **2008**, *4*, 232; (b) H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, J. W. Chin, *Molecular Cell* **2009**, *36*, 153-163; (c) M. Lammers, H. Neumann, J. W. Chin, L. C. James, *Nature Chemical Biology* **2010**, *6*, 331.
- [48] (a) R. Serwa, I. Wilkening, G. Del Signore, M. Mühlberg, I. Claußnitzer, C. Weise, M. Gerrits, C. P. R. Hackenberger, *Angewandte Chemie International Edition* **2009**, *48*, 8234-8239; (b) J. Xie, L. Supekova, P. G. Schultz, *ACS Chemical Biology* **2007**, *2*, 474-478; (c) X. Luo, G. Fu, R. E. Wang, X. Zhu, C. Zambaldo, R. Liu, T. Liu, X. Lyu, J. Du, W. Xuan, A. Yao, S. A. Reed, M. Kang, Y. Zhang, H. Guo, C. Huang, P.-Y. Yang, I. A. Wilson, P. G. Schultz, F. Wang, *Nature Chemical Biology* **2017**, *13*, 845; (d) M. S. Zhang, S. F. Brunner, N. Huguenin-Dezot, A. D. Liang, W. H. Schmied, D. T. Rogerson, J. W. Chin, *Nature Methods* **2017**, *14*, 729.
- [49] (a) S. Virdee, P. B. Kapadnis, T. Elliott, K. Lang, J. Madrzak, D. P. Nguyen, L. Riechmann, J. W. Chin, *Journal of the American Chemical Society* **2011**, *133*, 10708-10711; (b) X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angewandte Chemie International Edition* **2009**, *48*, 9184-9187; (c) M. Fottner, A.-D. Brunner, V. Bittl, D. Horn-Ghetko, A. Jussupow, V. R. I. Kaila, A. Bremm, K. Lang, *Nature Chemical Biology* **2019**, *15*, 276-284.
- [50] (a) N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *Journal of the American Chemical Society* **2004**, *126*, 14306-14307; (b) A. Deiters, D. Groff, Y. Ryu, J. Xie, P. G. Schultz, *Angewandte Chemie International Edition* **2006**, *45*, 2728-2731; (c) E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, *Nature Chemical Biology* **2007**, *3*, 769; (d) A. Gautier, A. Deiters, J. W. Chin, *Journal of the American Chemical Society* **2011**, *133*, 2124-2127.
- [51] (a) M. Cigler, T.-A. Nguyen, K. Lang, in *Oxidative Folding of Proteins: Basic Principles, Cellular Regulation and Engineering*, The Royal Society of Chemistry, **2018**, pp. 399-420; (b) T.-A. Nguyen, M. Cigler, K. Lang, *Angewandte Chemie International Edition* **2018**, *57*, 14350-14361.
- [52] I. Coin, V. Katritch, T. Sun, Z. Xiang, Fai Y. Siu, M. Beyermann, Raymond C. Stevens, L. Wang, *Cell* **2013**, *155*, 1258-1269.
- [53] M. Cigler, T. G. Müller, D. Horn-Ghetko, M.-K. von Wrisberg, M. Fottner, R. S. Goody, A. Itzen, M. P. Müller, K. Lang, *Angewandte Chemie International Edition* **2017**, *56*, 15737-15741.
- [54] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, *Science* **2013**, *342*, 357.
- [55] E. Scolnick, R. Tompkins, T. Caskey, M. Nirenberg, *Proceedings of the National Academy of Sciences* **1968**, *61*, 768.
- [56] K. A. Odoi, Y. Huang, Y. H. Rezenom, W. R. Liu, *PloS one* **2013**, *8*, e57035-e57035.
- [57] Y. Zeng, W. Wang, W. R. Liu, *Chembiochem* **2014**, *15*, 1750-1754.
- [58] *Angewandte Chemie International Edition* **2010**, *49*, 3211-3214.
- [59] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441.
- [60] (a) K. Wang, H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nature Biotechnology* **2007**, *25*, 770; (b) S. D. Fried, W. H. Schmied, C. Uttamapinant, J. W. Chin, *Angew Chem Weinheim Bergstr Ger* **2015**, *127*, 12982-12985; (c) W. H. Schmied, Z. Tnimov, C. Uttamapinant, C. D. Rae, S. D. Fried, J. W. Chin, *Nature* **2018**, *564*, 444-448.
-



- [61] H. Xiao, A. Chatterjee, S.-h. Choi, K. M. Bajjuri, S. C. Sinha, P. G. Schultz, *Angewandte Chemie International Edition* **2013**, *52*, 14080-14083.
- [62] Y. Zheng, P. S. Addy, R. Mukherjee, A. Chatterjee, *Chemical Science* **2017**, *8*, 7211-7217.
- [63] (a) C. P. Ramil, Q. Lin, *Chemical Communications* **2013**, *49*, 11007-11022; (b) D. M. Patterson, L. A. Nazarova, J. A. Prescher, *ACS Chemical Biology* **2014**, *9*, 592-605.
- [64] (a) A. Janny, *Berichte der deutschen chemischen Gesellschaft* **1882**, *15*, 2778-2783; (b) E. Fischer, in *Untersuchungen über Triphenylmethanfarbstoffe Hydrazine und Indole* (Eds.: E. Fischer, M. Bergmann), Springer Berlin Heidelberg, Berlin, Heidelberg, **1924**, pp. 679-682; (c) E. G. Sander, W. P. Jencks, *Journal of the American Chemical Society* **1968**, *90*, 6154-6162.
- [65] L. K. Mahal, K. J. Yarema, C. R. Bertozzi, *Science* **1997**, *276*, 1125.
- [66] (a) R. Sadamoto, K. Niikura, T. Ueda, K. Monde, N. Fukuhara, S.-I. Nishimura, *Journal of the American Chemical Society* **2004**, *126*, 3755-3761; (b) L. Wang, Z. Zhang, A. Brock, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2003**, *100*, 56.
- [67] A. Dirksen, P. E. Dawson, *Bioconjug Chem* **2008**, *19*, 2543-2548.
- [68] (a) Y. Zeng, T. N. Ramya, A. Dirksen, P. E. Dawson, J. C. Paulson, *Nat Methods* **2009**, *6*, 207-209; (b) J. Rayo, N. Amara, P. Krief, M. M. Meijler, *J Am Chem Soc* **2011**, *133*, 7469-7475.
- [69] (a) P. Crisalli, E. T. Kool, *Org Lett* **2013**, *15*, 1646-1649; (b) M. Rashidian, M. M. Mahmoodi, R. Shah, J. K. Dozier, C. R. Wagner, M. D. Distefano, *Bioconjug Chem* **2013**, *24*, 333-342.
- [70] G. A. Lemieux, C. L. De Graffenried, C. R. Bertozzi, *J Am Chem Soc* **2003**, *125*, 4708-4709.
- [71] H. Staudinger, J. Meyer, *Helvetica Chimica Acta* **1919**, *2*, 635-646.
- [72] E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org Lett* **2000**, *2*, 2141-2143.
- [73] R. Huisgen, *Angewandte Chemie International Edition in English* **1963**, *2*, 565-598.
- [74] (a) V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew Chem Int Ed Engl* **2002**, *41*, 2596-2599; (b) C. W. Tornøe, C. Christensen, M. Meldal, *The Journal of Organic Chemistry* **2002**, *67*, 3057-3064.
- [75] T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Organic Letters* **2004**, *6*, 2853-2855.
- [76] B. T. Worrell, J. A. Malik, V. V. Fokin, *Science* **2013**, *340*, 457-460.
- [77] L. Zhang, X. Chen, P. Xue, H. H. Sun, I. D. Williams, K. B. Sharpless, V. V. Fokin, G. Jia, *J Am Chem Soc* **2005**, *127*, 15998-15999.
- [78] J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, *Proceedings of the National Academy of Sciences* **2007**, *104*, 16793.
- [79] (a) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Journal of the American Chemical Society* **2002**, *124*, 9026-9027; (b) F. Tian, M.-L. Tsao, P. G. Schultz, *Journal of the American Chemical Society* **2004**, *126*, 15962-15963.
- [80] (a) A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *Journal of the American Chemical Society* **2003**, *125*, 11782-11783; (b) T. S. Young, I. Ahmad, A. Brock, P. G. Schultz, *Biochemistry* **2009**, *48*, 2643-2653; (c) W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, *Nature Methods* **2007**, *4*, 239.
-

- [81] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *Journal of the American Chemical Society* **2009**, *131*, 8720-8721.
- [82] T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angewandte Chemie International Edition* **2011**, *50*, 3878-3881.
- [83] (a) K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, J. W. Chin, *Journal of the American Chemical Society* **2012**, *134*, 10317-10320; (b) A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. M. Verkade, M. Wießler, C. Schultz, J. C. M. van Hest, F. L. van Delft, E. A. Lemke, *Chembiochem* **2012**, *13*, 2094-2099.
- [84] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew Chem Int Ed Engl* **2001**, *40*, 2004-2021.
- [85] (a) J.-F. Lutz, *Angewandte Chemie International Edition* **2007**, *46*, 1018-1025; (b) M. Meldal, C. W. Tornøe, *Chemical Reviews* **2008**, *108*, 2952-3015.
- [86] T. Chiara, P. Anna Maria, C. Michael, R. Paolo, *Current Topics in Medicinal Chemistry* **2018**, *18*, 591-610.
- [87] N. K. Devaraj, J. P. Collman, *QSAR & Combinatorial Science* **2007**, *26*, 1253-1260.
- [88] H. C. Kolb, K. B. Sharpless, *Drug Discov Today* **2003**, *8*, 1128-1137.
- [89] R. Breinbauer, M. Köhn, *Chembiochem* **2003**, *4*, 1147-1149.
- [90] N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo, C. R. Bertozzi, *ACS Chem Biol* **2006**, *1*, 644-648.
- [91] A. J. Link, D. A. Tirrell, *Journal of the American Chemical Society* **2003**, *125*, 11164-11165.
- [92] D. C. Kennedy, C. S. McKay, M. C. Legault, D. C. Danielson, J. A. Blake, A. F. Pegoraro, A. Stolow, Z. Mester, J. P. Pezacki, *J Am Chem Soc* **2011**, *133*, 17993-18001.
- [93] C. R. Becer, R. Hoogenboom, U. S. Schubert, *Angewandte Chemie International Edition* **2009**, *48*, 4900-4908.
- [94] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J Am Chem Soc* **2004**, *126*, 15046-15047.
- [95] K. J. Shea, J. S. Kim, *Journal of the American Chemical Society* **1992**, *114*, 3044-3051.
- [96] J. Dommerholt, S. Schmidt, R. Temming, L. J. Hendriks, F. P. Rutjes, J. C. van Hest, D. J. Lefeber, P. Friedl, F. L. van Delft, *Angew Chem Int Ed Engl* **2010**, *49*, 9422-9425.
- [97] (a) X. Ning, J. Guo, M. A. Wolfert, G.-J. Boons, *Angewandte Chemie International Edition* **2008**, *47*, 2253-2255; (b) N. E. Mbua, J. Guo, M. A. Wolfert, R. Steet, G.-J. Boons, *Chembiochem* **2011**, *12*, 1912-1921.
- [98] (a) J. C. Jewett, E. M. Sletten, C. R. Bertozzi, *Journal of the American Chemical Society* **2010**, *132*, 3688-3690; (b) C. G. Gordon, J. L. Mackey, J. C. Jewett, E. M. Sletten, K. N. Houk, C. R. Bertozzi, *Journal of the American Chemical Society* **2012**, *134*, 9199-9208.
- [99] M. Chigrinova, C. S. McKay, L.-P. B. Beaulieu, K. A. Udachin, A. M. Beauchemin, J. P. Pezacki, *Organic & Biomolecular Chemistry* **2013**, *11*, 3436-3441.
- [100] (a) K. E. Beatty, J. D. Fisk, B. P. Smart, Y. Y. Lu, J. Szychowski, M. J. Hangauer, J. M. Baskin, C. R. Bertozzi, D. A. Tirrell, *Chembiochem* **2010**, *11*, 2092-2095; (b) P. V. Chang, J. A. Prescher, E. M. Sletten, J. M. Baskin, I. A. Miller, N. J. Agard, A. Lo, C. R. Bertozzi, *Proceedings of the National Academy of Sciences* **2010**, *107*, 1821.
- [101] S. T. Laughlin, J. M. Baskin, S. L. Amacher, C. R. Bertozzi, *Science* **2008**, *320*, 664.
-

- [102] A. A. Poloukhine, N. E. Mbua, M. A. Wolfert, G. J. Boons, V. V. Popik, *J Am Chem Soc* **2009**, *131*, 15769-15776.
- [103] J. A. Codelli, J. M. Baskin, N. J. Agard, C. R. Bertozzi, *Journal of the American Chemical Society* **2008**, *130*, 11486-11493.
- [104] (a) X. Ning, R. P. Temming, J. Dommerholt, J. Guo, D. B. Ania, M. F. Debets, M. A. Wolfert, G.-J. Boons, F. L. van Delft, *Angewandte Chemie International Edition* **2010**, *49*, 3065-3068; (b) C. S. McKay, J. Moran, J. P. Pezacki, *Chemical Communications* **2010**, *46*, 931-933.
- [105] C. S. McKay, J. A. Blake, J. Cheng, D. C. Danielson, J. P. Pezacki, *Chemical Communications* **2011**, *47*, 10040-10042.
- [106] (a) J. E. Baldwin, R. G. Pudussery, A. K. Qureshi, B. Sklarz, *Journal of the American Chemical Society* **1968**, *90*, 5325-5326; (b) D. A. MacKenzie, A. R. Sherratt, M. Chigrinova, L. L. W. Cheung, J. P. Pezacki, *Current Opinion in Chemical Biology* **2014**, *21*, 81-88.
- [107] J. M. Palomo, *European Journal of Organic Chemistry* **2010**, *2010*, 6303-6314.
- [108] O. Diels, K. Alder, *Justus Liebigs Annalen der Chemie* **1928**, *460*, 98-122.
- [109] K. C. Nicolaou, S. A. Snyder, T. Montagnon, G. Vassilikogiannakis, *Angewandte Chemie International Edition* **2002**, *41*, 1668-1698.
- [110] S. Otto, B. F. N. Engberts Jan, in *Pure and Applied Chemistry, Vol. 72*, **2000**, p. 1365.
- [111] O. Boutureira, G. J. L. Bernardes, *Chemical Reviews* **2015**, *115*, 2174-2195.
- [112] (a) S. Mayer, K. Lang, *Synthesis* **2017**, *49*, 830-848; (b) B. L. Oliveira, Z. Guo, G. J. L. Bernardes, *Chemical Society Reviews* **2017**, *46*, 4895-4950.
- [113] (a) E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angewandte Chemie International Edition* **2012**, *51*, 4466-4469; (b) T. Plass, S. Milles, C. Koehler, J. Szymański, R. Mueller, M. Wießler, C. Schultz, E. A. Lemke, *Angewandte Chemie International Edition* **2012**, *51*, 4166-4170.
- [114] T. S. Elliott, F. M. Townsley, A. Bianco, R. J. Ernst, A. Sachdeva, S. J. Elsässer, L. Davis, K. Lang, R. Pisa, S. Greiss, K. S. Lilley, J. W. Chin, *Nature Biotechnology* **2014**, *32*, 465.
- [115] C. Uttamapinant, J. D. Howe, K. Lang, V. Beránek, L. Davis, M. Mahesh, N. P. Barry, J. W. Chin, *Journal of the American Chemical Society* **2015**, *137*, 4602-4605.
- [116] N. K. Devaraj, S. Hilderbrand, R. Upadhyay, R. Mazitschek, R. Weissleder, *Angewandte Chemie International Edition* **2010**, *49*, 2869-2872.
- [117] A. Darko, S. Wallace, O. Dmitrenko, M. M. Machovina, R. A. Mehl, J. W. Chin, J. M. Fox, *Chemical Science* **2014**, *5*, 3770-3776.
- [118] (a) J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox, R. A. Mehl, *Journal of the American Chemical Society* **2012**, *134*, 2898-2901; (b) R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li, R. A. Mehl, *Journal of the American Chemical Society* **2015**, *137*, 10044-10047.
- [119] (a) T. S. Elliott, A. Bianco, J. W. Chin, *Current Opinion in Chemical Biology* **2014**, *21*, 154-160; (b) T. S. Elliott, A. Bianco, F. M. Townsley, S. D. Fried, J. W. Chin, *Cell Chem Biol* **2016**, *23*, 805-815.
- [120] Y.-H. Tsai, S. Essig, J. R. James, K. Lang, J. W. Chin, *Nature Chemistry* **2015**, *7*, 554.
- [121] (a) A. Herner, Q. Lin, *Topics in Current Chemistry* **2015**, *374*, 1; (b) P. Kumar, S. T. Laughlin, in *Methods in Enzymology, Vol. 622* (Ed.: A. K. Shukla), Academic Press, **2019**, pp. 153-182.
-

- [122] (a) D. H. Wadsworth, B. A. Donatelli, *Synthesis* **1981**, 1981, 285-286; (b) A. Poloukhine, V. V. Popik, *The Journal of Organic Chemistry* **2003**, 68, 7833-7840.
- [123] (a) C. W. Riggsbee, A. Deiters, *Trends in Biotechnology* **2010**, 28, 468-475; (b) W. Szymański, J. M. Beierle, H. A. V. Kistemaker, W. A. Velema, B. L. Feringa, *Chemical Reviews* **2013**, 113, 6114-6178; (c) M. Zhu, H. Zhou, *Organic & Biomolecular Chemistry* **2018**, 16, 8434-8445; (d) C. Brieke, F. Rohrbach, A. Gottschalk, G. Mayer, A. Heckel, *Angewandte Chemie International Edition* **2012**, 51, 8446-8476; (e) G. Mayer, A. Heckel, *Angewandte Chemie International Edition* **2006**, 45, 4900-4921.
- [124] S. Nainar, M. Kubota, C. McNitt, C. Tran, V. V. Popik, R. C. Spitale, *Journal of the American Chemical Society* **2017**, 139, 8090-8093.
- [125] C. D. McNitt, H. Cheng, S. Ullrich, V. V. Popik, M. Bjerknes, *Journal of the American Chemical Society* **2017**, 139, 14029-14032.
- [126] (a) R. Huisgen, M. Seidel, G. Wallbillich, H. Knupfer, *Tetrahedron* **1962**, 17, 3-29; (b) J. S. Clovis, A. Eckell, R. Huisgen, R. Sustmann, *Chemische Berichte* **1967**, 100, 60-70.
- [127] Y. Wang, C. I. Rivera Vera, Q. Lin, *Organic Letters* **2007**, 9, 4155-4158.
- [128] (a) Y. Wang, W. J. Hu, W. Song, R. K. V. Lim, Q. Lin, *Organic Letters* **2008**, 10, 3725-3728; (b) Z. Yu, L. Y. Ho, Z. Wang, Q. Lin, *Bioorganic & Medicinal Chemistry Letters* **2011**, 21, 5033-5036.
- [129] (a) Y. Wang, Q. Lin, *Organic Letters* **2009**, 11, 3570-3573; (b) J. Wang, W. Zhang, W. Song, Y. Wang, Z. Yu, J. Li, M. Wu, L. Wang, J. Zang, Q. Lin, *Journal of the American Chemical Society* **2010**, 132, 14812-14818; (c) Z. Yu, Y. Pan, Z. Wang, J. Wang, Q. Lin, *Angewandte Chemie International Edition* **2012**, 51, 10600-10604; (d) Z. Yu, Q. Lin, *Journal of the American Chemical Society* **2014**, 136, 4153-4156; (e) Z. Yu, R. K. V. Lim, Q. Lin, *Chemistry – A European Journal* **2010**, 16, 13325-13329.
- [130] Y.-J. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Fang, K. L. Wooley, W. R. Liu, *ACS Chemical Biology* **2013**, 8, 1664-1670.
- [131] M. T. Taylor, M. L. Blackman, O. Dmitrenko, J. M. Fox, *Journal of the American Chemical Society* **2011**, 133, 9646-9649.
- [132] W. Song, Y. Wang, J. Qu, M. M. Madden, Q. Lin, *Angewandte Chemie International Edition* **2008**, 47, 2832-2835.
- [133] P. An, Z. Yu, Q. Lin, *Org Lett* **2013**, 15, 5496-5499.
- [134] Z. Yu, T. Y. Ohulchanskyy, P. An, P. N. Prasad, Q. Lin, *J Am Chem Soc* **2013**, 135, 16766-16769.
- [135] Y. Zhang, W. Liu, K. Z. Zhao, *Molecules* **2014**, 19.
- [136] J. L. Segura, N. Martín, *Chemical Reviews* **1999**, 99, 3199-3246.
- [137] (a) S. Arumugam, V. V. Popik, *Journal of the American Chemical Society* **2011**, 133, 5573-5579; (b) S. Arumugam, V. V. Popik, *Journal of the American Chemical Society* **2009**, 131, 11892-11899.
- [138] J. Li, H. Kong, L. Huang, B. Cheng, K. Qin, M. Zheng, Z. Yan, Y. Zhang, *Journal of the American Chemical Society* **2018**, 140, 14542-14546.
- [139] T. Posner, *Berichte der deutschen chemischen Gesellschaft* **1905**, 38, 646-657.
- [140] (a) A. B. Lowe, *Polymer Chemistry* **2010**, 1, 17-36; (b) C. E. Hoyle, C. N. Bowman, *Angewandte Chemie International Edition* **2010**, 49, 1540-1573.
- [141] D. P. Nair, M. Podgórski, S. Chatani, T. Gong, W. Xi, C. R. Fenoli, C. N. Bowman, *Chemistry of Materials* **2014**, 26, 724-744.
-

- [142] (a) B. D. Fairbanks, T. F. Scott, C. J. Kloxin, K. S. Anseth, C. N. Bowman, *Macromolecules* **2009**, *42*, 211-217; (b) A. B. Lowe, C. E. Hoyle, C. N. Bowman, *Journal of Materials Chemistry* **2010**, *20*, 4745-4750; cA. B. Lowe, *Polymer* **2014**, *55*, 5517-5549.
- [143] P. Jonkheijm, D. Weinrich, M. Köhn, H. Engelkamp, P. C. M. Christianen, J. Kuhlmann, J. C. Maan, D. Nüsse, H. Schroeder, R. Wacker, R. Breinbauer, C. M. Niemeyer, H. Waldmann, *Angewandte Chemie International Edition* **2008**, *47*, 4421-4424.
- [144] N. Floyd, B. Vijayakrishnan, J. R. Koepe, B. G. Davis, *Angewandte Chemie International Edition* **2009**, *48*, 7798-7802.
- [145] F. Li, A. Allahverdi, R. Yang, G. B. J. Lua, X. Zhang, Y. Cao, N. Korolev, L. Nordenskiöld, C.-F. Liu, *Angewandte Chemie International Edition* **2011**, *50*, 9611-9614.
- [146] Y. Li, M. Yang, Y. Huang, X. Song, L. Liu, P. R. Chen, *Chemical Science* **2012**, *3*, 2766-2770.
- [147] Y. Li, M. Pan, Y. Li, Y. Huang, Q. Guo, *Org Biomol Chem* **2013**, *11*, 2624-2629.
- [148] (a) H. Liu, H. Chung, *ACS Sustainable Chemistry & Engineering* **2017**, *5*, 9160-9168; (b) M. Teders, C. Henkel, L. Anhäuser, F. Strieth-Kalthoff, A. Gómez-Suárez, R. Kleinmans, A. Kahnt, A. Rentmeister, D. Guldi, F. Glorius, *Nature Chemistry* **2018**, *10*, 981-988.
- [149] M. V. Trivedi, J. S. Laurence, T. J. Siahaan, *Curr Protein Pept Sci* **2009**, *10*, 614-625.
- [150] H. J. Kang, E. N. Baker, *Trends Biochem Sci* **2011**, *36*, 229-237.
- [151] (a) W. H. Dunham, M. Mullin, A.-C. Gingras, *PROTEOMICS* **2012**, *12*, 1576-1590; (b) L. Trinkle-Mulcahy, *PROTEOMICS* **2012**, *12*, 1623-1638.
- [152] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Alcala, J. Lim, C. Fraughton, E. Llamasas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, M. Vidal, *Nature* **2005**, *437*, 1173-1178.
- [153] (a) M. Kim, C. V. Carman, T. A. Springer, *Science* **2003**, *301*, 1720; (b) R. A. Scheck, M. A. Lowder, J. S. Appelbaum, A. Schepartz, *ACS Chemical Biology* **2012**, *7*, 1367-1376.
- [154] Y. Yang, H. Song, P. R. Chen, *IUBMB Life* **2016**, *68*, 879-886.
- [155] J. Petschnigg, J. Snider, I. Stagljar, *Current Opinion in Biotechnology* **2011**, *22*, 50-58.
- [156] (a) A. A. Dombkowski, K. Z. Sultana, D. B. Craig, *FEBS Letters* **2014**, *588*, 206-212; (b) I. Kufareva, M. Gustavsson, L. G. Holden, L. Qin, Y. Zheng, T. M. Handel, in *Methods in Enzymology, Vol. 570* (Ed.: T. M. Handel), Academic Press, **2016**, pp. 389-420; (c) L. Qin, I. Kufareva, L. G. Holden, C. Wang, Y. Zheng, C. Zhao, G. Fenalti, H. Wu, G. W. Han, V. Cherezov, R. Abagyan, R. C. Stevens, T. M. Handel, *Science* **2015**, *347*, 1117; dH. Huang, R. Chopra, G. L. Verdine, S. C. Harrison, *Science* **1998**, *282*, 1669.
- [157] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2002**, *99*, 11020.
- [158] (a) E. M. Tippmann, W. Liu, D. Summerer, A. V. Mack, P. G. Schultz, *Chembiochem* **2007**, *8*, 2210-2214; (b) H.-w. Ai, W. Shen, A. Sagi, P. R. Chen, P. G. Schultz,

- Chembiochem* **2011**, *12*, 1854-1857; (c) C. Chou, R. Uprety, L. Davis, J. W. Chin, A. Deiters, *Chemical Science* **2011**, *2*, 480-483; dM. Zhang, S. Lin, X. Song, J. Liu, Y. Fu, X. Ge, X. Fu, Z. Chang, P. R. Chen, *Nature Chemical Biology* **2011**, *7*, 671.
- [159] (a) Y. Tanaka, M. R. Bond, J. J. Kohler, *Molecular BioSystems* **2008**, *4*, 473-480; (b) G. W. Preston, A. J. Wilson, *Chemical Society Reviews* **2013**, *42*, 3289-3301.
- [160] M. J. Long, J. R. Poganik, Y. Aye, *J Am Chem Soc* **2016**, *138*, 3610-3622.
- [161] A. J. Chmura, M. S. Orton, C. F. Meares, *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 8480-8484.
- [162] Z. Xiang, H. Ren, Y. S. Hu, I. Coin, J. Wei, H. Cang, L. Wang, *Nat Methods* **2013**, *10*, 885-888.
- [163] (a) Z. Xiang, V. K. Lacey, H. Ren, J. Xu, D. J. Burban, P. A. Jennings, L. Wang, *Angew Chem Int Ed Engl* **2014**, *53*, 2190-2193; (b) X. H. Chen, Z. Xiang, Y. S. Hu, V. K. Lacey, H. Cang, L. Wang, *ACS Chem Biol* **2014**, *9*, 1956-1961; (c) C. Hoppmann, V. K. Lacey, G. V. Louie, J. Wei, J. P. Noel, L. Wang, *Angew Chem Int Ed Engl* **2014**, *53*, 3932-3936; (d) C. Hoppmann, I. Maslennikov, S. Choe, L. Wang, *J Am Chem Soc* **2015**, *137*, 11218-11221; (e) T. Kobayashi, C. Hoppmann, B. Yang, L. Wang, *J Am Chem Soc* **2016**, *138*, 14832-14835; (f) B. Yang, S. Tang, C. Ma, S.-T. Li, G.-C. Shao, B. Dang, W. F. DeGrado, M.-Q. Dong, P. G. Wang, S. Ding, L. Wang, *Nature Communications* **2017**, *8*, 2240.
- [164] J. L. Furman, M. Kang, S. Choi, Y. Cao, E. D. Wold, S. B. Sun, V. V. Smider, P. G. Schultz, C. H. Kim, *J Am Chem Soc* **2014**, *136*, 8411-8417.
- [165] W. Xuan, J. Li, X. Luo, P. G. Schultz, *Angew Chem Int Ed Engl* **2016**, *55*, 10065-10068.
- [166] W. Xuan, S. Shao, P. G. Schultz, *Angew Chem Int Ed Engl* **2017**, *56*, 5096-5100.
- [167] N. Wang, B. Yang, C. Fu, H. Zhu, F. Zheng, T. Kobayashi, J. Liu, S. Li, C. Ma, P. G. Wang, Q. Wang, L. Wang, *Journal of the American Chemical Society* **2018**, *140*, 4995-4999.
- [168] J. Liu, S. Li, N. A. Aslam, F. Zheng, B. Yang, R. Cheng, N. Wang, S. Rozovsky, P. G. Wang, Q. Wang, L. Wang, *Journal of the American Chemical Society* **2019**, *141*, 9458-9462.
- [169] (a) H. Zhang, W. S. Trout, S. Liu, G. A. Andrade, D. A. Hudson, S. L. Scinto, K. T. Dicker, Y. Li, N. Lazouski, J. Rosenthal, C. Thorpe, X. Jia, J. M. Fox, *Journal of the American Chemical Society* **2016**, *138*, 5978-5983; (b) P. Kumar, T. Jiang, S. Li, O. Zainul, S. T. Laughlin, *Organic & Biomolecular Chemistry* **2018**, *16*, 4081-4085; (c) P. Kumar, O. Zainul, F. M. Camarda, T. Jiang, J. A. Mannone, W. Huang, S. T. Laughlin, *Organic Letters* **2019**, *21*, 3721-3725; (d) S. V. Mayer, A. Murnauer, M.-K. von Wrisberg, M.-L. Jokisch, K. Lang, *Angewandte Chemie International Edition* **2019**, *0*.
- [170] T. Gruending, K. K. Oehlenschlaeger, E. Frick, M. Glassner, C. Schmid, C. Barner-Kowollik, *Macromolecular Rapid Communications* **2011**, *32*, 807-812.
- [171] (a) T. Pauloehrl, G. Delaittre, V. Winkler, A. Welle, M. Bruns, H. G. Börner, A. M. Greiner, M. Bastmeyer, C. Barner-Kowollik, *Angewandte Chemie International Edition* **2012**, *51*, 1071-1074; (b) M. Winkler, J. O. Mueller, K. K. Oehlenschlaeger, L. Montero de Espinosa, M. A. R. Meier, C. Barner-Kowollik, *Macromolecules* **2012**, *45*, 5012-5019.
- [172] (a) P. Wan, W. Brousmiche Darryl, Z. Chen Christy, J. Cole, M. Lukeman, M. Xu, in *Pure and Applied Chemistry*, Vol. 73, **2001**, p. 529; (b) P. Wan, B. Barker, L. Diao, M. Fischer, Y. Shi, C. Yang, *Canadian Journal of Chemistry* **1996**, *74*, 465-475; (c) R.

- W. Van De Water, T. R. R. Pettus, *Tetrahedron* **2002**, *58*, 5367-5405; (d) D. C. Thompson, J. A. Thompson, M. Sugumaran, P. Moldéus, *Chemico-Biological Interactions* **1993**, *86*, 129-162; eB. Judy Louise, *Current Organic Chemistry* **2014**, *18*, 61-69.
- [173] (a) G. Gaudiano, M. Frigerio, P. Bravo, T. H. Koch, *Journal of the American Chemical Society* **1990**, *112*, 6704-6709; (b) I. Han, D. J. Russell, H. Kohn, *The Journal of Organic Chemistry* **1992**, *57*, 1799-1807; (c) M. Tomasz, A. Das, K. S. Tang, M. G. J. Ford, A. Minnock, S. M. Musser, M. J. Waring, *Journal of the American Chemical Society* **1998**, *120*, 11581-11593; (d) S. M. Kupchan, A. Karim, C. Marcks, *Journal of the American Chemical Society* **1968**, *90*, 5923-5924; (e) H. B. Bode, A. Zeeck, *Journal of the Chemical Society, Perkin Transactions 1* **2000**, 323-328.
- [174] (a) J. L. Goodman, K. S. Peters, P. M. Lahti, J. A. Berson, *Journal of the American Chemical Society* **1985**, *107*, 276-277; (b) M. I. Khan, J. L. Goodman, *Journal of the American Chemical Society* **1994**, *116*, 10342-10343.
- [175] (a) E. Dorrestijn, M. Kranenburg, M. V. Ciriano, P. Mulder, *The Journal of Organic Chemistry* **1999**, *64*, 3012-3018; (b) M. Yato, T. Ohwada, K. Shudo, *Journal of the American Chemical Society* **1990**, *112*, 5341-5342.
- [176] (a) D. A. Bolon, *The Journal of Organic Chemistry* **1970**, *35*, 3666-3670; (b) C. D. Cook, R. G. Inskeep, A. S. Rosenberg, E. C. Curtis, *Journal of the American Chemical Society* **1955**, *77*, 1672-1674; cC. D. Cook, B. E. Norcross, *Journal of the American Chemical Society* **1956**, *78*, 3797-3799.
- [177] K. E. O'Shea, M. A. Fox, *Journal of the American Chemical Society* **1991**, *113*, 611-615.
- [178] M. Toteva, J. Richard, *Advances in physical organic chemistry* **2011**, *45*, 39-91.
- [179] (a) K. Nakatani, N. Higashida, I. Saito, *Tetrahedron Letters* **1997**, *38*, 5005-5008; (b) E. A. Leo, J. Delgado, L. R. Domingo, A. Espinós, M. A. Miranda, R. Tormos, *The Journal of Organic Chemistry* **2003**, *68*, 9643-9647; (c) Y. Chiang, A. J. Kresge, Y. Zhu, *Journal of the American Chemical Society* **2000**, *122*, 9854-9855; (d) L. Diao, C. Yang, P. Wan, *Journal of the American Chemical Society* **1995**, *117*, 5369-5370.
- [180] (a) P. Wan, B. Chak, *Journal of the Chemical Society, Perkin Transactions 2* **1986**, 1751-1756; (b) B. Barker, L. Diao, P. Wan, *Journal of Photochemistry and Photobiology A: Chemistry* **1997**, *104*, 91-96.
- [181] (a) A. P. Kostikov, V. V. Popik, *Organic Letters* **2008**, *10*, 5277-5280; (b) A. P. Kostikov, V. V. Popik, *The Journal of Organic Chemistry* **2007**, *72*, 9190-9194.
- [182] A. Kulikov, S. Arumugam, V. V. Popik, *The Journal of Organic Chemistry* **2008**, *73*, 7611-7615.
- [183] (a) S. Arumugam, V. V. Popik, *Journal of the American Chemical Society* **2011**, *133*, 15730-15736; (b) S. Arumugam, J. Guo, N. E. Mbua, F. Friscourt, N. Lin, E. Nekongo, G.-J. Boons, V. V. Popik, *Chemical Science* **2014**, *5*, 1591-1598.
- [184] E. Modica, R. Zanaletti, M. Freccero, M. Mella, *The Journal of Organic Chemistry* **2001**, *66*, 41-52.
- [185] (a) P. G. McCracken, J. L. Bolton, G. R. J. Thatcher, *The Journal of Organic Chemistry* **1997**, *62*, 1820-1825; (b) J. L. Bolton, S. B. Turnipseed, J. A. Thompson, *Chemico-Biological Interactions* **1997**, *107*, 185-200; (c) D. Cabaret, S. A. Adediran, M. J. Garcia Gonzalez, R. F. Pratt, M. Wakselman, *J Org Chem* **1999**, *64*, 713-720.
- [186] W. F. Veldhuyzen, A. J. Shallop, R. A. Jones, S. E. Rokita, *Journal of the American Chemical Society* **2001**, *123*, 11126-11132.
-

- [187] (a) S. N. Richter, S. Maggi, S. C. Mels, M. Palumbo, M. Freccero, *Journal of the American Chemical Society* **2004**, *126*, 13973-13979; (b) D. Verga, S. N. Richter, M. Palumbo, R. Gandolfi, M. Freccero, *Organic & Biomolecular Chemistry* **2007**, *5*, 233-235; (c) W. Ping, S. Yang, Z. Lixia, H. Hanping, Z. Xiang, *Current Medicinal Chemistry* **2005**, *12*, 2893-2913.
- [188] (a) S. Arumugam, S. V. Orski, J. Locklin, V. V. Popik, *Journal of the American Chemical Society* **2012**, *134*, 179-182; (b) S. Arumugam, V. V. Popik, *Journal of the American Chemical Society* **2012**, *134*, 8408-8411.
- [189] D. Summerer, M. J. Schmidt, *Vol. EP3024823 (A2)*, **2015**.
- [190] (a) T. Itoh, *Progress in Polymer Science* **2001**, *26*, 1019-1059; (b) J. Dolenc, B. Sket, M. Strlic, *Tetrahedron Letters* **2002**, *43*, 5669-5671.
- [191] M. Ashram, S. Mizyed, P. E. Georghiou, *The Journal of Organic Chemistry* **2001**, *66*, 1473-1479.
- [192] R. A. Murphy, H. F. Kung, M. P. Kung, J. Billings, *Journal of Medicinal Chemistry* **1990**, *33*, 171-178.
- [193] M. Murata, T. Oyama, S. Watanabe, Y. Masuda, *The Journal of Organic Chemistry* **2000**, *65*, 164-168.
- [194] A. L. S. Thompson, G. W. Kabalka, M. R. Akula, J. W. Huffman, *Synthesis* **2005**, *2005*, 547-550.
- [195] (a) S. H. Schroeter, C. M. Orlando, *The Journal of Organic Chemistry* **1969**, *34*, 1181-1187; (b) K. Shima, H. Sakurai, *Bulletin of the Chemical Society of Japan* **1969**, *42*, 849-849; (c) T. H. Morris, E. H. Smith, R. Walsh, *Journal of the Chemical Society, Chemical Communications* **1987**, 964-965; (d) M. D'Auria, *Photochemical & Photobiological Sciences* **2019**.
- [196] (a) S. Sueur, M. Lagrenee, F. Abraham, C. Bremard, *Journal of Heterocyclic Chemistry* **1987**, *24*, 1285-1289; (b) D. L. Boger, S. M. Sakya, *The Journal of Organic Chemistry* **1988**, *53*, 1415-1423; (c) J. Sauer, D. K. Heldmann, J. Hetzenegger, J. Krauthan, H. Sichert, J. Schuster, *European Journal of Organic Chemistry* **1998**, *1998*, 2885-2896; (d) D. Che, T. Wegge, M. T. Stubbs, G. Seitz, H. Meier, C. Methfessel, *Journal of Medicinal Chemistry* **2001**, *44*, 47-57; (e) A. Hamasaki, R. Ducray, D. L. Boger, *The Journal of Organic Chemistry* **2006**, *71*, 185-193.
- [197] (a) H. Wu, S. C. Alexander, S. Jin, N. K. Devaraj, *Journal of the American Chemical Society* **2016**, *138*, 11429-11432; (b) E. Jiménez-Moreno, Z. Guo, B. L. Oliveira, I. S. Albuquerque, A. Kitowski, A. Guerreiro, O. Boutureira, T. Rodrigues, G. Jiménez-Osés, G. J. L. Bernardes, *Angewandte Chemie International Edition* **2017**, *56*, 243-247.
- [198] (a) Q. Li, T. Dong, X. Liu, X. Lei, *Journal of the American Chemical Society* **2013**, *135*, 4996-4999; (b) X. Zhang, T. Dong, Q. Li, X. Liu, L. Li, S. Chen, X. Lei, *ACS Chemical Biology* **2015**, *10*, 1676-1683.
- [199] A. Gubu, L. Li, Y. Ning, X. Zhang, S. Lee, M. Feng, Q. Li, X. Lei, K. Jo, X. Tang, *Chemistry – A European Journal* **2018**, *24*, 5895-5900.
- [200] Z. Song, M. Mertzman, D. L. Hughes, *Journal of Heterocyclic Chemistry* **1993**, *30*, 17-21.
- [201] (a) W. König, *Berichte der deutschen chemischen Gesellschaft (A and B Series)* **1923**, *56*, 1853-1855; (b) T. P. Forrest, G. A. Dauphinee, W. F. Miles, *Canadian Journal of Chemistry* **1969**, *47*, 2121-2122; (c) S. E. Denmark, S. Venkatraman, *The Journal of Organic Chemistry* **2006**, *71*, 1668-1676.
-



- [202] Y. Jiang, L. Xu, C. Zhou, D. Ma, in *C-H and C-X Bond Functionalization: Transition Metal Mediation*, The Royal Society of Chemistry, **2013**, pp. 1-45.
- [203] (a) K. Maruyama, M. Muraoka, Y. Naruta, *The Journal of Organic Chemistry* **1981**, *46*, 983-989; (b) K. Maruyama, T. Iwai, Y. Naruta, *Chemistry Letters* **1975**, *4*, 1219-1222; (c) Y. L. Chow, T. C. Joseph, H. H. Quon, J. N. S. Tam, *Canadian Journal of Chemistry* **1970**, *48*, 3045-3052; (d) T. Sasaki, K. Kanematsu, I. Ando, O. Yamashita, *Journal of the American Chemical Society* **1977**, *99*, 871-877.
- [204] F. W. Lichtenthaler, T. Weimer, S. Immel, *Tetrahedron: Asymmetry* **2004**, *15*, 2703-2709.
- [205] (a) Y. Zhang, L. Wang, M. Zhang, H.-K. Fun, J.-H. Xu, *Organic Letters* **2004**, *6*, 4893-4895; (b) L. Wang, Y.-C. Huang, Y. Liu, H.-K. Fun, Y. Zhang, J.-H. Xu, *The Journal of Organic Chemistry* **2010**, *75*, 7757-7768.
- [206] J.-H. Ho, T.-I. Ho, T.-H. Chen, Y. L. Chow, *Journal of Photochemistry and Photobiology A: Chemistry* **2001**, *138*, 111-122.
- [207] (a) T. Matsunaga, M. Arakaki, T. Kamiya, S. Endo, O. El-Kabbani, A. Hara, *Chemico-biological interactions* **2009**, *181*, 52-60; (b) T. Matsunaga, M. Hosogai, M. Arakaki, S. Endo, O. El-Kabbani, A. Hara, *Biological & pharmaceutical bulletin* **2012**, *35*, 1598-1602; (c) T. Matsunaga, K. Kamase, H. Takasawa, Y. Yamaji, S. Endo, O. El-Kabbani, A. Ikari, *Chemico-biological interactions* **2017**, *279*; (d) M. Yang, H. Ahmed, W. Wu, B. Jiang, Z. Jia, *Biomed Res Int* **2018**, *2018*, 9523968-9523968.
- [208] D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, *Science* **2002**, *296*, 913.
- [209] (a) H. Plutner, A. D. Cox, S. Pind, R. Khosravi-Far, J. R. Bourne, R. Schwaninger, C. J. Der, W. E. Balch, *The Journal of Cell Biology* **1991**, *115*, 31; (b) J. Cherfils, M. Zeghouf, *Physiological Reviews* **2013**, *93*, 269-309.
- [210] I. R. Vetter, A. Wittinghofer, *Science* **2001**, *294*, 1299.
- [211] R. R. Isberg, T. J. O'Connor, M. Heidtman, *Nature Reviews Microbiology* **2008**, *7*, 13.
- [212] T. Murata, A. Delprato, A. Ingmundson, D. K. Toomre, D. G. Lambright, C. R. Roy, *Nature Cell Biology* **2006**, *8*, 971-977.
- [213] S. Schoebel, L. K. Oesterlin, W. Blankenfeldt, R. S. Goody, A. Itzen, *Molecular Cell* **2009**, *36*, 1060-1072.
- [214] B. Schwer, J. Bunkenborg, R. O. Verdin, J. S. Andersen, E. Verdin, *Proceedings of the National Academy of Sciences* **2006**, *103*, 10224.
- [215] Q. Xiao, F. Zhang, B. A. Nacev, J. O. Liu, D. Pei, *Biochemistry* **2010**, *49*, 5588-5599.
- [216] D. Wiegandt, S. Vieweg, F. Hofmann, D. Koch, F. Li, Y.-W. Wu, A. Itzen, M. P. Müller, R. S. Goody, *Nature Communications* **2015**, *6*, 7773.
- [217] A. Meister, *Journal of Biological Chemistry* **1988**, *263*, 17205-17208.
- [218] (a) R. C. Fahey, W. C. Brown, W. B. Adams, M. B. Worsham, *Journal of Bacteriology* **1978**, *133*, 1126; (b) B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, J. D. Rabinowitz, *Nature Chemical Biology* **2009**, *5*, 593.
- [219] W. H. Schmied, S. J. Elsässer, C. Uttamapinant, J. W. Chin, *Journal of the American Chemical Society* **2014**, *136*, 15577-15583.
- [220] M. P. Müller, H. Peters, J. Blümer, W. Blankenfeldt, R. S. Goody, A. Itzen, *Science* **2010**, *329*, 946.
- [221] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Research* **2000**, *28*, 235-242.
- [222] L. Konermann, J. Pan, Y.-H. Liu, *Chemical Society Reviews* **2011**, *40*, 1224-1234.
-

- [223] (a) J. G. Kiselar, M. R. Chance, *J Mass Spectrom* **2010**, *45*, 1373-1382; (b) L. Wang, M. R. Chance, *Mol Cell Proteomics* **2017**, *16*, 706-716.
- [224] (a) A. Sinz, *Mass Spectrometry Reviews* **2006**, *25*, 663-682; (b) A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, *Mol Cell Proteomics* **2010**, *9*, 1634-1649; (c) J. Rappsilber, *J Struct Biol* **2011**, *173*, 530-540; (d) M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, G. Dollinger, *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 5802-5806; (e) E. V. Petrotchenko, C. H. Borchers, *Mass Spectrometry Reviews* **2010**, *29*, 862-876; (f) A. Leitner, M. Faini, F. Stengel, R. Aebersold, *Trends in Biochemical Sciences* **2016**, *41*, 20-32.
- [225] (a) A. Leitner, T. Walzthoeni, R. Aebersold, *Nature Protocols* **2013**, *9*, 120; (b) A. Sinz, *Expert Review of Proteomics* **2014**, *11*, 733-743.
- [226] C. Yu, L. Huang, *Anal Chem* **2018**, *90*, 144-165.
- [227] A. Sinz, *Anal Bioanal Chem* **2017**, *409*, 33-44.
- [228] K. L. Bennett, M. Kussmann, M. Mikkelsen, P. Roepstorff, P. Björk, M. Godzwon, P. Sørensen, *Protein Science* **2000**, *9*, 1503-1518.
- [229] L. Yang, X. Tang, C. R. Weisbrod, G. R. Munske, J. K. Eng, P. D. von Haller, N. K. Kaiser, J. E. Bruce, *Anal Chem* **2010**, *82*, 3556-3566.
- [230] (a) A. Kao, C.-l. Chiu, D. Vellucci, Y. Yang, V. R. Patel, S. Guan, A. Randall, P. Baldi, S. D. Rychnovsky, L. Huang, *Mol Cell Proteomics* **2011**, *10*, M110.002212-M002110.002212; (b) C. Iacobucci, M. Götze, C. H. Ihling, C. Piotrowski, C. Arlt, M. Schäfer, C. Hage, R. Schmidt, A. Sinz, *Nature Protocols* **2018**, *13*, 2864-2889.
- [231] (a) J. Mitchell Wells, S. A. McLuckey, in *Methods in Enzymology*, Vol. 402, Academic Press, **2005**, pp. 148-185; (b) L. Sleno, D. A. Volmer, *Journal of Mass Spectrometry* **2004**, *39*, 1091-1112.
- [232] (a) R. M. Kaake, X. Wang, A. Burke, C. Yu, W. Kandur, Y. Yang, E. J. Novitsky, T. Second, J. Duan, A. Kao, S. Guan, D. Vellucci, S. D. Rychnovsky, L. Huang, *Mol Cell Proteomics* **2014**, *13*, 3533-3543; (b) C. Yu, W. Kandur, A. Kao, S. Rychnovsky, L. Huang, *Anal Chem* **2014**, *86*, 2099-2106; (c) C. B. Gutierrez, C. Yu, E. J. Novitsky, A. S. Huszagh, S. D. Rychnovsky, L. Huang, *Anal Chem* **2016**, *88*, 8315-8322.
- [233] (a) M. Q. Müller, F. Dreiocker, C. H. Ihling, M. Schäfer, A. Sinz, *Anal Chem* **2010**, *82*, 6958-6968; (b) C. Hage, C. Iacobucci, A. Rehkamp, C. Arlt, A. Sinz, *Angewandte Chemie International Edition* **2017**, *56*, 14551-14555; (c) C. Hage, F. Falvo, M. Schäfer, A. Sinz, *Journal of The American Society for Mass Spectrometry* **2017**, *28*, 2022-2038.
- [234] M. Q. Müller, J. J. Zeiser, F. Dreiocker, A. Pich, M. Schäfer, A. Sinz, *Rapid Communications in Mass Spectrometry* **2011**, *25*, 155-161.
- [235] Y. Yang, H. Song, D. He, S. Zhang, S. Dai, S. Lin, R. Meng, C. Wang, P. R. Chen, *Nature Communications* **2016**, *7*, 12299.
- [236] S. Lin, D. He, T. Long, S. Zhang, R. Meng, P. R. Chen, *Journal of the American Chemical Society* **2014**, *136*, 11860-11863.
- [237] Q.-Q. Wang, R. A. Begum, V. W. Day, K. Bowman-James, *Organic & Biomolecular Chemistry* **2012**, *10*, 8786-8793.
- [238] N. Someshwar, C. R. Ramanathan, *Tetrahedron: Asymmetry* **2015**, *26*, 1209-1213.
- [239] S. K. Chaudhuri, S. Roy, S. Bhar, *Beilstein journal of organic chemistry* **2012**, *8*, 323-329.
- [240] S. J. Butler, *Chemistry – A European Journal* **2014**, *20*, 15768-15774.
-

- [241] (a) E. Mosettig, J. v. d. Kamp, *Journal of the American Chemical Society* **1930**, *52*, 3704-3710; (b) R. A. Urbanek, S. J. Suchard, G. B. Steelman, K. S. Knappenberger, L. A. Sygowski, C. A. Veale, M. J. Chapdelaine, *Journal of Medicinal Chemistry* **2001**, *44*, 1777-1793; (c) S. Lacombe, J.-P. Soumillion, A. El Kadib, T. Pigot, S. Blanc, R. Brown, E. Oliveros, C. Cantau, P. Saint-Cricq, *Langmuir* **2009**, *25*, 11168-11179.
- [242] D. T. Rogerson, A. Sachdeva, K. Wang, T. Haq, A. Kazlauskaitė, S. M. Hancock, N. Huguenin-Dezot, M. M. Muqit, A. M. Fry, R. Bayliss, J. W. Chin, *Nature chemical biology* **2015**, *11*, 496-503.
- [243] M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M. Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb, P. Mallick, *Nature biotechnology* **2012**, *30*, 918-920.
- [244] M. R. Hoopmann, A. Zelter, R. S. Johnson, M. Riffle, M. J. MacCoss, T. N. Davis, R. L. Moritz, *Journal of proteome research* **2015**, *14*, 2190-2198.
- [245] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, *Journal of proteome research* **2011**, *10*, 1794-1805.
- [246] L. Kall, J. D. Canterbury, J. Weston, W. S. Noble, M. J. MacCoss, *Nature methods* **2007**, *4*, 923-925.
- [247] M. R. Hoopmann, L. Mendoza, E. W. Deutsch, D. Shteynberg, R. L. Moritz, *Journal of the American Society for Mass Spectrometry* **2016**, *27*, 1728-1734.
- [248] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nature Methods* **2012**, *9*, 676-682.
-



## § 6. LIST OF ABBREVIATIONS

Å – angstrom	DSSO – disuccinimidyl sulfoxide
aaRS – aminoacyl-tRNA-synthetase	E – electrophile
abs – absorption (of light)	<i>Ec</i> – <i>Escherichia coli</i>
ACN – acetonitrile	EDC – 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
AMP – adenosine monophosphate	EDT – ethanedithiol
APS – ammonium persulfate	EDTA – ethylenediaminetetraacetic acid
BME – $\beta$ -mercaptoethanol	EGVyO – ethylene glycol vinyl ether
Boc – <i>tert</i> -butoxycarbonyl	(EG) <sub>2</sub> VyO – di(ethylene glycol) vinyl ether
bp – base pair(s)	em – emission (of light)
br – broad (NMR spectroscopy)	eq – equivalent(s)
BSA – bovine serum albumin	ESI – electrospray ionisation
°C – degree Celsius	ESIPT – excited state intramolecular proton transfer
CAM – ceric ammonium molybdate	ESPT – excited state proton transfer
CDCl <sub>3</sub> – deuterated chloroform	Et <sub>2</sub> O – diethylether
CDI – 1,1'-carbonyl diimidazole	Et <sub>3</sub> N – triethylamine
CHCl <sub>3</sub> – chloroform	EtOAc – ethyl acetate
CID – collision-induced dissociation	EtVyO – ethyl vinyl ether
cm – centimetre(s)	EtVyS – ethyl vinyl sulfide
conc. – concentrated	FBS – fetal bovine serum
CuAAC – copper(I)-catalysed azide-alkyne cycloaddition	Fmoc – 9-fluorenylmethoxycarbonyl
$\delta$ – chemical shift (NMR spectroscopy)	Fw – forward (primer)
d – doublet	g – gram(s)
D – dexter (“right”)	GDP – guanosine diphosphate
Da – Dalton	GEF – guanine nucleotide exchange factor
DA – Diels-Alder cycloaddition	GFP – green fluorescent protein
DCE – 1,2-dichloroethane	GSH – glutathione
DCM – dichloromethane	GTP – guanosine triphosphate
DDB – 1,4-dioxane dibromide	h – hour(s)
DIPEA – diisopropylethylamine	h $\nu$ – irradiation
DMF – dimethylformamide	Halo – haloalkane dehalogenase
DMSO – dimethylsulfoxide	HEPES – 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid
DMSO-d <sub>6</sub> – deuterated dimethylsulfoxide	HOBt – hydroxybenzotriazole
dppf – 1,1'-bis(diphenylphosphino)ferrocene	HOMO – highest occupied molecular orbital
DSBU – disuccinimidyl dibutyric urea	
DSC – <i>N,N'</i> -disuccinimidyl carbonate	

---

HPLC – high-performance liquid chromatography	NTA – nitrilotriacetic acid
Hz – Hertz	Nu – nucleophile
iEDDA – inverse electron-demand Diels-Alder cycloaddition	<i>o</i> – <i>ortho</i> (isomer)
IPTG – isopropyl $\beta$ -D-1-thiogalactopyranoside	<i>o</i> /N – overnight
<i>J</i> – coupling constant (NMR spectroscopy)	OD <sub>600</sub> – optical density at 600 nm
<i>k</i> – kilo	<i>o</i> NQM(s) – <i>ortho</i> -naphthoquinone methide(s)
<i>k</i> <sub>2</sub> – second-order rate constant	<i>o</i> NQMP(s) – <i>ortho</i> -naphthoquinone methide precursor(s)
<i>K</i> <sub>D</sub> – dissociation constant	<i>o</i> QQM(s) – <i>ortho</i> -quinolinone quinone methide(s)
$\lambda$ – wavelength	<i>p</i> – <i>para</i> (isomer)
L – laevus (“left”)	PAGE – polyacrylamide gel electrophoresis
LB – lysogeny broth	PBS – phosphate-buffered saline
LC – liquid chromatography	PCR – polymerase chain reaction
LEDs – light-emitting diodes	PDB – Protein Data Bank
LUMO – lowest unoccupied molecular orbital	PEG – polyethylene glycol
$\mu$ – micro	PEPC – proximity-enabled protein crosslinking
<i>m</i> – <i>meta</i> (isomer)	PMSF – phenylmethylsulfonyl fluoride
<i>m</i> – multiplet	ppm – parts-per-million
M – mega	PPTS – pyridinium <i>p</i> -toluenesulfonate
M – molar	PQ(s) – 9,10-phenanthrenequinone(s)
<i>Mb</i> – <i>Methanosarcina barkeri</i>	<i>R</i> <sub>f</sub> – retention factor
mCPBA – <i>meta</i> -chloroperbenzoic acid	RP – reverse phase
Me – methyl	RT – room temperature
MeOD – deuterated methanol	R <sub>v</sub> – reverse (primer)
MeOH – methanol	s – second(s)
mg – milligram	s – singlet
min – minute(s)	satur. – saturated
mL – millilitre(s)	SDS – sodium dodecylsulfate
<i>Mm</i> – <i>Methanosarcina mazei</i>	SE – succinimidyl ester
mmol – millimole(s)	sf – superfolder
MS – mass spectrometry	S <sub>N</sub> – nucleophilic substitution
Ms – methanesulfonyl, mesyl	SPAAC – strain-promoted azide-alkyne cycloaddition
Myo – myoglobin	t – triplet
NAM – nicotinamide	TAMRA – carboxytetramethylrhodamine
NEM – <i>N</i> -ethylmorpholine	TB – terrific broth
NHS – <i>N</i> -hydroxysuccinimide	TBAF – tetra- <i>n</i> -butylammonium fluoride
nm – nanometer(s)	TBDMS – <i>tert</i> -butyldimethylsilyl
NMR – nuclear magnetic resonance	
NOH – 3-hydroxymethyl-2-naphthol	

---

TBS – tris-buffered saline

TBST – tris-buffered saline with Tween 20

TC – tetracysteine

TEMED – tetramethylethylenediamine

TEV – Tobacco etch virus

Tf – trifluoromethanesulfonyl

TFA – trifluoroacetic acid

THF – tetrahydrofuran

TLC – thin layer chromatography

Ts – toluenesulfonyl, tosyl

q – quartet

QM – quinone methide

quint – quintet/pentet

UAAs – unnatural amino acid(s)

Ub – ubiquitin

UV – ultraviolet

Vy – vinyl

VyO – vinyl ether

W – watt

wb – western blot

---

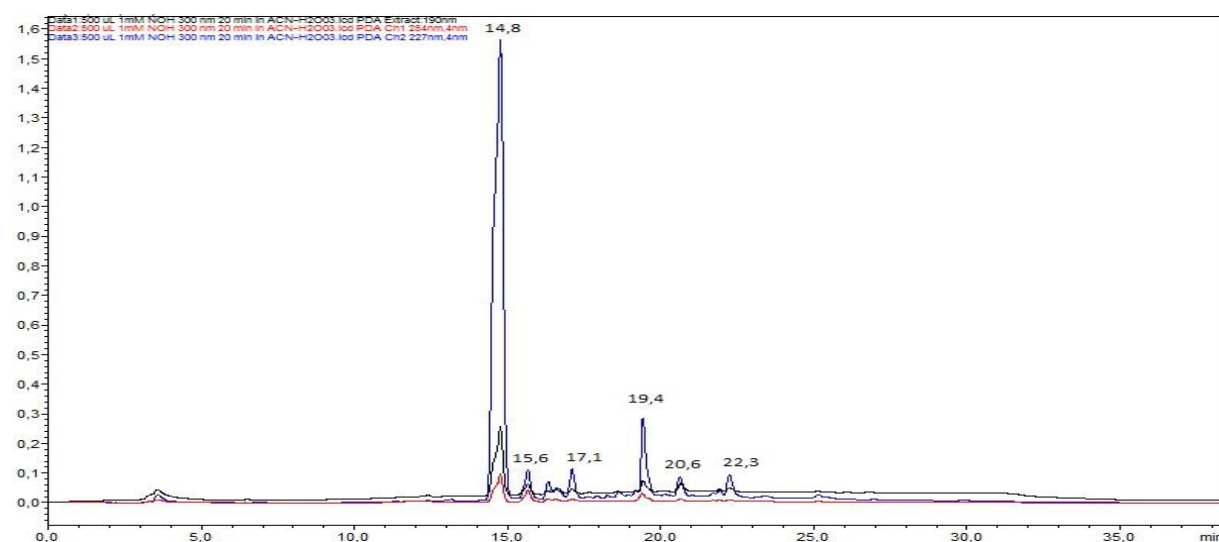




## § 7. APPENDIX

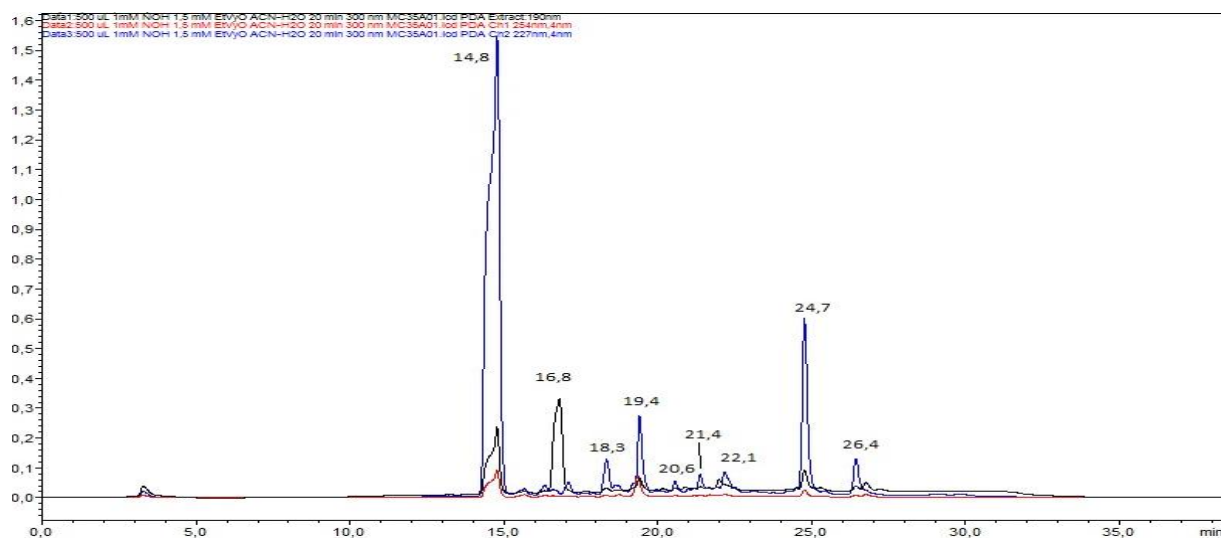
## 7.1. Supporting Figures – Chapter 2

*1 mM NOH, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



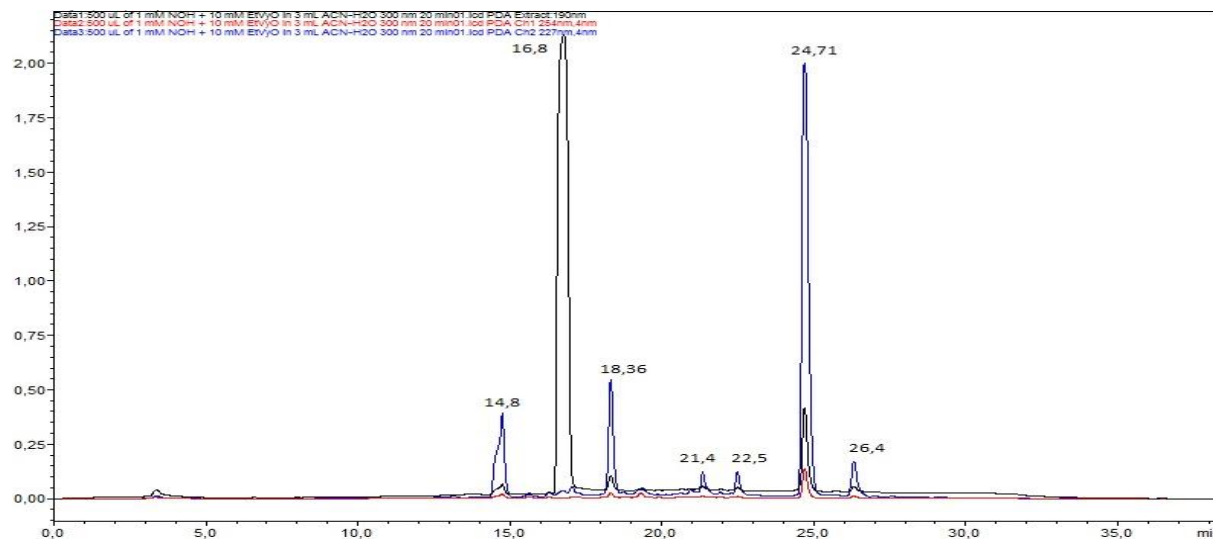
**Figure S2.1.** Irradiation of 1 mM NOH solution in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The retention time ( $t_R$ ) of NOH is 14.8 min. Secondary photoproducts ( $t_R$  above 15 minutes) are possibly the result of *o*NQM oligomer formation. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 1.5 mM EtVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



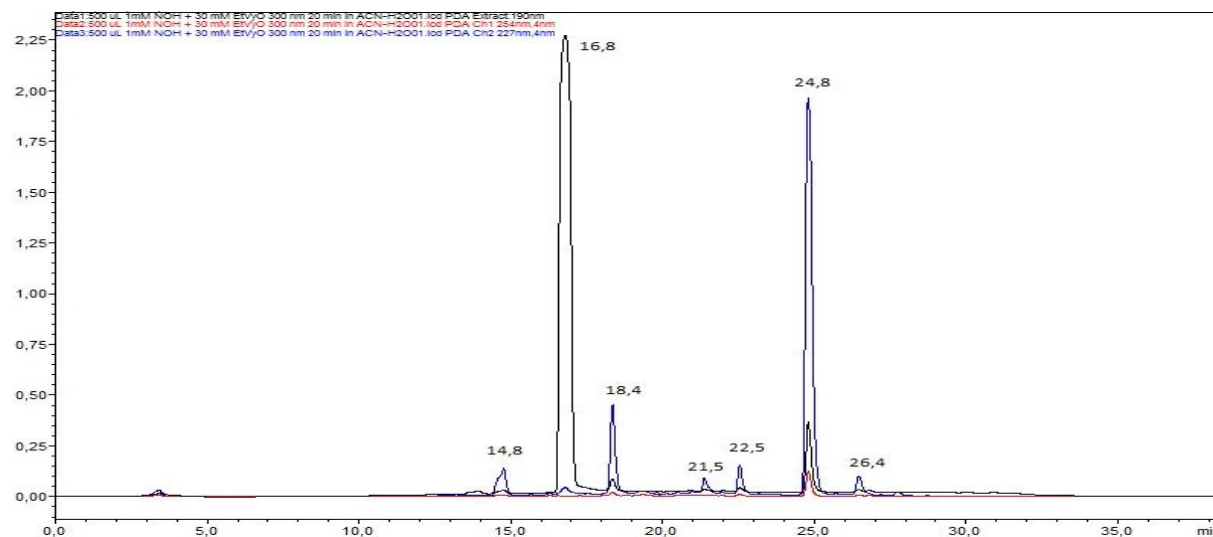
**Figure S2.2.** Irradiation of 1 mM NOH and 1.5 mM EtVyO in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The cycloaddition product formation can be observed at 24.7 min. Other peaks: NOH ( $t_R$  = 14.8 min), EtVyO ( $t_R$  = 16.8 min) and secondary photoproducts. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 10 mM EtVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



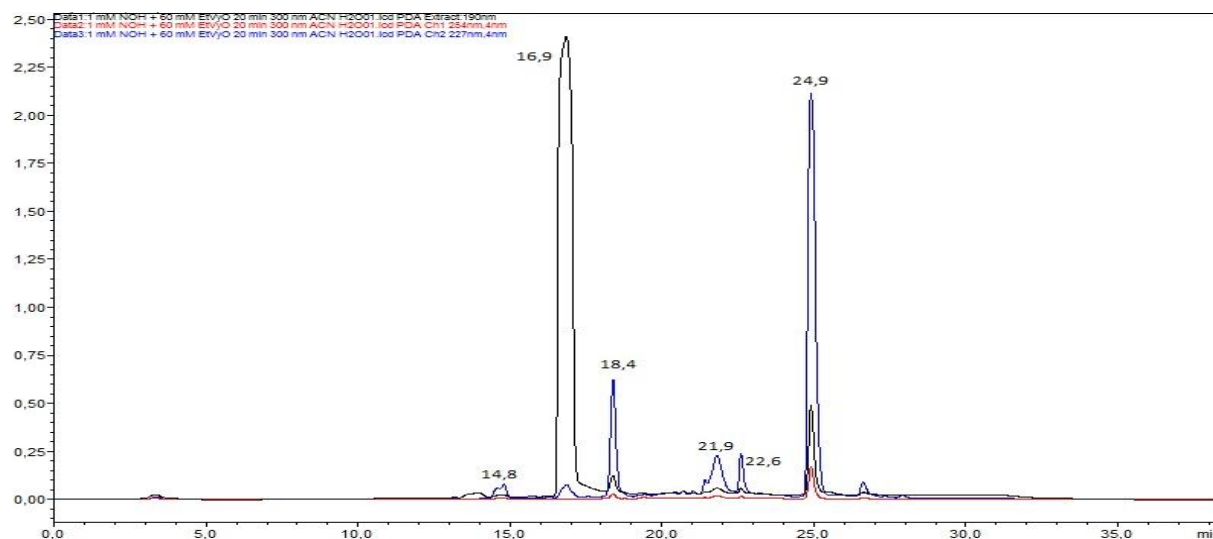
**Figure S2.3.** Irradiation of 1 mM NOH and 10 mM EtVyO in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The cycloaddition product formation can be observed at 24.7 min. Other peaks: NOH ( $t_R = 14.8$  min), EtVyO ( $t_R = 16.8$  min) and secondary photoproducts. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 30 mM EtVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



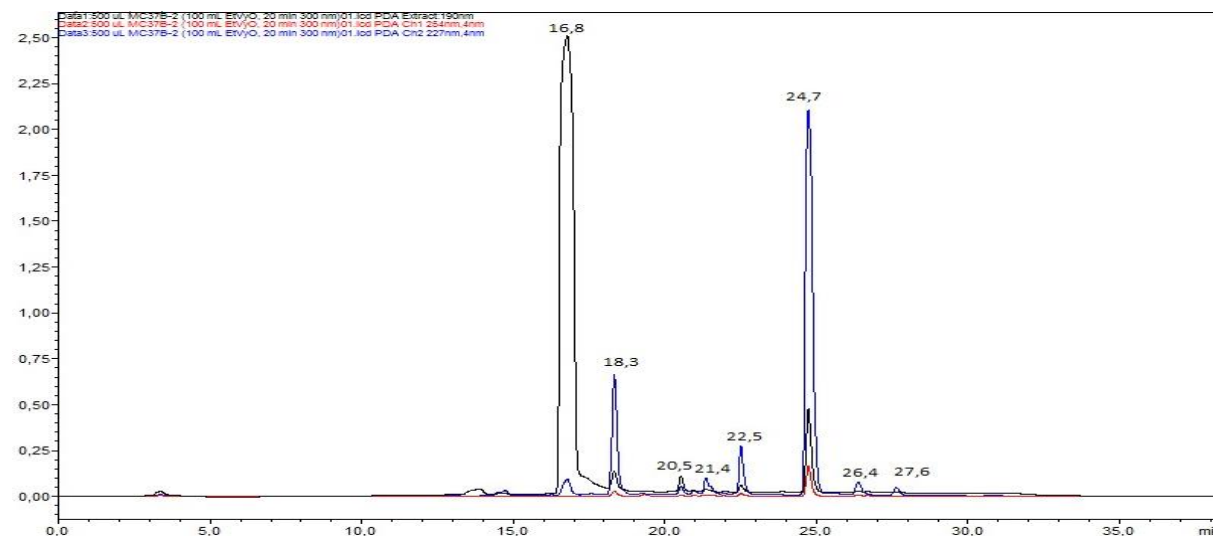
**Figure S2.4.** Irradiation of 1 mM NOH and 30 mM EtVyO in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The cycloaddition product formation can be observed at 24.8 min. Other peaks: NOH ( $t_R = 14.8$  min), EtVyO ( $t_R = 16.8$  min) and secondary photoproducts. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 60 mM EtVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



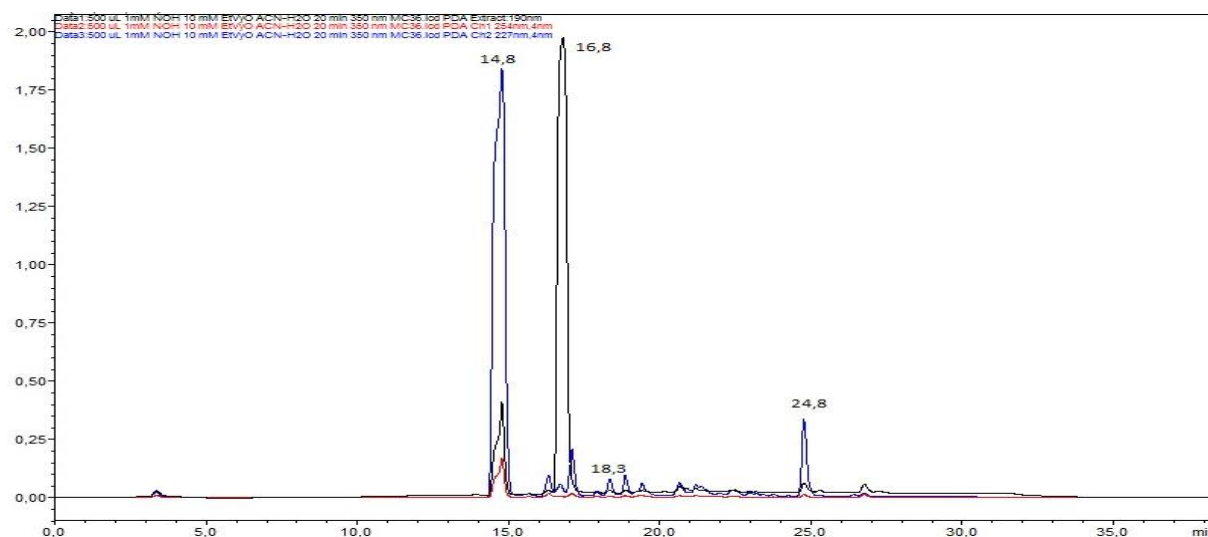
**Figure S2.5.** Irradiation of 1 mM NOH and 60 mM EtVyO in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The cycloaddition product formation can be observed at 24.9 min. Other peaks: NOH ( $t_R = 14.8$  min), EtVyO ( $t_R = 16.9$  min) and secondary photoproducts. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 100 mM EtVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



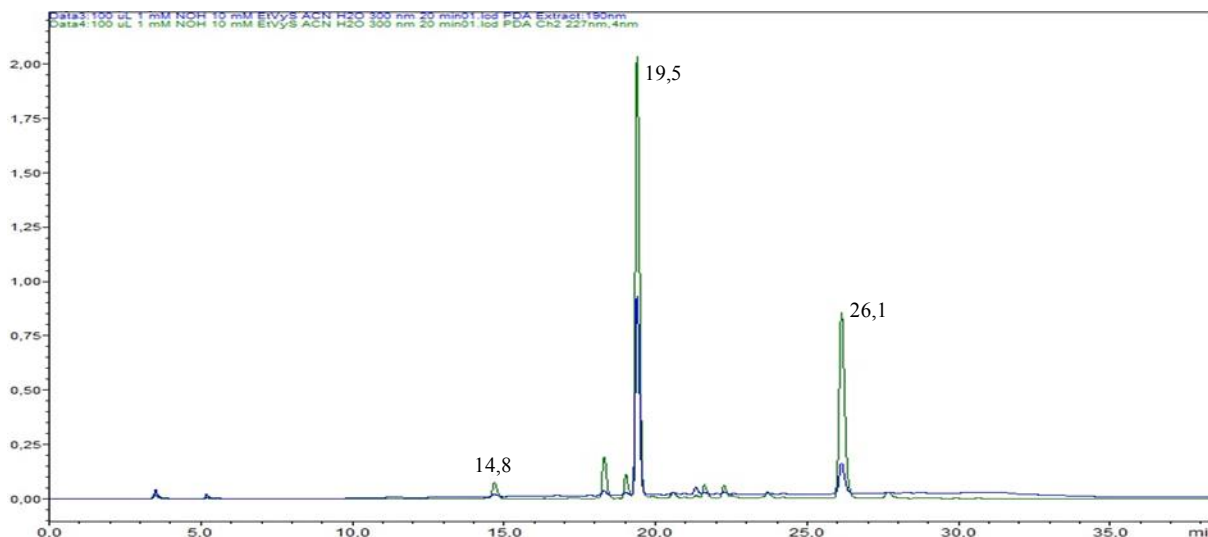
**Figure S2.6.** Irradiation of 1 mM NOH and 100 mM EtVyO in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The cycloaddition product formation can be observed at 24.7 min. Other peaks: EtVyO ( $t_R = 16.8$  min) and secondary photoproducts. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 10 mM EtVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 350 nm for 20 min*



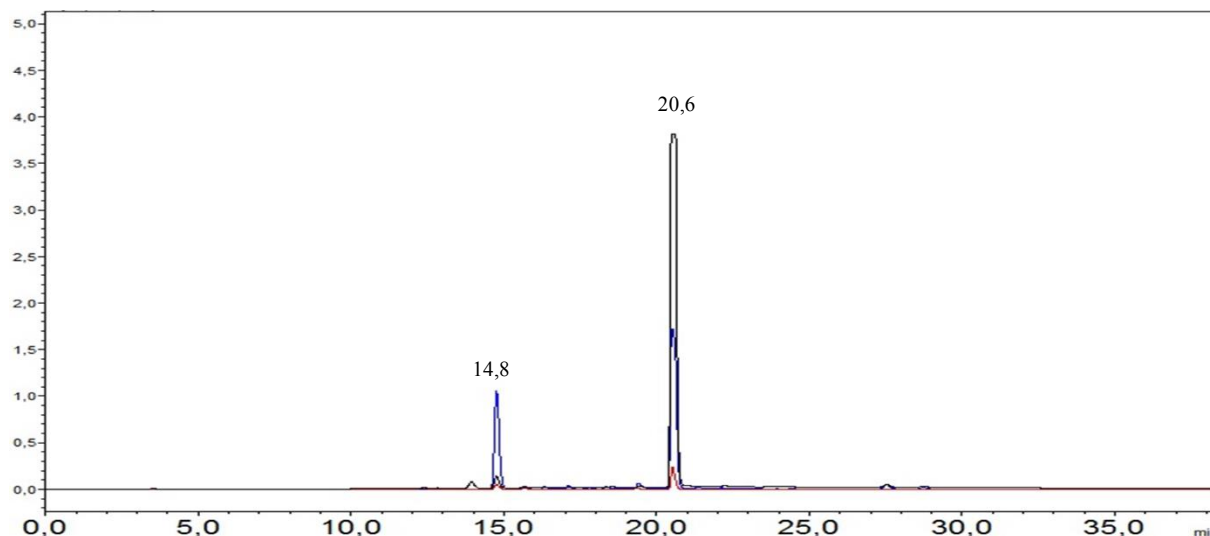
**Figure S2.7.** Irradiation of 1 mM NOH and 10 mM EtVyO in 50 % aqueous ACN for 20 min at 350 nm, analysed by HPLC. The cycloaddition product formation can be observed at 24.8 min. Other peaks: NOH ( $t_R = 14.8$  min), EtVyO ( $t_R = 16.8$  min) and secondary photoproducts. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 10 mM EtVyS, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



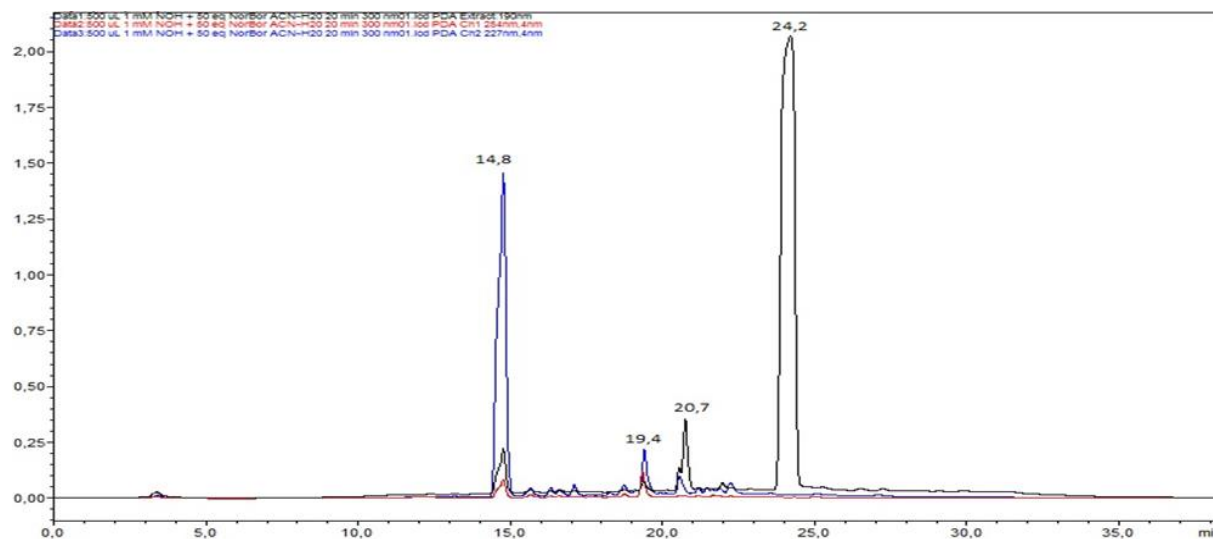
**Figure S2.8.** Irradiation of 1 mM NOH and 10 mM EtVyS in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The cycloaddition product formation can be observed at 26.1 min. Other peaks: NOH ( $t_R = 14.8$  min), EtVyS ( $t_R = 19.5$  min) and secondary photoproducts. UV absorbance is measured at 190 (blue) and 227 nm (green).

*1 mM NOH + 100 mM PhVyO, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



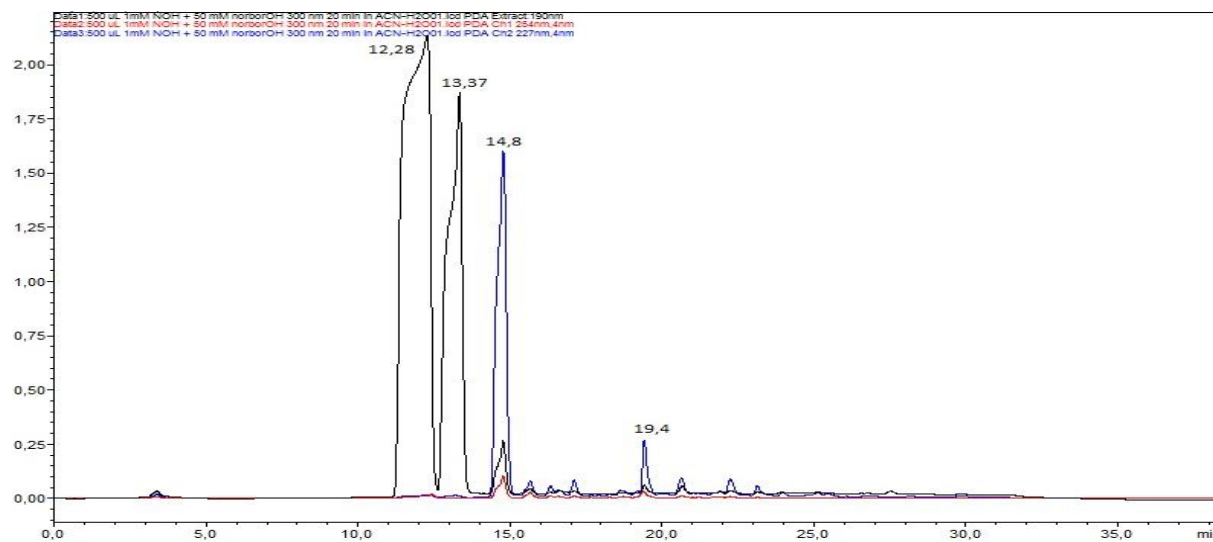
**Figure S2.9.** Irradiation of 1 mM NOH and 100 mM PhVyO in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. No cycloaddition product formation is observed. Instead, the following peaks are present: NOH ( $t_R = 14.8$  min) and PhVyO ( $t_R = 20.6$  min). UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

*1 mM NOH + 50 mM Nor, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*

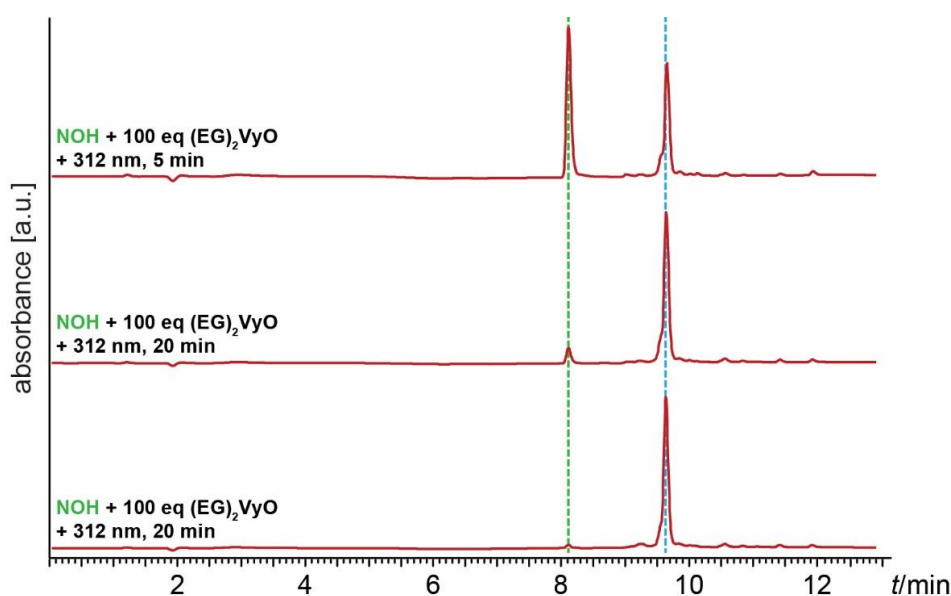


**Figure S2.10.** Irradiation of 1 mM NOH and 50 mM Nor in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. No cycloaddition product formation is observed. Instead, the following peaks are present: NOH ( $t_R = 14.8$  min), Nor ( $t_R = 20.7$  and 24.2 min) and secondary photoproducts (most prominent at 19.4 min). UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

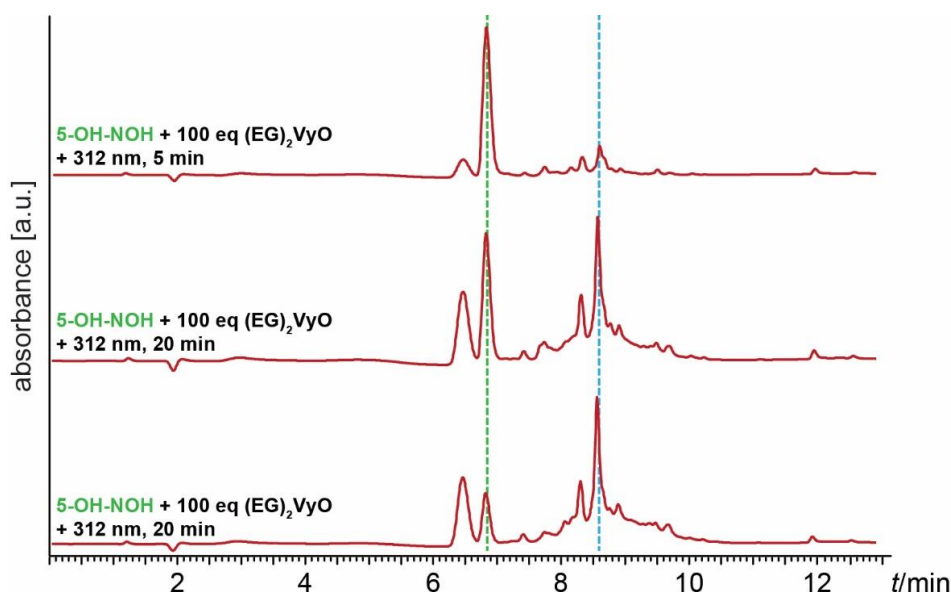
*1 mM NOH + 50 mM Nor-OH (endo/exo), ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min*



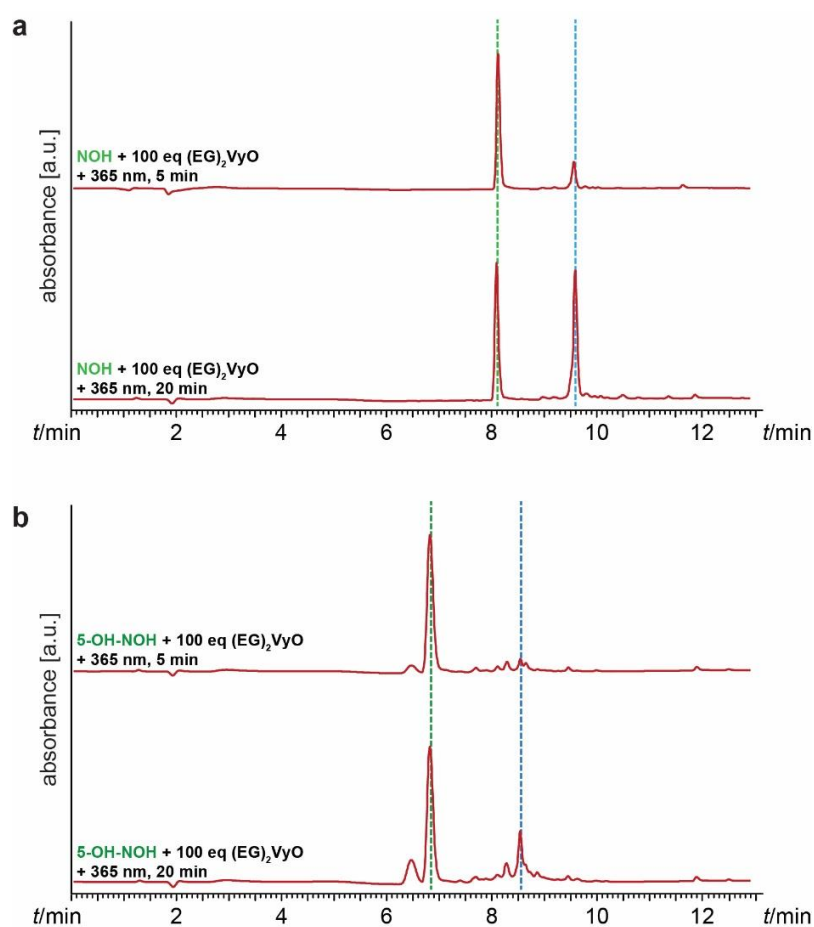
**Figure S2.11.** Irradiation of 1 mM NOH and 50 mM Nor-OH (*endo* and *exo* mixture) in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. No cycloaddition product formation is observed. Instead, the following peaks are present: NOH ( $t_R = 14.8$  min), Nor-OH ( $t_R = 12.3$  and 13.4 min) and secondary photoproducts (most prominent at 19.4 min). UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).



**Figure S2.12.** Irradiation of 1 mM NaOH and 100 mM (EG)<sub>2</sub>VyO in 50 % aqueous ACN at 312 nm, analysed by LC-MS. The cycloaddition product is represented by the blue dashed line. UV absorbance is measured at 280 nm. As a result, the peak corresponding to (EG)<sub>2</sub>VyO is not visible.

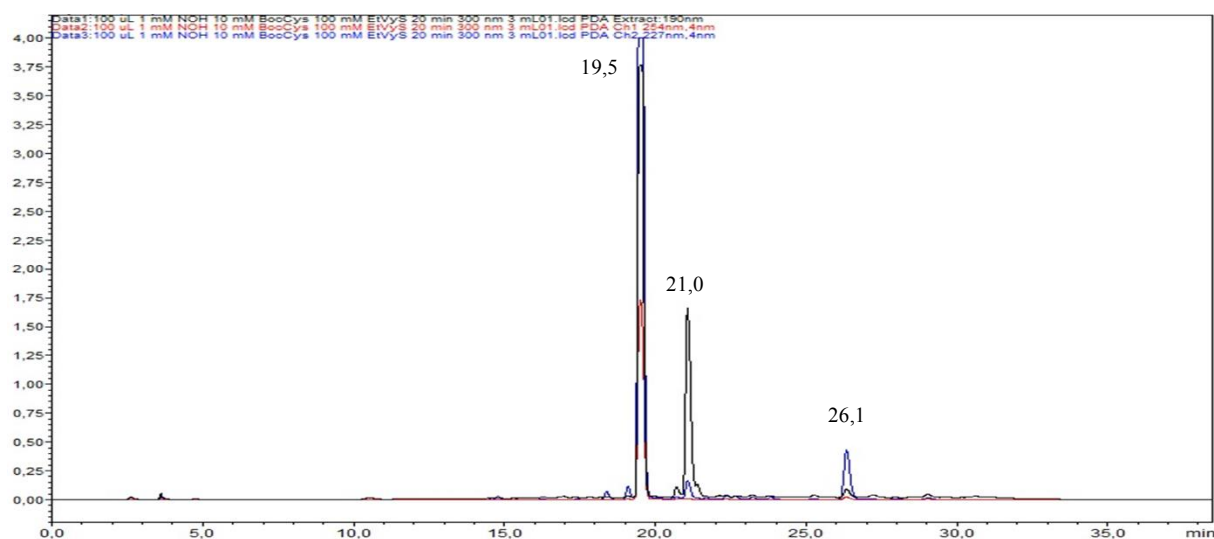


**Figure S2.13.** Irradiation of 1 mM 5-OH-NaOH and 100 mM (EG)<sub>2</sub>VyO in 50 % aqueous ACN at 312 nm, analysed by LC-MS. The cycloaddition product is represented by the blue dashed line. Significant amount of unidentified photolysis products is present as well. UV absorbance is measured at 280 nm. As a result, the peak corresponding to (EG)<sub>2</sub>VyO is not visible.



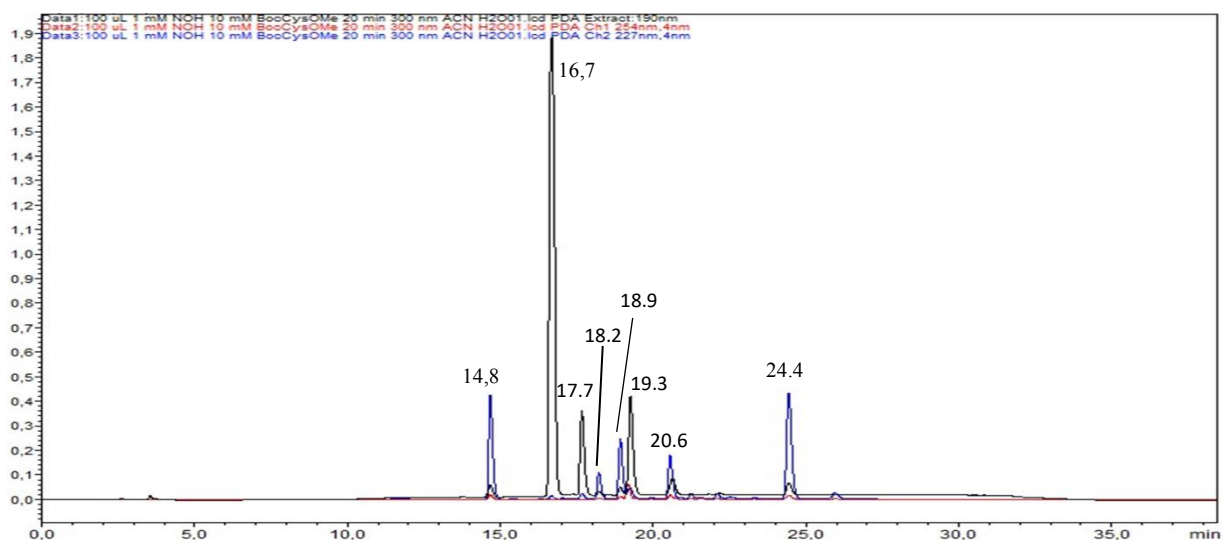
**Figure S2.14.** Irradiation of 1 mM *o*NQMP and 100 mM (EG)<sub>2</sub>VyO in 50 % aqueous ACN at 365 nm, analysed by LC-MS. (a) Reaction with NOH. (b) reaction with 5-OH-NOH. UV absorbance is measured at 280 nm. As a result, the peak corresponding to (EG)<sub>2</sub>VyO is not visible

1 mM NOH + 100 mM EtVyS + 10 mM Boc-Cys-OMe, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min



**Figure S2.15.** Irradiation of 1 mM NOH and 100 mM EtVyS in the presence of 10 mM Boc-Cys-OMe, in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The starting NOH (expected at  $t_R = 14.8$  min) is completely consumed. The peak corresponding to the cycloaddition product with EtVyS ( $t_R = 19.5$  min) can be observed at 26.1 min. Interestingly, Boc-Cys-OMe is completely consumed as well (expected at  $t_R = 17.7$  min). It underwent light-mediated thiol-ene reaction with EtVyS to afford the by-product observable at 21.0 min. UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).

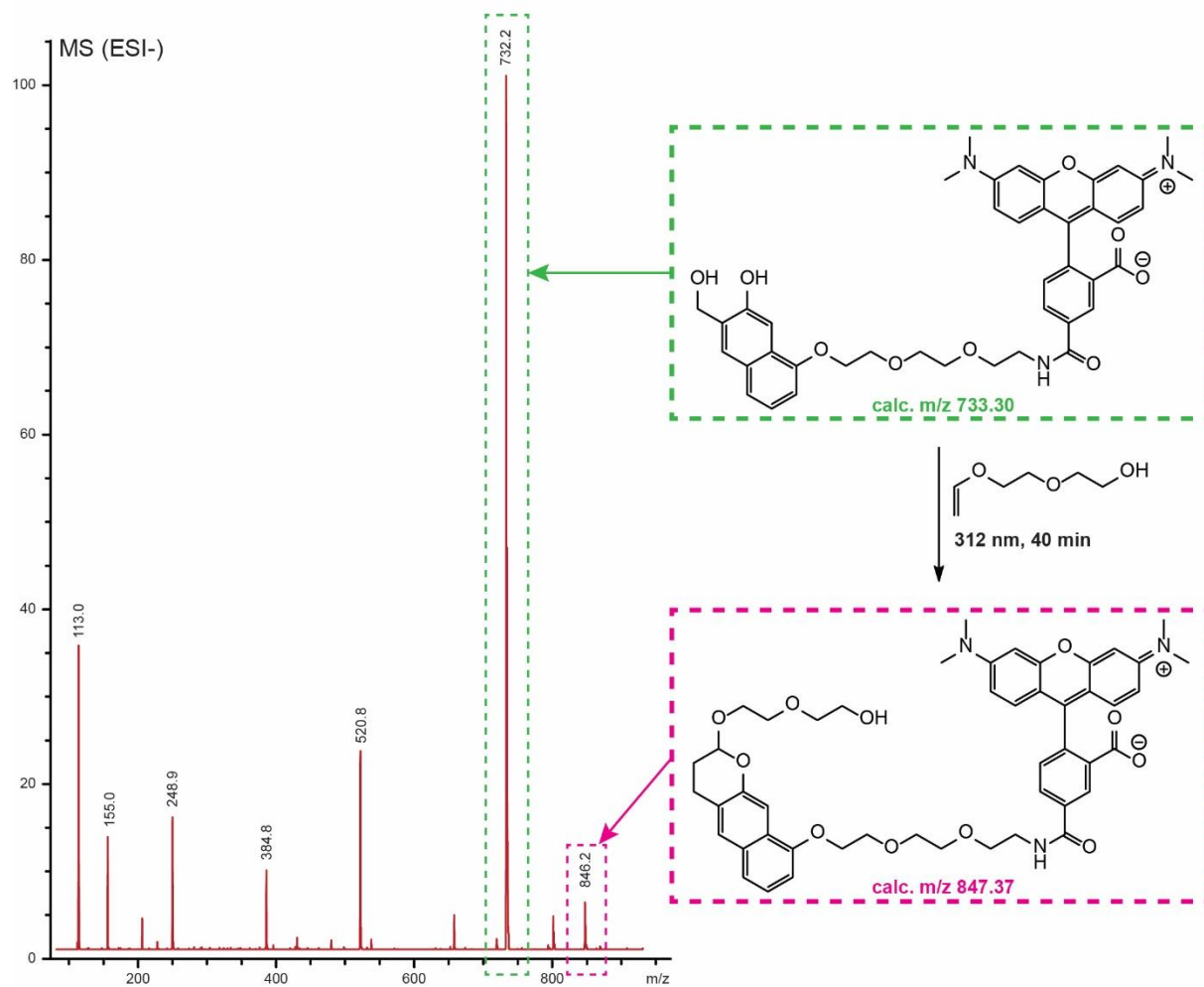
1 mM NOH + 100 mM EtVyO + 10 mM Boc-Cys-OMe, ACN/H<sub>2</sub>O (1:1), irradiated at 300 nm for 20 min



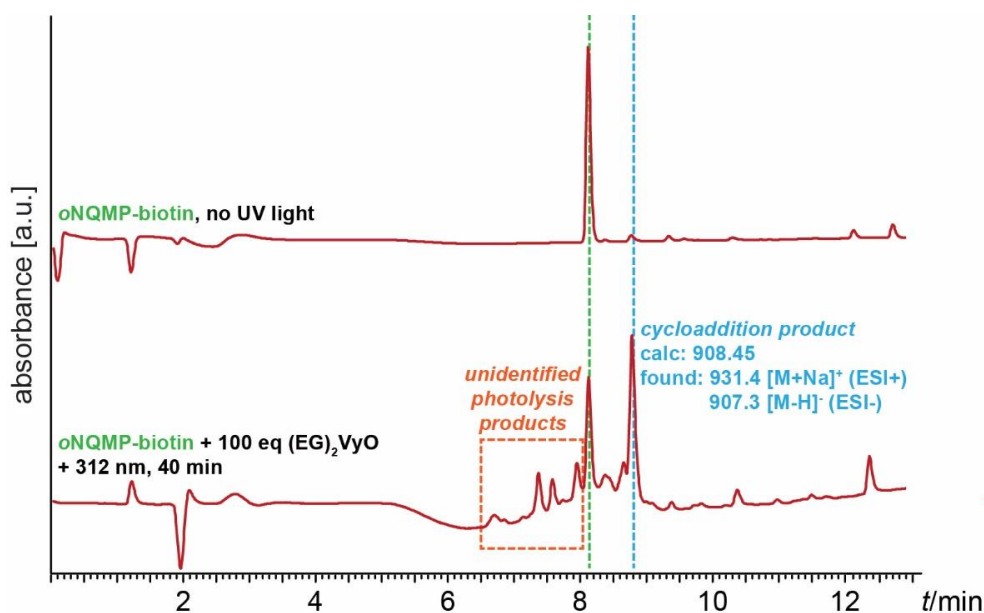
**Figure S2.16.** Irradiation of 1 mM NOH and 100 mM EtVyO in the presence of 10 mM Boc-Cys-OMe, in 50 % aqueous ACN for 20 min at 300 nm, analysed by HPLC. The peak corresponding to the cycloaddition product between NOH ( $t_R = 14.8$  min) with EtVyO ( $t_R = 16.7$  min) can be observed at 24.4 min. Other peaks include Boc-Cys-OMe ( $t_R = 17.7$  min), thiol-ene click by-product ( $t_R = 19.3$  min) and secondary photoproducts (*o*NQM oligomers at 18.2 and 18.9 min and *o*NQM-Cys adduct at 20.6 min which disappears over prolonged irradiation). UV absorbance is measured at 190 (black), 254 (red) and 227 nm (blue).



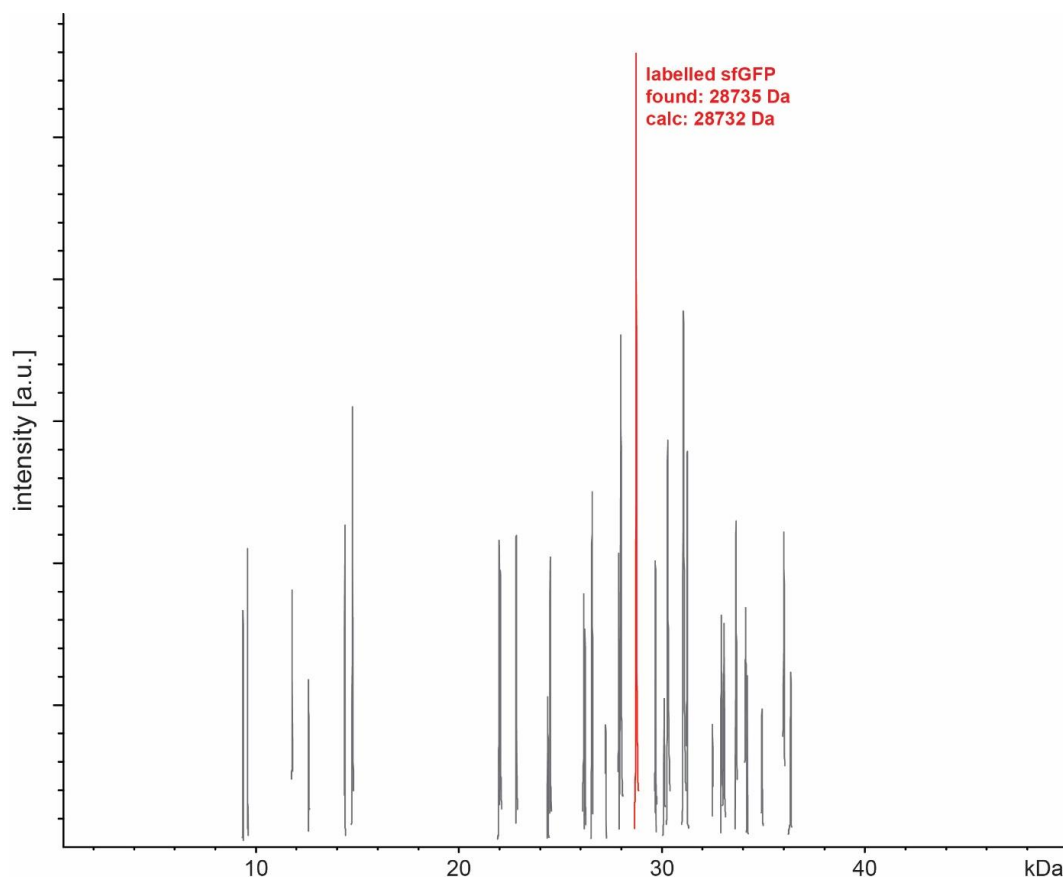
*1 mM oNQMP-TAMRA + 100 mM EGVyO, 40 min at 312 nm*



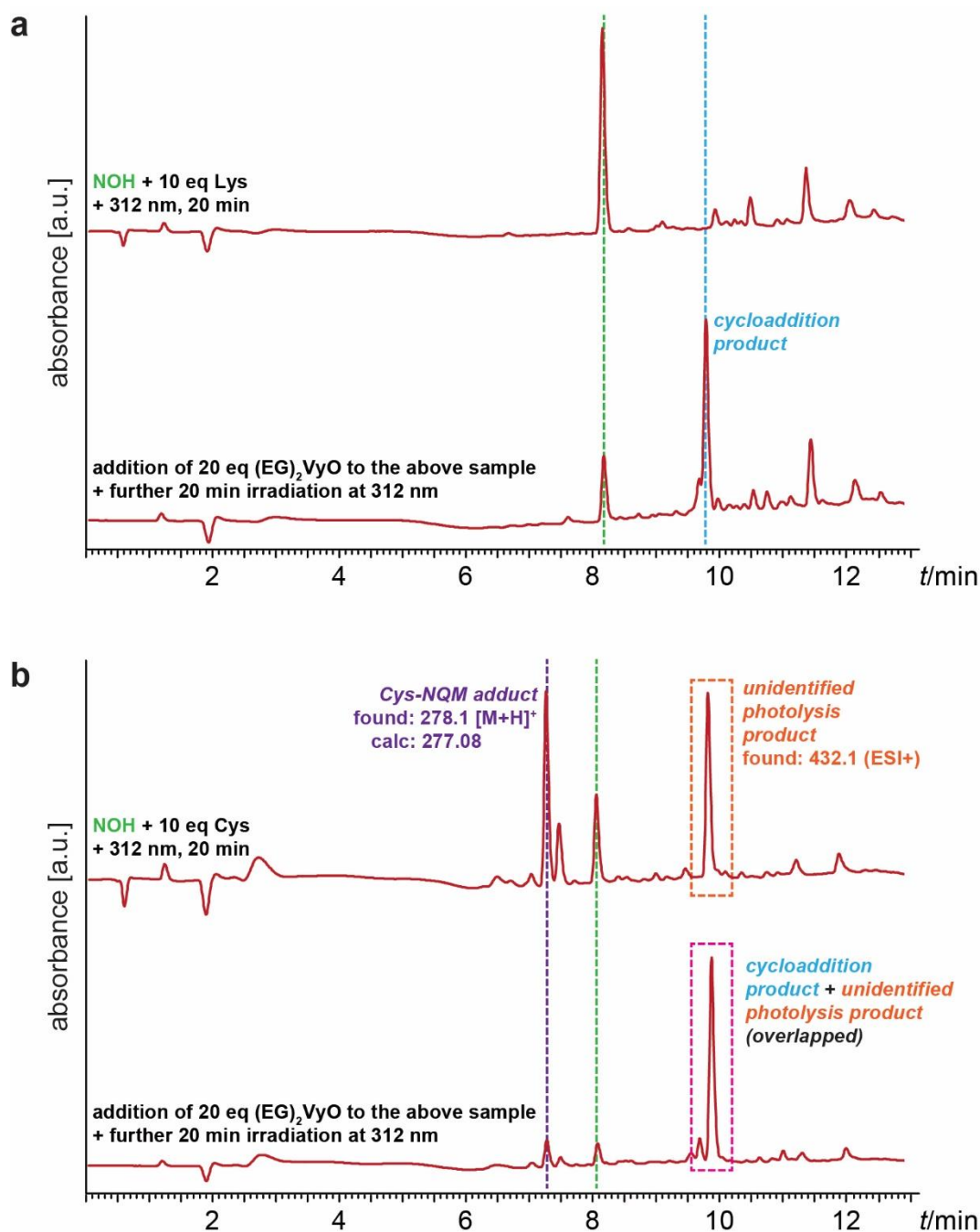
**Figure S2.17.** Irradiation of 1 mM *o*NQMP-TAMRA and 100 mM EGVyO for 40 min at 312 nm in 50 % aqueous ACN, analysed by LC-MS. Only a minor peak corresponding to the product was observed in the MS spectrum (ESI-). No difference in retention times between *o*NQMP-TAMRA and the product was observed and the chromatogram is thus not shown.



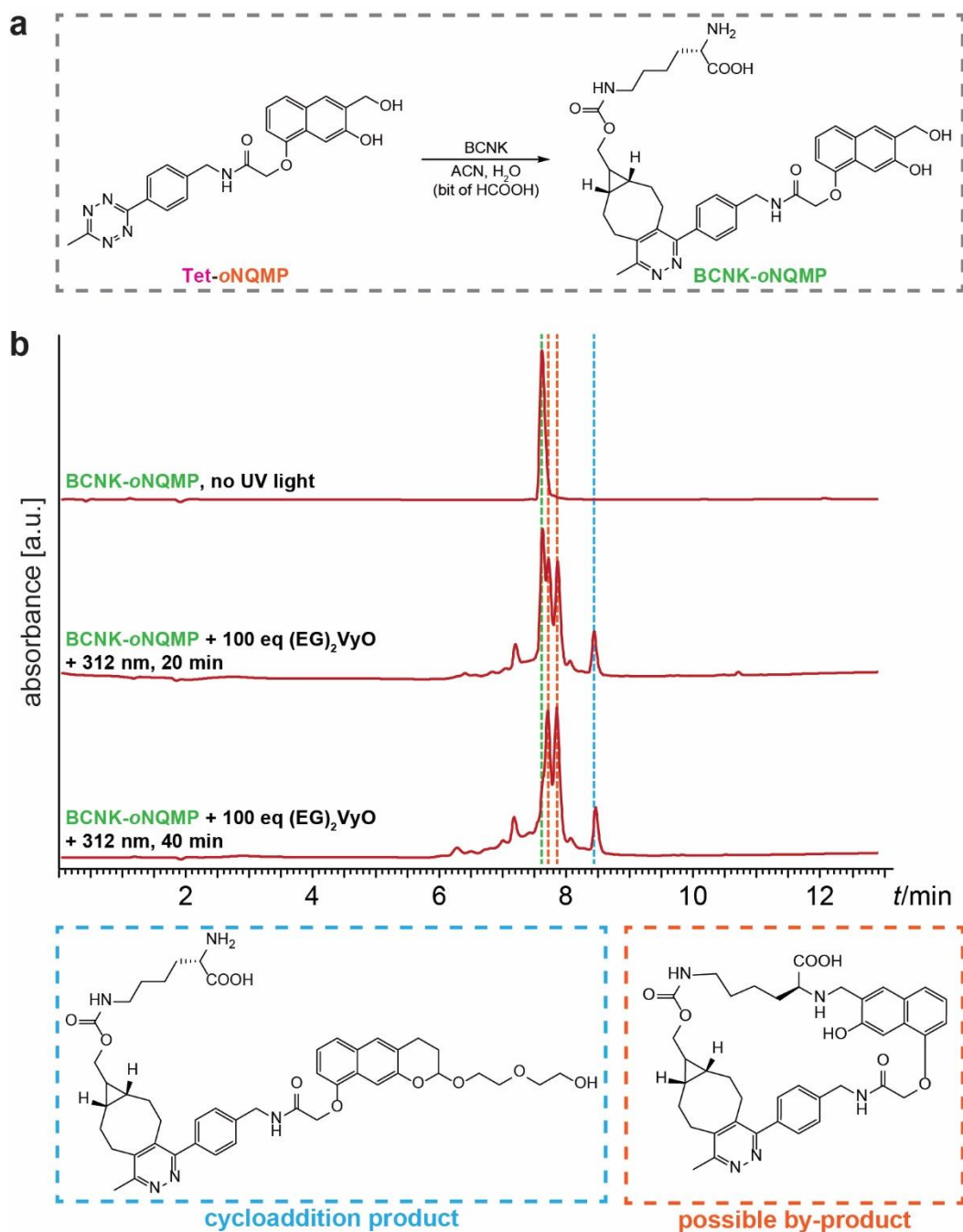
**Figure S2.18.** Light-induced reactivity of *o*NQMP-biotin with vinyl ethers, analysed by LC-MS. Irradiation of 1 mM *o*NQMP-biotin and 100 mM (EG)<sub>2</sub>VyO for 40 min at 312 nm in 50 % aqueous ACN affords the desired cycloaddition product (blue dashed line). Unidentified photolysis products (orange box) are observable as well. UV absorbance is measured at 280 nm. As a result, the peak corresponding to (EG)<sub>2</sub>VyO is not visible.



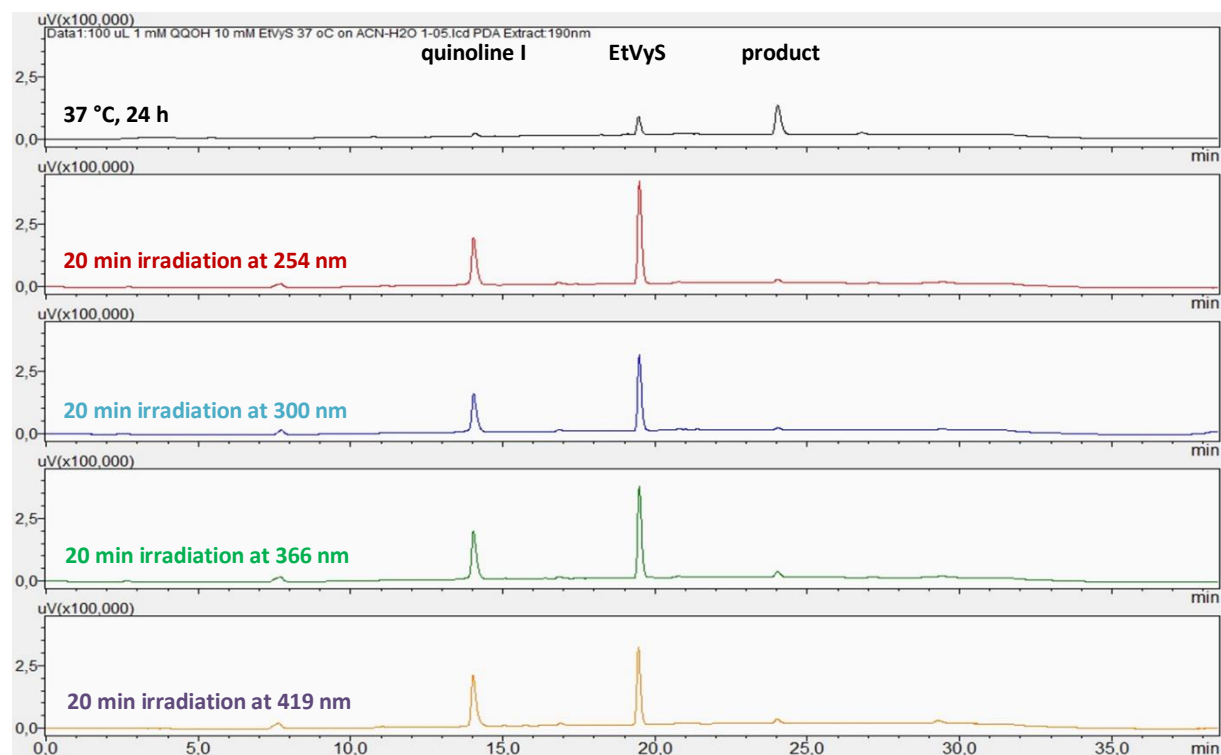
**Figure S2.19.** MS analysis of light-mediated sfGFP-N150VyOK-His6 labelling with *o*NQMP-biotin. Reaction conditions: 10  $\mu$ M sfGFP-VyOK, 100 eq *o*NQMP-biotin, 1 h irradiation at 312 nm, 1x PBS buffer pH 7.4, RT. In spite of difficulties during deconvolution, a mass corresponding to the biotinylated sfGFP was observed (red peak, 28735 Da). All the other observed masses (grey peaks) could not be rationally assigned; the mass of the unlabelled protein (27955 Da) was not observed.



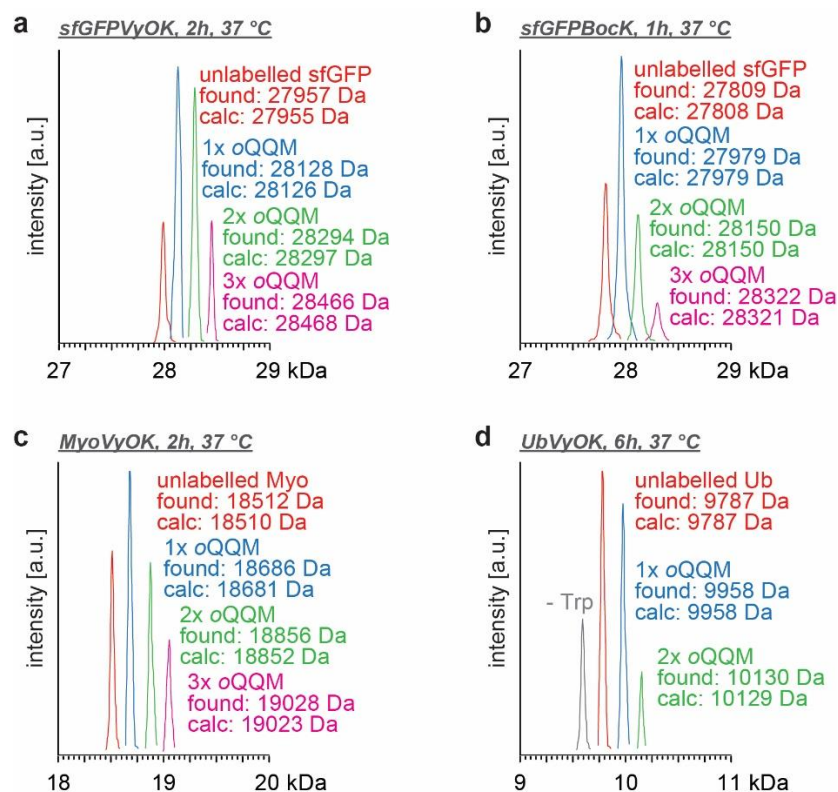
**Figure S2.20.** Light-induced reactivity of NOH with nucleophilic amino acids and subsequent light-mediated uncaging with vinyl ether, analysed by LC-MS. (a) After 20 min irradiation of 1 mM NOH (green dashed line) and 10 mM of lysine hydrochloride at 312 nm in ACN:1x PBS (pH 7.4), 1:1 (v/v) (upper chromatogram), 20 mM of (EG)<sub>2</sub>VyO were added and irradiation continued for 20 min longer (lower chromatogram). Formation of the cycloaddition product can be observed (blue dashed line). (b) Same as (a) for cysteine. In the absence of vinyl ether, a Cys-oNQM adduct can be observed (purple dashed line) and another unidentified photolysis product (orange box, upper chromatogram). The Cys-oNQM adduct disappears after prolonged irradiation in the presence of (EG)<sub>2</sub>VyO. Nevertheless, the peak of the cycloaddition product is overlapped with the unidentified photolysis product (pink box, lower chromatogram), judging by the masses observed for both species. UV absorbance is measured at 280 nm. As a result, peaks corresponding to lysine, cysteine and (EG)<sub>2</sub>VyO are not visible.



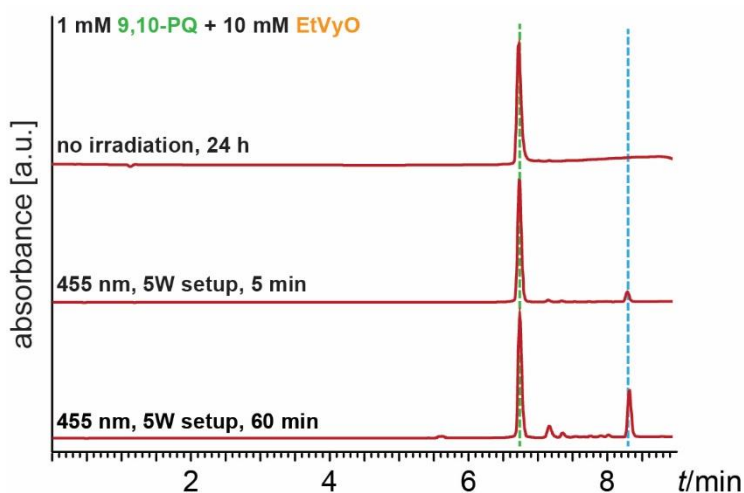
**Figure S2.21.** Light-induced reactivity of BCNK-*o*NQMP with vinyl ether. (a) BCNK-*o*NQMP was prepared via tetrazine ligation between BCNK and Tet-*o*NQMP. (b) Irradiation of 1 mM BCNK-*o*NQMP and 100 mM (EG)<sub>2</sub>VyO in ACN:1x PBS (pH 7.4), 1:1 (v/v) at 312 nm for 20 and 40 minutes. Although the formation of the cycloaddition product can be observed (blue dashed line), the photolysis reaction afforded two major products of the same mass (709.3 in ESI+ and 706.2 in ESI-), represented by orange dashed lines. Intramolecular nucleophilic addition of (possibly) the N-terminus of BCNK to the methylene group of *o*NQM might explain the formation of the by-product (calc. mass of the by-product is 707.33). UV absorbance is measured at 254 nm. As a result, peaks corresponding to (EG)<sub>2</sub>VyO are not visible.



**Figure S2.22.** Thermally- and photochemically-induced reactivity of quinoline I and EtVyS, analysed by HPLC. 1 mM of quinoline I and 10 mM EtVyS mixture in H<sub>2</sub>O:ACN, 5:1 (v/v) was either incubated at 37 °C for 24 h (black chromatogram) or irradiated in a Rayonet photoreactor for 20 minutes at different wavelengths: 254 (red), 300 (blue), 366 (green) or 419 (yellow chromatogram). The cycloaddition product was only observed when the respective *o*QQM-1 was thermally generated. Performed photolysis reactions, on the other hand, did not afford any product presumably due to relaxation of excited *o*QQM-1 through observed fluorescence.

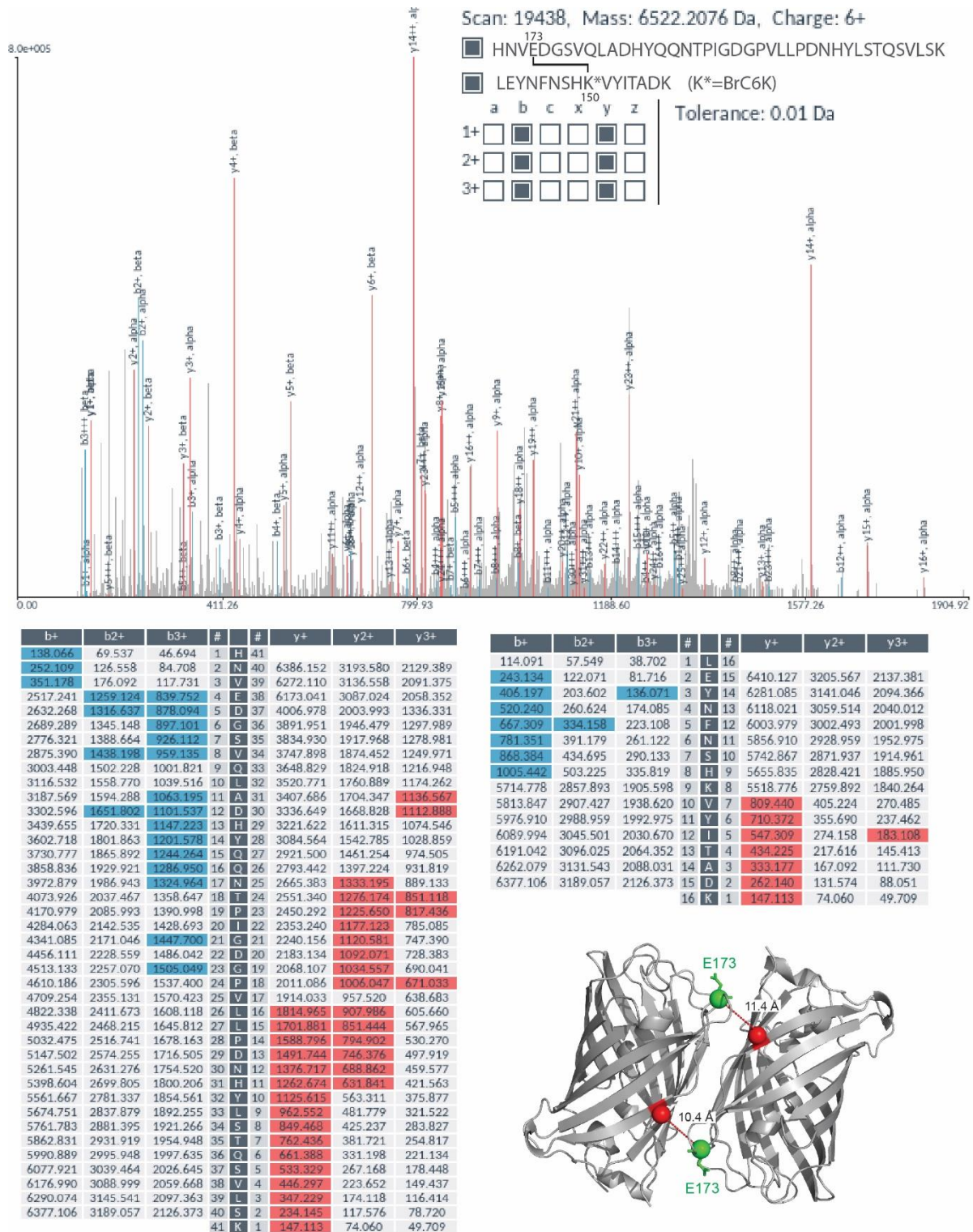


**Figure S2.23.** MS analysis of the *in vitro* labelling reaction between quinoline I and different proteins. 100  $\mu$ M of protein was incubated with an excess of quinoline I (50 eq, stock solution in DMSO) at pH 7.4 and 37  $^{\circ}$ C. The unspecific labelling was observed on: (a) sfGFP-N150VyOK-His6, (b) sfGFP-N150BocK-His6, (c) Myo-S4VyOK-His6 and (d) Ub-K6VyOK-His6.

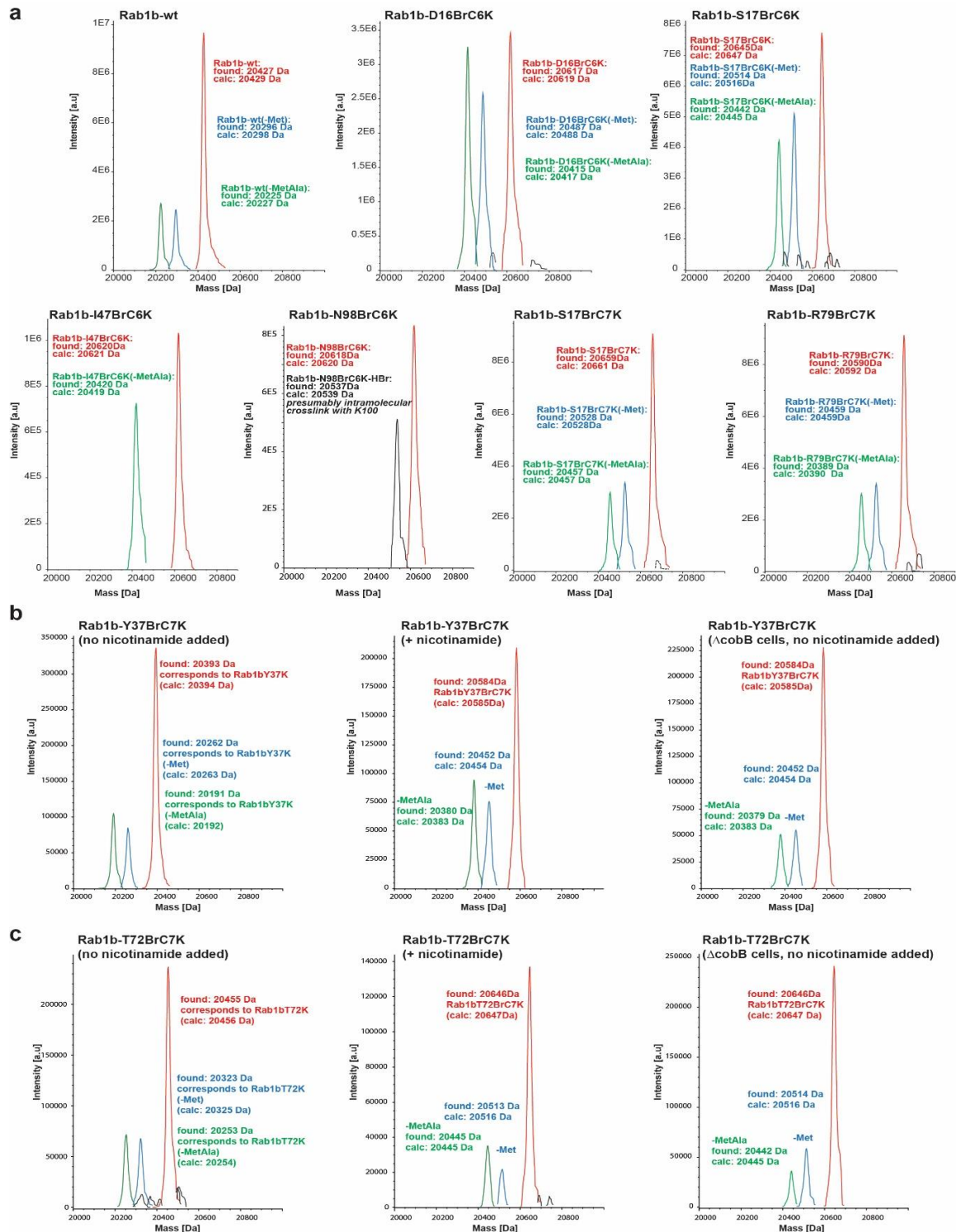


**Figure S2.24.** Small molecule visible-light-induced cycloaddition reaction between 9,10-PQ and EtVyO in DMSO. The efficiency of the reaction is dramatically diminished in DMSO, with approx. 20 % conversion achieved after one-hour irradiation at 455 nm. The cycloaddition product is represented by the blue dashed line. UV absorbance is measured at 254 nm. As a result, peaks corresponding to EtVyO are not visible.

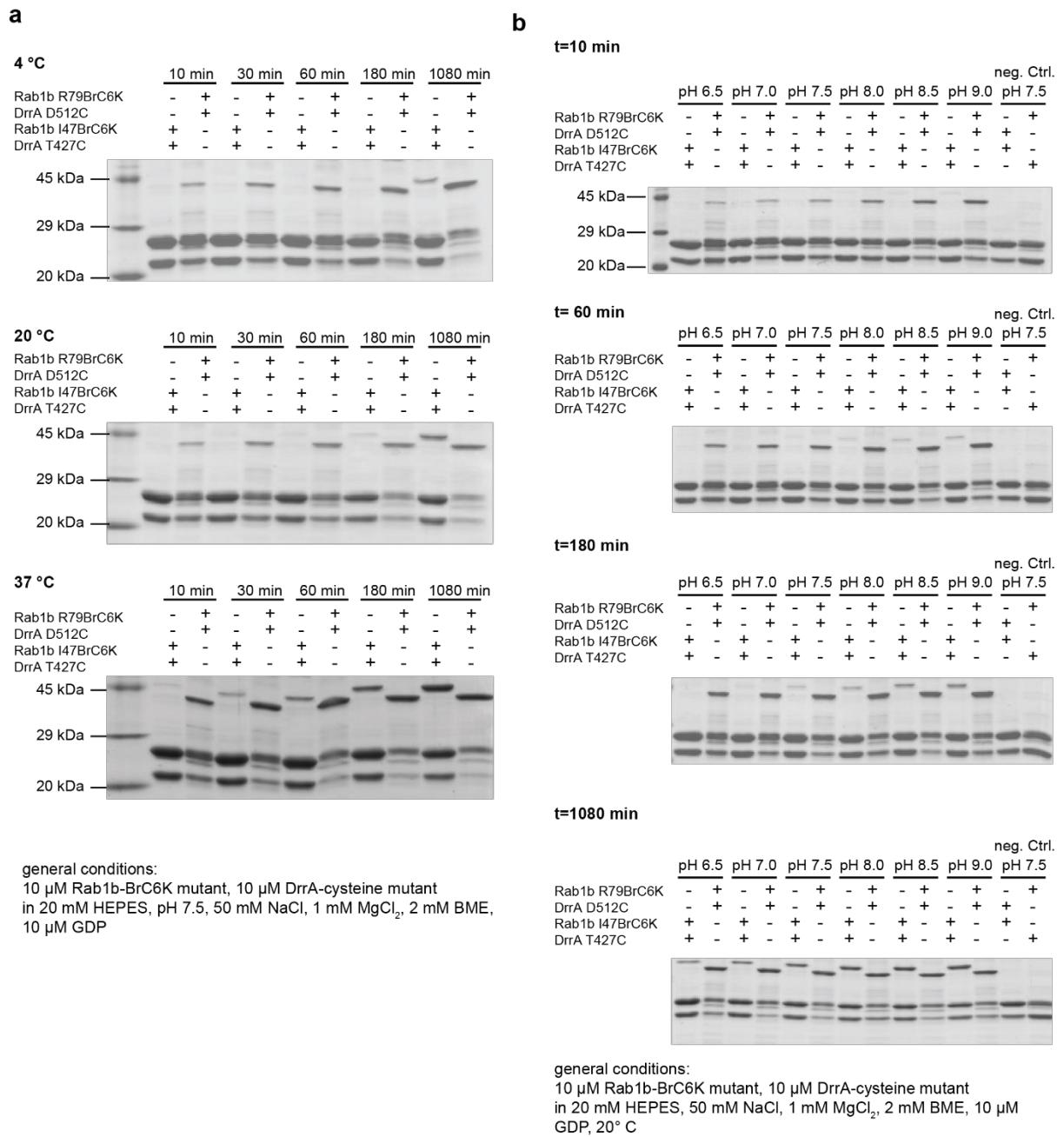
### 7.2. Supporting Figures – Chapter 3



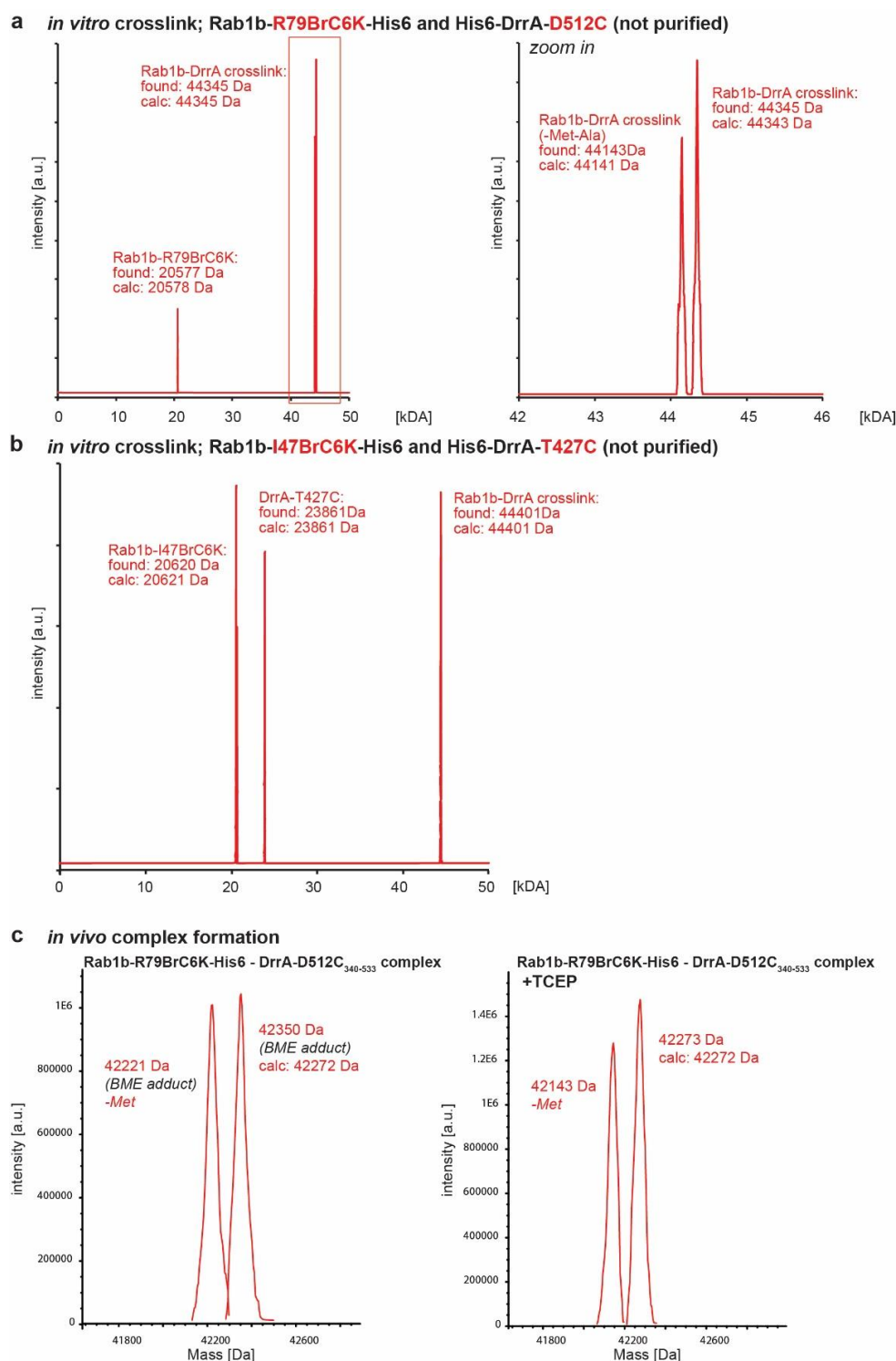
**Figure S3.1.** MS/MS analysis and identification of the crosslinked peptide in the sfGFP homodimer complex (cartoon representation in the bottom right corner, PDB: 1GFL) formed between E173 (green sphere in the cartoon representation) and BrC6K introduced at position 150 in sfGFP (red sphere in the cartoon representation). The crosslink was identified and validated using the software Kojak.<sup>[244]</sup> Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.



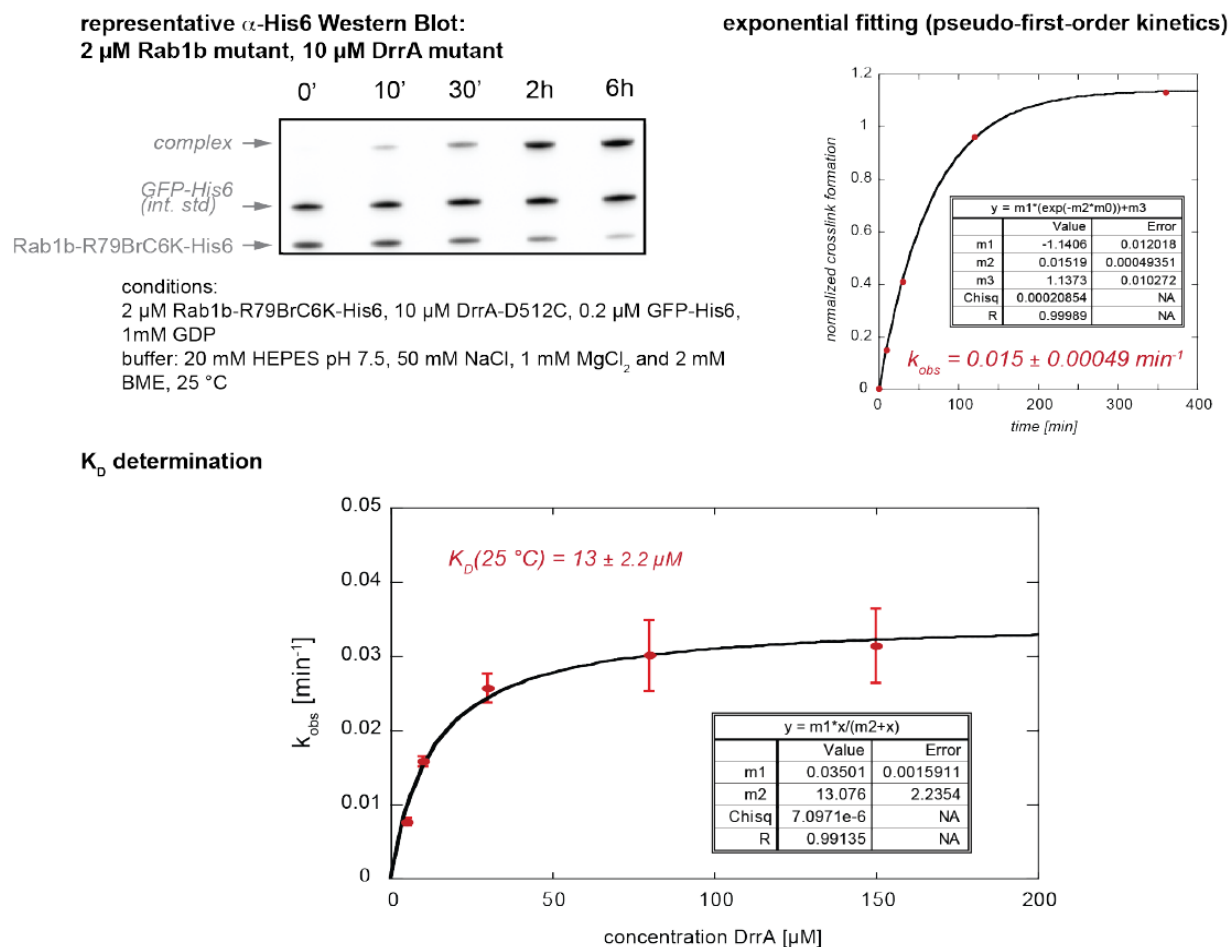




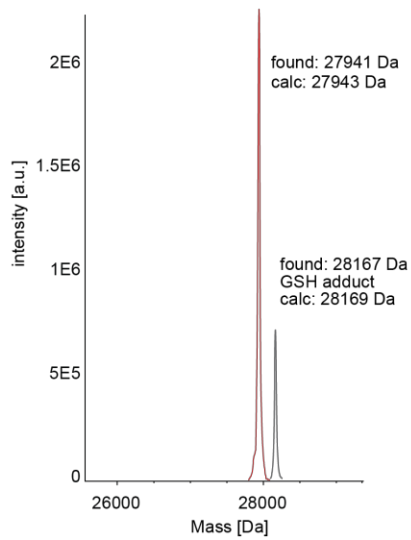
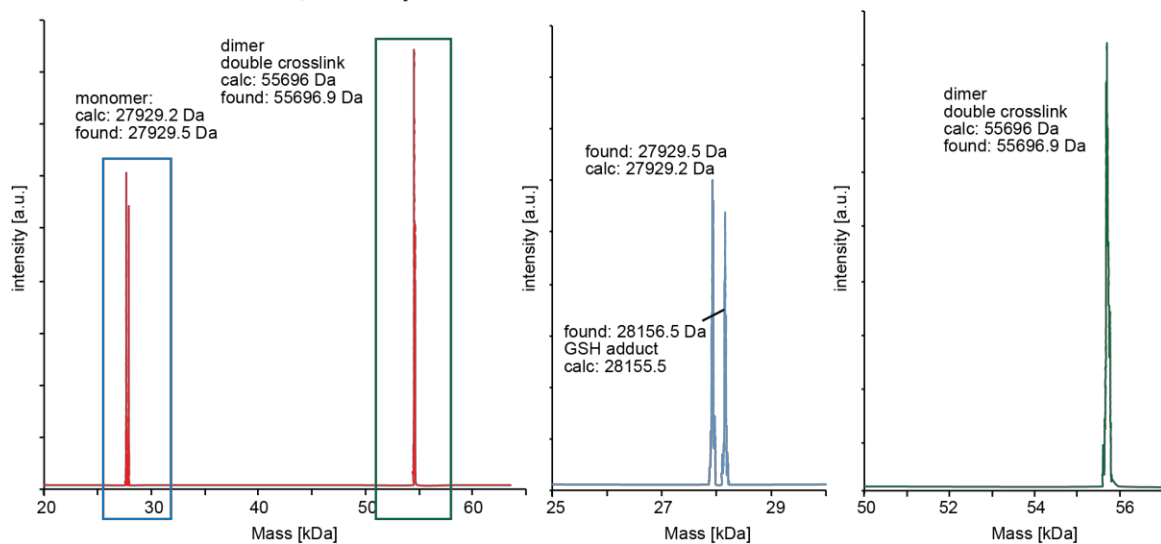
**Figure S3.3.** Screening of different conditions for *in vitro* crosslinking of complexes Rab1b-R79BrC6K:DrrA<sub>340-533</sub>-D512C and Rab1b-I47BrC6K:DrrA<sub>340-533</sub>-T427C. (a) Incubation of equimolar amounts of Rab1b and DrrA<sub>340-533</sub> variants (10 μM) at different temperatures. Samples taken at different time points were analysed by SDS-PAGE. (b) Incubation of equimolar amounts of Rab1b and DrrA<sub>340-533</sub> variants (10 μM) at 20 °C but at varying pH. Samples taken at different time points were analysed by SDS-PAGE. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.



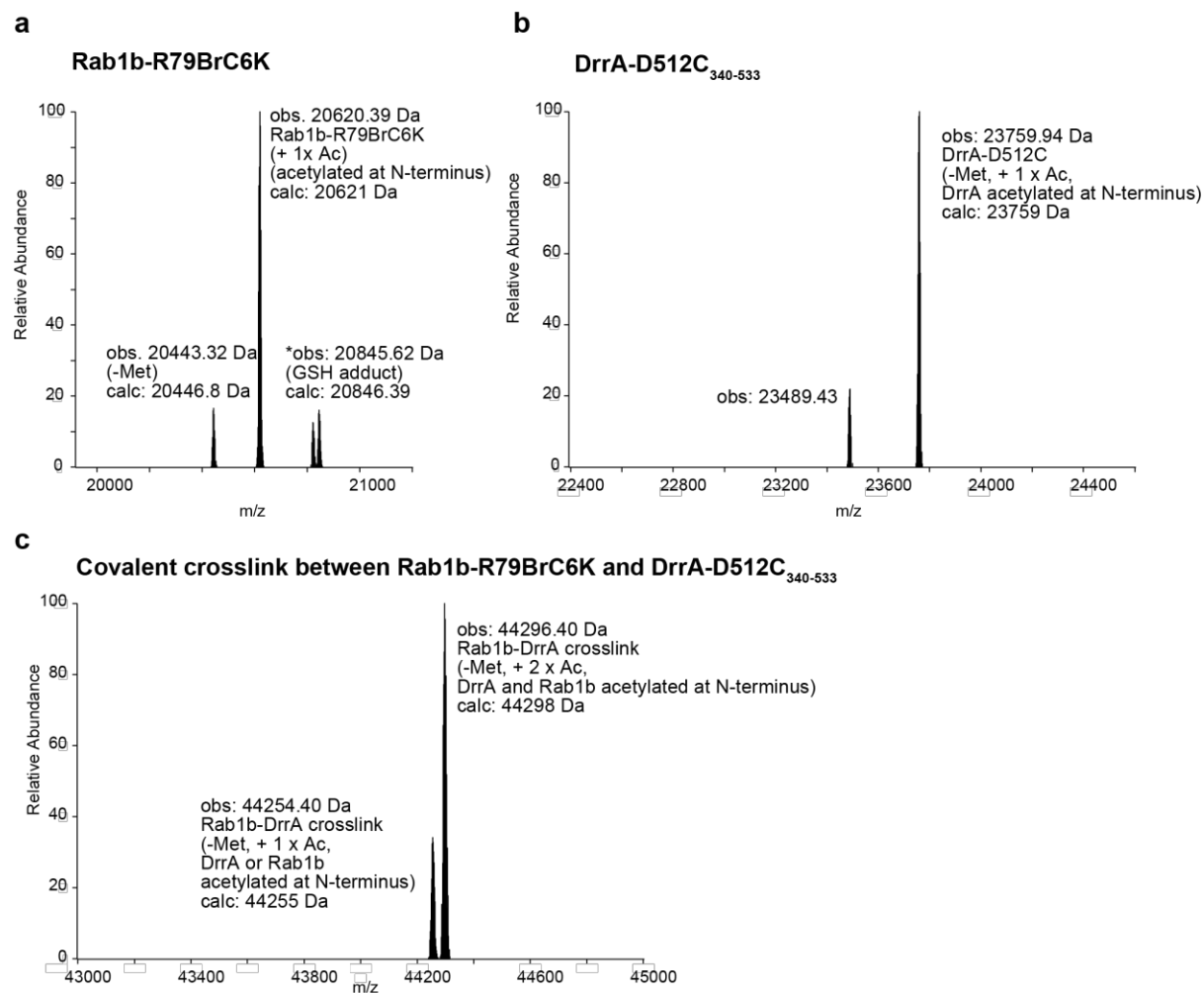
**Figure S3.4.** MS characterisation of the covalent crosslink formation between Rab1b-BrC6K and DrrA<sub>340-533</sub> variants *in vitro* and *in vivo*. (a) Incubation of Rab1b-R79BrC6K-His6 and His6-DrrA-D512C shows formation of the expected crosslink (observed mass: 44343 Da). As for Rab1b mutants, a second mass peak corresponding to the cleavage of the N-terminal Met-Ala dipeptide is visible. (b) Incubation of Rab1b-I47BrC6K-His6 and His6-DrrA-T427C shows formation of the expected crosslink (observed mass: 44401 Da). (c) Full-length protein mass spectrum of purified *in vivo* crosslinked Rab1b-R79BrC6K:DrrAD512C<sub>340-533</sub> complex. Left: The BME adduct of the covalently linked complex was detected in BME-containing storage buffer; right: Addition of TCEP to the protein sample affords only the mass of the covalently linked complex without the BME addition. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.



**Figure S3.5.** Determination of  $K_D$  of the crosslinking between Rab1b-R79BrC6K-His6 and DrrA<sub>340-533</sub>-D512C under pseudo-first order conditions. A fixed concentration of Rab1b-R79BrC6K (2  $\mu$ M) was incubated with increasing amounts of DrrA-D512C (5  $\mu$ M, 30  $\mu$ M, 80  $\mu$ M and 150  $\mu$ M) in the presence of 1 mM GDP. Furthermore, a fixed concentration of GFP-His6 (0.2  $\mu$ M) was added to the mixture as internal standard for subsequent quantification. The samples were incubated in crosslinking buffer at 25 °C, quenched at different time points and analysed via SDS-PAGE and anti-His western blot. The amount of crosslinked complex was normalised to the GFP signal, plotted against time and fitted with an exponential equation in order to determine the rate constant  $k_{obs}$ .  $k_{obs}$  values were plotted against DrrA concentration and the  $K_D$  was determined from the hyperbolic fit. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

**a sfGFP-N150BrC7K-his6, 24 h expression****b sfGFP-N150BrC6K-his6, 48 h expression**

**Figure S3.6.** MS analysis of purified sfGFP-N150BrC6K-His6 expressed in mammalian cells for (a) 24 and (b) 48 hours. Covalently crosslinked sfGFP homodimer is visible after 48-hour expression, while a GSH-adduct is also observable next to the monomeric sfGFP-BrC6K. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.



**Figure S3.7.** MS analysis full-length proteins expressed in HEK293T cells. (a) Rab1b-R79BrC6K; the observed mass (20620.39 Da) corresponds to N-terminal acetylated Rab1b-R79BrC6K-His6 while the peak at 20443.32 Da corresponds to the Rab1b-R79BrC6K-His6 mutant, where the N-terminal methionine has been cleaved off. A mass corresponding to a GSH adduct is denoted with the asterisk (\*). (b) His6-DrrA-D512C340-533; the observed mass (23759.94 Da) corresponds to the DrrA mutant where the N-terminal methionine has been cleaved off and the N-terminus is acetylated. (c) Covalently crosslinked complex that forms upon coexpression of Rab1b-R79BrC6K-His6 and His6-DrrA-D512C340-533: the observed mass (44296.40 Da) corresponds to the sum of the Rab1b and DrrA mutant and loss of HBr. Reprinted with permission from ref. 53. Copyright © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

## § 8. EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich eidesstattlich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen oder Hilfsmittel verwendet habe. Alle in dieser Arbeit sinngemäß oder wortwörtlich übernommenen Stellen habe ich gekennzeichnet. Diese Arbeit wurde für keinen anderen akademischen Grad eingereicht wie angegeben.

.....  
Ort, Datum

.....  
Unterschrift