Doctoral Thesis

Chemical protein synthesis by α-ketoacid-hydroxylamine ligation

Author(s):
Wucherpfennig, Thomas G.

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010666498

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
CHEMICAL PROTEIN SYNTHESIS BY
\(\alpha\)-KETOACID-HYDROXYLAMINE LIGATION

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH ZURICH)

Presented by

Thomas Georg Wucherpfennig

Master of Science in Chemistry, ETH Zurich,
Born on 07.12.1986
Citizen of Germany

Accepted on the recommendation of
Prof. Dr. Jeffrey W. Bode, examiner
Prof. Dr. Peter Kast, co-examiner
Prof. Dr. Pablo Rivera-Fuentes, co-examiner

2016
Abstract

Proteins are ubiquitous biomacromolecules in nature and fulfill diverse, important functions. Innumerable studies in life sciences require homogenous peptides and proteins of defined structure and composition. In many cases, the chemical synthesis of peptides and proteins offers unique advantages and possibilities compared to isolation from natural sources or recombinant protein production.

Direct linear chemical synthesis techniques such as solid-phase peptide synthesis routinely provide access to peptides and smaller proteins consisting of about 50–60 amino acid residues. However, most bioactive proteins are significantly larger, and alternative, convergent approaches such as the well-established native chemical ligation (NCL) are required. Despite its broad application, this method has several disadvantages, including the complicated synthesis of peptide thioesters and the requirement for a cysteine residue at the ligation site. Recently, several improvements have partially overcome these limitations, but NCL cannot yet provide a general solution to all protein targets and alternative approaches are desirable.

The Bode group developed the α-ketoacid-hydroxylamine (KAHA) ligation, a reaction using conditions and functional groups complementary to NCL. The key element of the KAHA ligation is the decarboxylative condensation of an α-ketoacid with a hydroxylamine to form an amide bond. Since its initial discovery in 2006, this reaction has undergone constant innovation, such as the development of 5-oxaproline, the best hydroxylamine derivative to date in terms of stability and reactivity for peptide and protein synthesis. These advances render the KAHA ligation a very flexible and powerful synthetic method as demonstrated in the synthesis of several smaller proteins.

This dissertation focuses on both technical improvements to the KAHA ligation and its application in the chemical synthesis of larger biologically interesting proteins. During this work, it was unexpectedly found that the primary products of the KAHA ligation with 5-oxaproline are esters rather than the anticipated amides. Mild conditions to rearrange these depsipeptides and depsiproteins to the natural amides were developed. The formation of esters provided insights into a possible mechanism of the KAHA ligation with 5-oxaproline.

A major limitation hampering the general applicability of the KAHA ligation – the cumbersome and limited synthesis of peptide α-ketoacids – was addressed. A novel protecting group for enantiopure α-ketoacids was developed and implemented in monomers for Fmoc-SPPS. This protocol delivers C-terminal peptide α-ketoacids directly upon resin cleavage without additional manipulation steps, and allows the inclusion of all canonical amino acids including cysteine, methionine and tryptophan. Further studies led to an orthogonal protecting group for α-ketoacids, which is stable to the acidic resin cleavage
conditions and can be selectively removed on demand. Both technologies provide a general, streamlined and scalable synthesis of peptide α-ketoacids and enable more flexible protein synthesis strategies.

KAHA ligations and the newly developed α-ketoacid protecting groups were utilized to synthesize milligram quantities of pure SUMO3, an important ubiquitin-like modifier protein involved in the regulation of numerous cellular processes. Despite containing two homoserine mutations, the synthetic SUMO3 protein retained full bioactivity and was readily conjugated to target proteins by the enzymatic SUMOylation machinery.

Derived from this synthetic protein, a chemical probe for cellular protein SUMOylation was developed. The probe overcomes the highly dynamic and reversible nature of protein SUMOylation by incorporating a C-terminal mutation retarding the deconjugation by endogenous SUMO-specific proteases. A biotin-tag and a fluorescence label on the probe allowed facile detection and affinity purification of SUMO–protein conjugates. The probe was applied in in vitro SUMOylation reactions using crude cell lysates as substrates and SUMOylated proteins were readily enriched and identified using MS-based proteomic experiments. Projecting forward, this technology could provide a versatile platform for identifying and profiling cellular protein SUMOylation under well-defined conditions such as cellular stress or pathogen infections to obtain further insights into the effects of SUMOylation on protein function.

The KAHA ligation technology was utilized in the synthesis of irisin, a myokine that is speculated to play an important role in modulating human fat metabolism. Milligram quantities of unlabeled and fluorescence labeled irisin were synthesized and used in cell binding studies, indicating the presence of a specific irisin-binding receptor on certain adipocytes. In future work, chemically synthesized and further modified irisin could be a valuable tool for the identification of the irisin receptor, shining light on the role of this controversial hormone in human physiology.
Zusammenfassung


Eine Hauptlimitierung, welche die breite Anwendbarkeit der KAHA-Ligation bein trächtigt – die umständliche und eingeschränkte Synthese der benötigten Peptid-α-ketosäuren – wurde behoben. Eine neue Schutzgruppe für enantiomerenreine

Die KAHA-Ligation und die neu entwickelten α-Ketosäure-Schutzgruppen wurden verwendet, um reines SUMO3 – ein wichtiges ubiquitin-artiges Modifikationsprotein, dass an der Regulation zahlreicher zellulärer Prozesse beteiligt ist – auf Milligrammskala zu synthetisieren. Obwohl das synthetische SUMO3-Protein zwei Homoserinmutationen enthielt, bewahrte es seine volle biologische Aktivität und wurde durch die enzymatische SUMOylierungsmaschinerie an Zielproteine konjugiert.


Acknowledgments

First, I wish to thank my supervisor and mentor, Prof. Dr. Jeffrey W. Bode, for providing me with the possibility to work on such exciting research projects. I am very grateful for his guidance through my PhD and for the all opportunities I got that allowed me to learn so many new aspects of chemistry and chemical biology.

Many thanks to my co-examiners, Prof. Dr. Peter Kast and Prof. Dr. Pablo Rivera-Fuentes, for taking their time to read this thesis and their valuable feedback.

I want to thank Mario Kessinger for helping me with so many aspects of PhD life and administrative issues.

I was very glad to work with so many knowledgeable and enthusiastic colleagues in the Bode group and at the LOC. Thank you all for your help with chemistry and for your company inside and outside the lab! In particular I want to thank all current and former members of the KAHA ligation subgroup for sharing ideas and inspiration: Dr. Ayodele Ogunkoya, Claudia Murar, Florian Rohrbacher, Dr. Frédéric Thuaud, Gábor Boross, Haewon Song, Dr. Ivano Pusterla, Dr. Javier Ruiz-Rodríguez, Dr. Satomi Shimura, Simon Baldauf, Thibault Harmand and Dr. Vijay Pattabiraman. Many thanks for all the innumerable discussions, coffee breaks and “Feierabendbiere” go to: André Zwicky, Benedikt Wanner, Florian Rohrbacher, Imants Kreituss, Felix Limberg, Dr. Ivano Pusterla, Dr. Johannes Teichert, Matthias Westphal, Dr. Michael Binanzer and Thibault Harmand!

Thank you for proofreading my thesis and your feedback: Dr. Christopher White, Florian Rohrbacher, Iain Stepek, Dr. Jonathan Hubert and Raphael Hofmann!

During my PhD, I had the pleasure to supervise two students in the lab, Felix Limberg and Estella Chin, who helped me a lot – thank you!

One of the most exciting experiences during my PhD was without doubt to join the Bode group at ITbM in Nagoya, Japan for three months. Many thanks to Prof. Jeffrey Bode for making this possible and to Dr. Shunsuke Oishi, Yoko Oishi, Dr. Chunmao He, Dr. Sameer Kulkarni and all other members of the ITbM lab in Nagoya for making my stay so pleasant and educative!

I am very grateful for the outstanding infrastructure and services at ETH that made my life and the work for my thesis so much easier and joyful. Many thanks to all hard working people at the HCI shop, the NMR and mass spectrometry services at the laboratory of organic chemistry at ETH.

My deepest gratitude for everything goes to my parents, Gertrud and Wolfgang, my girlfriend Yi-Lin and my family.
Table of Contents

CHAPTER 1: Introduction 2
1.1. Introductory remarks 2
1.2. Peptide chemistry 4
1.3. Coupling of protected protein segments 8
1.4. Protein synthesis by chemical ligation reactions 9
1.4.1. Native chemical ligation 10
1.4.2. Serine/threonine ligation 14
1.4.3. α-Ketoacid-hydroxylamine ligation 15
1.4.3.1. Mechanism of the KAHA ligation 15
1.4.3.2. Properties and synthesis of peptide α-ketoacids 16
1.4.3.3. Hydroxylamines for protein synthesis by KAHA ligation 18
1.4.3.4. Protein synthesis by KAHA ligation 20
1.5. Conclusions 22
1.6. References 23

CHAPTER 2: Formation of depsipeptides in the KAHA ligation with 5-oxaproline 32
2.1. Introductory remarks 32
2.2. Formation of depsipeptides in the KAHA ligation with 5-oxaproline 32
2.3. Mechanistic investigations 35
2.3.1. KAHA ligations with 5-oxaproline and different α-ketoacids 36
2.3.2. KAHA ligations with 5-oxaproline analogs 37
2.3.3. Influence of reaction solvent mixtures and epimerization analysis 38
2.3.4. Proposed reaction mechanism for KAHA ligations 40
2.4. Conclusions 44
2.5. References 45

CHAPTER 3: New protecting groups for the synthesis of peptide α-ketoacids 48
3.1. Introductory remarks 48
3.2. An acid-labile protecting group for α-ketoacids 49
3.2.1. Design of the protecting group 49
3.2.2. Screening of protecting groups and reaction optimization 50
3.2.3. Synthesis of a protected α-ketoacid monomer for SPPS 53
3.2.4. Synthesis of a C-terminal protein segment α-ketoacid using 41 54
3.3. An orthogonal protecting group for α-ketoacids 55
3.3.1. Design of the orthogonally protected α-ketoacid 56
3.3.2. Synthesis and studies of a orthogonally protected α-ketoacid monomer 58
### 3.3.3. Synthesis of a C-terminal protein segment with orthogonally protected α-ketoacid resin

50

### 3.4. Excursus: possible mechanism of the α-ketoacid protection reaction

62

### 3.5. Conclusions

63

### 3.6. References

64

#### CHAPTER 4: Chemical syntheses of SUMO3 proteins by KAHA ligation

68

### 4.1. Introductory remarks

68

### 4.2. Biological background

68

#### 4.2.1. SUMO and other ubiquitin-like proteins

68

#### 4.2.2. Mechanism of protein SUMOylation

69

#### 4.2.3. SUMO isoforms

70

#### 4.2.4. The enzymatic SUMOylation machinery

71

#### 4.2.5. Biological function of SUMO

73

### 4.3. Synthesis of SUMO3 protein and biochemical evaluation

74

#### 4.3.1. Retrosynthetic analysis of SUMO3

75

#### 4.3.2. Synthetic strategy

76

#### 4.3.3. Synthesis of protein segments

77

#### 4.3.4. Assembly of SUMO protein by sequential KAHA ligations

78

#### 4.3.5. Biochemical evaluation of synthetic SUMO3 (59)

82

#### 4.3.6. Conclusions

82

### 4.4. Modified SUMO3 proteins as probe for SUMOylation

83

#### 4.4.1. Design of the SUMOylation probe

84

#### 4.4.2. Synthetic strategy for modified SUMO proteins

87

#### 4.4.3. Synthesis of modified SUMO proteins

89

##### 4.4.3.1. Synthesis of the protein segments

89

##### 4.4.3.2. Synthesis of first generation modified SUMO3 proteins 60 and 61

91

##### 4.4.3.3. Synthesis of second generation modified SUMO3 proteins 62–65

94

##### 4.4.3.4. Synthesis of third generation modified SUMO3 protein 66

95

#### 4.4.4. Biochemical evaluation of modified SUMO3 proteins

97

##### 4.4.4.1. SUMOylation-deSUMOylation assays of first generation proteins 60 and 61

97

##### 4.4.4.2. SUMOylation-deSUMOylation assays of second generation proteins 62–65

99

#### 4.4.5. Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 as probe for cellular protein SUMOylation

100

### 4.5. Conclusions and outlook

104

### 4.6. References

105
# CHAPTER 5: Chemical synthesis of irisin protein

5.1. Introductory remarks 114
5.2. Biological background 114
5.3. Synthesis of irisin protein by KAHA ligation 116
5.3.1. Retrosynthetic analysis of irisin and synthetic strategy 116
5.3.2. Synthesis of protein segments 118
5.3.3. Assembly of irisin proteins by sequential KAHA ligations 119
5.4. Conclusions and outlook 122
5.5. References 123

# CHAPTER 6: Summary and outlook

6.1. References 126

# CHAPTER 7: Experimental Part

7.1. General Methods 132
7.2. Experimental for “CHAPTER 2: Formation of depsipeptides in the KAHA ligation with 5-oxaproline” 136
7.2.1. Preparation of starting materials 136
7.2.2. Studies with irisin model system 3 141
7.2.3. Studies with Fmoc-Ala-Leu-Hse-(4-F)-Phe-OH 6 tetrapeptide model system 144
7.2.4. NMR epimerization study 148
7.2.5. Model ligations with different $\alpha$-ketoacids 149
7.2.6. Model ligations with different hydroxylamines 151
7.2.7. HPLC study: KAHA ligation of H-Trp-Thr-Leu-$\alpha$-ketoacid 20 and H-Opr-(4-F)Phe-OH 5 156
7.2.8. Isotope labeling studies 158
7.3. Experimental for “CHAPTER 3: New protecting groups for the synthesis of peptide $\alpha$-ketoacids” 159
7.3.1. Synthesis of starting materials and relevant intermediates 159
7.3.2. Optimization of $\alpha$-ketoacid protection reaction using acid labile protecting groups 165
7.3.3. Synthesis of protected Fmoc-Leu $\alpha$-ketoacid 40 169
7.3.4. Synthesis of orthogonally protected Fmoc-leucine $\alpha$-ketoacid 49 172
7.3.5. Procedure for loading of protected Fmoc-Xaa $\alpha$-ketoacid onto resin 177
7.3.6. Model studies with orthogonally protected Fmoc-leucine $\alpha$-ketoacid 49 178
7.4. Experimental for “CHAPTER 4: Chemical synthesis of SUMO3 proteins by KAHA ligation”: Protein synthesis and characterization 184
7.4.1. Synthesis of relevant small molecule starting materials 184
7.4.2. Protein sequence of SUMO3 (2–92) 190
7.4.3. Synthesis of SUMO3 protein 190
7.4.3.1. Synthesis of protein segments 191
7.4.3.2. Sequential KAHA ligations for the synthesis of H-[SUMO3 (2–92)]-OH 194
7.4.4. First generation of modified SUMO3 protein 201
7.4.4.1. Synthesis of protein segments 201
7.4.4.2. Sequential KAHA ligations for the synthesis of modified SUMO3 proteins 60 and 61 206
7.4.5. Second generation of modified SUMO3 protein 212
7.4.5.1. Synthesis of protein segments 212
7.4.5.2. Sequential KAHA ligations for the synthesis of modified SUMO3 proteins 62–65-A/B 217
7.4.6. Third generation of modified SUMO3 protein 225
7.4.6.1. Synthesis of protein segments 225
7.4.6.2. Sequential KAHA ligations for the synthesis of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 228
7.5. Experimental for “CHAPTER 4: Chemical synthesis of SUMO3 proteins by KAHA ligation”: Biochemical experiments and SUMOylation assays with synthetic SUMO proteins 234
7.5.1. Materials and General Methods 234
7.5.2. SUMOylation of RanGAP1 with H-[SUMO3 (2–92)]-OH 59 239
7.5.3. SUMOylation studies of RanGAP1 using Biotin-[SUMO3 (2–92)]-OH 60 and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 240
7.5.4. SUMOylation studies of RanGAP1 with Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH proteins 62–65 242
7.5.5. SUMOylation studies of proteins from crude HEK293 cell lysates with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 244
7.6. Experimental for “CHAPTER 5: Chemical synthesis of irisin protein” 249
7.6.1. Synthesis of segments 250
7.6.2. Sequential KAHA ligations for the synthesis of irisin proteins 87 and 88 254

CHAPTER 8: Appendix 262
8.1. NMR spectra 262
8.2. Proteomics results 311
Publications and copyright permission notes

Part of this dissertation is reproduced with permission from:

Copyright © 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
License Number: 3852960718883

Copyright © 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
License Number: 3852960560210

Copyright © 2015 Springer International Publishing, Cham.
License Number: 3852960859426
## Symbols and abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\alpha]_D$</td>
<td>specific rotation at the wavelength of the sodium D line</td>
</tr>
<tr>
<td>$\delta$</td>
<td>chemical shift</td>
</tr>
<tr>
<td>$\tilde{\nu}$</td>
<td>wave number</td>
</tr>
<tr>
<td>Ac</td>
<td>acetyl</td>
</tr>
<tr>
<td>Boc</td>
<td>tert-butoxycarbonyl</td>
</tr>
<tr>
<td>Bt</td>
<td>benzotriazole</td>
</tr>
<tr>
<td>Bz</td>
<td>benzoyl</td>
</tr>
<tr>
<td>Cbz</td>
<td>carboxybenzyl</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>COMU</td>
<td>(1-cyano-2-ethoxy-2-oxo-ethylidenaminoxy)dimethylaminomorpholinocarbenium hexafluorophosphate</td>
</tr>
<tr>
<td>Cou</td>
<td>coumarin</td>
</tr>
<tr>
<td>d. r.</td>
<td>diastereomeric ratio</td>
</tr>
<tr>
<td>Dap</td>
<td>$(S)$-2,3-diaminopropionic acid</td>
</tr>
<tr>
<td>DIPEA</td>
<td>$N,N$-diisopropylethylamine</td>
</tr>
<tr>
<td>DMF</td>
<td>$N,N$-dimethylformamide</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-dimethylaminopyridine</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DODT</td>
<td>3,6-dioxa-1,8-octanediolthiol</td>
</tr>
<tr>
<td>E</td>
<td>entgegen (olefin geometry)</td>
</tr>
<tr>
<td>ee</td>
<td>enantiomeric excess</td>
</tr>
<tr>
<td>equiv</td>
<td>equivalent</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>Et</td>
<td>ethyl</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-fluorenlymethoxycarbonyl</td>
</tr>
<tr>
<td>FTICR</td>
<td>fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HATU</td>
<td>1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate</td>
</tr>
<tr>
<td>HCCA</td>
<td>$\alpha$-cyano-4-hydroxycinnamic acid</td>
</tr>
<tr>
<td>HCTU</td>
<td>1-[bis(dimethylamino)methylene]-5-chlorobenzotriazolium 3-oxide hexafluorophosphate</td>
</tr>
<tr>
<td>HMBC</td>
<td>heteronuclear multiple bond correlation</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>Hse</td>
<td>$(S)$-homoserine</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>i</td>
<td>iso</td>
</tr>
<tr>
<td>iPr</td>
<td>i-propyl</td>
</tr>
<tr>
<td>IR</td>
<td>infra-red</td>
</tr>
<tr>
<td>J</td>
<td>coupling constant</td>
</tr>
<tr>
<td>KAHA</td>
<td>$\alpha$-ketoacid hydroxylamine (ligation)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>m</td>
<td>multiplet (NMR), milli</td>
</tr>
<tr>
<td>M</td>
<td>molarity (mol/l)</td>
</tr>
<tr>
<td>m/z</td>
<td>mass to charge ratio</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix-assisted laser desorption/ionization</td>
</tr>
<tr>
<td>Me</td>
<td>methyl</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>Ms</td>
<td>methylsulfonyl</td>
</tr>
<tr>
<td>MTBE</td>
<td>methyl tert-butyl ether</td>
</tr>
<tr>
<td>n.r.</td>
<td>no reaction</td>
</tr>
<tr>
<td>NCL</td>
<td>native chemical ligation</td>
</tr>
<tr>
<td>NMM</td>
<td>N-methylmorpholine</td>
</tr>
<tr>
<td>NMP</td>
<td>N-methylpyrrolidinone</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>Opr</td>
<td>(S)-5-oxaproline</td>
</tr>
<tr>
<td>p</td>
<td>para</td>
</tr>
<tr>
<td>Pbf</td>
<td>2,2,4,6,7-pentamethyl dihydrobenzofuran-5-sulfonyl</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>Ph</td>
<td>phenyl</td>
</tr>
<tr>
<td>PMB</td>
<td>p-methoxybenzyl</td>
</tr>
<tr>
<td>PMP</td>
<td>p-methoxyphenyl</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>Pr</td>
<td>propyl</td>
</tr>
<tr>
<td>PTM</td>
<td>posttranslational modification</td>
</tr>
<tr>
<td>QTOF</td>
<td>quadrupole time-of-flight</td>
</tr>
<tr>
<td>R</td>
<td>general substituent</td>
</tr>
<tr>
<td>R</td>
<td>rectus (configuration)</td>
</tr>
<tr>
<td>r.t.</td>
<td>room temperature (22–25 ºC)</td>
</tr>
<tr>
<td>Rho</td>
<td>rhodamine</td>
</tr>
<tr>
<td>s</td>
<td>singlet</td>
</tr>
<tr>
<td>S</td>
<td>sinister (configuration)</td>
</tr>
<tr>
<td>SFC</td>
<td>supercritical fluid chromatography</td>
</tr>
<tr>
<td>SPPS</td>
<td>solid-phase peptide synthesis</td>
</tr>
<tr>
<td>t</td>
<td>triplet</td>
</tr>
<tr>
<td>TBAF</td>
<td>tetrabutylammonium fluoride</td>
</tr>
<tr>
<td>tBu</td>
<td>tert-butyl</td>
</tr>
<tr>
<td>tert</td>
<td>tertiary</td>
</tr>
<tr>
<td>Tf</td>
<td>trifluoromethylsulfonyl</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TIPS</td>
<td>triisopropylsilane</td>
</tr>
<tr>
<td>TLC</td>
<td>thin-layer chromatography</td>
</tr>
<tr>
<td>TMS</td>
<td>trimethylsilyl</td>
</tr>
<tr>
<td>Ubl</td>
<td>ubiquitin-like protein</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>Xaa</td>
<td>any α-amino acid</td>
</tr>
<tr>
<td>Z</td>
<td>zusammen (olefin geometry)</td>
</tr>
</tbody>
</table>
“Be naïve and do an experiment.”
Attributed to Feodor Lynen
CHAPTER 1

INTRODUCTION
CHAPTER 1: Introduction

1.1. Introductory remarks

Proteins\(^1\) are ubiquitous in nature and fulfill overwhelmingly diverse functions. For example, some proteins serve as structural elements of the cell and compose the cytoskeleton to give cells a defined 3-dimensional shape and anchor other cellular components. Transport proteins utilize chemical energy to relocate objects – ranging from ions and small molecules to macromolecules and vesicles – within the cell or to the outside, and are ultimately responsible for muscle contraction and movement of whole organisms. As enzymes, proteins catalyze the majority of chemical reactions in the cell and thereby play a key role in metabolism and energy turnover. Some proteins transmit information between cells and throughout the organism (e.g. hormones), and regulate numerous other processes such as DNA expression (e.g. transcription factors) and immune response (e.g. antibodies).

Proteins are polymers composed of – in most eukaryotes – 21 proteinogenic \(\alpha\)-amino acids (Figure 1), which are linked together through amide (or peptide) bonds. With the exception of glycine, all amino acids have side chains, most of them decorated with functional groups, mediating intra- and intermolecular interactions and reactivity. All amino acids possessing a side chain are chiral at the \(\alpha\)-carbon atom and natural (ribosomal) proteins contain exclusively (S)-\(\alpha\)-amino acids (except for (R)-cysteine). The length of these amino acid polymers can range from only a few amino acid residues (which are then often referred to as oligopeptides) to huge proteins containing thousands of amino acid residues.

![Amino Acids Diagram](image)

**Figure 1.** The 21 proteinogenic amino acids with one- and three-letter codes.

Yet, another level of complexity is added: amino acid residues in many proteins are chemically modified after translation with so called posttranslational modifications (PTMs), which can be very important for the structure and function of the protein and are often exploited by nature to regulate protein activity.\(^2\) These PTMs are highly diverse and can
involve oxidation of the amino acid side chain (e.g. hydroxylation or carboxylation) or modification of side chain functional groups via the attachment of small molecules (acylation, methylation, phosphorylation, glycosylation, and many others) or even other proteins (ubiquitination, SUMOylation, etc.) (Figure 2).

![Figure 2. Selected examples of amino acids modified with posttranslational modifications.](image)

Many properties and functions of proteins are dictated by their structure, since the protein structure provides a defined spatial arrangement of functional groups mediating the interactions and reactions of the protein with itself and other (macro-) molecules. It is not surprising that most proteins reveal a high level of organization in hierarchical structures (Figure 3). The linear sequence of amino acids in a polypeptide chain is called the primary structure. These polypeptide chains can locally fold into defined structures such as α-helices, β-strands and many others; these elements comprise the secondary structure. The overall structure ("fold") of a protein – which might contain a multitude of different secondary structure elements – is referred to as tertiary structure. Some proteins consist of more than one amino acid chain ("subunits") – with each its own tertiary structure – and the overall assembly of these subunits is termed quaternary structure.

![Figure 3. The hierarchical structure of proteins.](image)

Given the multitude of functions and processes proteins are involved in, it is obvious that proteins and their understanding on all kinds of different levels – such as structure, function and interactions with other molecules – play a central role in modern science. For many studies, access to pure and homogenous proteins is required.

For a long time, isolation from biological samples was the only access to proteins. In the last 50 years, the field of molecular biology has made significant advances and nowadays the recombinant production of many proteins in different organisms is routine. However, this approach has still limitations. First, in most cases only proteins composed of
the proteinogenic amino acids are readily accessible and despite recent advances expanding
the genetic code such as the amber codon suppression technology\textsuperscript{4} or the pyrrolysine
translational machinery,\textsuperscript{5} it is challenging to incorporate one or even multiple unnatural amino
acids in recombinant proteins. Second, the site-specific and homogenous incorporation of
PTMs in recombinant proteins is technically difficult and impossible to control precisely.\textsuperscript{6}

As an alternative to biologically produced material, chemical protein synthesis has
tremendously advanced in the last two decades and offers an attractive option to obtain pure
peptides and small- to mid-sized proteins for research and pharmaceutical applications.\textsuperscript{7}
Most importantly, the stepwise chemical synthesis provides absolute control over the primary
sequence of the synthesized peptides or proteins and makes possible the site-specific
incorporation of all kinds of non-proteinogenic amino acids, including many posttranslational
modifications.

1.2. Peptide chemistry

Following chemical intuition and the paradigm of biology, the chemical synthesis of
proteins relies on the amide bond-forming coupling of single amino acid building blocks. To
form an amide bond between two amino acids, synthetic chemists have to overcome two
significant challenges.

First, the formal dehydrative coupling of a carboxylic acid and an amine to form an
amide bond is thermodynamically unfavorable and has a high activation energy\textsuperscript{8} – therefore
some kind of “activating reagent” is required to render this process thermodynamically
favorable and feasible on a useful time scale (Scheme 1).

Second, the amino acids contain either reactive functional groups on the C- or the
N-terminus and possibly on the side chains that would react with the activated carboxylic
acid and yield undesired side products. To avoid side reactions, all functional groups that
should not participate in the coupling reaction have to be masked by suitable protecting
groups.

\begin{center}
\begin{tikzpicture}
\node at (0,0) (a) {\includegraphics[width=0.8\textwidth]{c1}};
\end{tikzpicture}
\end{center}

\textbf{Scheme 1. The requirement of an "activating agent" and protecting groups in amino acid coupling
reactions.}
Regarding the first challenge, it was promptly realized that activated carboxylic acid equivalents – such as acid chlorides – readily react with amines to form the desired amide bond, as illustrated by the landmark synthesis of a dipeptide by Emil Fischer in 1903 (Scheme 2).\(^9\)

![Scheme 2. Synthesis of a dipeptide via activation of a carboxylic acid as acid chloride.](image)

After more than hundred years of research, numerous different approaches to form amide bonds have been developed.\(^10\) For peptide and protein synthesis, modern chemists rely on “coupling reagents” to activate the carboxylic acid. A key feature of these coupling reagents is that they promote fast reactions under mild conditions, which ideally prevents epimerization of the amino acid stereocenter, a common problem in reactions using highly activated \(\alpha\)-amino acid derivatives such as acid chlorides.\(^11\) Over the years, a myriad of coupling reagents has been developed exhibiting differences regarding reactivity, solubility and safety properties.\(^12,13\) The most frequently used classes of coupling reagents include carbodiimides, uronium-based and phosphonium-based reagents (Figure 4).

![Figure 4. Examples of commonly used coupling reagents for amide bond formation.](image)

The second challenge – preventing unwanted side reactions of the activated amino acid with other functional groups – has been addressed by the development of several protecting group strategies. Each of these strategies must have two orthogonal sets of protecting groups: one temporary protecting group for the N-terminal amine; and one orthogonal, permanent protecting group for side chain functional groups and the C-terminus, which is only removed after complete synthesis. The two most commonly employed protecting group strategies are – named after the temporary backbone amine protecting group utilized – the “Boc”- and “Fmoc”-strategy (Figure 5).\(^14\)
Figure 5. Protecting group strategies for sequential amino acid couplings.

In the Boc-strategy, the N-terminal Boc protecting group can be removed using trifluoroactic acid (TFA), whereas the side chain functional groups are masked with protecting groups that are only removed under strongly acidic conditions such as liquid hydrogen fluoride (HF). In the Fmoc-strategy, the N-terminal amine is protected using the base-labile Fmoc-group and the side chain functional groups are masked with TFA-labile protecting groups (Figure 6).\(^\text{15}\)

Figure 6. Protecting groups frequently used in Fmoc-peptide synthesis.

A truly disruptive innovation in the field of peptide chemistry was the introduction of solid phase peptide synthesis (SPPS) by Merrifield in 1963.\(^\text{16}\) Previous solution-phase
techniques required the tedious isolation and purification of intermediates after every step and were in many cases further hampered by limited solubility of the growing peptide chain, thus restricting the length of accessible polypeptide chains. Synthesis on solid support addressed these constraints by mounting the growing peptide chain onto a solid polymeric support ("resin") via a linker attached to the C-terminus of the first amino acid. The immobilization on solid support prevents – in most cases – the aggregation of the growing peptide chain and ensures that the N-terminal end is exposed to the reagents in the solution phase. Reactants and reagents are introduced as solutions, which can be readily removed from the product by filtration and washings. This facile separation allows for the use of large excess of reagents, pushing chemical reactions to high conversion and improving the purity and yield of the desired product. Purification of intermediates is not required, as the desired product is always immobilized on the resin. After complete synthesis, the peptide is cleaved from the solid support and the side chain protecting groups are removed simultaneously to yield the desired peptide. If necessary, this can be further purified, typically by reversed-phase HPLC. Although both Fmoc- and Boc-protecting group strategies are compatible with SPPS, most standard syntheses are nowadays preferably performed using the Fmoc-strategy to avoid the safety concerns imposed by the need for more hazardous reagents such as HF in the Boc-strategy. In Scheme 3, a typical protocol for Fmoc-SPPS is illustrated.

Scheme 3. Workflow of Fmoc-solid phase peptide synthesis and examples of frequently utilized linkers.
There are few examples of larger proteins synthesized by linear SPPS, including ribonuclease A (124 residues) and human immunodeficiency virus Tat protein (86 residues). Despite elaborately optimized synthetic protocols and yields for individual steps often approaching the theoretical limit, longer polypeptides tend to increasingly aggregate on the solid support as they are built up, resulting in incomplete reactions. Additionally, statistics favor the accumulation of impurities (e.g. truncated peptides) with growing length of the peptide chains. Both factors lower the purity and yield of the desired product dramatically and limit in general the accessible length of polypeptides produced by routine and automated SPPS to 50–60 residues. However, most eukaryotic proteins are significantly larger – around 200–350 amino acid residues – and can clearly not be accessed by a single linear SPPS. To address this, several convergent strategies which assemble the final protein from individual segments have been developed.

1.3. Coupling of protected protein segments

A conceptually straightforward and, from historical perspective, the oldest solution to synthesize larger proteins is to divide the target into several smaller segments, which are easier to synthesize by SPPS, and assemble the segments using amide bond-forming coupling reagents (Scheme 4). However, this approach was initially hampered by the limited methods to access protected peptides with a free C-terminal carboxylic acid, but this issue has been addressed by the development of the highly acid labile 2-chlorotrityl linker, which allows for resin cleavage without side chain deprotection. As discussed before, the use of coupling reagents requires that the functional groups on the segments are – with the exception of the C- or N-termini to be coupled – fully protected.

Scheme 4. Coupling of fully protected protein segments in solution.

This convergent approach was used in the synthesis of several bioactive proteins, for example pleiotrophin (136 residues) and a precursor protein to the green-fluorescent protein (238 residues). However, the required protecting groups render the protein
segments highly hydrophobic, making their purification and characterization very challenging. Even more severely, most fully protected protein segments exhibit very limited solubility in commonly used organic solvents and segment coupling reactions thus have to be performed under highly dilute conditions, which slows down the reaction rate and promotes side reactions such as epimerization of the activated carboxylic acid.\textsuperscript{26}

These drawbacks stimulated the development of an alternative strategy for the synthesis of proteins, the so-called chemical ligation.

1.4. **Protein synthesis by chemical ligation reactions**

In general, a chemical ligation\textsuperscript{27} is a chemoselective reaction that covalently links two highly functionalized, unprotected molecules together; in most cases this term is used in the context of peptide and protein synthesis.

Chemical ligation reactions achieve their selectivity by using unique chemical functionalities introduced at the ligation site, which react together in a highly chemoselective fashion. Typically, a thermodynamic driving force for the ligation reaction is embedded in these metastable functional groups (such as labile, high-energy chemical bonds), which makes the addition of external reactants (e.g. coupling reagents in segment couplings) or even catalysts (e.g. enzymes in enzyme-mediated ligation reactions) unnecessary.\textsuperscript{28,29}

For most protein syntheses, it is desirable to have a chemical ligation reaction that yields a natural amide bond at the ligation site (Scheme 5), but there are chemical ligation reactions that form other bonds at the ligation site.\textsuperscript{30}

![Scheme 5. Concept of an amide bond-forming chemical ligation reaction.](image)

A chemical ligation reaction – especially in the context of chemical protein synthesis – has to meet several requirements. Such a reaction should be: a) *Chemoselective*. The two functional groups participating in the chemical ligation reaction have to exclusively react with each other and must tolerate the diverse functional groups present in the protein segments. b) *Rapid*. Chemical ligation reactions are bimolecular reactions and as such, the reaction rate strongly depends on the concentration of both reaction partners. The high molecular weight of the reaction partners – in protein synthesis this can be more than 10000 Daltons – restricts the achievable molar concentrations in the ligation reaction and the solubility of the protein segments is often limiting. Therefore, most chemical ligation reactions are performed under...
dilute conditions at low-millimolar concentrations. Additionally, the ligation partners are valuable molecules and it is desirable to use them in stoichiometric amounts. To achieve good conversion under these conditions within a reasonable reaction time, the ligation reaction has to be fast.\textsuperscript{31}c) \textit{Mild.} The chemical ligation reaction has to be performed under conditions (temperature, pH range, solvents) that do not impair the various functional groups or stereocenters present in the unprotected protein segment.

These considerations led to the development of three main classes of amide bond-forming chemical ligation reactions routinely used in chemical protein synthesis (Scheme 6).

1.4.1. Native chemical ligation

The foundation for the native chemical ligation is rooted in the finding of Wieland \textit{et al.} in 1953 that amino acid thioesters react with cysteine in aqueous buffer to form an amide bond.\textsuperscript{32} Their careful investigations showed that the reaction proceeds first via a transthioesterification reaction followed by an intramolecular $S$ to $N$ acyl transfer to give the amide bond (Scheme 7).

This acyl-shift principle was not utilized for a long time until Kemp \textit{et al.} developed the first ligation method for peptide segments using a template and an $O$ to $N$ acyl transfer in the so called “prior thiol capture strategy” in 1981 (Scheme 8).\textsuperscript{33} This method was further optimized and ultimately used to synthesize a 39-mer peptide from minimally protected peptide segments.\textsuperscript{34}
Nowadays, the most widely used method for chemical protein synthesis is native chemical ligation (NCL). Likely inspired by the initial findings of Wieland and Kemp, Kent et al. disclosed the native chemical ligation in 1994. In this reaction, unprotected C-terminal peptide thioesters react with N-terminal cysteine peptides under mild conditions to form the corresponding native amide bond. The reaction is highly chemoselective and tolerates all canonical amino acid side chain functionalities. Similar to Wieland’s work, the reaction mechanism comprises a reversible transthioesterification in the first step – which explains why the reaction tolerates unprotected internal cysteine residues – followed by an irreversible S to N acyl transfer (Scheme 9).

The reactions proceed optimally in buffered aqueous solutions at neutral pH range, which is limited by two factors: a) the reduced nucleophilicity of the cysteine thiol at low pH and b) the limited stability of thioesters towards hydrolysis and direct reaction with amines.
under more basic conditions. Importantly, the reaction proceeds well in denaturing solvent mixtures such as buffered solutions of GdmCl or urea that generally provide high solubility of the protein segments and exposes the reactive termini to the bulk solution. Systematic studies\textsuperscript{36,37} showed that the reaction benefits from the addition of aromatic thiols such as MPAA (4-mercaptopheynylacetic acid) for two reasons: the commonly used and more stable alkyl thioesters are converted \textit{in situ} to highly reactive aryl thioesters – thereby accelerating the overall reaction rate – and the excess thioles maintain reducing conditions and prevent cysteine oxidation.

The central advantages of the native chemical ligation include the use of denaturing conditions and a relatively fast reaction rate, which make the synthesis of proteins with sensitive posttranslational modifications such as glycosylation\textsuperscript{38} or phosphorylation possible.\textsuperscript{39} Additionally, the starting materials of NCL – protein segments with N-terminal cysteines or C-terminal thioesters – can be obtained by recombinant overexpression, which opens an avenue to the acquisition of large protein segments that could not be accessed by SPPS. Whereas segments with N-terminal cysteines can be readily produced via recombinant production of a precursor protein and processing with specific proteases to reveal the terminal cysteine,\textsuperscript{40} the access to thioesters is more laborious. Muir \textit{et al.} developed the so-called expressed protein ligation (EPL), which exploits a natural phenomenon of certain proteins (“Inteins”) to spontaneously undergo protein splicing, forming an intermediate thioester. Intein fusion proteins can be trapped with an exogenous thiol and the obtained thioesters used in a subsequent chemical ligation reaction.\textsuperscript{41}

One limitation of NCL is the non-trivial access to the required C-terminal thioesters. Whereas peptide thioesters are readily synthesized by Boc-SPPS, they are not compatible with the basic conditions used in the otherwise much milder Fmoc-SPPS, which precludes the synthesis of thioesters containing sensitive moieties such as many posttranslational modifications. Elaborate, but often not generally applicable protocols for the synthesis of peptide thioesters from precursor peptides that are accessible by Fmoc-SPPS have been developed. These include the thiolysis of \textit{N}-acyl-benzimidazolinones\textsuperscript{42} and the intramolecular \textit{N} to \textit{S} acyl transfer of tertiary amides\textsuperscript{43,44} or \textit{O} to \textit{S} acyl transfer approaches.\textsuperscript{45,46} Recently, Liu \textit{et al.} reported a broadly applicable synthetic method involving the mild thiolysis of readily accessible peptide hydrazides, which has the potential to become a standard method for the synthesis of peptide thioesters (Scheme 10).\textsuperscript{47,48} Nevertheless, all methods require additional chemical steps after the peptide synthesis, often leading to decreased purities and yields of the desired product.
Another significant disadvantage of NCL is the requirement for a thiol-containing amino acid such as cysteine or other unnatural amino acids at the ligation site. Cysteine is the least abundant amino acid and most proteins do not contain a cysteine residue at the desired ligation site. The introduction of cysteine mutations often affects the protein folding and activity due to its reactivity and potential to form intra- or intermolecular disulfide bonds. Numerous workarounds have been developed including the use of removable auxiliaries (e.g. N-oxyethanethiol), chemical modification after ligation such as alkylation – generating analogs of other amino acids – or reductive desulfurization leading to alanine or other natural amino acids at the ligation site (Scheme 11). However, all these methods have limitations (e.g. the incompatibility with other unprotected cysteines) and make additional chemical steps at late stage of the synthesis necessary.

**Scheme 10. Selected methods for the synthesis of C-terminal peptide thioester with Fmoc-SPPS.**

Mechanistically related variants of NCL include ligations at N-terminal histidine or selenocysteine or the use of C-terminal selenoesters (Scheme 12). Although these reactions offer distinct advantages such as new ligation sites or increased reaction rates,
they suffer from reduced chemoselectivity (His-ligation) or lability of the starting materials (Se-compounds), which limits their broader application.

In summary, native chemical ligation remains the current state of the art in chemical protein synthesis and has been applied in numerous syntheses of large and complex proteins. Selected examples include the total synthesis of tetraubiquitinated α-synuclein (444 residues),\(^{58}\) glycosylated human erythropoietin (166 residues)\(^{59}\), different glycoforms of the chemokine CCL1 (73 residues)\(^ {60}\) and the synthesis of both enantiomers of a chaperone-dependent protein, DapA (312 residues).\(^ {61}\) Yet, the limitations regarding suitable ligation sites for NCL render complementary ligation methods desirable.

### 1.4.2. Serine/threonine ligation

A relatively new chemical ligation using different functional groups is the serine/threonine ligation (STL) introduced by Li \textit{et al.} in 2010,\(^ {62}\) based on the elegant prior work of Tam \textit{et al.} in 1994.\(^ {63}\) Conceptually related to the template approach from Kemp, a C-terminal salicylaldehyde ester captures the N-terminal amine of a serine or threonine residue and forms a stable cyclic \(N,O\)-benzylidene acetal intermediate at the ligation site (Scheme 13). This can be readily cleaved under acidic conditions to yield a native amide bond at the ligation site.

The reaction is compatible with most amino acid residues at the C-terminus except lysine, aspartate and glutamate and provides more synthetic flexibility compared to NCL as it
relies on an N-terminal serine or threonine – amino acids that are significantly more abundant than cysteine. Current disadvantages include the requirement of relatively high substrate concentrations and the use of potentially reactive organic reaction solvents (mixtures of acetic acid and pyridine), which limit both the solubility of protein segments and the biocompatibility of the reaction. Nevertheless, the serine/threonine ligation was successfully applied in the synthesis of human erythrocyte acylphosphatase (98 residues), the semi-synthesis of peptoid-modified RNase A and several cyclic peptides.

1.4.3. \(\alpha\)-Ketoacid-hydroxylamine ligation

Initially reported by Bode et al. in 2006, the \(\alpha\)-ketoacid-hydroxylamine (KAHA) ligation involves the chemoselective reaction of fully unprotected peptide or protein segments decorated with a C-terminal \(\alpha\)-ketoacid and an N-terminal hydroxylamine to form an amide bond at the ligation site (Scheme 14). This reaction does not require any additional reagents or catalysts, and proceeds at slightly elevated temperatures in mixtures of water and organic solvent without the formation of problematic byproducts.

![Scheme 14. \(\alpha\)-Ketoacid-hydroxylamine ligation.](image)

Over the years, several hydroxylamines for KAHA ligation have been evaluated and differences were found not only regarding their stability and reactivity, but they also seem to operate under different reaction mechanisms. Based on the molecular structure and reaction mechanism, the ligation reactions are divided into two categories: type I ligation with \(O\)-unsubstituted peptide hydroxylamines and type II with \(O\)-substituted hydroxylamine derivatives.

1.4.3.1. Mechanism of the KAHA ligation

The mechanism of type I KAHA ligations has been extensively studied by the Bode group. Among other mechanistic experiments, \(^{18}\)O isotope labeling studies clearly
demonstrated that the oxygen atom of the newly formed amide originates from the hydroxylamine and neither from the \( \alpha \)-ketoacid nor from water in the reaction solvent. This can be explained by the reaction of the hydroxylamine and the \( \alpha \)-ketoacid forming a hemiaminal (Scheme 15). After dehydration, E- and Z-nitrones are formed, which are observable intermediates of the reaction. Subsequent attack of the carboxylate gives an \( \alpha \)-lactone, which can rearrange to an \( \alpha \)-oxaziridinyl carboxylate, followed by decarboxylation to afford the final amide product.

\[
\begin{align*}
\text{hemiaminal} & \quad \text{Z- and E-nitrones} \\
\text{amide} & \quad \text{oxaziridine} \\
\text{\( \alpha \)-lactone} & \quad + H^+/- H^+ \\
\end{align*}
\]

Scheme 15. Mechanism of KAHA type I ligations.

At the outset of this thesis, the mechanism of type II KAHA ligations was not investigated in full detail. Yet, several factors besides the obvious different molecular structure indicated that type II ligations would proceed via a mechanism distinct from type I ligations. For example, the presence of water in the reaction solvent significantly decreases the reaction rate of type I ligations, whereas type II ligations tolerate or even benefit from aqueous solvent mixtures. In contrast to type I ligations, isotope labeling studies showed almost full incorporation of \(^{18}\text{O}\) from labeled aqueous solvent in the KAHA ligation of \( O\text{-Bz} \) hydroxylamines.

In chapter 2, our investigations regarding the mechanism of KAHA ligations with 5-oxaproline, a type II ligation, are described and a possible reaction mechanism is discussed.

1.4.3.2. Properties and synthesis of peptide \( \alpha \)-ketoacids

Simple \( \alpha \)-ketoacids have been known for a long time: for example, pyruvic acid was prepared by Berzelius in 1835.\(^{71}\) They play important roles in different biological processes such as glycolysis (pyruvate), the Krebs-cycle (\( \alpha \)-ketoglutaric acid) and amino acid metabolism.

The most reactive position of \( \alpha \)-ketoacids is typically the highly electrophilic \( \alpha \)-carbonyl group. Simple \( \alpha \)-ketoacids are susceptible towards nucleophilic attack, often resulting in the formation of addition products (e.g. hydrates) or polymerization and cyclization.\(^{72}\) Under basic conditions, \( \alpha \)-ketoacids easily form enolates, resulting in epi-
merization of stereogenic centers in the β-position,\textsuperscript{73} which renders them, for example, incompatible with Fmoc-based peptide synthesis methods (Scheme 16).

Yet, peptide α-ketoacids are remarkably stable and tolerate fully unprotected amino acid side chains. Importantly, they do not epimerize in acidic aqueous solutions, rendering them stable towards standard peptide handling and purification methods, including reverse phase HPLC and lyophilization.

Scheme 16. Selected characteristics of α-ketoacids.

Small peptide α-ketoacids were first prepared by Wassermann \textit{et al.} by the oxidation of stable phosphorous ylides to acyl cyanides followed by hydrolysis (Scheme 17).\textsuperscript{74} However, this method has several disadvantages including long reactions times, which often leads to epimerization of the neighboring stereocenter, and the use of toxic ozone gas at low temperatures. Furthermore, the requirement for inert organic solvents often limits the solubility of the peptides.

Wasserman's phosphorus ylides

\[
\begin{align*}
\text{Cbz} & \quad \text{N} & \quad \text{O} & \quad \text{Ph} & \quad \text{N} & \quad \text{OH} \\
\text{Me} & \quad \text{H} & \quad \text{N} & \quad \text{O} & \quad \text{Cbz} & \quad \text{PPh}_3 \\
\text{Me} & \quad \text{H} & \quad \text{N} & \quad \text{O} & \quad \text{Cbz} & \quad \text{PPh}_3 \\
\text{CH}_2\text{Cl}_2, 36 \text{ h} & & 65 \% \text{ yield} \\
\end{align*}
\]

Bode's cyanosulfurylides

\[
\begin{align*}
\text{Cbz} & \quad \text{N} & \quad \text{O} & \quad \text{Ph} & \quad \text{N} & \quad \text{OH} \\
\text{Me} & \quad \text{H} & \quad \text{N} & \quad \text{O} & \quad \text{Cbz} & \quad \text{S} \\
\text{Me} & \quad \text{H} & \quad \text{N} & \quad \text{O} & \quad \text{Cbz} & \quad \text{S} \\
\text{CH}_2\text{Cl}_2, 5 \text{ min} & & 93 \% \text{ yield} \\
\end{align*}
\]

\text{CN} \quad \text{EDCI, HOBr} \\
\text{Oxone} \quad \text{H}_2\text{O/DMF, 10 min} \\
\text{stable under acidic resin cleavage} \\
\text{stable to HPLC purification} \\
\text{stable to lyophilization} \\
\text{storable at -20 °C for months} \\
\text{scalable and fast synthesis without epimerization} \\
\text{cheap and safe reagents} \\
\text{for protected amino acid monomers and small peptides}

Scheme 17. Synthesis of small peptide α-ketoacids using phosphorous ylides and cyanosulfurylides.\textsuperscript{75}

Inspired by Wassermann’s method, the Bode group developed an improved synthesis of peptide α-ketoacids using stable and storable cyanosulfurylides.\textsuperscript{76} These cyanosulfurylides can be readily obtained via coupling of carboxylic acids to sulfonium salts using standard peptide coupling reagents. In the second step, cyanosulfurylides are oxidized under mild conditions at ambient temperature using Oxone, a cheap commercially available mixed
Chapter 1: Introduction

Salt containing potassium peroxymonosulfate. The oxidation proceeds smoothly in mixtures of water and organic solvents, improving the solubility of the peptides significantly. It was demonstrated that most unprotected amino acid side chains are compatible with the oxidative conditions, with the notable exceptions of cysteine, methionine and (unprotected) tryptophan. Epimerization of the α-stereocenter does not occur throughout the reaction sequence and enantiopure α-ketoacids can be obtained. The method is easily scalable and has been applied to the decagram scale synthesis of α-ketoacids derived from Fmoc-protected amino acids.

A general method for the solid phase synthesis of peptide α-ketoacids was developed in the Bode group using an immobilized cyanosulfurylidyde (Scheme 18). This approach is compatible with standard, automated Fmoc SPPS and was used to synthesize numerous peptides with different C-terminal α-ketoacids, including alanine, arginine, glutamic acid, leucine, phenylalanine, tyrosine and valine. Despite its generality, this technique has two disadvantages: 1) it is not possible to access peptide α-ketoacids containing cysteine, methionine or unprotected tryptophan residues; and 2) the late stage oxidation of peptide cyanosulfurylides requires additional manipulation and purification steps after solid phase synthesis, which reduces the overall yield of precious peptide segments.


One goal of this thesis was to develop an alternative method for the synthesis of peptide α-ketoacids avoiding the limitations mentioned above and our efforts are described in chapter 3.

1.4.3.3. Hydroxylamines for protein synthesis by KAHA ligation

Several characteristics of hydroxylamines are important in the context of the KAHA ligation. First, hydroxylamines are much less basic compared to other amines found in proteins. For comparison, N,O-dimethylhydroxylamine has a pKₐ of 4.75, whereas the average pKₐ of N-terminal or side chain amino groups is 7.7 and 10.5 (lysine), respectively.
Under the acidic KAHA ligation conditions, a substantial amount of the hydroxylamine is in the more nucleophilic, non-protonated form, whereas all other amines are essentially quantitatively protonated. Second, hydroxylamines are good nucleophiles due to the $\alpha$-effect of the oxygen substituent and are probably the most nucleophilic group in the protein segment under acidic KAHA ligation reaction conditions, which enables a chemoselective reaction of the hydroxylamine with the $\alpha$-ketoacid.\textsuperscript{80}

Numerous hydroxylamine derivatives have been evaluated for the use in KAHA ligation (Figure 7).\textsuperscript{81}

![Figure 7. Hydroxylamines for chemical protein synthesis by KAHA ligation.](image)

Of all evaluated hydroxylamine derivatives, the cyclic 5-oxaproline\textsuperscript{82} and its aldehyde-forming derivative 4-ethoxy-5-oxaproline\textsuperscript{83} exhibit so far the best balance between high reactivity and sufficient stability. In combination with the ease of implementation in segment synthesis, 5-oxaproline is the most commonly employed hydroxylamine derivative in routine protein synthesis by KAHA ligation. It was later found – as described in full detail in chapter 2 – that the KAHA ligation with 5-oxaproline gives not the expected amide but rather an ester as major product. These ester products can be readily rearranged to the amides via an $O$ to $N$ acyl transfer under mildly basic conditions.

Recently, the Bode group disclosed the application of a serine-forming oxazetidine in the KAHA ligation, which exhibited much higher reactivity compared to 5-oxaproline.\textsuperscript{84} Yet, the routine application of this monomer in protein synthesis is hampered by the low stability of the unprotected monomer and its laborious synthesis. Nonetheless, this type of monomer has great potential and optimized protecting group strategies and protocols for handling and ligation might render this compound the preferred standard reaction partner in future KAHA ligations.

The KAHA ligation with 5-oxaproline gives an unnatural homoserine (Hse, $T^S$) residue at the ligation site (Scheme 19). If the ligation site is chosen properly – e.g. mutation of solvent exposed serine, threonine, methionine, or asparagine residues –, the impact of this rather conservative mutation on protein structure and function should be minimal. Yet, this has to be carefully investigated in every protein synthesis. A goal of this thesis was to evaluate, for the first time, the effect of homoserine mutations on the function of a biologically active protein as described in chapter 4.
From a chemical perspective, the properties of 5-oxaproline are similar to regular canonical amino acids. For example, 5-oxaproline can be readily protected at the N-terminus using common protecting groups such as Boc or Fmoc and it is fully compatible with standard conditions used in Fmoc-SPPS and resin cleavage. Conveniently, unprotected protein segments with protected or unprotected 5-oxaproline residues can be stored at -20 °C for months without decomposition.

In 2012, the Bode group reported a synthesis of 5-oxaproline using a chiral auxiliary, which was initially developed by Vasella in 1981 (Scheme 20). Both enantiopure Boc- and Fmoc-protected (S)-5-oxaproline can be obtained by this route. Recently, an organocatalytic route was developed in the Bode group, which is readily scalable, enabling the synthesis of 5-oxaproline on a kilogram-scale.

The first larger polypeptide synthesized by KAHA ligation was the bioactive form of human GLP (glucagon-like peptide) (7–36) (Scheme 21). Notably, a type I unsubstituted hydroxylamine was used in this synthesis, which required the use of purely organic solvent mixtures in the ligation step, limiting the solubility of the protein segments.
The requirement for organic solvents in type I KAHA ligations was overcome by the development of a well-performing type II hydroxylamine derivative, 5-oxaproline. This monomer was first applied in the synthesis of Pup (prokaryotic ubiquitin-like protein, 63 residues) and cspA (probable cold shock protein A, 67 residues, Scheme 22). Both proteins were synthesized using a two segment-one ligation strategy.

Scheme 22. Synthesis of Pup (2–63) by KAHA ligation.

For the synthesis of larger proteins, sequential ligations of more than two segments are required. To avoid side reactions such as polymerization or oligomerization, the bifunctional middle segments must be partially protected during the KAHA ligation. It was found that both Fmoc-protected N-terminal 5-oxaproline and C-terminal α-ketoacids masked as cyanosulfurylides are stable to the ligation conditions and can be employed in the synthesis of larger proteins using sequential ligations. A middle segment with Fmoc-protected N-terminal 5-oxaproline was utilized in the synthesis of variants of the important modifier protein UFM1 (ubiquitin fold modifier 1) (Scheme 23).

Scheme 23. Synthesis of UFM1 (2–83) by two sequential KAHA ligations.

Besides chemical protein synthesis, it should be noted that the KAHA ligation has found versatile applications in the synthesis of cyclic peptides and β-peptides.
1.5. Conclusions

Chemical protein synthesis has advanced tremendously over the last few decades and can provide specifically modified proteins in sufficient purity and quantity for biochemical studies. In addition to the already well-established native chemical ligation, the KAHA ligation developed in the Bode group has made significant progress over the last few years and demonstrated its capabilities in the synthesis of several smaller proteins. Yet, at the outset of the work described in this thesis, several goals remained to be addressed.

One aim of this thesis was to elaborate on the scope of the KAHA ligation and synthesize two larger and highly interesting bioactive proteins, the important modifier protein SUMO3 (chapter 4) and a protein believed to be involved in the regulation of fat metabolism, irisin (chapter 5). Supporting these synthetic endeavors, the two major limitations of the cyanosulfurylides strategy for the synthesis of C-terminal protein \(\alpha\)-ketoacids were addressed (chapter 3). In chapter 2, the surprising finding that the KAHA ligation with 5-oxaproline gives the corresponding ester rather than the expected amide and general strategies to convert the ester into the amide are described in full detail.
1.6. References

1. From Greek πρωτείος (proteios) meaning “primary” or “in the lead”. The term was introduced by the Dutch chemist Gerardus Johannes Mulder in 1839 to name the fundamental “molecule” (protein) that composes different animal substances such as egg white and tissues. Mulder, G. J.: Ueber die Zusammensetzung einiger tierischen Substanzen. *J. Prakt. Chem.* 1839, 16 (1), 129–152.


27 From latin: *ligare* – "to bind" or "to glue together".


Ogunkoya, A.: Total chemical protein synthesis with the $\alpha$-ketoacid hydroxylamine (KAHA) ligation. Diss. ETH No. 21846, ETH Zurich, **2014**


CHAPTER 2
FORMATION OF DEPSIPEPTIDES IN THE KAHA LIGATION WITH 5-OXAPROLINE

The work described in this chapter was done in collaboration with Florian Rohrbacher, ETH Zurich, who performed isotope labeling studies, trapping experiments and investigations related to the iminoether intermediates.
CHAPTER 2: Formation of depsipeptides in the KAHA ligation with 5-oxaproline

2.1. Introductory remarks

In this chapter the discovery that the major products of the KAHA ligation with 5-oxaproline are esters rather than the expected amides (Scheme 24) is described. Based on these findings, we developed mild conditions to convert these depsipeptides and depsiproteins into the corresponding amides. Mechanistic studies provided insights into the possible reaction mechanism as described in chapter 2.3.4.

Scheme 24. Formation of depsipeptides in the KAHA ligation with 5-oxaproline and O to N acyl transfer to form the amide at the ligation site.

2.2. Formation of depsipeptides in the KAHA ligation with 5-oxaproline

The initial observation that two products with the desired product mass are formed in the KAHA ligation with 5-oxaproline came from a project of Florian Rohrbacher towards the synthesis of cyclic peptides using the KAHA ligation with 5-oxaproline, which was later published separately.90

Alerted by this finding, we carefully reviewed the larger protein segment ligations we were utilizing in our protein syntheses. A representative example is a ligation we used in an earlier strategy towards the synthesis of irisin (see chapter 5.3). In the KAHA ligation of Fmoc-Opr-[Irisin (46–64)]-α-ketoacid 1 and H-Opr-[Irisin (66–112)]-OH 2, the formation of two peaks (3-A and 3-B) with the desired product mass in a ratio of about 4:1 was observed (Figure 8). Both peaks were isolated by preparative HPLC and exposed to basic conditions typically used to induce O to N acyl shifts (pH 10 buffer containing 6 M GdmCl).93 The minor product 3-B did not change under these conditions, but the major product 3-A cleanly converted into the minor product 3-B. Notably, cleavage of the polypeptide chain was not observed.
A plausible explanation for these observations would be that the major product 3-A of the KAHA ligation with 5-oxaproline is an ester (depsi-3) rather than an amide (amide-3). To test our hypothesis, we prepared a model ligation of two dipeptides 4 and 5 that would afford ligation products that could be readily characterized by NMR spectroscopy (Figure 9). Ligation of 4 and 5 under standard conditions led to formation of two peaks with the desired product mass in a ratio of about 9:1. Subjecting the crude ligation mixture directly to basic rearrangement conditions evoked clean conversion of the major product (6-A) into the minor product (6-B) within 45 min, a similar observation as before for the products 3-A and 3-B of the protein segment ligation of 1 and 2.
CHAPTER 2: FORMATION OF DEPSIPEPTIDES IN THE KAHA LIGATION WITH 5-OXAPROLINE

Both products were fully characterized by NMR spectroscopy (\(^1\)H, \(^{13}\)C, \(^1\)H-\(^{13}\)C and \(^1\)H-\(^{15}\)N HSQC and \(^1\)H-\(^{13}\)C HMBC), undoubtedly confirming that the major product 6-A of the ligation was the ester, which rearranged under basic conditions to the amide 6-B. Comparison with NMR spectra of independently synthesized diastereomers of 6-B showed that the ester product did not epimerize at the ligation site during ligation and rearrangement.

We corroborated our previous hypothesis and identified the major products of the KAHA ligation with 5-oxaproline as esters, which could be rearranged to the desired amides under mild basic conditions. Before we could adapt our synthesis protocols and continue with our protein synthesis – some involving multiple sequential KAHA ligations – we had to investigate whether the homoserine amide is stable under ligation conditions or if the depsipeptide is formed during the ligation via an \(N\) to \(O\) acyl transfer from the expected amide product (Scheme 25). Such \(N\) to \(O\) acyl shifts are occasional side reactions for serine\(^{94, 95}\) and threonine\(^{96}\) residues under acidic conditions, especially at elevated temperatures, which would – to some extent – resemble the conditions used in KAHA ligation. However, incubation of purified homoserine-containing peptide amide-6 or protein segment amide-3 under ligation conditions (8:2 DMSO:H\(_2\)O or 8:2 NMP:H\(_2\)O, 0.1 M oxalic acid, 60 °C) did not lead to the formation of depsi-6 or depsi-3, respectively (see Experimental for details).

Further control experiments showed that also the depsipeptides are stable under the ligation conditions. These results demonstrate that the formation of ester- and amide-product
during the ligation is an irreversible process, which is consistent with the observation that the ratio of ester- and amide-product remains constant during the course of the ligation (Scheme 26).

Scheme 26. Irreversible formation of ester and amide products in KAHA ligation with 5-oxaproline.

All investigated protein segment ligations with 5-oxaproline gave the esters at the ligation sites as initial major products, which were completely stable to ligation, HPLC purification and storage as lyophilized powder. The depsipeptides could be readily rearranged to the natural amides using basic buffers. In many syntheses, a one-pot ligation-rearrangement procedure could be applied, saving time and purification steps. In some large peptides or proteins such as \( \textit{3} \), the rearrangements were found to be somewhat slower, probably due to the presence of secondary structures. In these cases, the addition of 6 M GdmCl to the basic rearrangement buffer to establish denaturing conditions was found to significantly accelerate the rearrangement. Free cysteine residues readily formed intra- and intermolecular disulfides under the basic rearrangement conditions, which could be effectively suppressed by addition of TCEP (typically at 10 mM concentration), a commonly used reducing agent for disulfides in peptides.

With a reliable protocol for the rearrangement of depsipeptides and depsiproteins in hand, we could continue our protein syntheses. Separately, we conducted investigations to shine light on the reaction mechanism of KAHA ligations with 5-oxaproline, which will be discussed in the following chapter.

2.3. Mechanistic investigations

The formation of depsipeptides by KAHA ligations is solely observed for ligations with 5-oxaproline (and some analogs, vide infra); KAHA ligations with acyclic hydroxylamines such as \( \textit{O}\)-unsubstituted hydroxylamines (type I) and \( \textit{O}\)-Bz-hydroxylamines (type II) or the smaller homolog oxazetidine give exclusively the expected amide product (Scheme 27).
CHAPTER 2: FORMATION OF DEPSIPEPTIDES IN THE KAHA LIGATION WITH 5-OXAPROLINE

First, the investigation of several factors of which we speculated that they could induce or at least modulate the formation of esters in the KAHA ligation with 5-oxaproline is described. These factors included the type of the \(\alpha\)-ketoacid, neighboring amino acid residues close to the ligation site or substituents of the isoxazolidine ring. Later in this chapter, based on the observations made, a possible reaction mechanism discussed.

2.3.1. KAHA ligations with 5-oxaproline and different \(\alpha\)-ketoacids

Analysis of our protein segment ligations showed, that the formation of depsipeptides by KAHA ligations with 5-oxaproline is not restricted to the example shown in Figure 8 and does not appear to be affected by the type of \(\alpha\)-ketoacid or neighboring residues of the ligation site. The immediate products of ligations to form the proteins Pup (Leu \(\alpha\)-ketoacid), CspA (Tyr \(\alpha\)-ketoacid), UFM1 (Phe and Ala \(\alpha\)-ketoacid) and IFITM3 (Arg \(\alpha\)-ketoacid) are esters.\(^98\)

We also investigated the KAHA ligation of 5-oxaproline 5 with several small molecule \(\alpha\)-ketoacids and found that all gave approximately 9:1 mixtures of ester and amide products in good yield (Table 1). This further underlines that the substitution of the \(\alpha\)-ketoacid or neighboring amino acid residues seem not to have a significant impact on the formation of depsipeptides.
Table 1. KAHA ligation of 5 with different α-ketoacids.

<table>
<thead>
<tr>
<th>Entry</th>
<th>R1</th>
<th>Product</th>
<th>Ratio ester/amide during ligation (by UV)</th>
<th>Isolated yield (ester)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fmoc</td>
<td>depsi-8</td>
<td>93:7</td>
<td>61%</td>
</tr>
<tr>
<td>2</td>
<td>HOCOC</td>
<td>depsi-9</td>
<td>93:7</td>
<td>59%</td>
</tr>
<tr>
<td>3</td>
<td>Me</td>
<td>depsi-10</td>
<td>86:14</td>
<td>59%</td>
</tr>
<tr>
<td>4</td>
<td>Ph</td>
<td>depsi-11</td>
<td>87:13</td>
<td>55%</td>
</tr>
</tbody>
</table>

2.3.2. KAHA ligations with 5-oxaproline analogs

Although protein segment ligations indicated that the ester formation was independent from the nature of the residue preceding the 5-oxaproline, we aimed to investigate whether the amide bond connecting the 5-oxaproline to the preceding amino acid could play a role in the ester formation. We tested several analogues of 5-oxaproline under standard ligation conditions (Table 2), for example the regioisomer 12 (entry 1) and analogs where the carboxylic acid was replaced with a hydroxy group (14, entry 2) or completely removed (16, entry 3). In all cases examined, the major product of the ligation was an ester, independent of the substitution on the isoxazolidine ring.
**Table 2. KAHA ligation of oxaproline analogues**

<table>
<thead>
<tr>
<th>Entry</th>
<th>Hydroxylamine</th>
<th>R&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Major Product (as TFA salt)</th>
<th>Ratio Ester/Amide</th>
<th>Isolated Yield (Ester)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>depsi-13</td>
<td></td>
<td></td>
<td>40 %</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>depsi-15</td>
<td></td>
<td>92:8&lt;sup&gt;b&lt;/sup&gt;</td>
<td>43 %</td>
</tr>
<tr>
<td>3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>16</td>
<td>depsi-17</td>
<td></td>
<td>79:21&lt;sup&gt;d&lt;/sup&gt;</td>
<td>27 %</td>
</tr>
<tr>
<td>4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>18</td>
<td>depsi-19</td>
<td></td>
<td>58:42&lt;sup&gt;d&lt;/sup&gt;</td>
<td>14 %</td>
</tr>
</tbody>
</table>

<sup>a</sup> No amide formation observed during ligation by HPLC analysis;  
<sup>b</sup> By UV during ligation;  
<sup>c</sup> Ligation performed at 0.2 M concentration;  
<sup>d</sup> Based on isolated yields.

Notably, the KAHA ligation of 5-oxaprolinol 14 (entry 2) gives exclusively the ester *depsi*-15 with the hydroxy group in the γ position, indicating that the second hydroxy group in the molecule is not involved in the reaction. Removing all substituents from the isoxazolidine (16, entry 3) and increasing the ring size (18, entry 4) seems to slightly favor the formation of the amide compared to 5-oxaproline.

Overall, the comparable ester/amide ratios found in KAHA ligations of 5-oxaproline (ca. 9:1) and its analogs (4:1 to 9:1, omitting 18) suggests that the amide bond linking 5-oxaproline to the rest of the peptide and the neighboring amino acid residues are likely not involved in the reaction leading to the ester.

### 2.3.3. Influence of reaction solvent mixtures and epimerization analysis

Another important factor that we had to investigate was the role of the reaction solvent on the ester/amide ratio and on epimerization. We developed a model ligation (Scheme 28) that readily allowed the determination of the ester/amide ratio and the epimerization at the α-ketoacid-derived residue of the ligation products using a HPLC assay (Table 3).
Scheme 28. Model ligation of 20 and 5 yielding ester/amide products and epimers at homoserine separable by HPLC (top) and representative HPLC trace of a ligation in 7:3 DMSO:H2O + 0.1 M oxalic acid (bottom).

The identity of amide-21 and epi-amide-21 was confirmed by comparison with independently synthesized material. Depsi-21 and epi-depsi-21 were identified by O to N acyl shift to the known amide products. Importantly, no epimerization was observed under rearrangement conditions.

Table 3. KAHA ligation of 20 and 5 in different solvent mixtures.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Solvent mixture (^a)</th>
<th>% ester formed</th>
<th>% amide formed</th>
<th>% epimerization (^b) ester</th>
<th>% epimerization (^b) amide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7:3 DMSO:H2O</td>
<td>89</td>
<td>11</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>7:3 NMP:H2O</td>
<td>87</td>
<td>13</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>99:1 DMSO:H2O</td>
<td>97</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>99:1 NMP:H2O</td>
<td>90</td>
<td>10</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>99:1 NMPA:H2O</td>
<td>96</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>99:1 CH3CN:H2O</td>
<td>95</td>
<td>5</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>H2O</td>
<td>73</td>
<td>27</td>
<td>12</td>
<td>29</td>
</tr>
</tbody>
</table>

\(^a\) containing 0.1 M oxalic acid \(^b\) = epi-ester/ester \times 100.

The standard ligation conditions (entries 1 and 2) gave, as previously observed for protein segment ligations, a ratio of ester/amide of around 9:1. The amount of amide-21
formed in the ligation seems to correlate with the water content of the ligation solvent, although even in pure water only around 25 % amide and 75 % ester are formed.

The epimerization level of the major ester product *depsi-21* is low (≤ 5 %) under standard ligation conditions, which explains why this was not observed previously in the NMR assay (see chapter 2.2). Notably, the epimerization of the amide product formed in the ligation is consistently higher compared to the ester product, suggesting that the epimerization occurs not from a common intermediate but rather via two distinct pathways. Similar differences in epimerization were observed in studies with cyclic peptides and confirmed by chiral GC-MS analysis. This study also fortified the previous result, that the O to N acyl shift of the ester product under basic conditions does not increase the epimerization at the ligation site.

In summary, this model study showed that the extent of epimerization occurring during KAHA ligation with 5-oxaproline under standard conditions is small – especially regarding the major ester product – and does not affect the utility of the KAHA ligation for protein synthesis.

### 2.3.4. Proposed reaction mechanism for KAHA ligations

As discussed in chapter 1.4.3.3, there are two distinct mechanisms (type I and type II) for the KAHA ligation, depending on the hydroxylamine used in the reaction. Detailed studies on the mechanism of the type I ligation of *O*-unsubstituted hydroxylamines – which have water as leaving group – were previously reported. However, we have not fully elucidated the mechanism of the type II ligation of *O*-substituted variants, which includes the 5-oxaproline substrate. The formation of esters and the distinct extent of epimerization of the ester and the amide product in the KAHA ligation with 5-oxaproline provide some insights into the possible mechanism of the type II ligations.

Regarding the reaction equation, there is formally no water added or expelled in KAHA ligations of 5-oxaproline. However, an important observation is that KAHA ligations with 5-oxaproline performed in DMSO/H$_2$O give nearly complete $^{18}$O incorporation into both the ester and amide products, which suggests an elimination-addition mechanism. A representative example is the ligation of 4 and 5 shown in Scheme 29. Analysis of the crude reaction mixture showed no $^{18}$O incorporation in both unreacted starting materials, verifying that the label incorporation comes from the ligation. In contrast, no $^{18}$O label is incorporated in type I ligations (see chapter 1.4.3.1).
CHAPTER 2: FORMATION OF DEPSIPEPTIDES IN THE KAHA LIGATION WITH 5-OXAPROLINE

Scheme 29. Isotope labeling study in the KAHA ligation of 4 and 5.99

Given their intrinsic chemical reactivity, it is a reasonable assumption that the α-ketoacid \( a \) and the 5-oxaproline \( b \) form iminium \( d \) (Scheme 30). Ligation studies in \(^{18}\text{O}\)-labeled water (Scheme 29) suggest that the formation of imine \( d \) is not reversible under ligation conditions, as analysis of unreacted \( a \) and \( b \) showed no \(^{18}\text{O}\) incorporation after ligation. We hypothesize that iminium \( d \) undergoes rapidly a – probably concerted – decarboxylation and N-O bond cleavage sequence yielding nitrilium \( e \).100

Scheme 30. Proposed reaction mechanism for the KAHA ligation with 5-oxaproline.

A similar elimination/decarboxylation with simultaneous cleavage of an oxime N-O bond to yield an alcohol and a nitrile has been extensively studied by Kemp \textit{et al.} (Scheme 31).101
CHAPTER 2: FORMATION OF DEPSIPEPTIDES IN THE KAHA LIGATION WITH 5-OXAPROLINE

Scheme 31. Simultaneous decarboxylation and N-O-bond cleavage in a Kemp elimination.

Although direct epimerization of nitrilium $e$ to $epi$-$e$ via tautomerization (ketenimine $e'$) cannot be excluded, in this case it would be expected that ester and amide show a similar extent of epimerization. However, as the amide product shows in most cases a significantly higher extent of epimerization it seems likely that ester and amide epimerize via two independent mechanisms and not via $e'$.

The postulated highly reactive nitrilium $e$ could undergo different pathways. Nucleophilic attack of the $\gamma$-hydroxy group of the homoserine residue on nitrilium $e$ would lead to cyclic iminiumether $f$ (Path A), and subsequent acidic hydrolysis (Path A1) via $g$ would give the ester product. Examples are known for both the 6-endo attack of alcohols at an in situ generated nitrilium and the acidic hydrolysis of cyclic iminoethers to give the ester product (Scheme 32).

Scheme 32. Selected literature examples for a) 6-endo attack of an alcohol at a nitrilium$^{102}$ and b) acidic hydrolysis of a cyclic iminoether to give selectively the O-acyl product.$^{103}$

If the reaction proceeds via the hypothetical nitrilium $e$, the intramolecular attack of the $\gamma$-hydroxy group of the homoserine residue on nitrilium $e$ must be very fast, as all attempts to trap the nitrilium with an external nucleophile such as MeOH failed. Similarly, the fact that the ligation of 5-oxaprolinol 14 with Fmoc-leucine $\alpha$-ketoacid 7 yields exclusively the depsiprotein $depsi$-15 arising from attack of the $\gamma$-hydroxy group and not the one from the hydroxy group in $\beta'$-position (22, Scheme 33, Table 2) underlines that this 6-endo-dig attack is fast, at least much faster than the competing 5-endo-dig attack, which would be also favored by the Baldwin rules.$^{104}$
Scheme 33. Formation of a single depsiproduct in the KAHA ligation with 5-oxaprolinol 14 and 7.

In a KAHA ligation under strictly anhydrous conditions in MeOH, a single signal corresponding to the iminoether was observed in the MALDI-MS of the crude reaction mixture (Scheme 34). While being stable under anhydrous conditions, the iminoether was found to rapidly hydrolyze upon addition of water to give the ester product. Using an \(\alpha\)-ketoacid with a more stabilizing phenyl-substituent allowed for the isolation of the iminoether 23 and for the confirmation of its structure by NMR. Based on these findings, it is a plausible assumption that the reaction also proceeds via iminiumether \(f\) under aqueous ligation conditions.

Scheme 34. In situ observation by MS (a) and isolation (b) of iminoethers formed in the KAHA ligation under anhydrous conditions.

Iminiumethers such as \(f\) are prone to tautomerize to the protonated enammoniumether form \(h\), whereby chiral information is lost (Path A2). Acidic hydrolysis of \(h\) would give an epimeric mixture of the ester product.

The amide product could be formed via a direct addition of water to nitrilium \(e\) (Path C), which would explain the observed dependence of the ester/amide ratio on the water content of the reaction mixture (Table 3).

The epimerized amide product could be formed via another pathway (Path B), which is mechanistically related to the known epimerization of activated amino acids. Nucleophilic attack of the next amide carbonyl oxygen onto the highly electrophilic nitrilium carbon would lead to oxazoliminium \(k\), which could tautomerize to oxazole \(m\), whereby the
chiral information is lost. Acidic hydrolysis of m would give an epimeric mixture of the amide product.

Although the existence of nitrilium e could not be directly proven, the proposed reaction mechanism provides a reasonable explanation for the formation of ester and amide products in the KAHA ligation and the observed different extent of epimerization in both products.

2.4. Conclusions

We showed that the major products of the KAHA ligation of α-ketoacids and 5-oxaproline residues were esters, rather than the expected amides. Efficient protocols for the rearrangement of depsipeptides to natural amides under mild basic conditions were developed, making possible the streamlined synthesis of proteins by KAHA ligation with 5-oxaproline. Based on mechanistic investigations we proposed a reaction mechanism that could explain the formation of both depsipeptide- and amide-products and their distinct epimerization.

Although surprising and unexpected, the formation of depsipeptides by ester-forming chemoselective ligations has potential applications. Changing an amide bond to an ester is known to disrupt secondary structures in peptides and proteins, and has been exploited in different so-called “switch peptides” and “switch proteins”, which have applications in self-assembly and drug delivery. The KAHA ligation with 5-oxaproline could provide a new synthetic method for depsiproteins inaccessible by recombinant expression or other ligation methods.
2.5. References


99 Isotope labeling studies performed by Florian Rohrbacher (unpublished results and ref 98).

100 A similar nitrilium has been proposed for type I ligations (in contrast to our later mechanistic investigations, see ref. 70): Sanki, A. K.; Talan, R. S.; Sucheck, S. J.: Synthesis of small glycopeptides by decarboxylative condensation and insight into the reaction mechanism. J. Org. Chem. 2009, 74 (5), 1886–1896.


CHAPTER 2: FORMATION OF DEPSIPEPTIDES IN THE KAHA LIGATION WITH
5-OXAPROLINE

105 Pfau, M.; Felk, A.; Revial, G.: Iminoether - enaminoether tautomerism. Elaboration of quaternary

106 Antonovics, I.; Young, G. T.: Amino-acids and peptides. Part XXV. The mechanism of the
595–601.

dependent contributions to protein folding kinetics and thermodynamics: lessons from amide-

108 Hentschel, J.; Krause, E.; Börner, H. G.: Switch-peptides to trigger the peptide guided assembly
of poly(ethylene oxide)–peptide conjugates into tape structures *J. Am. Chem. Soc.* **2006**, *128*
(24), 7722–7723.

109 Saucède, L.; Santos, S. D.; Chandravarkar, A.; Mandal, B.; Mimna, R.; Murat, K.; Camus, M.-S.;
Bérard, J.; Grouzmann, E.; Adrian, M.; Dubochet, J.; Lopez, J.; Lashuel, H.; Tuchscherer, G.;
CHAPTER 3

NEW PROTECTING GROUPS FOR THE SYNTHESIS OF PEPTIDE $\alpha$-KETOACIDS

Part of the work described in this chapter was done in collaboration with a former postdoc in the Bode Group, Dr. Javier Ruiz-Rodriguez, who performed preliminary studies, and Felix Limberg, who screened several protecting groups and helped with reaction optimization as part of his master thesis, which was co-supervised by Thomas Wucherpfennig.
CHAPTER 3: New protecting groups for the synthesis of peptide α-ketoacids

3.1. Introductory remarks

Peptides with C-terminal α-ketoacids are essential components for the chemical protein synthesis using the KAHA ligation. However, unprotected α-ketoacids are not compatible with the basic conditions utilized in Fmoc-SPPS, as they can enolize and epimerize in the presence of base.\(^{73}\) As described in more detail in chapter 1.4.3.2, the previously developed route to peptide α-ketoacid circumvents this problem by using a stable cyanosulfurylide precursor, which is oxidized to the α-ketoacid after SPPS (Scheme 35 A). Although this oxidation is reliable and has been used in the synthesis of numerous proteins,\(^{110}\) it has two major limitations. First, the reaction conditions are not compatible with peptide segments containing residues that are sensitive to oxidation such as cysteine, methionine and (unprotected) tryptophan. Although these three amino acids are rather uncommon in mammalian proteins – they account for about 5 \% of all amino acid residues – this limits the protein space accessible by KAHA ligation considerably or makes the introduction of additional mutations necessary, often causing unforeseen effects on protein folding or function.\(^{49}\) Second, the late stage oxidation protocol introduces an additional manipulation and purification step after solid phase peptide synthesis, resulting in lowered overall yields of precious protein segments.

![Scheme 35. Synthetic methods for C-terminal peptide α-ketoacids.](image-url)
CHAPTER 3: NEW PROTECTING GROUPS FOR THE SYNTHESIS OF PEPTIDE α-KETOACIDS

Given the straightforward and scalable synthesis of amino acid derived α-ketoacids, it would be conceptually very compelling to start from a protected α-ketoacid monomer that has already the correct oxidation state, implement it in the SPPS and obtain directly the desired peptide α-ketoacid after resin cleavage without further manipulation (Scheme 35 B). However, at the outset of this project, no suitable protecting groups for α-ketoacids were known, facing us with the challenge to develop these.

Chapter 3.2 focuses on how the limitations of the cyanosulfurylride-based strategy were addressed by the development of a protecting group for α-ketoacids, that is suitable for SPPS and gives the C-terminal peptide α-ketoacid directly upon resin cleavage conditions. In chapter 3.3, this concept is advanced to an orthogonal α-ketoacid protecting group, which is stable to resin cleavage and can be selectively deprotected on demand under non-oxidizing conditions in solution.

3.2. An acid-labile protecting group for α-ketoacids

3.2.1. Design of the protecting group

The goal of this part of the project was to develop a traceless, acid-labile protecting group for α-ketoacids, implement it in an amino acid monomer suitable for SPPS of C-terminal peptide α-ketoacids and test it in the synthesis of a protein segment required for one of our protein synthesis. The following criteria were set for such a protecting group:

• it must be stable to the reagents and basic conditions used during Fmoc-SPPS;
• it must be removed under acidic resin-cleavage conditions to directly afford the C-terminal peptide α-ketoacid with unprotected side chains;
• it must undergo protection and deprotection through a mechanism that does not lead to epimerization of the α-ketoacid;
• it must be prepared on sufficient scale;
• the protected α-ketoacid monomer must be coupled to the solid support by a suitable linker.

For initial development of protected α-ketoacids, we focused on the protection of Fmoc-leucine α-ketoacid 7, as leucine is by far the most prevalent amino acid in proteins and frequently used as ligation site in KAHA ligations. Additionally, leucine has no side chain functional group that could interfere with the α-ketoacid protection.

At the outset of our studies we were particularly concerned with the issue of epimerization as we found that the simplest protection – masking the keto group as an acyclic dimethyl acetal – was feasible, but resulted in some cases in partial epimerization upon cleavage (Scheme 36 A).
We postulated that a protecting group which could be removed by a mechanism involving the formation of a hemiacetal, rather than an oxonium ion, could overcome this problem. 111 We envisioned that such a mechanism could be achieved by using a \( p \)-methoxybenzyl (PMB) acetal, in which the benzylic C-O bond is easily broken under acidic conditions to give the hemiacetal and a stabilized \( p \)-methoxybenzyl carbocation (Scheme 36 B).112

### 3.2.2. Screening of protecting groups and reaction optimization

Despite many attempts, we could not prepare the simplest implementation of this design with acyclic mono- or di-\( p \)-methoxybenzyl acetals. This prompted us to focus on a series of 1,2- and 1,3-diols that would give cyclic acetals (Table 4).

**Table 4. Evaluation of protecting groups for \( \alpha \)-ketoacids.**

<table>
<thead>
<tr>
<th>Protecting group</th>
<th>24[a]</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>product</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>protection yield</td>
<td>20 %</td>
<td>&lt; 5 %</td>
<td>&lt; 5 %</td>
<td>&lt; 5 %</td>
<td>29 %</td>
<td>28 % [b]</td>
</tr>
<tr>
<td>deprotection</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>epimerization</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

[a]synthesized from 7, trimethylorthofomate and sulfuric acid. [b]Initial, unoptimized yield. [c] Yield after optimization. PMP = \( p \)-methoxyphenyl.

Inspired by a protocol from Noyori et al., 114 we found in early experiments that 1,2-diol derivative 25 could protect Fmoc-leucine \( \alpha \)-ketoacid 7 to give 31, but only in poor yield. Hydroxythiol 26 also gave the protected \( \alpha \)-ketoacid 32 in similar low yield, but could not –
unlike the oxo derivative 31 – be deprotected under the desired conditions, which is presumably due to the higher stability of the O,S-acetal compared to the O,O-acetal.\textsuperscript{115} The 1,3-diol derivative 27 did not offer any advantage.\n
Modulating the electronic properties of the aromatic substituent by introducing a more electron withdrawing 4-bromo substituent (28) led to an increased yield of 34, but the deprotection was not feasible under the desired conditions. Attempts to derivatize 34 – for example to convert the 4-bromo substituent into an electron donating substituent – by Cu- or Pd-catalyzed reactions were unsuccessful, mostly due to the base lability of the N-terminal Fmoc-protecting group.\textsuperscript{116,117}\n
We were pleased to find that the \textit{gem}-dimethyl-1,3-diol derivative 29 provided the protected \( \alpha \)-ketoacid 35 in higher yield compared to the analogous 27 and allowed facile deprotection under resin cleavage conditions to return the \( \alpha \)-ketoacid. The increased yield of the protection reaction – especially compared to the unsubstituted derivative 34 – could be attributed to several factors. First, the Thorpe-Ingold effect of the \textit{gem}-dimethyl substitution should kinetically favor the cyclization to form the acetal.\textsuperscript{118} Second, the \textit{gem}-dimethyl substitution might stabilize a proposed reactive intermediate and slow down the decomposition of the protecting group, which is discussed in more detail in chapter 3.4.\n
To ensure that the protection/deprotection cycle did not epimerize the stereocenter of leucine \( \alpha \)-ketoacid, the retrieved 7 was analyzed by chiral SFC (Figure 10). Because the enantiomers of free Fmoc-leucine \( \alpha \)-ketoacid 7 were not separable by SFC, 7 was derivatized to amide 37 via KAHA ligation. Comparison with independently synthesized enantiomers of 37 demonstrated that the \( \alpha \)-ketoacid did not epimerize during protection and deprotection, corroborating our initial hypothesis that the deprotection of a more labile PMB-like acetal proceeds via a hemiacetal rather than via an oxonium ion.\n
\begin{figure}[h]
\centering
\includegraphics[width=\columnwidth]{figure.png}
\caption{Epimerization assay for protected \( \alpha \)-ketoacid 35 (top). SFC analysis of 37 obtained from epimerization assay and independently synthesized epimers (bottom).}
\end{figure}

Encouraged by these promising results, we selected 29 for further optimization of the protection step (Table 5).
Table 5. Optimization of the α-ketoacid protection reaction.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Equivalents of 29</th>
<th>Catalyst [loading]</th>
<th>Solvent</th>
<th>Temperature</th>
<th>yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>TMSOTf [20 mol %]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>28 %</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>TMSOTf [20 mol %]</td>
<td>CH₂Cl₂</td>
<td>-78 °C to 0 °C</td>
<td>23 %</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>TMSOTf [20 mol%]</td>
<td>THF</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Tf₂O [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>26 %</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>Ms₂O [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>SnCl₂ [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>BF₃•THF [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>TBAF [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>Tf(OiPr)₄ [100 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>31 %</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>50 °C</td>
<td>30 %</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>-20 °C</td>
<td>9 %</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td></td>
<td>18 %</td>
</tr>
<tr>
<td>14</td>
<td>1.5 [addition over 9h]</td>
<td>TMSOTf [10 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>66 %</td>
</tr>
</tbody>
</table>

[a] 0.25 M substrate concentration.

Different catalysts and loadings were investigated, whereby TMSOTf and Tf₂O performed comparably well (Table 5, entries 1 and 10), as opposed to other Lewis acids examined (entries 6–9), which gave no conversion. Although the exact reaction mechanism was not investigated, this observation suggests that the reaction proceeds likely not via electrophilic activation of the ketone but rather via “hidden Brönsted” catalysis of in situ generated TfOH. Control experiments using TfOH as catalyst showed formation of 35, along with decomposition of the starting materials. Lowering the temperature or increasing the reaction concentration led to decreased yields of 35 (entries 11 and 12). Most importantly, slow addition of 29 to the reaction mixture increased the yield to 66 %, even when the number equivalents of 29 was reduced from 3 to 1.5 (entry 14). A possible reaction mechanism that could explain these observations is discussed in chapter 3.4.

Although the protected Fmoc-leucine α-ketoacid 35 is formed as a complex mixture of diastereomers because of the racemic protecting group 29, we decided to proceed with it as the protecting group is removed in a traceless manner during the resin cleavage step to give enantiopure C-terminal leucine α-ketoacid. The easy availability of both Fmoc-leucine α-ketoacid 7 and racemic 29, along with the ability to perform the protection reaction on a decagram scale rendered this procedure suitable to produce enough material for routine use in protein synthesis. The method has been applied to other α-ketoacids derived from
phenylalanine and valine; unfortunately the protection reaction conditions were not compatible with acid labile side chain protecting groups, limiting the scope of utilizable \(\alpha\)-ketoacids.\(^{120}\) It should be noted that this limitation has been recently overcome by colleagues in the Bode group, who developed an acid-free protection protocol, providing access to more than ten protected \(\alpha\)-ketoacids derived from proteinogenic amino acids.\(^{121}\)

### 3.2.3. Synthesis of a protected \(\alpha\)-ketoacid monomer for SPPS

With a suitable protecting group and a reliable route to protected Fmoc-leucine \(\alpha\)-ketoacid 35 in hand, we proceeded to evaluate suitable ways to immobilize the protected \(\alpha\)-ketoacid on resin for SPPS. As first choice, 2-chlorotrityl chloride resin was tested, as loading is generally easy and efficient.\(^{23}\) However, it was found that the loading continuously decreased during the course of the SPPS of a test sequence, suggesting leaching from the rather labile solid support.

We decided to immobilize the protected \(\alpha\)-ketoacid via a chemically more robust Wang-type linker.\(^{122}\) Although resins with Wang-linker are commercially available, typically a high excess of the carboxylic acid – activated as a symmetrical anhydride – has to be used to ensure efficient loading, which would result in a considerable loss of precious material.\(^{123}\) For a more efficient use, the protected Fmoc-leucine \(\alpha\)-ketoacid 35 was coupled to a Wang-type linker by alkylation with benzylic chloride 38, followed by deprotection of the allyl ester 39 to afford the free carboxylic acid 40.

![Scheme 37. Synthesis of protected Fmoc-leucine \(\alpha\)-ketoacid 40 and immobilization on solid support.](image)

Loading 40 onto standard amino-functionalized polystyrene (or polyethyleneglycol) based solid support yielded a stable, storable resin 41 that could be readily used in automated SPPS and delivered the unprotected C-terminal \(\alpha\)-ketoacid upon resin cleavage.
Notably, a critical aspect in using the protected $\alpha$-ketoacid resin 41 was to find a suitable resin cleavage cocktail to prevent side reactions of scavengers with the $\alpha$-ketoacid. The commonly used scavenger triisopropylsilane (TIPS) lead to efficient reduction of the keto group.\textsuperscript{124, 125} Fortunately, 1,2-ethanediol (EDT) or the non-malodourous alternative 2,2’-(ethylenedioxy)-diethanethiol (DODT)\textsuperscript{126} were effective scavengers to give the fully unprotected peptide $\alpha$-ketoacids that could be purified by HPLC and stored as lyophilized powder for several months.

### 3.2.4. Synthesis of a C-terminal protein segment $\alpha$-ketoacid using 41

To test our new resin 41, we synthesized a protein segment, which was required for the synthesis of the SUMO3 protein (see chapter 4.3), and contained oxidation-sensitive methionine and cysteine residues – this protein segment would not have been accessible with the previously used cyanosulfurylide method. Using resin 41, the SPPS of Fmoc-Opr-[SUMO3 (31–52)]-$\alpha$-ketoacid 42 proceeded without problems. Cleavage using a cocktail of 2.5 % DODT in TFA, followed by purification with preparative HPLC yielded pure 42 in 7 % yield for SPPS and purification, which was later successfully used in the synthesis of SUMO3 protein by sequential KAHA ligations.

![Diagram](https://example.com/diagram.png)

Figure 11. SPPS of Fmoc-Opr-[SUMO3 (31–52)]-$\alpha$-ketoacid 42 using the protected leucine $\alpha$-ketoacid resin 41 (top). Analytical HPLC of the crude material obtained from resin cleavage (bottom, left) and purified 42 (bottom, right).

This example demonstrated the successful application of the protected leucine $\alpha$-ketoacid resin 41 in the synthesis of a C-terminal protein segment $\alpha$-ketoacid.
3.3. An orthogonal protecting group for α-ketoacids

Following our intended design, the acid-labile protected α-ketoacids gave directly C-terminal peptide α-ketoacids upon resin cleavage. However, this might not always be desirable, as the unprotected C-terminal α-ketoacid automatically determines the direction of protein assembly to be from C to N-terminus. Yet, the possibility to assemble (parts of) the protein in N to C-terminal direction would provide additional flexibility in the synthetic strategy, for example in order to access large proteins via convergent synthesis.

To enable a KAHA ligation at the N-terminus of a segment, the C-terminal α-ketoacid of the segment has to be temporarily protected, for example by a protecting group orthogonal to resin cleavage and ligation conditions. After KAHA ligation at the N-terminus, the C-terminal α-ketoacid would be deprotected and employed in a second KAHA ligation (Scheme 38).

It should be noted that cyanosulfurylides are stable to KAHA ligation conditions and have been used as masked α-ketoacids in N to C-terminal assembly strategies, but they preclude the presence of methionine, cysteine and tryptophan in the protein segments. The development of an orthogonal protecting group for α-ketoacids, enabling the N to C-terminal assembly of protein segments containing all canonical amino acids using KAHA ligation will be described in this chapter.
3.3.1. Design of the orthogonally protected $\alpha$-ketoacid

The goal of this part of the project was to develop an orthogonal protecting group for $\alpha$-ketoacids, implement it in a monomer suited for SPPS and evaluate it in context of a protein synthesis. An orthogonal protecting group for $\alpha$-ketoacids should meet the following criteria:

- it must be stable to the reagents and basic conditions used during Fmoc-SPPS;
- it must be stable to the acidic resin cleavage conditions;
- it must be stable to HPLC purification conditions;
- it must be stable to KAHA ligation conditions (pH 1–2, 60 °C);
- it must be selectively removed under mild conditions in the presence of unprotected amino acid side chains and without epimerization of the $\alpha$-ketoacid.

The key challenge was to identify suitable deprotection conditions to reveal the $\alpha$-ketoacid in presence of the numerous functional groups of the unprotected side chains. A simple solution would be to use acidic conditions for the deprotection, since these are well tolerated by free peptides. However, as the resin cleavage conditions are also acidic, this would require a switchable protecting group that could be easily triggered from being acid-stable during the resin cleavage to acid-labile before the $\alpha$-ketoacid deprotection.

From our previous studies (see chapter 3.2.2), we knew that protected $\alpha$-ketoacids with electron-poor aromatic substituents in the protecting group were stable to acidic resin cleavage conditions, whereas protecting groups with electron-rich aromatic substituents were readily cleaved. We hypothesized that a $p$-phenolester as aromatic substituent might be electron withdrawing enough to render the protecting group stable to resin cleavage conditions ("off") and could be later easily converted to a electron donating phenol, which would render the protecting group acid labile ("on") (Scheme 39).
At the outset of our studies, we had to identify a suitable phenolester that had to be stable to the conditions of Fmoc-SPPS including the repeated treatment with nucleophilic piperidine. In model studies, it was found that simple \( p \)-tolyl acetate was readily cleaved by 20 % piperidine in DMF, whereas sterically more hindered \( p \)-tolyl pivalate was stable under these conditions for more than 12 h (Scheme 40). However, we were concerned that such a sterically hindered ester might not be cleaved under mild conditions and decided to facilitate the cleavage by incorporating an intramolecular nucleophile using a reported 4-aminobutyrate derivative.\(^{128}\)

Scheme 39. Proposed concept of a switchable protecting group for \( \alpha \)-ketoacids.

At the outset of our studies, we had to identify a suitable phenolester that had to be stable to the conditions of Fmoc-SPPS including the repeated treatment with nucleophilic piperidine. In model studies, it was found that simple \( p \)-tolyl acetate was readily cleaved by 20 % piperidine in DMF, whereas sterically more hindered \( p \)-tolyl pivalate was stable under these conditions for more than 12 h (Scheme 40). However, we were concerned that such a sterically hindered ester might not be cleaved under mild conditions and decided to facilitate the cleavage by incorporating an intramolecular nucleophile using a reported 4-aminobutyrate derivative.\(^{128}\)
3.3.2. Synthesis and studies of a orthogonally protected α-ketoacid monomer

With a design in hand, we started the synthesis of the orthogonally protected Fmoc-leucine α-ketoacid 49 following a similar route as for the simple protected α-ketoacid 40 (Scheme 41). Fmoc-leucine α-ketoacid 7 was protected with 1,3-diol derivative 43 in 51% yield. Removal of the allyl protecting group from the phenol 44 and selective alkylation of the carboxylic acid 45 with 38 to install the linker gave 46, which was esterified with 47 in modest yield using COMU and catalytic DMAP. After deprotection of the allyl ester 48, 49 was immobilized on Rink-amide polystyrene to give resin 50. As in case of the simple protected Fmoc-leucine α-ketoacid 40, formed diastereomers arising from the racemic protecting group could not be separated by column chromatography at any stage during the synthesis.

Scheme 41. Synthesis and immobilization of orthogonally protected α-ketoacid 49.

To investigate the properties of the orthogonal protecting group, a part of the resin was used to synthesize the dipeptide Fmoc-Gln-Leu-(phenolester-protected)-α-ketoacid 51, which resembled the first two amino acid residues of a protein segment we ultimately aimed to synthesize. Resin cleavage under standard conditions using 2.5% DODT in TFA for 2 h gave the expected (phenolester-protected) α-ketoacid 51, which corroborated our hypothesis that the p-phenolester is sufficiently electron withdrawing to prevent deprotection of the α-ketoacid under acidic conditions (Figure 12). HPLC analysis of the crude mixture showed that two separable peaks (termed 51-A and 51-B) with the desired product mass are formed. Although the absolute configuration could not be assigned, NMR analysis confirmed that...
these two peaks were two (out of four possible) diastereomers arising from the racemic protecting group. As minor products, two peaks corresponding to the diastereomers of the (phenol-protected)-\(\alpha\)-ketoacid 52 were observed. Control experiments demonstrated that 52 was not formed during the resin cleavage, but rather on the heated HPLC column, suggesting that the phenolester 51 is stable under strongly acidic resin conditions but not at elevated temperatures.

We were pleased to find that the phenolester in 51 was readily cleaved under mild basic conditions to yield 52: crude 51 was dissolved in DMSO and treated with pH 8 buffer, evoking clean hydrolysis of the phenolester in less than 15 min to yield the (phenol-protected) \(\alpha\)-ketoacids 52-A and 52-B (Figure 13). Control experiments confirmed that the obtained (phenol-protected) \(\alpha\)-ketoacids 52-A and 52-B were stable to HPLC and KAHA ligation conditions (see Experimental). To avoid the previously observed partial cleavage of (phenolester-protected) \(\alpha\)-ketoacids (such as 51) under heated HPLC conditions, it was decided to directly convert these after resin cleavage into the stable (phenol-protected) \(\alpha\)-ketoacids (such as 52) before purification.

Figure 12. Cleavage of Fmoc-Gln-Leu-(phenolester protected)-\(\alpha\)-ketoacid 51 from solid support and HPLC analysis.

Figure 13. Hydrolysis of phenolester 51 (left) and HPLC monitoring (right).
Deprotection of (phenol-protected) α-ketoacids was achieved under acidic resin cleavage conditions. In a first test experiment, purified 52-A and 52-B were treated with TFA without scavenger for 1 h and HPLC analysis showed partial deprotection to yield the free α-ketoacid 53 (Figure 14). The unusually slow deprotection is presumably due to the absence of scavenger; in a similar deprotection of a protein segment (phenol-protected) α-ketoacid using 2.5 % DODT in TFA, clean deprotection to the α-ketoacid was achieved within 75 min (see Experimental).

![Figure 14. Acidic deprotection of 52-A and 52-B (left) and HPLC monitoring (right).](image)

It was clear that the formation of diastereomers from the racemic protecting group would complicate the purification of protein segments with C-terminal (phenol-protected) α-ketoacids. Based on the NMR analysis of 51-A and 51-B, it seemed likely that using an enantiopure protecting group 43 would give only one diastereomer of 51. To test our hypothesis, we prepared 49 from enantioenriched 43 (60 % ee) and synthesized the dipeptide Fmoc-Gln-Leu-(phenolester-protected)-α-ketoacid 51. Unfortunately, the crude HPLC traces showed an identical ratio of diastereomers, regardless whether 51 was prepared from racemic or enantioenriched 49. A plausible explanation would be that the protecting group was epimerized during the synthesis – presumably during the protection reaction – which is discussed in more detail in chapter 3.4.

![Figure 15. Preparation of 51 from racemic or enantioenriched 43 (left) and HPLC analysis showing identical ratios of diastereomers (right).](image)
3.3.3. Synthesis of a C-terminal protein segment with orthogonally protected α-ketoacid resin 50

Despite the formation of diastereomers – which would be problematic only during the preparation of the protein segment, as the traceless protecting group is removed in situ after the first ligation – we decided to use the orthogonally protected Fmoc-leucine α-ketoacid resin 50 in the synthesis of a SUMO3 fragment (see chapter 4.4.2). The bifunctional middle segment H-Opr-[SUMO3 (31–75)]-(phenol-protected)-α-ketoacid 54 was synthesized uneventfully by automated SPPS; Asp62-Gly63 was coupled as DMB-protected dipeptide to avoid aspartimide formation. Resin cleavage with 2.5 % DODT in TFA and a brief treatment under mildly basic conditions to reveal the free phenol on the α-ketoacid protecting group provided crude H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 (Scheme 42). The diastereomers in both desired product and impurities partially separated on the HPLC and made the purification cumbersome, resulting in lower purity of the product (Figure 16). Purification by preparative HPLC gave semi-pure H-Opr-[SUMO3 31-75]-(phenol-protected) α-ketoacid 54 in ca. 12 % yield for SPPS and purification, which was used without further purification in the following ligations (see chapter 4.4.3.2).

Scheme 42. Synthesis of H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54

Figure 16. HPLC analysis of crude (left) and for analytical purposes repurified (right) H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54. Note: Different HPLC conditions were used, for details see Experimental.
3.4. Excursus: possible mechanism of the $\alpha$-ketoacid protection reaction

In this short excursus, based on the results of the protection reaction of Fmoc-leucine $\alpha$-ketoacid 7 with 1,3-diol derivatives 29 and 43 (see chapters 3.2.2 and 3.3.2), a possible reaction mechanism is discussed.

The observation that slow addition of the protecting groups to the reaction mixture improved the yields dramatically, and the identification of a decomposition product of protecting group 29, $\beta,\beta$-dimethyl-4-methoxystyrene, led to the conclusion that 29 and 43 are not very stable under the reaction conditions. The detected decomposition product most likely forms via elimination from a stabilized benzylic cation. An analogous reaction of a similar, but less activated 1,3-diol under acidic conditions has been observed over 100 years ago (Scheme 43).

Scheme 43. Formation of $\beta,\beta$-dimethyl styrene from 2,2-dimethyl-1-phenylpropane-1,3-diol.

However, it is likely that the benzylic cation – generated by *in situ* formed TfOH, which induces elimination of TMS$_2$O – is also a productive intermediate in the protection reaction (Scheme 44). For example, the known PMB-protection of alcohols using $p$-methoxybenzyl alcohol and different Lewis and Brønsted-acids is believed to proceed via a PMB-carbocation.

Scheme 44. Proposed mechanism of the $\alpha$-ketoacid protection reaction with 29.

The hypothesis that the $\alpha$-ketoacid protection reaction proceeds via a carbocation is corroborated by the observation that identical ratio of diastereomers of 51 are observed, regardless if the utilized orthogonally protected $\alpha$-ketoacid 49 was synthesized from racemic or enantioenriched 43 (see Figure 15). This suggests that 49 is formed via a pathway involving epimerization, for example caused by the formation of a sp$^2$-hybridized carbocation during the protection reaction.
3.5. Conclusions

In this chapter, we described the development of 29, a protecting group for \( \alpha \)-ketoacids that allows the inclusion of all canonical amino acids, including cysteine, methionine and tryptophan in SPPS and delivers the C-terminal peptide \( \alpha \)-ketoacid without epimerization directly upon cleavage of the resin. The protected \( \alpha \)-ketoacids were readily immobilized on solid support, providing a stable and storable resin.

The concept was advanced to 43, a switchable, orthogonal protecting group for \( \alpha \)-ketoacids which provides protected \( \alpha \)-ketoacids upon resin cleavage that are stable to ligation conditions and can be readily deprotected after mild activation under well-tolerated acidic conditions. We demonstrated the application of both protected \( \alpha \)-ketoacid monomers (40 and 49) in the synthesis of two protein segments with C-terminal \( \alpha \)-ketoacids (42 and 54), which we utilized in the synthesis of SUMO3 proteins, described in the next chapter.

In summary, the new \( \alpha \)-ketoacid protecting groups allow the routine synthesis of (protected) peptide \( \alpha \)-ketoacids and enable protein synthesis by KAHA ligation in both C to N and N to C terminal direction, making convergent synthetic strategies possible.
3.6. References


120 *Unpublished results* of the Bode group.


Proteomic work described at the end of this chapter was done in collaboration with Nadine Sobotzki and Prof. Dr. Bernd Wollscheid, ETH Zurich.
CHAPTER 4: Chemical synthesis of SUMO3 proteins by KAHA ligation

4.1. Introductory remarks

In this chapter, the chemical synthesis of SUMO3, an important modifier protein in eukaryotes, is discussed. In chapter 4.2, the importance and the biological background of SUMO3 are highlighted and the mechanism of protein SUMOylation is reviewed. Chapter 4.3 focuses on the first chemical synthesis of SUMO3 protein, applying technologies developed in chapter 3 and demonstrating that the synthetic protein exhibits full bioactivity. In chapter 4.4, the development of an irreversibly conjugated version of the SUMO3 protein is presented and its application as probe for protein SUMOylation in biological samples such as cell lysates is discussed.

4.2. Biological background

4.2.1. SUMO and other ubiquitin-like proteins

SUMO (small ubiquitin-like modifier) proteins belong to the small but steadily growing protein family of ubiquitin-like proteins (Ubl); other members of this protein family include for example NEDD8, URM1, UFM1 and ISG15. Despite having a low amino acid sequence similarity throughout the protein family, all Ubls share a common β-grasp fold and most possess a conserved Gly-Gly motif at the C-terminus of the mature protein, as seen in ubiquitin (Ub).

Figure 17: Amino acid sequences of ubiquitin (Ub) and human SUMO 1–4 proteins. Underlined are propeptides, which are removed during protein maturation.

Ubls are modifier proteins that are posttranslationally conjugated to proteins and modulate different properties of their targets, such as the subcellular location, stability and three-dimensional shape. Ubls are conjugated to their targets and removed by deconjugation using distinct enzyme cascades that often resemble the ubiquitination machinery. The majority of these protein modifications with Ubls are reversible and very
dynamic, which explains the fact that in most cases only a small fraction of the target protein is modified by Ubls. In contrast to small posttranslational modifications such as phosphorylation, the attachment of Ubls introduces new large surfaces on the modified protein. These new interfaces provide an effective way to modulate the interactions of the modified protein with other macromolecules such as proteins and DNA – for example by sterically blocking existing binding sites or by creating additional binding sites. This might contribute to the crucial regulatory roles that modifications with Ubls play in many different cellular processes such as transcription, DNA repair, signal transduction, cell-cycle control and autophagy.

SUMO proteins are among the longest known and most well studied group of Ubls since the initial discovery of SUMO1 (also referred to as GMP1, Sentrin or SMT3C) in 1996. In humans and most mammals, four SUMO proteins have been discovered so far: SUMO1, SUMO2 (SMT3B or Sentrin-2), SUMO3 (SMT3A) and SUMO4. It should be noted that the nomenclature of SUMO2 and 3 is used inconsistently throughout the literature; we will follow the nomenclature used in the seminal publication and in the Uniprot database.

4.2.2. Mechanism of protein SUMOylation

The mechanism of protein SUMOylation has been reviewed in great detail in several reviews and will be briefly summarized here. SUMO proteins are conjugated to a huge variety of target proteins using a well-orchestrated enzymatic machinery, depicted in Figure 18. It should be noted that some target proteins can be SUMOylated multiple times at different positions (multiSUMOylation) or can be decorated with SUMO chains (polySUMOylation, vide infra).

Figure 18. Schematic depiction of a protein SUMOylation cycle. E1 = SUMO activating enzyme, E2 = SUMO conjugating enzyme, E3 = SUMO ligase, SENP = sentrin-specific protease (vide infra).
SUMO proteins are expressed in cells as C-terminally extended precursors. In the initial step, the freshly translated SUMO protein is processed by specific proteases (mostly sentrin-specific proteases, SENPs), which remove the C-terminal propeptide to reveal the distinctive C-terminal Gly-Gly motif. In the second step, the mature SUMO protein is activated using ATP to form a high-energy thioester bond between the C-terminal Gly residue of SUMO and cysteine 173 of the SUMO activating enzyme E1. In the third step, the activated SUMO is transferred in a transthioesterification reaction onto cysteine 93 of the SUMO conjugating enzyme E2. In the fourth step, the E2 enzyme transfers the SUMO to its target protein by conjugating the C-terminal glycine of SUMO onto the ε-amino group of a lysine residue under the formation of an isopeptide bond. In many cases – especially in *in vitro* experiments with high concentrations of UBC9 – the E2 enzyme is capable of conjugating the SUMO onto the target protein without the requirement for an E3 enzyme. Nevertheless, a number of SUMO E3 ligases were identified, facilitating the *in vivo* SUMOylation of a diverse set of substrate proteins. Finally, the protein SUMOylation is reversed (deSUMOylation) by specific proteases (e.g. SENPs or others) that cleave the isopeptide bond and unveil the unmodified target protein and free SUMO, which can be used in another modification cycle.

4.2.3. SUMO isoforms

As previously mentioned, there are four isoforms of SUMO proteins found in humans. SUMO2 and 3 are 96 % identical; SUMO2 and 3 are each ca. 45 % identical to SUMO1 and ca. 83 % identical to SUMO4. Despite the differences in the primary amino acid sequences (see Figure 17), all SUMO isoforms adopt a β-grasp fold similar to ubiquitin, which is crucial for their recognition by the SUMOylation machinery (Figure 19).

![Figure 19. Tertiary structure of ubiquitin (left) and SUMO3 (right).](image)

The SUMO proteins can be divided in three groups based on their properties and functions: SUMO1, SUMO2/3 and SUMO4. As already seen from the primary amino acid sequence, SUMO2 and 3 are very similar (in most contexts indistinguishable), and are
therefore often referred to as SUMO2/3. Selected distinctions between the SUMO isoforms include examples described below.

Not surprisingly, SUMO isoforms have different sets of target proteins. Some target proteins are conjugated only to SUMO1, whereas some are conjugated only to SUMO2/3 and others to all SUMO isoforms.\textsuperscript{156}

The cellular concentrations of free and conjugated SUMOs vary among the SUMO isoforms. SUMO2/3 is in general more abundant than SUMO1 and there is a significant pool of unconjugated SUMO2/3 present in the cells, whereas there is almost no unconjugated SUMO1.\textsuperscript{157} This is also reflected in different conjugation dynamics: SUMO1 conjugation is much less dynamic than the other SUMO isoforms and the response of the SUMOylation machinery to protein damaging stimuli is different.\textsuperscript{157,158}

SUMO2/3 can be SUMOylated itself at a single, conserved lysine residue (Lys11) that is embedded in a consensus motif (\textit{vide infra}), resulting in the formation of poly-SUMO chains.\textsuperscript{159} SUMO1 posses a similar lysine residue, but it is not part of a consensus motif and SUMOylation does not appear to occur. Nevertheless, SUMO1 can be used to terminate poly-SUMO2/3 chains.\textsuperscript{160}

SUMO4 has a proline residue (Pro90) close to the C-terminus, rendering it resistant to maturation by regular SENP proteases.\textsuperscript{161} The biological role of SUMO4 is not very well understood, especially as the literature regarding the functionality of SUMO4 is ambiguous.\textsuperscript{161,162}

In summary, SUMO1 and SUMO2/3 have distinct properties and at least to some degree different functions. SUMOylation in general is essential for most organisms. Depending on the organism, loss of function of one SUMO isoform can be compensated by other isoforms, for example in mouse, where SUMO1 is not essential.\textsuperscript{163} Protein modification with SUMO4 is much less understood and seems to be limited to certain tissues (e.g. kidney, lymph nodes).\textsuperscript{164}

4.2.4. The enzymatic SUMOylation machinery

The SUMOylation machinery consists of a surprisingly small number of enzymes, especially compared to the high number of enzymes found in the ubiquitination machinery (Table 6).

\textbf{Table 6. Comparison of enzymes involved in ubiquitination and SUMOylation.}\textsuperscript{165,166}

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Deconjugating enzymes</th>
<th>E1 activating enzymes</th>
<th>E2 conjugating enzymes</th>
<th>E3 ligases</th>
</tr>
</thead>
<tbody>
<tr>
<td>ubiquitin</td>
<td>ca. 100 \textsuperscript{Ref.167}</td>
<td>2 (UBE1, UBE1L2)</td>
<td>&gt; 37 \textsuperscript{Ref.168}</td>
<td>&gt; 1000</td>
</tr>
<tr>
<td>SUMO</td>
<td>9 \textsuperscript{Ref.169}</td>
<td>1 (SAE1/SAE2)</td>
<td>1 (UBC9)</td>
<td>&gt; 7 \textsuperscript{Ref.131}</td>
</tr>
</tbody>
</table>
SUMO-activating enzyme E1

Remarkably, all SUMO isoforms are activated by the same E1 enzyme, consisting of SUMO-activating enzyme subunit 1 (SAE1, also known as AOS1) and SUMO-activating enzyme subunit 2 (SAE2, UBA2).

SUMO-conjugating enzyme E2

Similarly, one single, highly conserved E2 enzyme, UBC9, is used to conjugate SUMO to the target protein. In addition, the E2 enzyme is involved in selecting the protein target and acceptor lysine residue. At least three different mechanisms or a combination thereof are used to select the acceptor lysine residue (Figure 20).

Figure 20. Prototypic mechanisms of acceptor lysine selection in the SUMOylation process.

In the simplest case, the acceptor lysine residue is embedded in a short consensus motif (ΨKXE; Ψ, a bulky aliphatic residue) that is recognized by UBC9 directly. Recently, several related motifs were identified, such as the phosphorylation-dependent SUMOylation motif (PDSM), where SUMOylation is regulated by phosphorylation of a serine near the consensus motif. In most proteomes investigated to date, the majority of all SUMOylated lysine residues are part of the consensus motif or a variant of it. The direct, E3-independent binding of proteins containing the consensus motif to UBC9 is rather weak in most cases, resulting in low efficient SUMOylation. However, this weak binding is believed to be responsible for the SUMOylation of target proteins in in vitro experiments in the absence of E3 ligases with unphysiologically high concentrations of UBC9.

Alternatively, the target protein contains a hydrophobic SUMO-interacting motif (SIM) that binds to the SUMO of a SUMO-UBC9 complex and brings one or more lysine residues in close proximity to the UBC9 active site.

SUMO ligase E3

In the third mechanism, an E3 ligase acts as non-covalent scaffold, which binds to both the target protein and the SUMO-UBC9 complex and orients the target lysine – which is often but not necessarily embedded in a consensus motif – towards the UBC9 active site. Several E3 ligases have been identified in humans, including members from the PIAS (protein inhibitor of activated STAT) family, RanBP2 (Ran-binding protein 2) and PC2 (polycomb protein 2). Although a huge number of proteins can be SUMOylated in vitro
without E3 ligases (see above), E3-mediated SUMOylation is most likely the predominant pathway in vivo.\textsuperscript{179}

**SUMO-specific proteases**

SUMO-specific proteases have two fundamental roles regarding protein SUMOylation. First, they are required to remove the propeptide from the C-terminus of the precursor protein to reveal the characteristic Gly-Gly motif (peptidase activity). Second, and more importantly, they are responsible for deSUMOylation by cleaving the isopeptide bond between SUMO and the lysine of the target protein (isopeptidase activity), rendering them a central factor in regulating protein SUMOylation. So far, a surprisingly small number or SUMO-specific proteases have been identified; all of them cysteine proteases. They are divided into three groups: sentrin-specific proteases (SENPs), deSUMOylating isopeptidases (DeSIs),\textsuperscript{180} and the SUMO-specific isopeptidase USPL1 (also known as ubiquitin-specific peptidase-like protein 1).\textsuperscript{181}

SUMO-specific proteases vary in their peptidase activity, isoform preference and subcellular localization. For example, the six SENPs (SENP1-3 and SENP5-7) expressed in humans have both isopeptidase and – with the exception of SENP6 and -7 – peptidase activity, however they exhibit some isoform preference in vivo.\textsuperscript{153} SENP2, which is localized in the cytoplasm – unlike other SENPs that are mostly found in the nucleus – prefers SUMO2/3 to SUMO1, whereas SENP1 processes both isoforms. The recently found DeSI1/-2 and USPL1 exhibit mostly isopeptidase activity.

**4.2.5. Biological function of SUMO**

Protein SUMOylation is involved in the regulation of numerous biological processes including transcription, DNA repair, nuclear transport and several signal transduction pathways, which has been extensively reviewed in the literature.\textsuperscript{151,179} Two examples are briefly discussed below.

A classic, yet not fully understood example of SUMOylation regulating nuclear transport is RanGAP1 (RanGTPase activating protein 1).\textsuperscript{182} RanGAP1 is one of the oldest known SUMOylated proteins and widely used as a target protein in in vitro SUMOylation assays. Once SUMOylated, RanGAP1 forms a tight complex with UBC9 and RanBP2 at the cytoplasmic side of the nuclear pore complex, where it activates Ran, a GTPase, which plays a central role in nucleocytoplasmic transport processes.

Many transcription factors are SUMOylated, affecting their activity.\textsuperscript{183} In many cases, SUMOylation of a transcription factor results in downregulation of transcription. For example, the transcription factor c-Myb can be SUMOylated on two lysine residues, reducing the transactivation capacity of c-Myb.\textsuperscript{184} In contrary, the stress-induced SUMOylation of heat
shock transcription factor 1 (HSF1) was found to have a positive effect on transcription of heat shock proteins.\textsuperscript{185}

Notably, there is an extensive crosstalk between SUMOylation and other posttranslational modifications. A large number of lysine residues that are SUMOylated, can also be modified by ubiquitination, acetylation or methylation.\textsuperscript{186} The finding of phosphoserine-containing consensus motifs for SUMOylation (see above) shows that the crosstalk between different PTMs goes far beyond a simple competition for modifiable lysine residues, but rather adds a mutual regulatory role between the PTMs. This is corroborated by the existence of SUMO-targeted ubiquitin ligases (STUbLs), which facilitate the ubiquitination of SUMOylated proteins and often lead to degradation of the target protein.\textsuperscript{187}

Given the broad involvement in a large variety of cellular processes, it is not surprising that SUMOylation seems to play an important role in the pathogenesis of several diseases,\textsuperscript{188} for example Huntington’s disease,\textsuperscript{189} or in the infection by pathogens such as bacteria or viruses.\textsuperscript{190} Several therapeutic approaches targeting the SUMOylation of relevant proteins have been proposed, for example inhibiting the SUMOylation of Myc, a dysregulated transcription factor in certain human cancers.\textsuperscript{191}

In summary, the family of SUMO proteins is biologically very important and we envisioned that our KAHA ligation technology could provide a good synthesis route to chemically well defined SUMO3 protein, which could be used in further biological studies.

4.3. Synthesis of SUMO3 protein and biochemical evaluation

In continuation of our interest in the synthesis and study of modifier proteins, we decided to chemically synthesize SUMO3 using the KAHA ligation. To the best of our knowledge, SUMO2 or -3 have not been chemically synthesized before; SUMO1 has been chemically synthesized by Melnyk \textit{et al.} in 2014 using a NCL strategy.\textsuperscript{192}

The overall goal of this part of the project was to establish a reliable route for the total chemical synthesis of mature SUMO3 protein and to evaluate whether the synthetic protein would show natural bioactivity – in other words whether the synthetic SUMO3 would be recognized by the enzymatic SUMOylation machinery and conjugated to target proteins.

Besides this, we identified several questions we aimed to answer with this synthetic endeavor:

- Can the protected $\alpha$-ketoacid resin 41 (see chapter 3.2.4) be used to reliably synthesize large protein segment $\alpha$-ketoacids?
- Can a sterically more demanding $\alpha$-ketoacid – valine $\alpha$-ketoacid – be used in KAHA ligations?
• Is the KAHA ligation viable in the presence of unprotected cysteine residues, which could potentially cause side reactions due to their nucleophilic side chain?

4.3.1. Retrosynthetic analysis of SUMO3

Being synthetic organic chemists, our first step towards the synthesis of SUMO3 was a retrosynthetic analysis, starting with a careful examination of the amino acid sequence and the structure of SUMO3, summarized in Figure 21.

Figure 21. Primary, secondary and tertiary structure of SUMO3, derived from Uniprot entry P55854, SUMO3 (14–92).

At 92 amino acids, the sheer size of SUMO3 would have rendered the direct synthesis by SPPS quite challenging. A reasonable alternative was to use a more convergent protein segment ligation approach. Among the available ligation techniques, two possibilities seemed plausible – native chemical ligation using C-terminal thioesters and N-terminal cysteine residues or the α-ketoacid-hydroxylamine ligation using C-terminal α-ketoacids and N-terminal hydroxylamines.

SUMO3 contains a cysteine residue located in the middle of the protein, which might have been suited as a ligation site in native chemical ligation. However, this would have required the synthesis of two segments of about 45 amino acids length, which might have imposed synthetic challenges, especially considering the limited technologies for the synthesis of large protein thioester segments that were available at the time. The introduction of another ligation site for NCL seemed not to be a promising option, as this would have led to an additional unnatural cysteine residue that could have had unforeseen consequences for protein folding and activity, for example by forming intramolecular disulfides. Removal of the unnatural cysteine residue would have required a multistep, late stage desulfurization protocol, complicating the synthesis.

Given these limitations, we focused on using the KAHA ligation for the synthesis of SUMO3. In contrast to NCL, the KAHA ligation does not depend on cysteine residues at the ligation site, providing us with more flexibility in the synthetic strategy. To make the protein segments readily accessible by SPPS, we investigated the possibility of a strategy using two sequential ligations of three protein segments each of about 30 residues. Naturally, we
wanted to introduce as few mutations as possible in the final protein, so the search for suitable ligation sites was governed by two factors: a) which protein segment α-ketoacids were accessible and b) which homoserine mutations introduced by the KAHA ligation with 5-oxaproline could be expected to cause little disturbance to protein structure and function.

With these strategic considerations in mind, we chose two ligation sites, Val29–Gln30 and Leu52–Ser53 (Figure 22), which would result in two mutations after KAHA ligation with 5-oxaproline, Gln30Hse and Ser53Hse.

At first glance, the ligation site Val29–Gln30 might seem not so obvious as it is located in a relatively ordered β-strand structure, imposing the risk of the Hse mutation altering the fold. However, the mutated Gln30 is on the solvent exposed side of the β-strand, which encouraged us that the alteration imposed by the Hse mutation would be tolerated. From a chemical point of view, this ligation site offered us the exciting possibility to evaluate the performance of a valine α-ketoacid in the KAHA ligation, an α-ketoacid that has not been tested before. The second ligation site, Leu52–Ser53, required a well-established KAHA ligation with leucine α-ketoacid. It is embedded in a flexible loop region and considering the quite conservative Ser53Hse mutation, we were confident that mutations at this ligation site would have little impact on the overall protein fold and function.

4.3.2. Synthetic strategy

With the ligation sites set, we focused on designing a suitable strategy for the assembly of the protein segments (Scheme 45). Dictated by the KAHA ligation chemistry, protein segments 1 (55) and 2 (42) (counted from the N-terminus) possess a C-terminal α-ketoacid, whereas segment 2 (42) and 3 (56) have an N-terminal 5-oxaproline. A closer look at the amino acid sequence revealed that the bifunctional segment 2 (42) contains two oxidation sensitive residues, Met43 and Cys47, which had two consequences: first, it precluded the use of a protein segment cyanosulfurylide precursor, as Met and Cys residues are generally not compatible with the oxidative conditions used for transforming the cyanosulfurylide into the α-ketoacid. Instead, this provided us with the perfect scenario to
CHAPTER 4: CHEMICAL SYNTHESIS OF SUMO3 PROTEINS

evaluate the protected \( \alpha \)-ketoacid resin 41 we developed in chapter 3.2. Second, as the orthogonally protected \( \alpha \)-ketoacid resin 50 (see chapter 3.3) – which would make an assembly in N to C direction possible – was not yet developed, the use of protected \( \alpha \)-ketoacid resin 41 determined the assembly direction of SUMO3 in this synthesis to be from C to N-terminus. Consequently, we decided to orthogonally protect the N-terminal oxaproline of segment 2 (42) during the first ligation using the established Fmoc protection.

Scheme 45. Synthetic strategy for SUMO3.

4.3.3. Synthesis of protein segments

Having designed a synthetic approach to SUMO3, we proceeded with the solid phase synthesis of the three segments shown in Scheme 45. The segment syntheses are outlined below; details and characterization data can be found in the Experimental.

The N-terminal segment H-[SUMO3 (2–29)]-\( \alpha \)-ketoacid 55 was prepared uneventfully by oxidation of the corresponding cyanosulfurylde, which was synthesized using our previously reported linker on polystyrene resin (Scheme 46).76

Scheme 46. Synthesis of H-[SUMO3 (2-29)]-\( \alpha \)-ketoacid 55.

As a test case for the newly developed resin, the bifunctional middle segment Fmoc-Opr-[SUMO3 (31–52)]-\( \alpha \)-ketoacid 42 was synthesized by automated SPPS on protected \( \alpha \)-ketoacid resin 41 as described in more detail in chapter 3.2.4.
In the first synthesis of the C-terminal segment H-Opr-[SUMO3 (54–92)]-OH 56, significant aspartimide formation was observed, lowering the crude purity significantly. To prevent this, two dipeptide building blocks were introduced during the synthesis: Asp70–Thr71 was coupled manually as pseudoproline dipeptide (Fmoc-Asp(OtBu)-Thr(ΨMe,Mepro)-OH) and Asp62–Gly63 as DMB-backbone protected dipeptide (Fmoc-Asp(OtBu)-N(Dmb)-Gly-OH). With these optimizations, H-Opr-[SUMO3 (54–92)]-OH 56 could be readily prepared by automated synthesis on polystyrene resin functionalized with a Wang-linker (Scheme 47).

Scheme 47. Synthesis of H-Opr-[SUMO3 (54-92)]-OH 56.

4.3.4. Assembly of SUMO protein by sequential KAHA ligations

Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42 and H-Opr-[SUMO3 (54–92)]-OH 56 were found to be readily soluble in a mixture of 7:3 NMP:H2O containing 0.1 M oxalic acid, and the ligation proceeded smoothly at 20 mM protein segment concentration at 60 °C within 13 h (Figure 23). No side reactions from the unprotected cysteine residue in 42 were observed during the ligation, demonstrating that the KAHA ligation is viable in presence of free thiols. After purification by preparative HPLC, Fmoc-Opr-[SUMO3 (31–92)]-OH depsipeptide 57 was obtained in 28 % yield.

Figure 23. KAHA ligation of 42 and 56 (left) and HPLC monitoring (right).

In previous protein syntheses, our solvents of choice were mixtures of DMSO and H2O containing 0.1 M oxalic acid, mainly attributed to the high solubility of most protein
segments in these mixtures. However, alerted by literature precedence,\textsuperscript{196} we found that cysteine and methionine residues were readily oxidized by DMSO under acidic ligation conditions at elevated temperatures. Gratifyingly, solvent mixtures of distilled and degassed NMP and H\textsubscript{2}O containing 0.1 M oxalic acid were identified as suitable ligation solvents and exhibited similar protein segment solubility as DMSO:water mixtures, but avoided the oxidation of cysteine and methionine residues.

Prior to the next ligation, two steps had to be performed: the N-terminal 5-oxaproline of \textit{depsi-57} had to be Fmoc-deprotected, and we opted to convert the ester-linkage at the ligation site in \textit{depsi-57} to the natural amide bond. For a streamlined procedure we aimed to develop a one-pot procedure. However, attempting the solution-phase Fmoc-deprotection using our previously developed conditions – treating \textit{depsi-57} with 5 \text{%} HNE\textsubscript{2} in DMSO at a peptide concentration of typically 5–10 mM for 5 min – resulted in the significant formation of a byproduct, which turned out to be a covalent adduct between the protein segment and \textit{in situ} generated dibenzofulvene. The most plausible explanation for this would be the formation of a thioether between the cysteine thiol and the electrophilic dibenzofulvene (Scheme 48).\textsuperscript{197}

\begin{center}
\textbf{Scheme 48. Formation of thioether side product during Fmoc-deprotection of 57 under standard conditions.}
\end{center}

We were pleased to find that simply performing the deprotection under more dilute conditions (< 1 mM protein segment concentration) suppressed the thioether formation and gave clean conversion to H-Opr-[SUMO3 (31–92)]-OH \textit{depsi-58}. After completion of the Fmoc-deprotection, the reaction mixture was directly diluted with rearrangement buffer (pH 9.5, 0.2 M carbonate buffer, 10 mM TCEP) and complete O to N acyl shift was observed after 2 h at room temperature (Figure 24). The addition of TCEP was crucial to avoid oxidation of cysteine and methionine residues by air or DMSO. Purification gave H-Opr-[SUMO3 (31–92)]-OH 58 in 47 \text{%} yield for Fmoc deprotection, O to N acyl shift and purification.
With H-Opr-[SUMO3 (31–92)]-OH 58 in hand, we proceeded to the final ligation. H-[SUMO3 (2–29)]-α-ketoacid 55 and H-Opr-[SUMO3 (31–92)]-OH 58 readily dissolved in 7:3 NMP:H₂O (15 mM peptide concentration) containing 0.1 M oxalic acid and were reacted at 60 °C. Despite the sterically more demanding valine α-ketoacid, the ligation was complete within 9 h (Figure 25) and provided H-[SUMO3 (2–92)]-OH depsi-59 in 23 % yield after purification by preparative HPLC.

The final O to N acyl shift to convert depsi-59 to 59 was performed in a modified rearrangement buffer containing 6 M GdmCl to provide denaturing conditions and potentially facilitate the rearrangement. Analysis by analytical HPLC indicated complete conversion to H-[SUMO3 (2–92)]-OH 59. However, the difference in retention time of depsi-59 and 59 was small (< 0.1 min) under various HPLC conditions examined (Figure 26). The final synthetic SUMO protein H-[SUMO3 (2–92)]-OH 59 was obtained in 87 % yield and in milligram-quantities after purification by preparative HPLC.
CHAPTER 4: CHEMICAL SYNTHESIS OF SUMO3 PROTEINS

Figure 26. \( O \) to \( N \) acyl shift of \( \text{depsi-59 to 59} \) (left) and HPLC monitoring (right).

The identity and purity of SUMO3 \( H-[\text{SUMO3 (2–92)}]-\text{OH} \) 59 was confirmed by analytical HPLC and ESI-QTOF MS analysis (Figure 27).

Figure 27. Analytical HPLC (left) and deconvoluted ESI-QTOF MS (right) of \( H-[\text{SUMO3 (2–92)}]-\text{OH} \) 59.

A CD spectrum of SUMO3 \( H-[\text{SUMO3 (2–92)}]-\text{OH} \) 59 was recorded in PBS. It compares well with a spectrum reported in the literature\(^ {198} \) of recombinant SUMO3 protein suggesting that the synthetic SUMO3 readily adopts a similar fold.

Figure 28. CD spectra of synthetic \( H-[\text{SUMO3 (2–92)}]-\text{OH} \) 59 (left) in comparison to a CD spectra for recombinant SUMO3 reported in the literature.
4.3.5. Biochemical evaluation of synthetic SUMO3 (59)

With the final SUMO3 (59) in hand, we were eager to investigate whether the two homoserine mutations in the synthetic protein would disturb the protein function. We performed in vitro SUMOylation reactions to evaluate whether the SUMOylation enzymes E1 and E2 would recognize our synthetic SUMO3 (59) and conjugate it to a well-established SUMOylation target, RanGAP1. Separating the products of the enzymatic reaction by SDS-PAGE followed by Western blotting using a SUMO2/3 antibody revealed that the synthetic SUMO3 protein (59) was readily accepted by the SUMOylation enzymes to generate a synthetic SUMO3-RanGAP1 conjugate (Figure 29). This result demonstrates that the synthetic SUMO3 (59) containing two homoserine mutations exhibits full biological activity regarding the enzymatic conjugation compared to natural SUMO3.

![Figure 29. In vitro conjugation of synthetic SUMO3 (59) to RanGAP1 using SUMOylation enzymes (left). Analysis of the SUMOylation reaction by Western blot using anti SUMO2/3 antibody (right).](image)

As negative controls, reactions lacking MgATP were utilized. Recombinant SUMO3 contained an impurity migrating below the band corresponding to the RanGAP1-SUMO3 conjugate.

4.3.6. Conclusions

We developed a reliable and scalable synthesis of SUMO3 using the KAHA ligation. For the synthesis of one segment, we successfully utilized our newly developed protected α-ketoacid resin. This not only considerably streamlined the synthesis compared to our previously used cyanosulfurylilide method but also enabled for the first time the presence of oxidation sensitive cysteine and methionine residues in the protein α-ketoacid segments. The KAHA ligation in the synthesis of SUMO3 (59) were the first examples of an unprotected cysteine residue being present during KAHA ligation and we were pleased to observe that the free thiol did not lead to side reactions. A sterically more demanding α-ketoacid – valine α-ketoacid – was utilized in the KAHA ligation of protein segments for the first time and exhibited similar reactivity compared to sterically less demanding α-ketoacids such as leucine α-ketoacid.
In vitro assays demonstrated that the synthetic SUMO3 protein (59) was readily processed by the enzymatic SUMOylation machinery and conjugated to a SUMOylation target, RanGAP1. These results disclosed that – despite two homoserine mutations introduced by the KAHA ligation with 5-oxaproline – the synthetic SUMO3 protein (59) had full bioactivity regarding the enzymatic conjugation compared to natural SUMO3.

4.4. Modified SUMO3 proteins as probe for SUMOylation

The goal of this project was to develop and chemically synthesize a modified SUMO3 protein (“SUMOylation probe”) that would allow for enriching and identifying SUMOylated proteins in biological samples such as cell lysates. Given the broad involvement of protein SUMOylation in the regulation of various biological processes and disease pathogenesis, it is of great interest from a biological perspective to investigate and further understand protein SUMOylation.

Different methods relying on the overexpression of genes encoding for affinity-tagged SUMO proteins have been developed and successfully used to identify a huge number of SUMOylated proteins and specific SUMOylation sites. However, these methods require the complicated and time consuming genetic transformation of cells, whereas an in vitro assay using a chemically well defined SUMOylation probe could be used to quickly screen different cell lysates or tissue homogenates. Additionally, the analysis of cellular protein SUMOylation – especially regarding low abundant SUMOylated proteins – is complicated by two factors: the highly dynamic and often transient nature of protein SUMOylation, and the typically small fraction of a given target protein that is in fact SUMOylated. We envisioned that chemically synthesized proteins could provide a unique opportunity to address these challenges and set two major goals for our synthetic protein (Figure 30).

Figure 30. Comparison of cellular protein SUMOylation with natural and modified, irreversibly conjugated SUMO3 and application as probe for protein SUMOylation.

First, the synthetic SUMO3 protein should be irreversibly conjugated to target proteins, rendering the modification with the SUMO3 permanent. Eliminating the dynamic
equilibrium between conjugation and deconjugation should increase the concentration of SUMOylated proteins in general, and potentially allow for the interception of targets that are SUMOylated only transiently or in low ratio. Furthermore, this would provide a convenient way to avoid the general and rapid deSUMOylation often observed upon cell lysis. Second, the synthetic SUMO3 protein should be decorated with an affinity tag to enable specific enrichment and purification, facilitating downstream protein identification by reducing the sample complexity.

In the course of this project, three generations of modified SUMO proteins were synthesized. In chapter 4.4.3, the chemical synthesis of the modified SUMO proteins is discussed. Chapter 4.4.4 focuses on the biochemical evaluation of these modified SUMO proteins and the application of the third generation probe to enrich and identify SUMOylated proteins from HEK293 lysates is demonstrated. While we initially aimed to employ our SUMOylation probe in cellulo, preliminary studies indicated that it would be very challenging to transfec sufficient quantities of our protein-based probe into the cells and we decided to focus on cell lysates for the proof of concept.

4.4.1. Design of the SUMOylation probe

To begin, we had to solve two problems: 1) we had to find a way to render our synthetic protein resistant towards deconjugation – without affecting the recognition and processing by the SUMOylation machinery, and 2) we had to decide on a suitable affinity purification tag that would be reasonably facile to implement in the synthesis, would be compatible with SPPS and KAHA ligation conditions and would not interfere with the protein function.

Literature precedence indicated that the C-terminal propeptide of SUMO4 – unlike the propeptides of SUMO1-3 – would not be removed by the cellular SENP enzymes. This was attributed to a mutation (Gln90Pro) at the P4 position that is present exclusively in SUMO4; the rest of the C-terminal region is highly conserved among all SUMO isoforms. An in vitro study profiling the substrate preference of different SENP enzymes using a tetrapeptide library resembling the C-terminal region of SUMO also found that a proline residue at the P4 position effectively prevents processing through the SENPs. In a study focused on the dynamics of SUMO chains, Salvesen et al. overexpressed a gene encoding for a mature SUMO2 protein containing a proline mutation at the P4 position in HEK293 cells and observed protein conjugates of mutant SUMO2 that were resistant to cleavage by SENP enzymes. Taken together, these precedents suggested that the proline mutation would be a good candidate to develop an irreversibly conjugated synthetic SUMO.

As affinity purification tag, biotin seemed a well-suited choice to us for several reasons. Biotin is routinely used in combination with immobilized streptavidin or avidin –
proteins binding biotin very tightly – in pulldown experiments and the required components for affinity purification are commercially available. The carboxylic acid of biotin can be readily exploited as synthetic handle to link the biotin for example via amide formation to the protein. Compared to other affinity purification tags, biotin is quite small and likely would not affect the properties and folding of the SUMO protein. Additionally, a biotin tag would be orthogonal to other frequently used, genetically encodable tags such as His- or FLAG-tags. Convinced by these advantages, we decided to use biotin as an affinity tag. Although there was literature precedence for the solid phase synthesis of biotinylated peptides, the stability of the biotin during cleavage and ligation had to be carefully evaluated.

First generation of modified SUMO proteins

The simplest implementation of our general design was realized in the first generation of modified SUMO proteins (Figure 31). In this protein (61), a Gln89Pro mutation and a biotin affinity tag, directly linked to the N-terminus, were incorporated. As a control, a similar protein (60) containing the natural Gln89 was synthesized.

Figure 31: First generation of modified SUMO proteins.

Second generation of modified SUMO proteins

Although the results of the first generation were promising (see chapter 4.4.4.1), we seized the unique opportunity that chemical synthesis offered and evaluated a small library of modified proteins (Figure 32, 62–65) that could potentially be used as probes for SUMOylation and that would be very challenging to obtain by recombinant overexpression.

The first modification we wanted to test aimed at the C-terminal Gly-Gly motif, which should be replaced with a carbon analog missing the amide bond between Gly91 and Gly92 (64) – it would be very interesting to see if the SUMOylation machinery tolerated this modification in the highly conserved C-terminus. In the second modification, we decided to N-methylate the amide bond between Gln88 and Gln89, mimicking the tertiary amide bond of the Gln89Pro mutation. The change in secondary structure caused by the Gln89Pro mutation is believed to be responsible for the resistance towards isopeptidases and it would be interesting to see if a simple backbone N-methylation had a similar effect.

We additionally introduced two adaptions to the design of our modified proteins. First, we realized that a hydrophilic, flexible linker between the biotin affinity tag and the protein
might be beneficial for binding to (immobilized) streptavidin or avidin, and selected a commonly used PEG linker for this purpose. Second, as we aimed to perform these explorative syntheses on the smallest scale possible, we opted to introduce a highly UV active chromophore in the protein, which should facilitate monitoring of the ligations. The chromophore would have to tolerate the acidic reaction media and elevated temperatures during resin cleavage and KAHA ligation, which led us to choose a chemically robust and easily introduced coumarin-based dye. For synthetic convenience and to use previously obtained resins, both modifications were linked to an additional amino acid introduced on the N-terminus of protein segment 1.

Figure 32: Second generation of modified SUMO proteins.

The results from the second generation of modified SUMO3 proteins as probes for protein SUMOylation fortified our selection of using a Gln89Pro mutation to render the conjugation irreversible (see chapter 4.4.4.2).

Third generation of modified SUMO proteins

At this point of the project, we needed to restart and scale up our probe synthesis to obtain sufficient quantities for the evaluation of SUMOylation of proteins in cell lysates and optimize protocols for the enrichment of SUMOylated proteins. We seized the opportunity to make another change to our probe design, aiming to substitute the Western blots we used to visualize probe-labeled proteins with an alternative method, in-gel fluorescence scanning. In our application, Western blots had two drawbacks: first, the procedure for Western blotting was quite time consuming – it took around one day to develop the Western blots. Second, we found in preliminary studies using Western blots of HEK293 cell lysates that a considerable number of proteins are naturally biotinylated and SUMOylated (even after cell lysis), which would render it very challenging to differentiate the natural background from proteins that would be labeled with our exogenous SUMOylation probe.

In order to make our probe readily detectable by in-gel fluorescence with the available equipment, we had to exchange the coumarin dye, which would not be very well suited for
in-gel fluorescence due to its short wavelength absorption at ca. 380 nm, with a rhodamine dye, which has a more convenient absorption around 570 nm and is commonly used for fluorescence detection (Figure 33).

Figure 33: Third generation of modified SUMO protein.

4.4.2. Synthetic strategy for modified SUMO proteins

Already at the beginning of the project, it was clear that we had to synthesize multiple versions of our SUMOylation probe, which differed in the amino acid sequence in the C-terminal region. To minimize the synthetic effort for the C-terminal protein segments, we decided to move the ligation site between segment 2 and 3 towards the C-terminus. In the synthesis of SUMO2 – which was performed by a colleague, Dr. Vijay Pattabiraman, in parallel to our previous synthesis of SUMO3 – an additional ligation site between Leu76 and Glu77 was used (corresponding to Leu75 and Glu76 in SUMO3). The synthesized SUMO2 protein with a third homoserine mutation at position 77 was readily conjugated to RanGAP1, demonstrating that this mutation – at least in SUMO2 – was well tolerated by the enzymatic SUMOylation machinery. Given the high structural similarity between SUMO2 and SUMO3, and considering the fact that Glu76 in SUMO3 is solvent exposed, we assumed that introducing a homoserine mutation at this position would likely not affect the bioactivity of the protein (Figure 34).

Figure 34. Disconnection of SUMO3 into three segments with moved ligation site (left). Residues mutated to Hse through KAHA ligation with 5-oxaproline highlighted in the protein structure (right).

With the ligation sites set, we revisited our strategy to assemble the protein segments. The previous synthetic strategy included a protein assembly in C to N direction –
consequently, any variation in the C-terminal segment would require a completely independent synthesis cycle for each variant involving two ligations and several purifications. On the contrary, a synthetic strategy including an assembly in N to C-terminal direction would offer the possibility to have a common synthetic route for all C-terminal variants of each generation until the last ligation and reduce the overall synthetic effort significantly.

As previously discussed, an N to C-terminal assembly requires that the $\alpha$-ketoacid in the middle segment is either masked as cyanosulfurylde precursor or is otherwise orthogonally protected. Gratifyingly, we had developed the orthogonally protected $\alpha$-ketoacid resin 50 (see chapter 3.3) at that time and were eager to evaluate this concept in the setting of a larger protein synthesis.

For the first and second generation of modified SUMO3 proteins (60–65), in which a flexible strategy to incorporate the different C-termini was important, we decided to utilize an assembly in N to C direction (Scheme 49).

Scheme 49: Synthetic strategy for first and second generation modified SUMO proteins 60–65 using an assembly in N to C direction.

For the synthesis of the third generation protein (66), we decided to go back to an assembly in C to N-terminal direction (Scheme 50), as this would avoid the tedious purification of the middle (phenol-protected) $\alpha$-ketoacid segment (see chapter 3.3.3).
4.4.3. Synthesis of modified SUMO proteins

4.4.3.1. Synthesis of the protein segments

Selected examples for the synthesis of the required protein segments for all three generations of modified SUMO proteins are discussed here, full detail and characterization data can be found in the Experimental.

N-terminal segments

The N-terminal segment of the first generation protein, Biotin-[SUMO3 (2–29)]-α-ketoacid 67, differs from the previously prepared H-[SUMO3 (2–29)]-α-ketoacid 55 only by the N-terminal biotin. However, for synthetic convenience and to rule out incompatibilities of the sulfur-containing biotin with the oxidative conditions used for cyanosulfonylide oxidation, we decided to synthesize 67 on protected valine α-ketoacid resin. Unlike before, significant aspartimide formation was observed during the synthesis, which was avoided by coupling Asp25-Gly26 as DMB-protected dipeptide. Biotin was readily introduced using regular amide coupling conditions (Scheme 51).
Scheme 51. Synthesis of Biotin-[SUMO3 (2–29)]-α-ketoacid 67.

For the second and third generation proteins, the N-terminal modifications were introduced via in-house made building blocks 79 and 80 by coupling onto the same resin as for the first generation (Scheme 51). We were pleased to observe that all modifications were stable to resin cleavage and purification conditions.

Figure 35: Synthesis of Biotin-PEG-Dap(Cou)-[SUMO3 2–29]-α-ketoacid 68. The synthesis of Biotin-PEG-Dap(Rho)-[SUMO3 2–29]-α-ketoacid 75 followed a similar protocol, see Experimental for details.

Bifunctional middle segments

The first and second generation proteins utilize the same bifunctional middle segment with a C-terminal (phenol-protected) α-ketoacid. The segment H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 was synthesized by automated SPPS on orthogonally protected α-ketoacid resin 50 as a test case for the newly developed resin, as described in more detail in chapter 3.3.3.

The middle segment H-Opr-[SUMO3 (31–75)]-α-ketoacid 76 of the third generation protein was synthesized uneventfully on resin 41.
C-terminal segments

The C-terminal segments for all proteins were synthesized routinely on polystyrene resin functionalized either with a Wang or a 2-chlorotrityl linker.

A notable exception was the synthesis of H-Opr-[SUMO3 (77–92) Q88\textsubscript{NMe}Q89]-OH \textbf{74}. Two separate isomers of the desired protein segment were obtained in approximately equal amounts (Figure 36). To differentiate these isomers, the earlier eluting isomer of H-Opr-[SUMO3 (77–92) Q88\textsubscript{NMe}Q89]-OH is named \textbf{74-A} and the later eluting isomer \textbf{74-B}.

![Figure 36. Synthesis (left) of 74 and crude analytical HPLC (right) revealing two isomers 74-A and 74-B.](image)

Given the complexity of the molecule, we could only speculate about the exact nature of these isomers. Theoretically, they could be unusually stable E- and Z-isomers at the tertiary amide bond between Gln88 and Gln89.\textsuperscript{207} Such tertiary amide bonds typically isomerize between E- and Z-form on time scales of seconds to minutes at room temperature.\textsuperscript{208} However, no interconversion between \textbf{74-A} and \textbf{74-B} was observed even at elevated temperature (70 °C, 30 min), which makes it unlikely that the observed isomers are E- and Z-isomers of the amide bond between Gln88 and Gln89. Alternatively, the isomers could come from epimerization at Gln89. Gln88\textsubscript{NMe}Gln89 was introduced as an in-house made dipeptide building block (\textbf{81}), and NMR analysis of the dipeptide indicated a d.r. of > 9:1 suggesting that no significant epimerization occurred. However, we cannot exclude that the epimerization occurred during the SPPS. Nevertheless, both isomers \textbf{74-A} and \textbf{74-B} were separated by preparative HPLC and used independently for the synthesis of the modified SUMO3 proteins \textbf{65-A} and \textbf{65-B}.

4.4.3.2. Synthesis of first generation modified SUMO3 proteins 60 and 61

From a protein synthesis perspective, we identified several questions we aimed to answer with the synthesis of the first generation of modified SUMO proteins:

- Can orthogonally protected \(\alpha\)-ketoacids such as \textbf{50} be employed in the N to C-terminal assembly of larger protein segments?
• Can the orthogonally protected α-ketoacid be deprotected in the setting of highly functionalized protein segments?
• Is the biotin affinity label stable to ligation conditions?

Biotin-[SUMO3 (2–29)]-α-ketoacid 67 and H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 reacted smoothly within 7 h under ligation conditions similar to those used in the synthesis of 57 (Figure 37). As in the model peptide studies (see chapter 3.3.2), the phenol-protected α-ketoacid was found to be stable to ligation conditions.

For synthetic efficiency, we employed an one-pot deprotection protocol for the (phenol-protected) α-ketoacid. The crude ligation mixture was treated with 2.5 % DODT in TFA for 2 h and Biotin-[SUMO3 (2–75)]-α-ketoacid depsi-69 was obtained in 23 % yield after purification by HPLC. To avoid epimerization of the C-terminal α-ketoacid under the basic rearrangement conditions, the ester-linkage at the ligation site was retained.

![Figure 37. KAHA ligation of 67 and 54 (top) and HPLC monitoring of KAHA ligation and in situ α-ketoacid deprotection (bottom).](image)

With Biotin-[SUMO3 (2–75)]-α-ketoacid depsi-69 in hand, we continued to the final ligations with the C-terminal segments 71 and 72 (Figure 38). The ligations proceeded uneventfully with both segments within 8 h. The crude ligation mixtures were subjected to
rearrangement conditions to convert the two esters at the ligation sites to the corresponding amides. The final proteins, Biotin-[SUMO3 (2–92)]-OH 60 and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 were obtained in 35 % and 32 % yield, respectively.

Figure 38. KAHA ligations of depsi-69 and 71 or 72 (top). HPLC monitoring of the assembly of 61 (bottom, left) and deconvoluted FTICR MS of 61 (bottom, right). Data for 60 can be found in the Experimental.

In summary, we synthesized two modified SUMO3 proteins 60 and 61 using an assembly in N to C direction, which was enabled by a new orthogonally protected α-ketoacid. While the overall synthetic strategy was successful, the purification of the (phenol-protected) α-ketoacid segment 54 was hampered by the presence of diastereomers arising from the racemic protecting group. The (phenol-protected) α-ketoacid was found to be stable under ligation conditions and could be selectively deprotected under acidic conditions in a one-pot procedure after ligation, keeping the number of necessary purification steps to a minimum. The biotin affinity label on the N-terminus of 60 and 61 was stable under all reaction conditions and no side reactions were observed.
4.4.3.3. Synthesis of second generation modified SUMO3 proteins 62–65

Following the same synthetic strategy as the first generation, the second generation proteins were assembled in N to C direction. The first KAHA ligation between Biotin-PEG-Dap(Cou)-[SUMO3 (2–29)]-α-ketoacid 68 and H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 followed by one-pot deprotection of the α-ketoacid using previously developed conditions proceeded uneventfully and gave Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70 in 38 % yield over two steps (Figure 39).

The second ligations between Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70 and the C-terminal variants 71 – 74 were performed in a parallel fashion and, after global O to N acyl shift, furnished the desired SUMO3 proteins 62 – 65 in 22–37 % isolated yield (Figure 40).
In conclusion, we synthesized five modified SUMO3 proteins varying in their C-terminal region in a streamlined and parallelized fashion using our previously designed N to C-terminal assembly route. Two newly introduced modifications at the N-terminus – a coumarin-labeled amino acid and a PEG-linker between the proteins and the biotin affinity label – were found to be well tolerated both by the chemical synthesis and the enzymatic SUMOylation machinery (see chapter 4.4.4.2).

4.4.3.4. Synthesis of third generation modified SUMO3 protein 66

For the third generation of modified SUMO proteins, an assembly in C to N direction was utilized. The first KAHA ligation between Fmoc-Opr-[SUMO3 (31–75)]- α-ketoacid 76 and H-Opr-[SUMO3 (77–92 Gln89Pro)-OH 72 gave Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH depsi-77 in 57 % yield (Figure 41).
The N-terminus of Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH depsi-77 was deprotected and the ester at the ligation site converted to an amide using the previously developed one-pot protocol, giving H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78 in 49 % yield over two steps (Figure 42).

The final ligation between Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]- α-ketoacid 75 and H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78 followed by a one-pot O to N acyl shift gave the final product Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 in 23 % yield (Figure 43).
Figure 43. KAHA ligation of 75 and 78 and one-pot O to N acyl shift

The identity and purity of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 was confirmed by analytical HPLC and ESI-QTOF MS analysis (Figure 44).

Figure 44. Analytical HPLC (left) and deconvoluted ESI-QTOF MS (right) of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66.

4.4.4. Biochemical evaluation of modified SUMO3 proteins

4.4.4.1. SUMOylation-deSUMOylation assays of first generation proteins 60 and 61

With the synthetic SUMO3 proteins 60 and 61 available, we investigated whether they would be conjugated to our model protein RanGAP1 and whether a SENP enzyme would deconjugate them.

The in vitro SUMOylation of RanGAP1 was performed in a similar fashion as detailed in chapter 4.3.5. To investigate the deconjugation reaction, different concentrations of a representative deSUMOylation enzyme, SENP2, were added to the crude SUMOylation
reactions and incubated at 37 °C for 1 h. Separation of the reactions by SDS-PAGE followed by Western blotting using an anti RanGAP1 antibody revealed that both synthetic SUMO3 proteins 60 and 61 were accepted by the SUMOylation enzymes to generate synthetic SUMO3-RanGAP1 conjugates (Figure 45, lane 2 and 7). The presence of unconjugated RanGAP1 in the SUMOylation with the SUMO Gln89Pro mutant 61 could be an indication that SUMOylation with SUMO3 Gln89Pro 61 is slower or less efficient compared to 60; however, this incomplete conversion was not observed in other experiments using similar proteins (see chapter 4.4.4.2). Analysis of the deSUMOylation reactions clearly showed that the SENP enzyme readily cleaved the RanGAP1-conjugate with synthetic SUMO3 containing the natural Gln89 residue (60), even at the lowest tested SENP concentration (lane 5). In contrast, the RanGAP1-conjugate with synthetic SUMO3 containing the Gln89Pro mutation (61) was found to be much more stable towards deconjugation and was only minimally cleaved at non-physiologically high concentrations of SENP enzyme (lane 9).

Figure 45. In vitro conjugation and deconjugation of synthetic SUMO3 proteins 60 and 61 to RanGAP1 (top) and Western blot analysis using anti RanGAP1 antibody (bottom). * = impurity from RanGAP1.
4.4.4.2. SUMOylation-deSUMOylation assays of second generation proteins 62-65

The synthetic SUMO3 proteins 62 - 65 were subjected to a SUMOylation–deSUMOylation assay similar as described in chapter 4.4.4.1. Separating the crude reactions by SDS-PAGE followed by Western blotting showed significant differences between the synthetic proteins 62 - 65 (Figure 46).

The Western blot (using an anti SUMO2/3 antibody) of the crude SUMOylation reactions indicated that the Gln89Pro mutant 63 (lane 2) and one of the two backbone N-methylated proteins (65-B, lane 5) were conjugated to RanGAP1 with similar efficiency as the natural Gln89 protein 62 (lane 1). The other two mutants, 64 and 65-A (lanes 3 and 4), were conjugated to RanGAP1 to a significantly smaller extent, suggesting that they are not very well recognized by the SUMOylation enzymes. However, it is noteworthy that the SUMOylation enzymes recognize to some extent the Gly91Ava mutant 64, even though it lacks the characteristic and highly conserved C-terminal Gly-Gly motif.

Analysis of the deSUMOylation reactions (performed with 1 µM SENP2 enzyme) by Western blot (using an anti RanGAP1 antibody) showed that no unconjugated RanGAP1 is present in case of the Gln89Pro mutant 63 (lane 7). This finding – in accordance with the previous results of 65 in chapter 4.4.3.2– suggests that proteins with a Gln89Pro mutation are conjugated to RanGAP1 very efficiently and that the conjugates are highly resistant towards cleavage. For all other proteins, unconjugated RanGAP1 is observed, but it cannot be distinguished if this is from incomplete conjugation or partial deconjugation. Qualitatively, for 65-B (lane 10) it appears like the conjugate would be somewhat more resistant towards hydrolysis compared to the natural Gln89 protein 62, however it cannot be concluded if this is due to the N-methylation or a potential epimerization.
In summary, the in vitro SUMOylation and deSUMOylation reactions further underlined the potential utility of modified SUMO3 proteins containing a Gln89Pro mutation such as Biotin-PEG-Dap(Cou)-[SUMO3 (2–92) Gln89Pro] \(63\) as irreversibly conjugated probed for protein SUMOylation. The other tested mutant proteins were either less efficiently conjugated \((64\text{ or } 65-A)\) or did not provide an advantage \((65-B)\) compared to \(63\).

4.4.5. Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH \(66\) as probe for cellular protein SUMOylation

With milligram quantities of the third generation modified SUMO3 protein \(66\) in hand, we investigated the SUMOylation of crude HEK293 lysates.

In vitro SUMOylation reactions were performed as before using recombinant RanGAP1 protein for the positive control or crude HEK293 cell lysates as substrates, followed by separation by SDS PAGE and in-gel fluorescence detection (Figure 47). A control experiment demonstrated that the cell lysate alone exhibited no fluorescence (lane 2), thus the fluorescent bands observed in the SUMOylation experiments were only coming from the labeled probe and its conjugates. As positive control (lane 3), the SUMOylation of RanGAP1 was used, showing that – as before – the SUMOylation enzymes readily accept the synthetic protein as substrate.

![Figure 47. In-gel fluorescence image of SDS PAGE of SUMOylation reactions using 66 and HEK293 cell lysates.](image)

In the SUMOylation reactions of crude cell lysates (containing different concentrations of protease inhibitors) as substrates (lanes 4–6), a multitude of high molecular weight bands corresponding to cellular proteins labeled with the SUMOylation probe were observed. This demonstrated that proteins from crude cell lysates can be readily SUMOylated with \(66\) and detected by in-gel fluorescence. The observed band patterns and intensity distributions were
very similar in all three samples, showing that the amount of protease inhibitor present in the lysis buffer does not affect the SUMOylation machinery.

In order to optimize the conditions of the in vitro SUMOylation reactions, we focused on evaluating the activity of the endogenous SUMOylation enzymes in cell lysates (Figure 48).

The control experiment – the in vitro SUMOylation of RanGAP1 (lanes 4–6) – showed that in the absence of cell lysate, both E1 and E2 enzyme (SAE1/SAE2 and UBC9, respectively) have to be added to observe protein SUMOylation. In contrast, SUMOylation experiments of crude cell lysates in the presence or the absence of exogenous E1 or E2 enzyme indicated that the endogenous E1 enzyme displayed high activity and the addition of recombinant E1 enzyme (lane 1 vs. 3) had no effect on protein SUMOylation. However, only very little protein SUMOylation was observed when no E2 enzyme was added (lane 2). One possible explanation could be that the endogenous E2 enzyme is either inactive or not abundant after cell lysis. Alternatively, the addition of E2 enzyme could result in changing the SUMOylation mechanism from E3-dependent at lower E2 concentrations to E3-independent at higher E2 concentrations, which could also result in an overall increased protein SUMOylation. As a consequence of these results, we decided to only add E2 enzyme to in vitro SUMOylation reactions of cell lysates.

Next, we wanted to evaluate potential applications of our in vitro SUMOylation assay and qualitatively investigated the influence of external stimuli such as heat shock or addition of E3 enzymes on protein SUMOylation (Figure 49). Incubating the cell cultures at elevated temperatures (43 °C vs. 37 °C for 30 min) prior to cell lysis resulted in an increased protein SUMOylation in the in vitro SUMOylation reaction (e.g. lanes 3 vs. 4 or 7 vs. 8), which reflects the well-known upregulation of the SUMOylation machinery as response to heat shock. This result indicated that our in vitro SUMOylation assay is in principle capable of detecting
modulations of the SUMOylation machinery that occurred prior to the cell lysis and \textit{in vitro} SUMOylation.

In a second experiment, we added two known SUMO E3 ligases, PIAS1 and RanBP2, to the SUMOylation reactions. Whereas the addition of PIAS1 had no obvious effect on protein SUMOylation (e.g. lanes 1 vs. 2), the addition of RanBP2 – which itself is being SUMOylated – enhanced the general protein SUMOylation significantly (e.g. lanes 1 vs. 3). These results indicated that our \textit{in vitro} SUMOylation could be used, for example, to investigate the SUMOylation targets of different E3 ligases.

In a proof of principle study, the next goal was to identify at least some of the proteins SUMOylated with our probe 66 using shotgun proteomics. To provide enough material for such an analysis, it was necessary to scale up the SUMOylation reactions and enrich the SUMOylated proteins (Figure 50). We optimized an affinity purification protocol using monomeric avidin immobilized on agarose beads and were able to enrich SUMOylated proteins from \textit{in vitro} SUMOylation reactions of HEK293 and HeLa lysates as illustrated in Figure 51 (e.g. lanes 1 and 2 vs. 7 and 8).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure49.png}
\caption{In-gel fluorescence image of SDS PAGE of \textit{in vitro} SUMOylation reactions of HEK293 or HeLa cell lysates and response to heat shock (left) prior to cell lysis or presence of E3 ligases during SUMOylation (right).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure50.png}
\caption{Workflow for enrichment of SUMOylated proteins and LC-MS/MS analysis.}
\end{figure}
CHAPTER 4: CHEMICAL SYNTHESIS OF SUMO3 PROTEINS

Figure 51. In-gel fluorescence image of SDS PAGE of in vitro SUMOylation reactions of HEK293 or HeLa cell lysates and affinity purification of SUMOylated proteins.

LC-MS/MS analysis of enriched SUMOylated proteins from a HEK293 cell lysate was performed by our collaborators, Nadine Sobotzki and Prof. Dr. Bernd Wollscheid, ETH Zurich. In preliminary results, 127 proteins could be identified that were enriched after SUMOylation and affinity purification, compared to a control experiment in which the SUMOylation probe was omitted (Figure 52, see Appendix for a comprehensive list of identified proteins).

Figure 52. Identification of proteins from HEK293 lysates labeled with 66.

Unfortunately, peptides containing the SUMOylation site and bearing the SUMO peptide fragment could not be identified, presumably due to the highly complicated fragmentation pattern of the relatively large SUMO-derived peptide (32 amino acid residues), a well-known problem in the proteomic analysis of SUMOylated proteins. In future studies, this could be improved by introducing an additional trypsin cleavage site (e.g. a Lys or Arg mutation) close to the C-terminus, creating smaller SUMO-tags that would give a simpler fragmentation pattern and would be easier to identify.

A comparison with proteins reported in the literature showed that at least 37 of the 127 identified proteins were known to be SUMOylated, including well known SUMOylation problems: SUMO-derived peptide is 32 AA residues large
targets such as RanGAP1,\textsuperscript{182} UBC9\textsuperscript{211} and EEF2,\textsuperscript{212} which further supports our belief that the enriched proteins are SUMOylated. However, the ultimate confirmation that a specific protein is SUMOylated would be the unambiguous identification of the SUMOylation site.

In summary, these results demonstrate that our synthetic modified SUMO3 protein\textsuperscript{66} can be used as probe for protein SUMOylation and used to enrich the SUMOylated proteins for downstream proteomic analysis.

4.5. Conclusions and outlook

In this chapter, the synthesis of several versions of SUMO3 proteins using different synthetic strategies based on the KAHA ligation was described. We showed that the homoserine mutations introduced by the KAHA ligation did not affect recognition by the enzymatic SUMOylation machinery, and the synthetic proteins were successfully conjugated to a model protein, RanGAP1, and numerous proteins in crude cell lysates. In order to develop an irreversibly conjugated version of SUMO3, we investigated several mutations in the C-terminal region of SUMO3 and found that a Gln89Pro mutation effectively prevents deconjugation of SUMO3, while not interfering with recognition by the SUMOylation enzymes required for conjugation.

As a proof of concept, we demonstrated that using our probe\textsuperscript{66}, proteins from cell lysates could be readily SUMOylated, visualized using in-gel fluorescence, enriched with an affinity purification step and identified using shotgun proteomics.

This technology platform could open avenues for further application in chemical biology. Besides the identification of SUMOylated proteins, the probe could be used in combination with easily introduced photoaffinity tags to identify new enzymes involved with the SUMOylation machinery – for example only one E2 enzyme and a very small number of E3 ligases and SUMO-specific proteases have been identified so far, and given the complexity of the enzymatic ubiquitination machinery it is likely that there are more to be found. Alternatively, the probe could be used to profile global protein SUMOylation or the SUMOylation of a specific protein target under different conditions, for example stress conditions, cancer or pathogen infections. Furthermore, it would be interesting to study protein SUMOylation in living cells. Such studies would allow investigating the crosstalk of SUMO with other posttranslational modifications in more detail and provide an opportunity to interrogate protein modification under E3-mediated SUMOylation conditions. Introduction of proteins into living cells is difficult, but not impossible. A possible solution could be to chemically conjugate the SUMOylation probe to different cell-penetrating peptides, which could facilitate the cellular uptake.\textsuperscript{213}
4.6. References


CHAPTER 5
CHEMICAL SYNTHESIS OF
IRISIN PROTEIN
CHAPTER 5: Chemical synthesis of irisin protein

5.1. Introductory remarks

In this chapter the application of the KAHA ligation in the chemical synthesis of irisin is described. Irisin is a protein hormone that has received significant attention in recent years due to its putative function in connecting physical exercise with fat metabolism, described in more detail in chapter 5.2. Chapter 5.3 focuses on the synthetic strategy, the synthesis of the protein segments and the assembly of two versions of the irisin protein, one unlabeled and one labeled with a rhodamine fluorescence dye.

5.2. Biological background

In 2012, irisin was described as a myokine found both in mice and humans, which is secreted from skeletal muscle upon physical exercise and stimulates brown-fat-like development in certain white fat tissues. Irisin is highly conserved between mice and humans. It consists of 112 amino acid residues and is proteolytically cleaved from the N-terminal region of a precursor protein, FNDC5 (fibronectin domain containing protein 5, a type 1 transmembrane protein), and is secreted into the extracellular space. Irisin was also found to be glycosylated; however, the exact glycosylation pattern or its function remained unclear.

While a receptor for irisin is yet to be identified, it has been shown in a mouse model that irisin has significant effects on the browning of specific white adipose tissues by stimulating the development of brite (brown-in-white, also termed beige) adipocytes (Figure 53). One proposed mechanism for this effect is the upregulation of UCP1 (uncoupling protein 1, thermogenin) expression, a protein involved in thermogenesis and characteristic for brown fat cells. The conversion of white fat into beige or brown fat enhances metabolic uncoupling and caloric expenditure and has been shown to reduce obesity and improve glucose homeostasis in vivo. It was suggested that irisin – at least partially – conveys the positive effects of physical exercise on health. Consequently, it was believed that irisin would possess a significant therapeutical potential for the treatment of obesity and other metabolism-related diseases.
However, since the initial discovery in 2012, the claimed effects and relevance of irisin in humans has been highly controversial. Several studies found contradictory correlations between physical exercise, the observed health benefits and measured irisin levels, disputing the role of irisin as the “human exercise gene”.  

The significance of many studies correlating irisin with health benefits was questioned as the reported irisin levels in human serum varied by several orders of magnitude, even in healthy subjects. In most published studies, the levels of irisin were determined using similar, commercially available ELISA assays. It was later demonstrated that at least four irisin antibodies used in these ELISA assays were not properly validated and cross-reacted with non-specific proteins in human serum, thus severely distorting the measured irisin levels and calling the drawn conclusions into question.

Additionally, the relevance of irisin in humans and its exact sequence was debated, as it was found that the methionine-encoding start codon in human FNDC5 is mutated from normal ATG – as found in all other investigated species – to ATA, which is not a usual start codon. Although there are examples of proteins produced from genes containing unusual start codons, model studies in transformed HEK293 cells showed that human FNDC5 with ATA as start codon resulted in extremely low protein levels, suggesting that humans – unlike mice – might have an effective gene knockout of FNDC5. Alternatively, it was speculated that the gene expression could start from a downstream ATG start codon (encoding for Met76), which would result in a truncated FNDC5 protein and a different constitution of irisin.

In contrast to these observations, irisin was recently unequivocally identified and quantified in human blood using quantitative mass spectrometry with isotope-labeled peptides used as internal standards, demonstrating that FNDC5 is indeed translated from the non-canonical ATA start codon and thus confirming its proposed primary structure. Furthermore, this study provided evidence that irisin plasma levels are elevated in individuals undergoing aerobic interval training.
In summary, the presence of irisin in human serum has been confirmed, but the exact role in human metabolism and the connection to physical exercise is not yet fully understood. Access to well-defined and homogenously modified irisin would allow additional investigations to be undertaken, potentially helping to shine light on its biochemical role and physiological function.

5.3. Synthesis of irisin protein by KAHA ligation

Given the remarkable interest regarding the biological function of irisin, especially shortly after its discovery, and the limited access to homogenous protein samples, we elected to synthesize irisin using our KAHA ligation technology.

The overall goal of this project was to establish a synthesis to access milligram quantities of irisin protein. We also decided to synthesize a protein version labeled with a rhodamine fluorescence dye for biochemical studies.

5.3.1. Retrosynthetic analysis of irisin and synthetic strategy

The first step towards the synthesis of irisin was an analysis of the amino acid sequence and the structure of irisin, summarized in Figure 54. The fold of irisin is dominated by a fibronectin type 3 protein domain consisting of a small globule formed by 7 \( \beta \)-strands.\(^{225}\)

![Figure 54. Primary, secondary and tertiary structure of irisin, derived from Uniprot entry QNAU1, irisin (1–112). Rendering is based on a X-ray structure of irisin (2–98).\(^{226}\)](image)

Irisin consists of 112 amino acid residues. In order to bring the size of the individual segments down to around 35 – 40 amino acid residues each, which we expected to be readily accessible by SPPS, it was clear that a two ligation-three segment strategy would be required. To minimize the potential effect of the homoserine mutations introduced by KAHA ligation on the protein structure, we selected ligation sites between Leu26–Glu27 and Leu64–Glu65, as they are both located in solvent exposed, flexible loop regions and not in well structured \( \beta \)-strands (Figure 55).
Figure 55. Disconnection of irisin into three segments (left). Residues mutated to Hse through KAHA ligation with 5-oxaproline highlighted in the protein structure (right).

With the ligation sites set, we decided to assemble both labeled and unlabeled irisin in C to N-terminal direction. This approach would allow us to use a common first ligation for both targets, followed by a second ligation with either a fluorescent-labeled or unlabeled N-terminal segment, which would improve overall efficiency.

The N-terminal 5-oxoproline residue in the bifunctional segment 2 (84) had to be orthogonally protected during the first ligation with segment 3 (85). Although we could have used the well-established, base-labile Fmoc-group for protection of the 5-oxoproline, we decided to employ a recently developed photolabile 1-NPEOC (1-(2-nitrophenyl)ethyl oxycarbony) protecting group, as this could be easily removed in the diluted ligation mixture. In Scheme 52, the synthetic strategy toward irisin is summarized.

Scheme 52. Synthetic strategy for synthetic irisin proteins 87 and 88.
5.3.2. Synthesis of protein segments

With a suitable synthetic strategy in hand, we proceeded with the solid phase synthesis of the four protein segments shown in Scheme 52. The segment syntheses are outlined below; details and characterization data can be found in the Experimental.

The N-terminal segment H-[Irisin (1–26)]-α-ketoacid 82 was readily synthesized by automated SPPS using protected α-ketoacid resin 41 and double couplings at every step (Scheme 53). H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83 was synthesized from the same resin via manually coupling Fmoc-Dap(Rho)-OH 89 using COMU and NMM, followed by Fmoc deprotection and resin cleavage.

The bifunctional middle segment 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 was synthesized by automated SPPS up to Asp28 using protected α-ketoacid resin 41 and double couplings at every step (Scheme 54). 1-NPEOC-Opr-OH 90 was manually coupled using HATU and NMM. Although a mixture of diastereomers of 90 – arising from the racemic 1-NPEOC group – was used, a single peak of 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 was observed under all investigated HPLC conditions.
CHAPTER 5: CHEMICAL SYNTHESIS OF IRISIN PROTEIN

Scheme 54. Synthesis of the middle irisin segment 84.

The C-terminal segment H-Opr-[Irisin (66–112)]-OH 85 was readily synthesized by SPPS on 2-chlorotrityl polystyrene resin using double couplings at every step (Scheme 55).

Scheme 55. Synthesis of C-terminal irisin segment 85.

5.3.3. Assembly of irisin proteins by sequential KAHA ligations

With all required protein segments ready, we proceeded to assemble the two irisin proteins by sequential KAHA ligations. In the first ligation, 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 and H-Opr-[Irisin (66–112)]-OH 85 were reacted at a protein segment concentration of 16 mM in a mixture of 9:1 NMP:H₂O with 0.1 M oxalic acid at 60 ºC (Figure 56). Upon dissolution, the protein segments instantly formed a gel, which may have affected the mixing efficiency and achievable conversion of the reaction. Unfortunately, gel formation was also observed under more dilute conditions (10 mM protein concentration). After 7 h, the mixture was diluted with 1:1 H₂O:CH₃CN + 0.1 % TFA and irradiated with a regular laboratory bench top UV lamp at a wavelength of 365 nm for 30 min to remove the 1-NPEOC protecting group. After purification by preparative HPLC, H-Opr-[Irisin (28–112)]-OH deps-86 was obtained in 18 % yield over the ligation and deprotection steps. In order to increase the overall efficiency of the synthesis and potentially improve the solubility of 86, we decided to proceed with the ester linkage at the ligation site and perform a global O to N acyl shift after the second ligation.
With H-Opr-[Irisin (28-112)]-OH despsi-86 in hand, we proceeded to the final ligation steps with the N-terminal segments 82 and 83 (Figure 57). Unfortunately, gel formation, together with incomplete solvation of starting materials was observed in both ligations, even under more dilute conditions (11 mM). Accounting for this and to push the reactions to maximum conversion, the ligations were allowed to react for 18 h at 60 °C. The crude ligation mixtures were subsequently subjected to rearrangement conditions to convert the two esters at the ligation sites to the corresponding amides. The final irisin proteins, H-[Irisin (1–112)]-OH 87 and H-Dap(Rho)-[Irisin (1–112)]-OH 88, were obtained after HPLC purification in 17 % and 15 % yield, respectively.
CHAPTER 5: CHEMICAL SYNTHESIS OF IRISIN PROTEIN

Figure 57. KAHA ligations of 86 and 82 or 83 (top). HPLC monitoring of the assembly of 87 (bottom, left) and deconvoluted FTICR MS of 87 (bottom, right). Data of 88 can be found in the Experimental.

The purity and identity of 87 and 88 were confirmed by analytical HPLC and FTICR MS. CD-spectra indicated the presence of β-sheets in both proteins, which would be in accordance with the structure reported in the literature (Figure 58).

Figure 58. CD-spectra of 87 (left) and 88 (right). 10 µM protein concentration in 20 % TFE/10 mM phosphate buffer (pH 6.0) at 25 ºC.

In a preliminary study in collaboration with the Wolfrum Group at ETH, the binding of the synthetic irisin proteins to inguinal white adipose tissue derived stromal-vascular fraction (SVF) cells was evaluated. SVF contains a variety of different cell types, including endothelial, nerve and precursor cells. Analysis by fluorescence-activated cell sorting (FACS) showed that the synthetic irisin proteins bind specifically to a subset of cells in SVF, which
could be an indication for the presence of the – yet unidentified – irisin receptor on these cells.\textsuperscript{228} These promising results suggest that the synthetic irisin proteins could be suitable tools for further investigations into the biological function of irisin.

5.4. Conclusions and outlook

We synthesized milligram quantities of labeled and unlabeled irisin protein \textsuperscript{87} and \textsuperscript{88} using two sequential KAHA ligations. The conversions and isolated yields of both ligations were unusually low, presumably due to limited mixing and solubility during the reaction caused by gel formation. The extent of gel formation could potentially be reduced by introducing a solubility tag\textsuperscript{229} to increase the hydrophilicity of the protein or by introducing isoacyl-peptides\textsuperscript{230} to temporarily break the secondary structure of the protein during synthesis.

In a preliminary cell binding study, the synthetic proteins were found to bind to a subset of cells derived from white adipose tissue, suggesting that these cells might possess a specific receptor for irisin. The synthetic route established could provide access to additionally modified irisin proteins which could be used as probes to identify the irisin receptor, for example by incorporating photoaffinity probes such as benzophenones or diazirines in the protein.\textsuperscript{231}
5.5. References


Unpublished results in collaboration with Sebastian Müller, ETH Zurich, who performed the cell binding assay. Manuscript in preparation.


CHAPTER 6
SUMMARY AND OUTLOOK
CHAPTER 6: Summary and outlook

The work described in this dissertation covers both technical improvements of the KAHA ligation and its application in the synthesis of two important proteins.

Formation of depsipeptides in the KAHA ligation with 5-oxaproline

It was demonstrated that the primary products of KAHA ligations with 5-oxaproline are ester-containing depsipeptides, rather than the anticipated amides. Mild conditions to convert the depsipeptides and depsiproteins into the natural amides were developed, making possible the streamlined synthesis of proteins by sequential KAHA ligations with 5-oxaproline. It should be noted that due to their increased polarity, depsipeptides often improve solubility and handling of unfolded synthetic proteins, which can be very useful for the synthesis of hydrophobic proteins.

New protecting groups for the synthesis of peptide α-ketoacids

Two new protecting groups for α-ketoacids were developed, which are fully stable to Fmoc-SPPS conditions and enable the inclusion of all canonical amino acids in the peptide segments, including cysteine, methionine and tryptophan. The acid-labile protecting group 29 provides peptide α-ketoacids directly upon standard acidic resin cleavage without the need for further manipulation. The switchable, orthogonal protecting group 43 gives protected peptide α-ketoacids upon resin cleavage. These protected peptide α-ketoacids are stable under KAHA ligation conditions and can be selectively activated and deprotected on demand, which facilitates bidirectional and convergent segment ligation strategies. Both protecting groups expedite the synthesis of peptide α-ketoacids and broaden the scope of accessible proteins by KAHA ligation. For future synthesis, additional orthogonal protecting groups for α-ketoacids (e.g. photolabile protecting groups) 232 would enable one-pot synthesis strategies for the efficient assembly of larger proteins.

Future directions in KAHA ligation technology

From the perspective of protein synthesis technology, a major challenge is to develop new reaction partners that provide a faster ligation reaction. 233 This would allow for more dilute reaction conditions at ambient temperature, addressing problems arising from the often limited solubility of large protein segments and their tendency to form gels. However, it should be noted that the current KAHA ligation technology is already well suited especially for the synthesis of hydrophobic proteins or proteins with low solubility, as the acidic reaction conditions often provide better solubilization than reactions performed at neutral pH, even under denaturing conditions. Yet, with a faster version of the KAHA ligation in hand, the synthesis of very challenging protein targets, such as highly hydrophobic membrane proteins or proteins larger than 250 amino acid residues could be facilitated. One solution could be
the recently disclosed amide bond-forming reaction between potassium trifluoroborates (KATs) and substituted hydroxylamines, which exhibits reaction rates that are orders of magnitude faster than the KAHA ligation. Although this reaction was employed so far only for side-chain modifications of proteins, it is a promising candidate for the development of monomers suited for protein synthesis.

Furthermore, it is of great interest to develop more reaction partners for KAHA ligation that give natural amino acid residues at the ligation site (such as the recently reported serine-forming oxazetidine), which would reduce the number of mutations introduced into the synthetic protein. Alternatively, monomers yielding unique functional at the ligation site (e.g. the aldehyde-forming 4-ethoxy-5-oxaprole) are desirable, as they provide new opportunities to introduce chemical handles for subsequent site-specific modification of proteins.

Taken together, these technical improvements will help to further establish the KAHA ligation as a premier technique for protein synthesis.

**Chemical synthesis of SUMO3 proteins**

Sequential KAHA ligations and protected α-ketoacids were used to synthesize milligram quantities of the important modifier protein SUMO3. Despite containing two homoserine mutations, the synthetic protein was readily recognized by the enzymatic SUMOylation machinery and conjugated to target proteins. Based on this synthetic protein, a chemical probe for cellular protein SUMOylation was developed. The probe was functionalized with a rhodamine fluorescence dye and a biotin tag for affinity purification. By incorporating a mutation in the C-terminal region of the protein, the probe overcomes the highly dynamic and reversible nature of protein SUMOylation, which is often a problem in the analysis of cellular protein SUMOylation. The probe was applied in *in vitro* SUMOylation reactions using crude cell lysates as substrates and SUMOylated proteins were readily enriched and identified using MS-based proteomics.

This technology platform could find numerous applications in chemical biology. For example, the probe could be used to profile the global cellular protein SUMOylation or the SUMOylation of a specific protein target in specific cell types (e.g. cancer cells) or in cells grown under well-defined conditions, such as cellular stress or infection by pathogens such as bacteria or viruses. Alternatively, the probe could be further functionalized with easily introduced photoaffinity tags to identify new members of the – so far surprisingly small – enzymatic SUMOylation machinery. Especially the identification of specialized SUMO E3 ligases might be of great interest, as misregulated protein SUMOylation is associated with several diseases. Ultimately, if would be interesting to study protein SUMOylation in living cells. Such studies would allow for investigating the crosstalk of SUMO with other posttranslational modifications in more detail and provide an opportunity to interrogate
protein modification under E3-mediated SUMOylation conditions. A major challenge would be the introduction of the SUMO probe into living cells. A possible solution could be to chemically conjugate the SUMOylation probe to different cell-penetrating peptides, which could facilitate the cellular uptake of the protein.\(^{213}\)

**Chemical synthesis of irisin protein**

The KAHA ligation technology was utilized to synthesize irisin, a myokine that is speculated to play an important role in the regulation of human fat metabolism. Although milligram quantities of unlabeled and fluorescence labeled irisin proteins \(^{87}\) and \(^{88}\) were obtained, the conversions and isolated yields of the ligations were unusually low, presumably due to limited mixing and solubility during the reaction caused by gel formation. Yet, it is remarkable that the KAHA ligation succeeded to provide the desired proteins under these unfavourable conditions. In future syntheses, the extent of gel formation could be reduced by incorporating solubilizing tags to increase the hydrophilicity of the protein segments or by implementing isoacyl-dipeptides that temporarily break the secondary structure of the protein during the synthesis.

A cell-binding study demonstrated that the synthetic irisin protein could specifically bind to a subset of cells in the SVF of inguinal adipose tissue, which is an indication for the presence of a specific irisin receptor. For future studies, the established synthetic route to modified irisin proteins could be utilized to produce precisely modified probes containing photoaffinity tags (e.g. benzophenones or diazirines), targeting the identification of the yet unknown receptor for irisin. This strategy could be of great use to elucidate the still controversial role of irisin in mouse and human physiology.
6.1. References


CHAPTER 7

EXPERIMENTAL
CHAPTER 7: Experimental Part

7.1. General Methods

Reagents and solvents

Fmoc-amino acids with suitable side chain protecting groups, HCTU and HATU were purchased from Peptides International (Louisville, KY, USA) and ChemImpex (Wood Dale, IL, USA). Dipeptide building blocks were either purchased as in case of Fmoc-Asp(OtBu)-N(Dmb)-Gly-OH from Bachem (Bubendorf, Switzerland) or synthesized according to modified literature procedures as for the pseudoproline dipeptides. COMU was obtained from Luxembourg Bio Technologies (Rehovot, Israel). Solvents for column chromatography (hexane, EtOAc, MeOH, MTBE, MeOH, CH₂Cl₂) were of technical grade and distilled prior to use. HPLC grade CH₃CN from Sigma-Aldrich was used for analytical and preparative HPLC purification. DMF (> 99.8%) or NMP (> 99%) from Sigma-Aldrich were directly used without further purification for solid phase peptide synthesis. Other commercially available reagents and solvents were purchased from Sigma-Aldrich (Buchs, Switzerland), Acros Organics (Geel, Belgium), ACR (Karlsruhe, Germany) and TCI Europe (Zwijndrecht, Belgium). 1-NPEOC-², Boc-³ and Fmoc-Opr-OH⁴ were prepared as previously reported by our group.

Characterization

¹H and ¹³C NMR spectra were recorded on Bruker DRX400, Bruker AVIII400 or Bruker AVIII600 spectrometers. Chemical shifts for ¹H NMR (400 or 600 MHz) and ¹³C NMR (101 or 150 MHz) are expressed in parts per million and are referred to residual undeuterated solvent signals. Coupling constants are reported in Hertz (Hz) and corresponding splitting patterns are indicated as follows: s, singlet; d, doublet; dd, doublet of doublet; td, triplet of doublet; t, triplet; m, multiplet. High-resolution mass spectra were recorded by the mass spectrometry service of the Laboratory of Organic Chemistry at ETH Zurich either with a Bruker maXis instrument (ESI-QTOF MS) equipped with an ESI source and a Qq-TOF detector or with a Bruker solariX instrument (ESI or MALDI FTICR MS) with a FTICR detector. 4-hydroxy-α-cyanocinnamic acid was used as matrix for MALDI analysis. MS-MS analysis was performed by the Functional Genomics Center Zurich, Switzerland. Low

² Unpublished results of Dr. F. Thuaud, manuscript in preparation.
resolution mass spectra were recorded on a Bruker Microflex MALDI (with HCCA as matrix), a Surveyor MSQ Plus Spectrometer coupled to a Thermo Fisher/Dionex Ultimate 3000 LC or on a Thermo Scientific™ Exactive™ Plus Orbitrap Mass Spectrometer coupled to a LC auto sampler. Infrared spectra were measured on a JASCO FT:IR-4100 spectrophotometer and are reported as wavenumber (cm⁻¹) of the absorption maxima (w = weak, m = medium, s = strong) for the range between 4000 cm⁻¹ and 1000 cm⁻¹. Optical rotations were measured on a JASCO P-2000 instrument at ambient temperature operating at the sodium D-line (λ = 589.3 nm) in a cell with 100 mm path length. CD-spectra were recorded on a Aviv Model 430 instrument.

Reactions and purification
All reactions utilizing air- or moisture sensitive reagents were performed in flame-dried glassware under an atmosphere of nitrogen. Reactions and fractions from column chromatography were monitored by thin layer chromatography using pre-coated glass plates (Merck, silica 60 F254) and visualized by fluorescence quenching at 254 nm, by staining with potassium permanganate or cerium ammonium molybdate. Column chromatography was performed on Silicycle SiO₂ type F60 (230-400 mesh) using a forced flow of air at 0.5-1.0 bar.

HPLC: Peptides and protein fragments were analyzed and purified by reversed phase high performance liquid chromatography (RP-HPLC) on Jasco analytical and preparative instruments equipped with dual pumps, mixer and in-line degasser, a variable wavelength UV detector (simultaneous monitoring of the eluent at 220 nm, 254 nm and 301 nm) and a Rheodyne injector fitted with a 20 or 1000 µl injection loop or on a Gilson preparative instrument. If required, the columns were heated using an Alltech column heater or a water bath (preparative HPLC). The mobile phase for RP-HPLC were Milipore-H₂O containing 0.1 % (v/v) TFA and HPLC grade CH₃CN containing 0.1 % (v/v) TFA. In the described HPLC analysis and purifications, TFA was always used as solvent modifier unless explicitly noted. Analytical HPLC was performed on a Shiseido Capcell Pak C18 UG120 (5 µm, 120 Å pore size, 4.6 mm I.D. x 250 mm) column, on a Shiseido Capcell Pak C18 UG 80 (5 µm, 120 Å pore size, 4.6 mm I.D. x 250 mm) column or on a Shiseido MGII C18 column (5 µm, 4.6 mm I.D. x 250 mm) columns at a flow rate of 1 ml/min. Preparative HPLC was performed on a Shiseido Capcell Pak MGII (5 µm, 100 Å pore size, 20 mm I.D. x 250 mm) column, on Shiseido Capcell Pak C4 or C18 columns (5 µm, 80 Å pore size, 50 x 250 mm) or on a Phenomenex Jupiter C4 column (5 µm, 300 Å pore size, 30 mm I.D. x 250 mm) at indicated flow rates (typically 10, 20 or 40 ml/min, unless otherwise noted).

The following type of solvent gradient were used: The column was pre-equilibrated at 5 % CH₃CN for typically 10 min. After injection of the sample, the solvent composition was held at
5 % CH₃CN for 5 min (10 min for preparative HPLC) and then ramped up within 5 min to the starting solvent composition of the gradient (in the example depicted below 30 % CH₃CN). Within the gradient run time, a linear solvent gradient was run to the final solvent composition (e.g. 50 % CH₃CN). After the gradient run time, the solvent composition was changed to 95 % CH₃CN within 1 min and the column was flushed for 7 min. Within 1 min, the solvent composition was changed back to 5 % CH₃CN and the run ended. For the sake of simplicity, only the gradient time and the starting and end composition of the eluent will be stated at the individual experiments, although all experiments included the full cycle as described above.

![General solvent gradient used in this work. In this example, the gradient would be termed: 30 to 50 % CH₃CN in 20 min.]

**Solid phase peptide synthesis**

Peptides were synthesized on a Multisyntech Syro I parallel synthesizer or CS Bio 136X synthesizers using Fmoc SPPS chemistry. The following Fmoc amino acids with side-chain protection groups were used: Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-His(1-Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH. SPPS was performed on Rink-amide polystyrene resin, Wang-polystyrene resin or 2-chlorotrityl polystyrene resin.

Manual loading of the first amino acid residue on the resin and subsequent Fmoc-SPPS followed established standard protocols.⁵,⁶ A brief summary of the utilized synthesis protocols: Fmoc deprotections were performed with 20 % piperidine in DMF (2 x 6 min). Couplings were performed with Fmoc amino acid (4.0 equiv relative to resin substitution), HCTU (3.9 equiv) and NMM (8.0 equiv) in DMF for 30 min. If required, the coupling step was repeated once (double coupling) and LiCl washes (0.8 M LiCl in DMF) were performed.

---


before Fmoc deprotection and coupling. After coupling, unreacted free amine was capped by treatment with 20 % acetic anhydride and 10 % NMM (v/v) in DMF for 2 x 5 min.

Amino acid residues prone to epimerization such as cysteine and histidine were coupled using preformed 6-Cl-HOBt esters. In a typical procedure, Fmoc-Cys(Trt)-OH or Fmoc-His(Trt)-OH (5 equiv relative to resin loading) were dissolved in a minimal amount of CH₂Cl₂ and 6-Cl-HOBt (5.5 equiv) and DIC (2.5 equiv) were added. The mixture was stirred for 1 h at room temperature, the solvent removed under reduced pressure and the residue dissolved in a minimal amount of DMF was added to the resin and was allowed to react for 1-2 h.

**Manual coupling of special amino acids**

Valuable non-standard monomers (e.g. protected 5-oxaproline) were coupled manually. The monomer (2.5 equiv) was dissolved in a minimal amount of DMF or NMP (minimal concentration of monomer: 0.1 M) and COMU or HATU (2.48 equiv) and NMM (5 equiv) were added. After a brief period of preactivation (1 min), the solution was added to the resin and allowed to react for 2-4 h.

**Resin cleavage procedures**

**Method A**: General cleavage protocol for peptide segments synthesized on Rink-amide polystyrene resin, Wang-polystyrene resin or 2-chlorotrityl polystyrene resin. The dry resin was placed in a glass vial, a mixture of 95:5 TFA:TIPS (15 ml/g resin) was added and the suspension shaked for 2 h. The resin was removed by filtration and washed with TFA (5 ml/g resin), the filtrate was placed in a plastic centrifugal tube (40 ml) and volatiles removed under reduced pressure. The residue was triturated with Et₂O (ca. 15 ml/g resin), centrifuged (2500 x g, 4 min) and the supernatant was removed by decantation. This trituration/washing step was repeated once. The crude material was dried and dissolved in a suitable solvent (DMSO or 1:1 H₂O:MeCN + 0.1 % TFA or mixtures thereof) for RP-HPLC purification.

**Method B**: Cleavage protocol for peptide α-ketoacid segments synthesized on α-ketoacid resins. The dry resin was placed in a glass vial, a mixture of 2.5 % DODT in TFA (15 ml/g resin) was added and the suspension shaked for 2 h. The resin was removed by filtration and washed with TFA (5 ml/g resin), the filtrate was placed in a plastic centrifugal tube (40 ml) and volatiles removed under reduced pressure. The residue was triturated with Et₂O (ca. 15 ml/g resin), centrifuged (2500 x g, 4 min) and the supernatant was removed by decantation. This trituration/washing step was repeated once. The crude material was dried and dissolved in a suitable solvent (DMSO or 1:1 H₂O:MeCN + 0.1 % TFA or mixtures thereof) for RP-HPLC purification.
7.2. Experimental for “CHAPTER 2: Formation of depsipeptides in the KAHA ligation with 5-oxaproline”

7.2.1. Preparation of starting materials

**Fmoc-Ala-Leu-α-ketoacid 4**

Ketoacid 4 was prepared from protected leucine α-ketoacid resin 41 at a 0.10 mmol scale (0.29 g resin with a substitution capacity of 0.34 mmol/g) as described in the General Methods. The resin was cleaved according to Method B and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 20 to 90 % CH3CN in 40 min, flow rate 40 ml/min) to give 4 (37.1 mg, 82.1 µmol, 82 %) as a white solid.

**1H NMR** (400 MHz, d6-DMSO) δ = 8.32 (d, J=6.9, 1H), 7.96 – 7.86 (m, 2H), 7.78 – 7.64 (m, 2H), 7.52 (d, J=7.7, 1H), 7.47 – 7.37 (m, 2H), 7.37 – 7.28 (m, 2H), 4.80 – 4.69 (m, 1H), 4.31 – 4.14 (m, 3H), 4.14 – 4.04 (m, 1H), 1.75 – 1.58 (m, 1H), 1.56 – 1.41 (m, 2H), 1.21 (d, J=7.1, 3H), 0.89 (d, J=6.6, 3H), 0.86 (d, J=6.5, 3H).

**13C NMR** (101 MHz, d6-DMSO) δ = 195.64, 172.93, 163.13, 155.59, 143.91, 143.76, 140.69, 127.61, 127.05, 125.30, 120.09, 65.60, 53.47, 49.49, 46.63, 37.88, 24.32, 22.95, 21.19, 18.13.


**H-Opr-(4F)Phe-OH 5**

The dipeptide 5 was synthesized on 2-chlorotrityl resin at a 0.20 mmol scale (0.30 g resin with a substitution capacity of 1.5 mmol/g) as described in the General Methods and cleaved from the resin following Method A. The residue was dissolved in a mixture of 1:1 water and acetonitrile containing 0.1 % TFA and directly purified by HPLC (Shiseido Capcell Pak MGII (20 x 250 mm), heated to 60 °C, 20 to 60 % CH3CN in 20 min, flow rate 10 ml/min). Product containing fractions were pooled together and lyophilized to give 5 (27.6 mg, 55.2 µmol, 28 %) as white solid.

**1H NMR** (400 MHz, d6-DMSO) δ = 8.17 (d, J=8.3, 1H), 7.25 – 7.16 (m, 2H), 7.13 – 6.96 (m, 2H), 4.48 (td, J=8.3, 4.9, 1H), 3.99 (dd, J=9.4, 4.5, 1H), 3.90 – 3.76 (m, 1H), 3.71 – 3.58 (m, 1H), 3.08 (dd, J=13.9, 4.9, 1H), 2.94 (dd, J=13.8, 8.4, 1H), 2.47 – 2.36 (m, 1H), 2.17 – 2.05 (m, 1H).

**13C NMR** (101 MHz, d6-DMSO) δ = 172.29, 170.12, 161.10 (d, J=241.9), 133.18 (d, J=3.0), 131.09 (d, J=8.0), 114.81 (d, J=21.1), 69.15, 60.32, 52.94, 35.82, 33.55.

**HR-MS** (ESI): calculated for C13H16FN2O4 [M+H]+: 283.1089, found: 283.1092.
Diastereomers of Fmoc-Ala-Leu-Hse-(4-F)Phe-OH 6

Three diastereomers of Fmoc-Ala-Leu-Hse-(4-F)Phe-OH 6, Fmoc-Ala-(S)-Leu-(S)-Hse-(4-F)Phe-OH 6-A, Fmoc-Ala-(R)-Leu-(S)-Hse-(4-F)Phe-OH 6-B and Fmoc-Ala-(S)-Leu-(R)-Hse-(4-F)Phe-OH 6-C were synthesized by SPPS on 2-chlorotrityl resin employing commercially available enantiopure Fmoc-(R)-Leu-OH and Fmoc-(S or R)-Hse(OTrt)-OH. The peptides were cleaved from the resins with a mixture of 5 % TFA and 1 % triisopropyl silane in CH₂Cl₂ for 30 min. The resin was removed by filtration and the solution concentrated under reduced pressure. The crude peptides were precipitated with an ice-cold mixture of 1:1 pentane:Et₂O and the solvent removed after centrifugation by decantation. The crude materials were dissolved in a mixture of 1:1 H₂O:CH₃CN containing 0.1 % TFA and directly purified by HPLC (Shiseido Capcell Pak MGII (20 x 250 mm), heated to 60 °C, 40 to 85 % CH₃CN in 20 min, flow rate 10 ml/min). Product containing fractions were combined and lyophilized to give the products as white solids.

To allow easy comparison with the obtained ligation products of 4 and 5, the characterization data of 6-A, 6-B and 6-C is shown on page 140.

Synthesis of Fmoc-Leu-α-ketoacid 7

(9H-fluoren-9-yl)methyl (S)-(1-cyano-5-methyl-2-oxo-1-(tetrahydro-1H-4-thiophen-1-ylidene)hexan-3-yl)carbamate (Fmoc-(S)-Leu cyanosulfurylide) S1

Fmoc-(S)-leucine (8.84 g, 25.0 mmol, 1.00 equiv) and 1-cyanomethyltetrahydro-thiophenonium bromide (6.50 g, 31.3 mmol, 1.25 equiv) were dissolved in dry DMF (100 mL) and cooled to 0 °C. DIPEA (9.69 g, 13.1 mL, 75.0 mmol, 3.00 equiv) and T3P® (50% in EtOAc, 18.9 mL, 31.3 mmol, 1.25 equiv) were added and the mixture stirred for 45 min at 0°C, after which LC-MS analysis showed complete consumption of the starting material. The reaction mixture was diluted with MTBE (200 ml) and H₂O (200 ml). The phases were separated and the organic phase washed with H₂O, saturated aqueous NaHCO₃ and brine, dried over anhydrous Na₂SO₄, filtrated and concentrated under reduced pressure to give S1 as a off-white solid (10.9 g, 23.6 mmol, 95 %) which was used in the next step without further purification.

¹H NMR (400 MHz, CDCl₃) δ = 7.76 (d, J=7.5, 2H), 7.65 – 7.58 (m, 2H), 7.42 – 7.35 (m, 2H), 7.35 – 7.28 (m, 2H), 5.44 (d, J=8.8, 1H), 4.72 (td, J=9.4, 3.9, 1H), 4.42 – 4.29 (m, 2H), 4.22
(t, J=7.2, 1H), 3.45 – 3.27 (m, 4H), 2.67 – 2.52 (m, 2H), 2.15 – 2.00 (m, 2H), 1.81 – 1.65 (m, 1H), 1.59 – 1.42 (m, 2H), 1.03 (d, J=6.5, 3H), 0.96 (d, J=6.6, 3H).

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ = 190.62, 156.07, 144.27, 144.08, 141.40, 127.74, 127.16, 125.38, 120.05, 119.32, 66.95, 54.93, 47.41, 45.03, 43.38, 28.69, 28.66, 25.08, 23.64, 21.79.

$[\alpha]_{D}^{26} = +36.8^\circ$ (c = 1.0, CHCl$_3$).

IR (thin film, cm$^{-1}$): $\tilde{\nu}$ = 3282 (w), 2954 (m), 2169 (s), 1714 (s), 1590 (s), 1247 (s).

MS (ESI) calculated for C$_{27}$H$_{31}$N$_2$O$_3$S$_1$ [M+H]$^+$: 463.2050, found: 463.2052.

$(S)$-3-(((9H-fluoren-9-yl)methoxy) carbonyl)amino)-5-methyl-2-oxohexanoic acid (Fmoc-(S)-Leu-\(\alpha\)-ketoacid) 7

To a solution of Fmoc-(S)-Leu cyanosulfurylide S1 (7.00 g, 15.1 mmol, 1.00 equiv) in THF (200 mL) and H$_2$O (200 mL), Oxone (Acros, min. 4.5 % active oxygen, 18.6 g, 30.3 mmol, 2.00 equiv) was added in one portion. The biphasic reaction mixture was vigorously stirred for 1 h at room temperature. After completion (as judged by LC-MS) the reaction mixture was quenched by addition of Me$_2$S (1.86 g, 30.3 mmol, 2.00 equiv). The layers were separated and the aqueous phase was extracted with MTBE. The combined organic phases were washed with an aqueous solution of citric acid (10% w/v), H$_2$O, brine and dried over anhydrous Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure to afford the product 7 (5.20 g, 13.6 mmol, 90 %) as a white solid, which was used without further purification in the next step.

$^{1}$H NMR (400 MHz, d$_6$-DMSO) $\delta$ = 7.95 – 7.87 (m, 2H), 7.71 (s, 2H), 7.45 – 7.37 (m, 2H), 7.36 – 7.29 (m, 2H), 4.71 – 4.57 (m, 1H), 4.33 – 4.28 (m, 2H), 4.26 – 4.19 (m, 1H), 1.76 – 1.58 (m, 1H), 1.53 – 1.43 (m, 2H), 0.9 (d, J=7.8, 3H), 0.88 (d, J=7.8, 3H).

$^{13}$C NMR (101 MHz, d$_6$-DMSO) $\delta$ = 196.05, 163.30, 156.09, 143.67, 140.74, 127.65, 127.08, 127.05, 125.21, 120.14, 65.71, 55.28, 46.66, 37.45, 24.40, 23.04, 20.92.

$[\alpha]_{D}^{26} = +22.8^\circ$ (c = 1.0, CHCl$_3$).

IR (thin film, cm$^{-1}$): $\tilde{\nu}$ = 3350 (m), 3066 (w), 2958 (m), 1724 (s), 1710 (s), 1528 (m), 1447 (m), 1334 (m), 1260 (s).

MS (ESI): calculated for C$_{22}$H$_{22}$N$_1$O$_5$ [M-H]$^-$: 380.1503, found: 380.1497.
H-Trp-Thr-Leu-α-ketoacid 20

H-Trp-Thr-Leu-α-ketoacid 20 was prepared on protected leucine α-ketoacid resin 40 at a 0.5 mmol scale (using an excess of 1.0 g resin with a substitution capacity of 1.5 mmol/g). The resin was cleaved according to Method B and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 25 to 60 % CH3CN in 30 min, flow rate 40 ml/min) to give 20 (0.15 g, 0.34 mmol, 67 %) as white solid.

\[ ^1H \text{NMR} \ (400 \text{ MHz, d}_6\text{-DMSO}) \delta = 11.00 \ (d, J=2.5, 1H), 8.74 \ (d, J=8.1, 1H), 8.33 \ (d, J=7.2, 1H), 7.97 \ (s, 4H), 7.78 – 7.72 \ (m, 1H), 7.39 – 7.33 \ (m, 1H), 7.25 – 7.17 \ (m, 1H), 7.12 – 7.06 \ (m, 1H), 7.05 – 6.93 \ (m, 1H), 4.86 \ (dd, J=9.9, 7.2, 4.4, 1H), 4.30 \ (dd, J=8.1, 5.3, 1H), 4.21 – 4.11 \ (m, 1H), 4.01 – 3.91 \ (m, 1H), 3.26 \ (dd, J=14.8, 4.8, 1H), 3.02 \ (dd, J=14.9, 8.8, 1H), 1.75 – 1.64 \ (m, 1H), 1.60 – 1.42 \ (m, 2H), 1.12 \ (d, J=6.2, 3H), 0.91 \ (d, J=6.5, 3H), 0.88 \ (d, J=6.5, 3H). \]

\[ ^{13}C \text{NMR} \ (101 \text{ MHz, d}_6\text{-DMSO}) \delta = 196.44, 169.76, 168.73, 163.71, 136.37, 127.11, 125.15, 121.12, 118.68, 118.32, 111.45, 106.85, 66.75, 58.68, 53.50, 52.42, 38.26, 27.47, 24.42, 23.05, 21.25, 19.80. \]

HR-MS (ESI): calculated for C_{22}H_{31}N_{4}O_{6} [M+H]^+: 447.2238, found 447.2236.

di-tert-butyl isoxazolidine-2,5-dicarboxylate S2

Following a literature procedure for similar compounds.\(^7\) tert-Butyl N-hydroxycarbamate (0.366 g, 2.75 mmol, 1.10 equiv) and tert-butyl 2,4-dibromobutanoate (0.887 g, 2.50 mmol, 1.00 equiv) were dissolved in EtOH (10 ml) and KOH (0.280 g, 5.00 mmol, 2.00 equiv) was added. The solution was heated to reflux for 1 h, cooled to room temperature and the precipitate removed by filtration. Volatiles were removed under reduced pressure and the residue purified by column chromatography (SiO\(_2\), hexane:EtOAc 95:5 to 8:2) to give S2 (0.364 g, 1.33 mmol, 53 %) as colorless oil.

\[ ^1H \text{NMR} \ (400 \text{ MHz, CDCl}_3) \delta = 4.53 – 4.43 \ (m, 1H), 3.73 – 3.62 \ (m, 2H), 2.49 – 2.31 \ (m, 2H), 1.50 \ (s, 9H), 1.48 \ (s, 9H). \]

\[ ^{13}C \text{NMR} \ (101 \text{ MHz, CDCl}_3) \delta = 169.31, 157.38, 82.43, 82.23, 46.73, 31.94, 28.32, 28.10. \]

HR-MS (ESI): calculated for C_{13}H_{24}NO_{5} [M+H]^+: 274.1649, found: 274.1652.

CHAPTER 7: EXPERIMENTAL

tert-butyl (S)-3-(hydroxymethyl)isoxazolidine-2-carboxylate S3

To a solution of Boc-Opr-OH (0.11 g, 0.54 mmol, 1.0 equiv) in dry THF (2.5 ml) at 0 °C was added BH$_3$•SMe$_2$ (0.20 g, 0.22 mL, 2.6 mmol, 4.9 equiv) dropwise. The reaction mixture was warmed to room temperature and stirred for 8 hours. The reaction was quenched by the addition of saturated NH$_4$Cl, diluted with H$_2$O and extracted with ethyl acetate. The combined organic layers were washed with H$_2$O, brine, dried over Na$_2$SO$_4$, filtered, concentrated under reduced pressure and purified by column chromatography (SiO$_2$, hexane:EtOAc 9:1 to 2:3) to give S3 as a colorless oil (0.056 g, 0.28 mmol, 51 %).

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ = 4.32 – 4.18 (m, 1H), 4.12 – 4.03 (m, 1H), 3.80 – 3.70 (m, 1H), 3.67 (ddd, $J$=11.4, 3.8, 1.4, 1H), 3.57 (dd, $J$=11.2, 7.3, 1H), 2.46 – 2.29 (m, 1H), 2.00 (ddddd, $J$=12.1, 9.1, 7.9, 5.7, 1H), 1.51 (s, 9H).

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ = 158.27, 82.77, 69.09, 65.09, 61.38, 31.15, 28.34.

$[\alpha]_D^{26} = -$ 90.2° (c = 1.62, CHCl$_3$).

IR (thin film, cm$^{-1}$): $\tilde{\nu}$ = 3393, 2932, 2360, 1700, 1456, 1367, 1251, 1071, 993.


tert-butyl isoxazolidine-2-carboxylate S4

Tert-butyl N-hydroxycarbamate (0.366 g, 2.75 mmol, 1.10 equiv) and 1,3-dibromopropane (0.505 g, 2.50 mmol, 1.00 equiv) were dissolved in EtOH (10 ml) and KOH (0.280 g, 5.00 mmol, 2.00 equiv) was added. The solution was stirred at room temperature for 6 h and the precipitate removed by filtration. Volatiles were removed under reduced pressure and the residue purified by column chromatography (SiO$_2$, hexane:EtOAc 99:1 to 9:1) to give S4 (0.305 g, 1.76 mmol, 70 %) as colorless oil.

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ = 3.91 (t, $J$=7.1, 2H), 3.64 – 3.54 (m, 2H), 2.28 – 2.13 (m, 2H), 1.49 (s, 9H).

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ = 157.84, 82.00, 68.53, 46.91, 28.35, 28.02.

---

8 Kindly prepared by Estella Chin as part of her research project.
tert-butyl 1,2-oxazinan-2-carboxylate S5

\[
\text{Boc}_2\text{O} + \text{KOH} \xrightarrow{\text{EtOH, 60ºC, 6 h}} \text{Boc}_2\text{O}
\]

tert-butyl \(N\)-hydroxycarbamate (0.366 g, 2.75 mmol, 1.10 equiv) and 1,4-dibromobutane (0.540 g, 2.50 mmol, 1.00 equiv) were dissolved in EtOH (10 ml) and KOH (0.280 g, 5.00 mmol, 2.00 equiv) was added. The solution was heated to 60 ºC for 6 h, cooled to room temperature and the precipitate removed by filtration. Volatiles were removed under reduced pressure and the residue purified by column chromatography (SiO\(_2\), hexane:EtOAc 99:1 to 9:1) to give \(\text{S5}\) (0.300 g, 1.60 mmol, 64 %) as colorless oil. The spectroscopic data matched with the reported literature.\(^9\)

\(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta = 3.99 \text{--} 3.92 (\text{m, 2H}), 3.68 \text{--} 3.61 (\text{m, 2H}), 1.83 \text{--} 1.74 (\text{m, 2H}), 1.73 \text{--} 1.64 (\text{m, 2H}), 1.52 (\text{s, 9H}).\)

\(^13\)C NMR (101 MHz, CDCl\(_3\)) \(\delta = 155.25, 81.33, 71.56, 46.95, 28.48, 24.52, 22.74.\)

7.2.2. Studies with irisin model system 3

KAHA ligation for the preparation of Fmoc-Opr-[Irisin (46–112)]-OH 3

For the synthesis of the protein segments see pages 252 and 253.

a) Ligation: Fmoc-Opr-[Irisin (46–64)]-\(\alpha\)-ketoacid 1 (9.3 mg, 3.5 µmol, 1.5 equiv) and H-Opr-[Irisin (66–112)]-OH 2 (13.0 mg, 2.40 µmol, 1.00 equiv) were weighed into a glass vial and dissolved in 8:2 NMP:H\(_2\)O (0.20 ml, 12 mM) containing 0.1 M oxalic acid. The mixture was

heated to 60 °C and the ligation monitored by analytical HPLC (Shiseido Capcell Pak C18 MGII (4.6 mm x 250 mm), 25 to 75 % CH₃CN in 25 min). During the ligation, the formation of two peaks (eluting after \( t_R = 27.8 \) min, \textbf{3-A}, and \( t_R = 29.3 \) min, \textbf{3-B}) with the expected mass of 7961 Da was observed. The first eluting peak corresponds to the depsipeptide at the Leu-Hse ligation site (Fmoc-Opr-[Irisin (46–112)]-OH \textit{depsi-3}), whereas the second peak has the desired amide connection at this position (Fmoc-Opr-[Irisin (46–112)]-OH \textit{amide-3}). After 17 h, the mixture was diluted with 1:1 CH₃CN/H₂O (800 µl) containing 0.1% TFA and directly purified by preparative HPLC (Shiseido Capcell Pak MGII heated to 65 °C, 20 mm x 250 mm; 35 to 55 % CH₃CN with 0.1 % TFA in 20 min). Both peaks were collected separately to give Fmoc-Opr-[Irisin (46–112)]-OH \textit{depsi-3} (5.6 mg, 0.70 µmol, 30 %) and Fmoc-Opr-[Irisin (46–112)]-OH \textit{amide-3} (1.5 mg, 0.19 µmol, 8 %).

**HPLC monitoring of the KAHA ligation of Fmoc-Opr-[Irisin (46–64)]-\( \alpha \)-ketoacid \textbf{1} and H-Opr-[Irisin (66–112)]-OH \textbf{2}**. Shiseido Capcell Pak C18 MGII (4.6 x 250 mm); 25 to 75 % CH₃CN in 25 min.

**Fmoc-Opr-[Irisin (46–112)]-OH \textit{depsi-3}**:
Analytical HPLC: \( t_R = 27.8 \) min.
MALDI FTICR MS: calculated molecular weight (C₃₅₂H₅₅₁N₉₁O₁₁₃S₃) [M+H]^+: 7957.9402; found: 7961.9508 [A+4].

**MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of \textit{depsi-3}**.
Fmoc-Opr-[Irisin (46–112)]-OH amide-3:

Analytical HPLC: $t_R = 29.3$ min.

MALDI FTICR MS: calculated molecular weight (C$_{352}$H$_{551}$N$_{91}$O$_{113}$S$_{3}$) [M+H]$^+$: 7957.9402; found: 7961.8838 [A+4].

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of amide-3.

b) O to N acyl shift: Fmoc-Opr-[Irisin (46–112)]-OH depsi-3 (1.1 mg, 0.14 µmol) was dissolved in rearrangement buffer (0.20 ml, 0.2 M Na$_2$CO$_3$ buffer pH 10, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 75 min. The rearrangement was monitored by analytical HPLC (Shiseido Capcell Pak C18 MGII (4.6 x 250 mm), 25 to 75 % CH$_3$CN in 25 min) and quenched by addition of 1:4:4 TFA:CH$_3$CN:H$_2$O (50 µl). The mixture was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 ºC, 35 to 55 % CH$_3$CN in 20 min, flow rate 10 ml/min). The product containing fractions were pooled and lyophilized to give pure Fmoc-Opr-[Irisin (46–112)]-OH amide-3 (0.8 mg, 0.11 µmol, 81 %) with identical analytical properties as the amide product directly obtained from the ligation.

HPLC monitoring of the O to N acyl shift of Fmoc-Opr-[Irisin (46–112)]-OH depsi-3 to Fmoc-Opr-[Irisin (46–112)]-OH amide-3. Shiseido Capcell Pak C18 MGII (4.6 x 250 mm), 25 to 75 % CH$_3$CN in 25 min.

Stability study:

Purified Fmoc-Opr-[Irisin (46–112)]-OH depsi-3 and Fmoc-Opr-[Irisin (46–112)]-OH amide-3 were dissolved separately in 8:2 NMP:H$_2$O containing 0.1 M oxalic acid and heated to 60 ºC for 2 h (ligation conditions). No interconversion between both products was observed,
suggesting that the formation of depsi- and amide-product is irreversible under ligation conditions. This is consistent with the observation that the ratio of formed depsipeptide and amide remains constant during the course of the ligation.

Analytical HPLC trace of purified Fmoc-Opr-[Irisin (46–112)]-OH depsi-3 (bottom trace) and Fmoc-Opr-[Irisin (46–112)]-OH amide-3 after incubation under ligation conditions at 60 °C for 2 h. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 75 % CH₃CN in 25 min.

7.2.3. Studies with Fmoc-Ala-Leu-Hse-(4-F)-Phe-OH 6 tetrapeptide model system

KAHA ligation of Fmoc-Ala-Leu-α-ketoacid 4 and H-Opr-(4-F)-Phe-OH 5

Fmoc-Ala-Leu-α-ketoacid 4 (4.5 mg, 10 mmol, 1.2 equiv) and HN-Opr-(4-F)Phe-OH 5 (2.3 mg, 8.2 mmol, 1.0 equiv) were weighed into a 2 ml glass vial and were dissolved in 8:2 DMSO/H₂O (1.0 ml, 20 mM) with 0.1 M oxalic acid and allowed to react at 60 °C. The ligation was monitored by analytical HPLC (Shiseido Capcell Pak C18 MGII (4.6 x 250 mm); heated to 60 °C, 30 to 80 % CH₃CN in 20 min). Two new peaks were observed after 45 min reaction time, whereby the major peak eluting after $t_R = 23.3$ min corresponds to Fmoc-Ala-Leu-(depsi)-Hse-(4-F)Phe-OH depsi-6 and the minor peak eluting after $t_R = 25.5$ min corresponds to the desired Fmoc-Ala-Leu-Hse-(4-F)Phe-OH amide-6. After 4 h, the ligation mixture was cooled to room temperature, diluted with 1:1 H₂O:CH₃CN containing 0.1 % TFA and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 35 to 70 % CH₃CN in 20 min, flow rate 10 ml/min) and lyophilized to yield depsi-6 (2.4 mg, 3.5 mmol, 35 %) and amide-6 (< 0.2 mg) as white solids. NMR-analysis confirmed the identity of the depsipeptide.
HPLC monitoring of the KAHA ligation of Fmoc-Ala-Leu-α-keto acid 4 and H-Opr-(4-F)-Phe-OH 5. Shiseido Capcell Pak C18 MGII (4.6 x 250 mm), heated to 60 °C, 30 to 80 % CH3CN in 20 min.

Fmoc-Ala-Leu-(depsi)-Hse-(4-F)Phe-OH depsi-6:

$^1$H NMR (600 MHz, d$_6$-DMSO) $\delta =$ 8.74 (d, J=7.7, 1H), 8.19 (d, J=7.8, 1H), 8.16 (s, 4H), 7.89 (d, J=7.5, 2H), 7.72 (t, J=8.0, 2H), 7.50 (d, J=7.6, 1H), 7.44 – 7.40 (m, 2H), 7.32 (dd, J=8.0, 6.9, 2H), 7.31 – 7.27 (m, 2H), 7.11 (t, J=8.9, 2H), 4.49 (ddd, J=9.1, 7.7, 4.9, 1H), 4.34 – 4.28 (m, 1H), 4.28 – 4.23 (m, 2H), 4.20 (t, J=7.0, 1H), 4.15 – 4.10 (m, 2H), 4.10 – 4.05 (m, 1H), 3.91 – 3.85 (m, 1H), 3.10 (dd, J=14.1, 4.9, 1H), 2.93 (dd, J=14.1, 9.1, 1H), 2.12 – 1.96 (m, 2H), 1.68 – 1.60 (m, 1H), 1.58 – 1.47 (m, 2H), 1.22 (d, J=7.1, 3H), 0.88 (d, J=6.6, 3H), 0.83 (d, J=6.5, 3H).

$^{13}$C NMR (151 MHz, d$_6$-DMSO) $\delta =$ 172.87, 172.16, 172.14, 168.16, 161.96, 160.35, 157.95, 157.75, 155.64, 143.90, 143.76, 140.72, 133.30, 133.28, 131.02, 130.97, 127.65, 127.06, 125.29, 125.27, 120.13, 115.09, 114.95, 65.61, 60.46, 53.88, 50.25, 49.71, 49.66, 46.65, 40.06, 35.48, 30.01, 24.18, 22.81, 21.31, 18.05.

HR-MS (ESI): calculated for C$_{37}$H$_{44}$FN$_4$O$_8$ [M+H]$^+$: 691.3138, found 691.3133.

KAHA ligation of Fmoc-Ala-Leu-α-ketoacid 4 and H-Opr-(4-F)-Phe-OH 5 and one-pot O to N acyl transfer

In a 4 ml glass vial, Fmoc-Ala-Leu-α-ketoacid 4 (6.3 mg, 14 mmol, 1.0 equiv) and H-Opr-(4-F)-Phe-OH 5 (3.9 mg, 14 mmol, 1.0 equiv) were dissolved in 8:2 DMSO:H$_2$O (1.4 ml, 10 mM) with 0.1 M oxalic acid and allowed to react at 60 °C for 24 h. Reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak C18 MGII (4.6 x 250 mm); heated to 60 °C, 30 to 80 % CH3CN in 20 min). After completion of the ligation, Na$_2$CO$_3$/NaHCO$_3$ buffer (pH 9.5, 1.5 ml, 0.2 M) was added to the ligation mixture and the mixture adjusted to pH ≈ 10.
by addition of Na$_2$CO$_3$ solution (0.2 M) and incubated at room temperature. After 45 min, analytical HPLC showed complete conversion of Fmoc-Ala-Leu-(depsi)-Hse-(4-F)Phe-OH amide-6 to Fmoc-Ala-Leu-Hse-(4-F)Phe-OH amide-6. The solution was acidified to pH 1 by addition of 1:4:4 TFA:CH$_3$CN:H$_2$O and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 40 to 85 % CH$_3$CN in 20 min, flow rate 10 ml/min) and lyophilized to yield amide-6 (3.4 mg, 4.9 mmol, 36 %) as white solid.

HPLC monitoring of the KAHA ligation (bottom trace) and one pot O to N acyl transfer (upper trace) of Fmoc-Ala-Leu-α-keto acid 4 and H-Opr-(4-F)Phe-OH 5. Shiseido Capcell Pak MGII C18 column (4.6 x 250 mm); heated to 60 °C, 30 to 80 % CH$_3$CN in 20 min.

Fmoc-Ala-Leu-Hse-(4-F)Phe-OH amide-6:

$^1$H NMR (600 MHz, d$_6$-DMSO) δ = 12.77 (s, 1H), 8.00 (d, $J$=7.9, 1H), 7.92 – 7.90 (m, 1H), 7.90 – 7.87 (m, 3H), 7.84 (d, $J$=8.0, 1H), 7.75 – 7.68 (m, 2H), 7.52 (d, $J$=7.6, 1H), 7.41 (t, $J$=7.4, 2H), 7.36 – 7.29 (m, 2H), 7.25 – 7.20 (m, 2H), 7.10 – 7.02 (m, 2H), 4.46 – 4.43 (m, 1H), 4.42 – 4.37 (m, 1H), 4.33 – 4.29 (m, 1H), 4.29 – 4.27 (m, 0H), 4.27 – 4.24 (m, 2H), 4.21 (t, $J$=7.1, 1H), 4.09 – 4.00 (m, 1H), 3.41 – 3.35 (m, 2H), 3.02 (dd, $J$=14.0, 6.6, 1H), 1.84 – 1.76 (m, 1H), 1.67 – 1.60 (m, 1H), 1.60 – 1.53 (m, 1H), 1.45 – 1.35 (m, 2H), 1.20 (d, $J$=7.3, 3H), 0.83 (d, $J$=6.6, 3H), 0.80 (d, $J$=6.5, 3H).

$^{13}$C NMR (151 MHz, d$_6$-DMSO) δ = 172.53, 172.36, 171.78, 171.30, 161.81, 160.21, 155.68, 143.91, 143.75, 140.69, 133.44, 133.42, 130.98, 130.92, 127.62, 127.07, 127.06, 125.30, 125.26, 120.10, 114.89, 114.75, 65.63, 57.51, 53.23, 50.86, 49.97, 49.74, 46.63, 40.66, 35.76, 35.29, 24.07, 23.04, 21.54, 18.10.

HR-MS (ESI): calculated for C$_{37}$H$_{44}$FN$_4$O$_8$ [M+H]$^+$: 691.3138, found 691.3136.

Characterization of independently synthesized 6-A, 6-B and 6-C:

Fmoc-Ala-(S)-Leu-(S)-Hse-(4-F)Phe-OH 6-A

$^1$H NMR (400 MHz, d$_6$-DMSO) δ = 12.77 (s, 1H), 8.00 (d, $J$=7.8, 1H), 7.90 (d, $J$=8.2, 1H), 7.95 – 7.86 (m, 2H), 7.84 (d, $J$=8.0, 1H), 7.75 – 7.65 (m, 2H), 7.52 (d, $J$=7.6, 1H), 7.44 – 7.38 (m, 2H), 7.36 – 7.29 (m, 2H), 7.27 – 7.18 (m, 2H), 7.11 – 6.96 (m, 2H), 4.45 – 4.36 (m, 1H), 4.34 – 4.28 (m, 1H), 4.30 – 4.26 (m, 1H), 4.27 – 4.23 (m, 2H), 4.23 – 4.18 (m, 1H), 4.11 – 4.00 (m, 1H), 3.46 – 3.32 (m, 2H), 3.03 (dd, $J$=14.0, 5.1, 1H), 2.88 (dd,
$J$=14.0, 8.6, 1H), 1.86 – 1.72 (m, 1H), 1.68 – 1.59 (m, 1H), 1.58 – 1.50 (m, 1H), 1.48 – 1.31 (m, 2H), 1.20 (d, $J$=7.1, 3H), 0.83 (d, $J$=6.5, 3H), 0.80 (d, $J$=6.5, 3H).

$^{13}$C NMR (101 MHz, $d_6$-DMSO) $\delta$ = 172.51, 172.35, 171.76, 171.28, 162.20, 159.80, 155.66, 143.89, 143.74, 140.68, 133.43, 133.40, 130.98, 130.90, 127.61, 127.06, 125.26, 120.08, 114.91, 114.71, 65.62, 57.50, 53.21, 50.85, 49.96, 49.72, 46.63, 40.66, 35.75, 35.28, 24.06, 23.03, 21.53, 18.09.

HR-MS (ESI): calculated for C$_{37}$H$_{44}$FN$_4$O$_8$ [M+H]$^+$: 691.3138, found 691.3140.

Fmoc-Ala-(R)-Leu-(S)-Hse-(4-F)Phe-OH 6-B

$^1$H NMR (400 MHz, $d_6$-DMSO) $\delta$ = 8.07 (d, $J$=8.2, 1H), 7.93 (d, $J$=8.0, 1H), 7.98 (d, $J$=8.2, 1H), 7.89 (d, (m, 2H), 7.53 (d, $J$=7.2, 1H), 7.45 – 7.37 (m, 2H), 7.35 – 7.28 (m, 2H), 7.25 – 7.19 (m, 2H), 7.10 – 7.00 (m, 2H), 4.41 – 4.34 (m, 1H), 4.33 – 4.29 (m, 1H), 4.28 – 4.24 (m, 1H), 4.26 – 4.23 (m, 2H), 4.22 – 4.18 (m, 1H), 4.13 – 4.03 (m, 1H), 3.34 – 3.25 (m, 2H), 3.05 (dd, $J$=13.8, 4.7, 1H), 2.87 (dd, $J$=13.8, 9.4, 1H), 1.77 – 1.64 (m, 1H), 1.65 – 1.57 (m, 1H), 1.57 – 1.49 (m, 1H), 1.46 – 1.40 (m, 2H), 1.22 (d, $J$=7.1, 3H), 0.85 – 0.77 (m, 6H).

$^{13}$C NMR (101 MHz, $d_6$-DMSO) $\delta$ = 172.70, 172.56, 171.78, 171.35, 162.20, 159.80, 155.78, 143.97, 143.72, 140.70, 140.67, 133.54, 133.51, 131.02, 130.94, 127.61, 127.06, 125.32, 125.21, 120.08, 114.89, 114.68, 65.74, 57.46, 53.39, 50.83, 50.22, 49.96, 46.61, 40.58, 35.85, 34.97, 24.20, 23.11, 21.27, 18.24.

Fmoc-Ala-(S)-Leu-(R)-Hse-(4-F)Phe-OH 6-C

$^1$H NMR (400 MHz, $d_6$-DMSO) $\delta$ = 8.07 (d, $J$=8.3, 1H), 8.00 (d, $J$=8.3, 1H), 7.96 (d, $J$=7.7, 1H), 7.91 – 7.86 (m, 2H), 7.74 – 7.66 (m, 2H), 7.52 (d, $J$=7.2, 1H), 7.44 – 7.38 (m, 2H), 7.35 – 7.29 (m, 2H), 7.24 – 7.17 (m, 2H), 7.09 – 7.00 (m, 2H), 4.45 – 4.37 (m, 1H), 4.35 – 4.29 (m, 1H), 4.29 – 4.26 (m, 1H), 4.27 – 4.23 (m, 2H), 4.23 – 4.16 (m, 1H), 4.11 – 4.01 (m, 1H), 3.34 – 3.15 (m, 2H), 3.06 (dd, $J$=13.8, 4.7, 1H), 2.87 (dd, $J$=13.8, 9.4, 1H), 1.79 – 1.65 (m, 1H), 1.58 (d, $J$=7.6, 1H), 1.55 – 1.49 (m, 1H), 1.48 – 1.42 (m, 2H), 1.19 (d, $J$=7.2, 3H), 0.86 (d, $J$=6.5, 3H), 0.82 (d, $J$=6.5, 3H).

$^{13}$C NMR (101 MHz, $d_6$-DMSO) $\delta$ = 172.58, 172.49, 171.84, 171.28, 162.20, 159.80, 155.72, 143.94, 143.72, 140.68, 133.63, 133.60, 131.07, 130.99, 127.61, 127.06, 125.30, 125.23, 120.08, 114.88, 114.67, 65.65, 57.41, 53.33, 51.37, 50.00, 49.76, 46.62, 40.74, 35.97, 35.10, 24.13, 22.91, 21.68, 17.94.
Stability study:
Purified Fmoc-Ala-Leu-(depsi)-Hse-(4-F)Phe-OH depsi-6 and Fmoc-Ala-Leu-Hse-(4-F)Phe-OH amide-6 were separately incubated under ligation conditions (8:2 DMSO:H₂O, 60 °C, 2 h) and analyzed by analytical HPLC. Similarly as for the protein segments depsi-3 and amide-3, no interconversion between both products was observed, further supporting that the formation of depsi- and amide-product is irreversible under ligation conditions.

7.2.4. NMR epimerization study

To further confirm the identity of the rearranged Fmoc-Ala-Leu-Hse-(4-F)Phe-OH amide-6 product and to investigate if epimerization either at the formed homoserine residue or at the α-ketoacid-derived leucine residue occurred, a comparative study with separately synthesized amide-6 (6-A) and its diastereomers containing (R) amino acids at the Leu (6-B) or Hse (6-C) position was performed. It was found that the retention time of all three diastereomers on HPLC was very similar (< 5 s difference). However, comparing the NMR spectra of the rearranged ligation product amide-6 and the SPPS-synthesized diastereomers shows that the ligation and the rearrangement yields exclusively the expected Fmoc-Ala-(S)-Leu-(S)-Hse-(4-F)Phe-OH amide-6.

The ¹H-NMR spectrum of the rearranged ligation product amide-6 (top spectrum) is identical to a ¹H spectra of an authentic sample synthesized by SPPS from all-(S) amino acids (6-A, second spectrum). Diasteromers containing (R) amino acid at Leu (6-B), third spectra and Hse (6-B, bottom spectra) have clearly distinguishable ¹H NMR spectra, for example in the region of the amide protons (8.2 – 7.5 ppm). This comparison demonstrates that the ligation and rearrangement does not alter the stereoconfiguration of residues at the ligation site.
7.2.5. Model ligations with different $\alpha$-ketoacids

Different $\alpha$-ketoacids were screened in ligation reactions with H-Opr-(4-F)Phe-OH 5. In all cases, the depsipeptide was obtained as major product.

In a 2 ml vial, $\alpha$-ketoacid (24 mmol, 1.2 equiv) and H-Opr-(4F)-Phe-OH 5 (20 mmol, 1.0 equiv) were dissolved in 8:2 DMSO:H$_2$O (1.0 ml, 20mM) and heated to 60$^\circ$C overnight. The products were purified by preparative HPLC (Shiseido Capcell Pak MGII (20 x 250 mm), heated to 65 $^\circ$C, different gradients) and lyophilized.

<table>
<thead>
<tr>
<th>Entry</th>
<th>$R^1$</th>
<th>Product</th>
<th>ratio ester/amide during ligation (by UV)</th>
<th>Isolated yield (ester)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>depsi-8</td>
<td>93:7</td>
<td>61 %</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>depsi-9</td>
<td>93:7</td>
<td>59 %</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>depsi-10</td>
<td>86:14</td>
<td>59 %</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>depsi-11</td>
<td>87:13</td>
<td>55 %</td>
</tr>
</tbody>
</table>

Fmoc-Leu-(depsi)-Hse-(4F)-Phe-OH 8

$^1$H NMR (400 MHz, d$_6$-DMSO) $\delta$ = 13.04 (s, 1H), 8.74 (d, $J$=7.7, 1H), 8.18 – 8.11 (m, 2H), 7.93 – 7.87 (m, 2H), 7.76 (d, $J$=8.1, 1H), 7.74 – 7.68 (m, 2H), 7.47 – 7.38 (m, 2H), 7.36 – 7.31 (m, 2H), 7.31 – 7.25 (m, 2H), 7.16 – 7.04 (m, 2H), 4.49 (ddd, $J$=9.1, 7.5, 4.9, 1H), 4.39 – 4.27 (m, 2H), 4.27 – 4.16 (m, 1H), 4.15 – 4.10 (m, 2H), 4.10 – 4.06 (m, 1H), 3.93 – 3.78 (m, 1H), 3.10 (dd, $J$=14.1, 4.8, 1H), 2.93 (dd, $J$=14.1, 9.0, 1H), 2.14 – 1.96 (m, 2H), 1.69 – 1.60 (m, 1H), 1.60 – 1.53 (m, 1H), 1.52 – 1.43 (m, 1H), 0.90 (d, $J$=6.4, 3H), 0.85 (d, $J$=6.4, 3H).

$^{13}$C NMR (101 MHz, d$_6$-DMSO) $\delta$ = 172.56, 172.11 , 168.10 , 161.13 (d, $J$=241.9), 156.14 , 143.81 , 143.67 , 140.73 , 133.26 (d, $J$=3.1), 130.97 (d, $J$=8.0), 127.65 , 127.03 , 125.20 , 125.16 , 120.15 , 120.12 , 114.99 (d, $J$=212), 60.47 , 53.84 , 52.08 , 49.64 , 46.66 , 35.47 , 29.98 , 24.22 , 22.85 , 21.07.

HR-MS (ESI): calculated for C$_{34}$H$_{35}$F$_3$N$_3$O$_7$ [M+H]$^+$: 620.2767, found 620.2767.
Succinyl-(depsi)-Hse-(4F)-Phe-OH 9

\[ ^1H \text{ NMR (400 MHz, } d_6\text{-DMSO) } \delta = 13.01 \text{ (s, 1H), 12.27 \text{ (s, 1H),} \]
\[ 8.76 \text{ (d, } J=7.8, \text{ 1H), 8.15 \text{ (s, 2H), 7.33} - 7.24 \text{ (m, 2H), 7.20} - 7.07 \text{ (m, 2H), 4.49} \text{ (ddd, } J=9.1, 7.7, 4.9, \text{ 1H), 4.22} - 4.01 \text{ (m, 2H), 3.87} \text{ (t, } J=6.4, \text{ 1H), 3.10} \text{ (dd, } J=14.1, 4.9, \text{ 1H), 2.93} \text{ (dd, } J=14.1, 9.0, \text{ 1H), 2.55} - 2.51 \text{ (m, 2H), 2.48} - 2.45 \text{ (m, 2H), 2.19} - 1.94 \text{ (m, 2H).} \]

\[ ^{13}C \text{ NMR (101 MHz, } d_6\text{-DMSO) } \delta = 173.38, 172.13, 171.98, 168.25, 161.14 \text{ (d, } J=242.5, \text{ 133.28 \text{ (d, } J=3.1), 130.98 \text{ (d, } J=8.0), 115.01 \text{ (d, } J=21.0), 59.93, 53.83, 49.72, 35.50, 30.05, 28.60.} \]

\[ \text{HR-MS (ESI): calculated for C}_{17}H_{22}FN_2O_7 \text{ [M+H]+: 385.1406, found 385.1411.} \]

Propionyl-(depsi)-Hse-(4F)-Phe-OH 10

\[ ^1H \text{ NMR (400 MHz, } d_6\text{-DMSO) } \delta = 8.76 \text{ (d, } J=7.8, \text{ 1H), 7.29 (dd, } J=8.5, 5.7, \text{ 2H), 7.12} \text{ (t, } J=8.9, \text{ 2H), 4.53} - 4.44 \text{ (m, 1H), 4.07} \text{ (ddd, } J=11.3, 7.7, 4.3, \text{ 2H), 3.86} \text{ (dd, } J=7.2, 5.4, \text{ 1H), 3.09} \text{ (dd, } J=14.1, 4.9, \text{ 1H), 2.93} \text{ (dd, } J=14.0, 8.9, \text{ 1H), 2.31 (q, } J=7.5, \text{ 2H), 2.13} - 1.93 \text{ (m, 2H), 1.02 (t, } J=7.5, \text{ 3H).} \]

\[ ^{13}C \text{ NMR (101 MHz, } d_6\text{-DMSO) } \delta = 173.47, 172.14, 168.41, 162.34, 159.93, 133.31, 133.28, 131.03, 130.95, 114.88, 59.73, 53.81, 49.80, 35.56, 30.18, 26.63, 8.86. \]

\[ \text{HR-MS (ESI): calculated for C}_{16}H_{21}FN_2NaO_5 \text{ [M+Na]+: 363.1327 found 363.1332.} \]

Benzoyl-(depsi)-Hse-(4F)-Phe-OH 11

\[ ^1H \text{ NMR (400 MHz, } d_6\text{-DMSO) } \delta = 8.82 \text{ (d, } J=7.8, \text{ 1H), 8.05} - 7.94 \text{ (m, 2H), 7.71} - 7.62 \text{ (m, 1H), 7.57} - 7.49 \text{ (m, 2H), 7.35} - 7.24 \text{ (m, 2H), 7.18} - 7.07 \text{ (m, 2H), 4.51} \text{ (td, } J=8.2, 4.9, \text{ 1H), 4.41} - 4.24 \text{ (m, 2H), 3.97} \text{ (dd, } J=7.5, 5.4, \text{ 1H), 3.10} \text{ (dd, } J=14.1, 5.0, \text{ 1H), 2.95} \text{ (dd, } J=14.1, 8.7, \text{ 1H), 2.28} - 2.19 \text{ (m, 1H), 2.19} - 2.09 \text{ (m, 1H).} \]

\[ ^{13}C \text{ NMR (101 MHz, } d_6\text{-DMSO) } \delta = 172.12, 168.60, 165.53, 162.34, 159.94, 133.37, 133.30, 133.27, 131.05, 130.97, 129.59, 129.30, 128.65, 115.09, 114.88, 60.58, 53.79, 49.90, 35.61, 30.37. \]

\[ \text{HR-MS (ESI): calculated for C}_{20}H_{21}FN_2NaO_5 \text{ [M+Na]+: 411.1327 found 411.1336.} \]
7.2.6. Model ligations with different hydroxylamines

KAHA ligation with (rac)-isoxazolidine-5-carboxylic acid 12

\[
\text{\textbf{S2}} \quad \begin{array}{c}
\text{\textit{H}} \\
\text{TFA}
\end{array} \quad \begin{array}{c}
\text{\textit{H}} \\
\text{\textit{H}}
\end{array} \quad \begin{array}{c}
\text{HO} \\
\text{\textit{H}}
\end{array} \quad \begin{array}{c}
\text{\textit{H}} \\
\text{\textit{H}}
\end{array} \quad \begin{array}{c}
\text{OH} \\
\text{\textit{H}}
\end{array} \quad \begin{array}{c}
\text{HO} \\
\text{\textit{H}}
\end{array} \quad \begin{array}{c}
\text{\textit{H}} \\
\text{\textit{H}}
\end{array} \quad \begin{array}{c}
\text{OH} \\
\text{\textit{H}}
\end{array} \\
\text{8:2 DMSO:H}_2\text{O} \\
\text{0.1 M oxalic acid} \\
\text{60 ºC, 5 h}
\end{array} \quad \text{(TFA salt)} \quad \text{depsi-13}
\]

di-tert-butyl isoxazolidine-2,5-dicarboxylate \(\text{S2}\) (15.1 mg, 52.5 \(\mu\)mol, 1.00 equiv) was dissolved in TFA (1.0 ml) and incubated at room temperature for 1 h. Volatiles were removed under reduced pressure. 4-Hydroxyphenylpyruvic acid (10.9 mg, 60.8 \(\mu\)mol, 1.10 equiv) and a mixture of 8:2 DMSO:H\(\text{2O}\) (1.0 ml, 55 mM) with 0.1 M oxalic acid was added and the solution heated to 60 °C for 5 h. The crude reaction mixture was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, room temperature, 20 to 60 % CH\(_3\)CN in 20 min, flow rate 10 ml/min) to give \textit{depsi-13} (5.6 mg, 22.1 \(\mu\)mol, 40 %) as white solid.

HPLC monitoring of the KAHA ligation of 4-hydroxyphenyl pyruvic acid and isoxazolidine-5-carboxylic acid \(12\).

\* = Impurities from 4-hydroxyphenyl pyruvic acid. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 20 to 65 % CH\(_3\)CN in 25 min.

4-amino-2-(2-(4-hydroxyphenyl)acetoxy)butanoic acid \textit{depsi-13}

\[\text{\textbf{(TFA salt)}}\]

\(\text{\textbf{1H NMR}}\) (400 MHz, d\(_{6}\)-DMSO) \(\delta = 13.34\) (s, 1H), 9.35 (s, 1H), 7.79 (s, 3H), 7.06 (d, \(J=8.5\), 2H), 6.70 (d, \(J=8.5\), 1H), 4.94 (dd, \(J=8.8\), 4.3, 1H), 3.68 – 3.52 (m, 2H), 2.98 – 2.77 (m, 2H), 2.16 – 1.92 (m, 2H).

\(\text{\textbf{13C NMR}}\) (101 MHz, d\(_{6}\)-DMSO) \(\delta = 170.94, 170.51, 156.27, 130.33, 123.93, 115.09, 69.83, 39.02, 35.56, 28.48.\)

\(\text{HR-MS}\) (ESI): calculated for C\(_{12}\)H\(_{16}\)NO\(_5\) [M+H]\(^+\): 254.1023; found 254.1018.

151
KAHA ligation with (S)-isoxazolidin-3-ylmethanol 14

tert-butyl (S)-3-(hydroxymethyl)isoxazolidine-2-carboxylate S3 (9.9 mg, 49 µmol, 1.0 equiv) was dissolved in TFA (1 ml) and the solution stirred at room temperature for 1 h. Volatiles were removed under reduced pressure and Fmoc-leucine α-ketoacid 7 (36.9 mg, 97.0 µmol, 2.00 equiv) was added and the mixture dissolved in 8:2 DMSO:H2O (0.50 ml, 0.10 M) with 0.1 M oxalic acid and heated to 60 °C for 2 h. Reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 30 to 70 % CH3CN in 20 min). The ratio of depsi-15 to amide-15 in the crude reaction was 92:8 by UV. The crude reaction was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, room temperature, 40 to 90 % CH3CN with 0.1% TFA in 20 min, flow rate 10 ml/min) to give depsi-15 (9.2 mg, 21 µmol, 43 %). NMR analysis confirmed the identity of depsi-15.

Purified depsi-15 (9.2 mg, 21 µmol) was dissolved in pH 9.5 buffer (0.2 ml, 0.2 M Na2CO3 buffer) and incubated at room temperature. After 15 min, analysis by analytical HPLC indicated complete conversion to amide-15. The crude mixture was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, room temperature, 40 to 90 % CH3CN with 0.1% TFA in 20 min, flow rate 10 ml/min) to give amide-15 (7.8 mg, 18 µmol, 85 %) as white solid.

Analytical HPLC of the crude ligation of 12 and 7 (bottom trace), the purified depsi-15 (middle trace) and the O to N acyl shift (top trace). * = Impurities from 7. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 30 to 70 % CH3CN in 20 min.
(S)-3-amino-4-hydroxybutyl (((9H-fluoren-9-yl)methoxy)carbonyl)-L-leucinate depsi-15:

\[ ^1H \text{NMR} \text{ (400 MHz, d}_6\text{-DMSO)} \delta = 7.94 – 7.86 \text{ (m, 2H)}, 7.86 \text{ (s, 2H)}, 7.79 \text{ (d, J=7.9, 1H)}, 7.75 – 7.68 \text{ (m, 2H)}, 7.52 – 7.38 \text{ (m, 2H)}, 7.38 – 7.28 \text{ (m, 2H)}, 4.44 – 4.26 \text{ (m, 2H)}, 4.26 – 4.18 \text{ (m, 1H)}, 4.18 – 4.10 \text{ (m, 2H)}, 4.10 – 4.01 \text{ (m, 1H)}, 3.65 – 3.53 \text{ (m, 1H)}, 3.48 – 3.35 \text{ (m, 1H)}, 3.22 – 3.09 \text{ (m, 1H)}, 1.89 – 1.75 \text{ (m, 2H)}, 1.68 – 1.60 \text{ (m, 1H)}, 1.60 – 1.51 \text{ (m, 1H)}, 1.51 – 1.40 \text{ (m, 1H)}, 0.96 – 0.75 \text{ (m, 6H)}. \]

\[ ^1H \text{NMR} \text{ (101 MHz, , CDCl}_3) \delta = 172.70, 156.13, 143.83, 143.70, 140.75, 127.66, 127.05, 125.21, 120.14, 65.61, 60.86, 60.43, 52.21, 52.10, 49.74, 46.67, 28.24, 22.81, 21.13. \]

HR-MS (ESI): calculated for C\text{25H33N2O5 [M+H]}^+: 441.2384 found 441.2387.

(9H-fluoren-9-yl)methyl ((S)-1-(((S)-1,4-dihydroxybutan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate amide-15:

\[ ^1H \text{NMR} \text{ (400 MHz, d}_6\text{-DMSO)} \delta = 7.92 – 7.86 \text{ (m, 2H)}, 7.78 – 7.69 \text{ (m, 2H)}, 7.57 \text{ (d, J=8.4, 1H)}, 7.47 – 7.42 \text{ (m, 2H)}, 7.42 – 7.37 \text{ (m, 1H)}, 7.37 – 7.27 \text{ (m, 2H)}, 4.33 – 4.18 \text{ (m, 3H)}, 4.06 – 3.95 \text{ (m, 1H)}, 3.81 – 3.70 \text{ (m, 1H)}, 3.47 – 3.37 \text{ (m, 2H)}, 3.37 – 3.31 \text{ (m, 1H)}, 3.28 – 3.22 \text{ (m, 1H)}, 1.76 – 1.64 \text{ (m, 1H)}, 1.59 \text{ (s, 1H)}, 1.52 – 1.43 \text{ (m, 1H)}, 1.42 \text{ (s, 2H)}, 0.91 – 0.81 \text{ (m, 6H)}. \]

\[ ^{13}C \text{NMR} \text{ (101 MHz, d}_6\text{-DMSO)} \delta = 172.19, 155.82, 143.95, 143.72, 140.70, 127.61, 127.04, 125.31, 120.09, 65.47, 63.18, 57.93, 53.16, 48.06, 46.70, 41.02, 34.14, 24.19, 23.07, 21.48. \]

HR-MS (ESI): calculated for C\text{25H33N2O5 [M+H]}^+: 441.2384 found 441.2383.

**KAHA ligation with isoxazolidine 16**

Boc-isoxazolidine S4 (70.0 mg, 0.404 mmol, 1.00 equiv) was dissolved in TFA (1.0 ml) and incubated at room temperature for 1 h. Volatiles were removed under a stream of nitrogen and the residue dissolved in 1:1 H\text{2O}:CH\text{3CN} containing 0.1 % TFA and lyophilized. Phenylpyruvic acid (79.6 mg, 0.485 mmol, 1.20 equiv) was added to the crude isoxazolidine salt and dissolved in a mixture of 8:2 DMSO:H\text{2O} (1.6 ml, 0.25 M) containing 0.1 M oxalic acid and heated to 60 °C for 14 h. The crude mixture was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, heated to 60 °C, 25 to 55 % CH\text{3CN} in 30 min) to give depsi-17 (21.6 mg, 0.117 mmol, 27 %) and amide-17 (5.7 mg, 0.029 mmol, 7 %).
HPLC monitoring of the KAHA ligation of phenylpyruvic acid and isoxazolidine 16 (top) and pure phenylpyruvic acid (bottom). Unlike in most other cases, the amide is eluting prior to the ester under these HPLC conditions.

Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 10 to 90 % CH3CN in 25 min.

3-aminopropyl 2-phenylacetate _depsi_17:

\[
\text{\textbf{1H NMR (400 MHz, d6-DMSO)} } \delta = 7.78 (s, 3H), 7.38 - 7.02 (m, 5H), 4.10 (t, J=6.3, 2H), 3.68 (s, 2H), 2.84 (t, J=7.6, 2H), 1.86 (dq, J=8.1, 6.4, 2H).
\]

\[
\text{\textbf{13C NMR (101 MHz, d6-DMSO)} } \delta = 171.11, 134.30, 129.32, 128.35, 126.84, 61.50, 40.21, 36.11, 26.32.
\]

HR-MS (ESI): calculated for C11H16NO2 [M+H]^+: 194.1176; found: 194.1175.

N-(3-hydroxypropyl)-2-phenylacetamide _amide_17:

\[
\text{\textbf{1H NMR (400 MHz, d6-DMSO)} } \delta = 8.00 (t, J=5.6, 1H), 7.33 - 7.14 (m, 5H), 4.40 (t, J=5.2, 1H), 3.41 - 3.35 (m, 2H), 3.38 (s, 2H), 3.14 - 3.02 (m, 2H), 1.61 - 1.45 (m, 2H).
\]

\[
\text{\textbf{13C NMR (101 MHz, d6-DMSO)} } \delta = 170.01, 136.56, 128.90, 128.16, 126.25, 58.36, 42.41, 35.82, 32.39.
\]


KAHA ligation with 1,2-oxazinane 15

Boc-1,2-oxazinane S5 (67.0 mg, 0.358 mmol, 1.00 equiv) was dissolved in TFA (1.0 ml) and incubated at room temperature for 1 h. Volatiles were removed under a stream of nitrogen and the residue dissolved in 1:1 H2O:CH3CN containing 0.1 % TFA and lyophilized. Phenylpyruvic acid (70.5 mg, 0.430 mmol, 1.20 equiv) was added to the crude 1,2-oxazinane salt and dissolved in 8:2 DMSO:H2O (0.1 M oxalic acid 60 °C, 14 h). The crude mixture was then directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, heated to 60 °C, 25 to 55 % CH3CN in 30 min) to give _depsi_19 (10.6 mg, 51.1 µmol, 14 %) and _amide_19 (7.6 mg, 0.037 mmol, 10 %).
HPLC monitoring of the KAHA ligation of phenylpyruvic acid and 1,2-oxazinane 18 (top) and pure phenylpyruvic acid (bottom). Unlike in most other cases, the amide is eluting prior to the ester under these HPLC conditions.

Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 10 to 90 % CH3CN in 25 min.

*coeluting on analytical HPLC with impurities of phenylpyruvic acid.

Purified depsi-19 (0.5 mg, 2.4 µmol) was dissolved in pH 9.5 buffer (50 µl, 0.2 M Na2CO3 buffer) and incubated at room temperature. After 14 h, analysis by analytical HPLC indicated almost complete rearrangement to amide-19 and a minor amount of phenylacetic acid was observed, suggesting hydrolysis of the ester bond.

HPLC monitoring of O to N acyl shift of purified depsi-19. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 10 to 90 % CH3CN in 25 min. *coeluting with impurities of phenylpyruvic acid.

4-aminobutyl 2-phenylacetate depsi-19:

\[
{^1}H\text{ NMR} (400 \text{ MHz, } d_6-\text{DMSO}) \delta = 7.74 (s, 3H), 7.42 – 7.14 (m, 5H), 4.05 (t, J=6.2, 2H), 3.66 (s, 2H), 2.90 – 2.69 (m, 2H), 1.72 – 1.48 (m, 4H).
\]

\[
{^{13}}C\text{ NMR} (101 \text{ MHz, } d_6-\text{DMSO}) \delta = 171.14, 134.33, 129.29, 128.36, 126.83, 63.64, 40.30, 38.45, 25.10, 23.68.
\]

HR-MS (ESI): calculated for C\(_{12}\)H\(_{18}\)NO\(_2\) [M+H]\(^+\): 208.1332; found: 208.1335.

N-(4-hydroxybutyl)-2-phenylacetamide amide-19:

\[
{^1}H\text{ NMR} (400 \text{ MHz, } d_6-\text{DMSO}) \delta = 8.00 (t, J=5.7, 1H), 7.40 – 7.00 (m, 5H), 4.37 (t, J=5.1, 1H), 3.37 (s, 2H), 3.40 – 3.34 (m, 2H), 3.09 – 2.97 (m, 2H), 1.49 – 1.29 (m, 4H).
\]

\[
{^{13}}C\text{ NMR} (101 \text{ MHz, } d_6-\text{DMSO}) \delta = 169.81, 136.60, 128.89, 128.14, 126.23, 60.38, 42.43, 38.54, 29.88, 25.80.
\]

HR-MS (ESI): calculated for C\(_{12}\)H\(_{17}\)NNa\(_1\)O\(_2\) [M+Na]\(^+\): 230.1151; found: 230.1150.
7.2.7. HPLC study: KAHA ligation of H-Trp-Thr-Leu-α-ketoacid 20 and H-Opr-(4-F)Phe-OH 5

The KAHA ligation of H-Trp-Thr-Leu-α-ketoacid 20 and H-Opr-(4-F)Phe-OH 5 was studied using different reaction solvents. The ligation products depsi-21 and amide-21, and the corresponding epimers at the Leu residue, epi-depsi-21 and epi-amide-21 were well separated on HPLC, allowing a straightforward analysis of the ratio of ester/amide formed in the reaction and potential epimerization of the products. The identity of the amide products was confirmed by comparison with samples independently synthesized by SPPS; the identity of the depsi-products was confirmed by O to N acyl transfer to the known amide-product.

In a typical procedure, H₂N-Trp-Thr-Leu-α-ketoacid 20 (2.3 mg, 5.2 µmol, 1.0 equiv) and H-Opr-(4-F)Phe-OH 5 (1.5 mg, 5.2 µmol, 1.0 equiv) were weighed into a 1 ml glass vial and dissolved in the solvent mixture (0.25 ml, 21 mM peptide concentration). The solution was heated to 60 °C for 12 h and analyzed by analytical HPLC (Shiseido Capcell Pak C18 MGII (4.6 x 250 mm); 25 to 45 % CH₃CN in 13 min).

Representative HPLC monitoring of KAHA ligation of 20 and 5. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 25 to 45 % CH₃CN in 13 min.
Table S1: KAHA ligation of 20 and 5 in different solvent mixtures.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Solvent mixture&lt;sup&gt;a&lt;/sup&gt;</th>
<th>% ester formed</th>
<th>% amide formed</th>
<th>% epimerization&lt;sup&gt;b&lt;/sup&gt; ester</th>
<th>% epimerization amide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7:3 DMSO:H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>89</td>
<td>11</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>7:3 NMP:H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>87</td>
<td>13</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>99:1 DMSO:H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>97</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>99:1 NMP:H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>90</td>
<td>10</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>99:1 NMPA:H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>96</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>99:1 CH&lt;sub&gt;3&lt;/sub&gt;CN:H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>95</td>
<td>5</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>73</td>
<td>27</td>
<td>12</td>
<td>29</td>
</tr>
</tbody>
</table>

<sup>a</sup> containing 0.1 M oxalic acid  
<sup>b</sup> = epi-ester/ester x 100

Analytical HPLC of purified epi-depsi-21, depsi-21, epi-amide-21 and amide-21 (lower four traces) and O to N acyl shift of epi-depsi-21 and depsi-21 (upper two traces). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 25 to 45 % CH<sub>3</sub>CN in 13 min.

**epi-depsi-21:**
Analytical HPLC:  
MALDI FTICR MS: calculated molecular weight C<sub>34</sub>H<sub>46</sub>F<sub>1</sub>N<sub>6</sub>O<sub>8</sub>: 685.3356;  
observed: 685.3364

**depsi-21:**
Analytical HPLC:  
MALDI FTICR MS: calculated molecular weight C<sub>34</sub>H<sub>46</sub>F<sub>1</sub>N<sub>6</sub>O<sub>8</sub>: 685.3356;  
observed: 685.3363

**epi-amide-21:**
Analytical HPLC:  
MALDI FTICR MS: calculated molecular weight C<sub>34</sub>H<sub>46</sub>F<sub>1</sub>N<sub>6</sub>O<sub>8</sub>: 685.3356;  
observed: 685.3355

**amide-21:**
Analytical HPLC:  
MALDI FTICR MS: calculated molecular weight C<sub>34</sub>H<sub>46</sub>F<sub>1</sub>N<sub>6</sub>O<sub>8</sub>: 685.3356;  
observed: 685.3357
7.2.8. Isotope labeling studies

The isotope labeling experiments were performed by Florian Rohrbacher (ETH Zurich) as part of our joint publication.\textsuperscript{10}

KAHA ligation of Fmoc-Ala-Leu-\(\alpha\)-ketoacid and H-Opr-(4-F)-Phe-OH in \(^{18}\)O labeled water

\[
\text{Fmoc-Ala-Leu-}\alpha\text{-ketoacid} \quad 4 \quad (0.8 \text{ mg, 1.8 } \text{\(\mu\text{mol, 1.0 equiv)}\)} \quad \text{and oxalic acid} \quad (0.8 \text{ mg, 8.9 } \text{\(\mu\text{mol, 5.0 equiv, final concentration: 0.10 M)}}\)
\]
were dissolved in a mixture of DMSO (70.8 \(\mu\)L) and \(H_2^{18}\)O (97 atom % \(^{18}\)O, 17.7 \(\mu\)L, 984 \(\mu\)mol, 556 equiv). The mixture was shaken in a sealed vial for 2 h at 60 °C. Analysis of the reaction mixture by LC-MS showed no detectable amounts of \(^{18}\)O incorporation in both starting materials and full incorporation of \(^{18}\)O in the product. For high-resolution MS, an aliquot of the reaction mixture was separated by analytical HPLC and the product-containing fraction was analyzed.

\textit{depsi-6}: \\
HR-MS (FTICR MALDI) m/z calculated for \(C_{37}H_{44}FN_4^{16}O_7^{18}O \quad [M+H]^+ \) 693.3180, found 693.3183, 95 % incorporation of \(^{18}\)O; 98 % incorporation yield. Similar observed for \textit{amide-6}.

\begin{center}
\textbf{MALDI FTICR MS of purified \textit{depsi-6} from the isotope labeling study.}
\end{center}

7.3. Experimental for “CHAPTER 3: New protecting groups for the synthesis of peptide α-ketoacids”

7.3.1. Synthesis of starting materials and relevant intermediates

Protecting groups for α-ketoacids

1-(4-methoxyphenyl)-2,2-dimethylpropane-1,3-diol S6

S6 was synthesized according to a procedure previously reported by Wynberg et al.\textsuperscript{11} KOH (28.1 g, 0.500 mol, 1.00 equiv) was dissolved in cooled EtOH (450 ml) and a mixture of p-anisaldehyde (68.1 g, 0.500 mol, 1.00 equiv) and isobutyraldehyde (72.1 g, 1.00 mol, 2.00 equiv) was added over 30 min, so that the reaction temperature did not exceed 30 °C. The yellow solution was heated to 55 °C overnight. After cooling to room temperature, H\textsubscript{2}O (500 ml) was added and the mixture extracted with CH\textsubscript{2}Cl\textsubscript{2} (3 x 200 ml). The combined extracts were washed with brine, dried over anhydrous Na\textsubscript{2}SO\textsubscript{4} and concentrated under reduced pressure. The crude brown solid was purified by two subsequent recrystallizations (cyclohexane:toluene 2:1) to give the product S6 (37.9 g, 0.180 mol, 36 %) as a colorless solid with identical spectroscopic properties as reported.

1-(4-(allyloxy)phenyl)-2,2-dimethylpropane-1,3-diol S7

S7 was synthesized according to a procedure previously reported by Wynberg et al.\textsuperscript{11} KOH (1.56 g, 0.0277 mol, 1.00 equiv) was dissolved in cooled EtOH (25 ml) and a mixture of 4-(allyloxy)benzaldehyde (4.50 g, 0.0277 mol, 1.00 equiv) and isobutyraldehyde (1.55 g, 0.0554 mol, 2.00 equiv) was added over 30 min, so that the reaction temperature did not exceed 30 °C. The yellow solution was heated to 60 °C for 24 h. After cooling to room temperature, H\textsubscript{2}O (500 ml) was added and the mixture extracted with CH\textsubscript{2}Cl\textsubscript{2} (3 x 200 ml). The combined extracts were washed with brine, dried over anhydrous Na\textsubscript{2}SO\textsubscript{4} and concentrated under reduced pressure. Purification by column chromatography (SiO\textsubscript{2}, cyclohexane:EtOAc 95:2 to 70:30) gave S7 (2.80 g, 11.8 mmol, 43 %) as colorless solid.

\textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) \( \delta = 7.25 – 7.21 \) (m, 2H), 6.92 – 6.84 (m, 2H), 6.06 (ddt, \( J = 17.3, 10.6, 5.3, 1H \)), 5.42 (dq, \( J = 17.2, 1.6, 1H \)), 5.29 (dq, \( J = 10.5, 1.4, 1H \)), 4.61 (s, 1H), 4.54 (dt, \( J = 5.3, 1.5, 2H \)), 3.61 – 3.46 (m, 2H), 0.86 (s, 3H), 0.83 (s, 3H).

\textsuperscript{13}C NMR (101 MHz, CDCl\textsubscript{3}) \( \delta = 158.20, 133.94, 133.41, 128.73, 117.84, 114.12, 82.06, 72.28, 68.97, 39.39, 22.86, 19.16. \)

IR (thin film, cm⁻¹): υ = 3349 (m), 2961 (m), 2872 (m), 1610 (m), 1509 (s), 1241 (s), 1035 (s).


4-(4-methoxyphenyl)-2,2,5,5,8,8-hexamethyl-3,7-dioxa-2,8-disilanonane 29

1-(4-methoxyphenyl)-2,2-dimethylpropane-1,3-diol S₆ (31.5 g, 0.150 mol, 1.00 equiv) was dissolved in dry CH₂Cl₂ (500 ml) and imidazole (30.6 g, 0.450 mol, 3.00 equiv) was added. Distilled TMSCl (40.8 g, 0.375 mol, 2.50 equiv) was added at 0 °C over 30 min and the reaction allowed to stir at room temperature for 1 h. Hexanes (250 ml) were added and the resulting suspension filtered. The filtrate was concentrated and purified by column chromatography on silica (hexane:EtOAc 9:1 with 0.1% NEt₃) to give 29 (43.3 g, 0.122 mol, 81%) as a colorless liquid.

¹H NMR (400 MHz, CDCl₃) δ = 7.22 (d, J = 8.5 Hz, 2H), 6.83 (d, J = 8.8 Hz, 2H), 4.60 (s, 1H), 3.83 (s, 3H), 3.44 (d, J = 9.4 Hz, 1H), 3.15 (d, J = 9.3 Hz, 1H), 0.84 (s, 3H), 0.69 (s, 3H), 0.13 (s, 9H), -0.02 (s, 9H).

¹³C NMR (101 MHz, CDCl₃) δ = 158.49, 134.87, 129.10, 112.58, 77.26, 69.21, 55.24, 40.79, 20.84, 19.55, 0.19, -0.36.

IR (thin film, cm⁻¹): υ = 2958 (s), 2903 (m), 1613 (m), 1511 (s), 1249 (s), 1078 (s), 876 (s).


4-(4-(allyloxy)phenyl)-2,2,5,5,8,8-hexamethyl-3,7-dioxa-2,8-disilanonane 43

1-(4-(allyloxy)phenyl)-2,2-dimethylpropane-1,3-diol S₇ (2.80 g, 0.0118 mol, 1.00 equiv) was dissolved in dry CH₂Cl₂ (50 ml) and imidazole (2.42 g, 35.4 mmol, 3.00 equiv) was added. Distilled TMSCl (2.82 g, 26.0 mmol, 2.20 equiv) was added at 0 °C over 30 min and the reaction allowed to stir at room temperature for 1 h. Hexanes (250 ml) were added and the resulting suspension filtered. The filtrate was concentrated and purified by column chromatography on silica (hexane:EtOAc 95:5 with 0.1% NEt₃) to give 43 (4.35 g, 0.0114 mol, 96%) as a colorless liquid.

¹H NMR (400 MHz, CDCl₃) δ = 7.22 – 7.11 (m, 2H), 6.84 – 6.75 (m, 2H), 6.08 (dtt, J=17.3, 10.6, 5.4, 1H), 5.42 (dq, J=17.2, 1.6, 1H), 5.28 (dq, J=10.5, 1.4, 1H), 4.57 (s, 1H), 4.52 (dt, J=5.3, 1.5, 2H), 3.41 (d, J=9.4, 1H), 3.12 (d, J=9.4, 1H), 0.81 (s, 3H), 0.66 (s, 3H), 0.10 (s, 9H), -0.05 (s, 9H).

¹³C NMR (101 MHz, CDCl₃) δ = 157.53, 135.06, 133.68, 129.09, 117.66, 113.36, 77.25, 69.19, 68.91, 40.78, 20.84, 19.55, 0.19, -0.36.

IR (thin film, cm⁻¹): υ = 2958 (s), 1611 (m), 1509 (s), 1222 (s), 1083 (s), 888 (s), 839 (s).

Enantioenriched \textit{S7} was synthesized according to a modified procedure from Ryu \textit{et al.}\textsuperscript{12} \((S)\)-diphenyl(pyrrolidin-2-yl)methanol (0.186 g, 0.600 mmol, 0.200 equiv) and phenylboroxine (0.0636 g, 0.204 mmol, 0.0667 equiv) were dissolved in 60 ml toluene and activated molecular sieves (4 Å, ca. 2 g) were added. The mixture was heated to reflux for 14 h. The molecular sieves were removed by filtration and the toluene removed under reduced pressure. Toluene (20 ml) was added to the residue and the mixture concentrated to a volume of ca. 5 ml. The clear solution was cooled to -40 °C and a solution of TfOH (0.2 M in toluene, 90.1 mg, 0.480 mmol, 0.160 equiv) was added dropwise over 15 min. After complete addition, the clear solution was stirred for another 30 min at -40 °C. A solution of triphenylphosphine oxide (0.418 g, 1.50 mmol, 0.500 equiv) in toluene (5 ml) was added over 10 min, followed by addition of a solution of 4-(allyloxy)benzaldehyde (0.487 g, 3.00 mmol, 1.00 equiv) in toluene (1 ml). After complete addition, ((1-methoxy-2-methylprop-1-en-1-yloxy)trimethylsilane (0.628 g, 3.60 mmol, 1.20 equiv) was added and the mixture stirred for 14 h at -40 °C. The reaction was quenched by addition of a saturated aqueous solution of NaHCO\textsubscript{3} (5 ml), warmed to room temperature and diluted with EtOAc (20 ml). The reaction mixture was washed with water, brine and an aqueous solution of citric acid (10 % w/v). The organic phase was dried over MgSO\textsubscript{4} and concentrated. The residual oil was used directly without purification in the next step.

\textit{LiAlH}_4 (0.227 g, 6.00 mmol, 2.00 equiv) was suspended in dry THF (10 ml) at 0 °C and a solution of of crude methyl \((R)\)-3-(4-(allyloxy)phenyl)-3-hydroxy-2,2-dimethylpropanoate (0.790 g) in THF (5 ml) was added dropwise. The reaction was stirred for 30 min at 0 °C, after which analysis by TLC indicated complete conversion. Excess \textit{LiAlH}_4 was quenched by the careful and slow addition of 0.25 ml water, 0.25 ml NaOH (15 % in water) and 0.75 ml water. Solids were removed by filtration and the filtrate washed with an aqueous solution of citric acid (10 % w/v), brine and the organic phase were dried over MgSO\textsubscript{4} and concentrated. Purification by column chromatography (SiO\textsubscript{2}, cyclohexane:EtOAc 9:1 to 3:2) gave \textit{S7}

(0.550 g, 2.33 mmol, 78 % over two steps) as colorless solid with identical spectroscopic properties as rac-S7.

Analysis by SFC (Daicel Chiralpak AD-H, 4.6 x 250 mm, gradient: 5 % iPrOH in CO2 to 50 % iPrOH in CO2 over 10 min; flow rate: 3.0 ml/min;) showed a ratio of enantiomers of 80:20 (60 % ee). The absolute configuration was assigned to (R) by comparing similar substrates in the literature reference.

SFC traces of racemic (bottom trace) and enantioenriched S7 (top trace, 60 % ee).

**Synthesis of linker 38**

1. allyl alcohol, PTSA, benzene, reflux
2. 4-hydroxybenzaldehyde, K2CO3, DMF, 70 °C
3. NaBH4, MeOH, 0 °C to rt

44 % yield over 3 steps

allyl 5-(4-(hydroxymethyl)phenoxy)pentanoate S8

In a 500 ml round bottom flask equipped with a Dean-Stark trap, 5-bromopentanoic acid (50.0 g, 0.276 mol, 1.00 equiv), allyl alcohol (24.1 g, 0.414 mol, 1.50 equiv) and p-toluenesulfonic acid (0.475 g, 2.76 mmol, 0.0100 equiv) were dissolved in benzene (275 ml) and heated to reflux for 3.5 h, after which analysis by TLC indicated complete conversion. The solvent was removed under reduced pressure, the residual oil redissolved in MTBE, washed with a saturated aqueous solution of NaHCO3 and brine, dried over MgSO4 and concentrated under reduced pressure.

The crude allyl 5-bromopentanoate (54.1 g, 0.245 mol) and p-hydroxybenzaldehyde (32.9 g, 0.269 mol, 1.10 equiv) were dissolved in DMF (250 ml). K2CO3 (71.1 g, 0.515 mol, 2.10 equiv) was added and the resulting suspension stirred at 70 °C for 2 h, after which analysis by TLC indicated complete conversion. The mixture was cooled to room temperature, diluted with H2O (750 ml), extracted with MTBE (3 x 200 ml) and the combined extracts washed with H2O and brine, dried over MgSO4 and concentrated.
Crude allyl 5-(4-formylphenoxy)pentanoate (61.3 g, 0.233 mol) was dissolved in MeOH (500 ml) and NaBH₄ (4.42 g, 0.117 mol, 0.500 equiv) was added in portions at 0 °C. After complete addition, the mixture was warmed to room temperature and stirred for 30 min after which analysis by TLC indicated complete conversion. Excess NaBH₄ was quenched by the careful addition of ice and the mixture was acidified to pH 2–3 by the addition of an aqueous solution of citric acid (10% w/v). MeOH was removed under reduced pressure. The remaining aqueous phase was extracted with MTBE and the combined extracts washed with brine, dried and concentrated. The crude oil was purified by column chromatography (SiO₂, hexane:MTBE 1:2:1 to 1:1) to give S₈ as a colorless oil (32.0 g, 0.121 mol, 44 % over 3 steps) that solidified in the freezer.

**¹H NMR** (400 MHz, CDCl₃) δ = 7.42 – 7.10 (m, 2H), 6.95 – 6.74 (m, 2H), 6.05 – 5.79 (m, 1H), 5.41 – 5.06 (m, 2H), 4.61 (s, 2H), 4.58 (dd,  J=5.8, 1.5, 2H), 4.04 – 3.91 (m, 2H), 2.47 – 2.35 (m, 2H), 1.91 – 1.76 (m, 4H).

**¹³C NMR** (101 MHz, CDCl₃) δ = 173.21, 158.73, 133.22, 132.37, 128.77, 118.38, 114.68, 67.55, 65.24, 65.22, 33.99, 28.81, 21.78.

**IR** (thin film, cm⁻¹): ν = 3413 (m), 2941 (m), 2872 (m), 1731 (s), 1611 (m), 1510 (s), 1242 (s), 1167 (m), 826 (m).


**allyl 5-(4-(chloromethyl)phenoxy)pentanoate 38**

To an ice-cold solution of allyl 5-(4-(hydroxy-methyl)phenoxy)pentanoate S₈ (6.61 g, 25.0 mmol) in dry CH₂Cl₂ (50 ml) was added SOCl₂ (3.87 g, 32.5 mmol, 1.30 equiv) slowly via a syringe. After complete addition, a few drops of DMF were added and the mixture was stirred at room temperature for 1 h, after which TLC indicated complete conversion. To quench excess SOCl₂, ice was added to the mixture and stirring continued for 10 min. The mixture was washed with H₂O and brine, dried over MgSO₄ and concentrated. Pure allyl 5-(4-(chloromethyl)phenoxy)pentanoate 38 (6.79 g, 24.0 mmol, 96 %) was obtained as colorless oil that solidified in the freezer.

**¹H NMR** (400 MHz, CDCl₃) δ = 7.32 – 7.27 (m, 2H), 6.89 – 6.82 (m, 2H), 5.92 (ddt,  J=17.2, 10.4, 5.7, 1H), 5.39 – 5.20 (m, 2H), 4.58 (dt,  J=5.7, 1.4, 2H), 4.56 (s, 2H), 4.02 – 3.88 (m, 2H), 2.46 – 2.37 (m, 2H), 1.87 – 1.81 (m, 4H).

**¹³C NMR** (101 MHz, CDCl₃) δ = 173.15, 159.17, 132.35, 130.16, 129.76, 118.36, 114.78, 67.55, 65.20, 46.45, 33.95, 28.74, 21.74.

**IR** (thin film, cm⁻¹): ν = 2945 (s), 2873 (s), 1736 (s), 1611 (s), 1584 (m), 1513 (s), 1242 (s), 1174 (s).

Synthesis of 4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethylbutanoic acid 47

Diisopropylamine (5.21 g, 7.24 ml, 51.5 mmol, 1.15 equiv) was dissolved in THF (25 ml) and cooled to -78 °C. A solution of nBuLi (1.6 M, 30.8 ml, 49.3 mmol, 1.10 equiv) was added and stirred for 10 min at -78 °C and 30 min at 0 °C. The slightly yellowish solution of LDA was cooled to -78 °C and NMP (4.44 g, 4.31 ml, 4.48 mmol, 1.00 equiv) was added over 5 min. The reaction was stirred for 30 min at -78 °C and MeI (6.99 g, 3.08 ml, 49.3 mmol, 1.10 equiv) was added dropwise over 30 min and stirred for 1 h. A solution of LDA (made as before: diisopropylamine (5.21 g, 7.24 ml, 51.5 mmol, 1.15 equiv) was dissolved in THF (25 ml) and cooled to -78 °C. A solution of nBuLi (1.6 M, 30.8 ml, 49.3 mmol, 1.10 equiv) was added and stirred for 10 min at -78 °C and 30 min at 0 °C.) was added to the reaction mixture over 10 min at -78 ºC and stirred for 1 h. Mel (6.99 g, 3.08 ml, 49.3 mmol, 1.10 equiv) was added dropwise over 30 min and stirred for 2 h. The reaction was allowed to warm to 0 ºC and quenched by the addition of 1 M HCl. The reaction mixture was diluted with CH₂Cl₂ (200 ml), washed with H₂O (2 x 100 ml), brine, dried over MgSO₄ and concentrated under reduced pressure. Destillation under high vacuum yielded 1,3,3-trimethylpyrrolidin-2-one (3.35 g, 26.3 mmol, 59 %) as colorless oil, which was used directly in the next step.

1,3,3-trimethylpyrrolidin-2-one (2.40 g, 18.8 mmol, 1.00 equiv) was dissolved in conc. HBr (47 % in H₂O, 15 ml) and heated to reflux for 6 days. Volatiles were carefully removed under reduced pressure and the obtained brownish crystals were dissolved in H₂O (20 ml). The pH of the solution was adjusted to 7.5 by addition of solid Na₂CO₃ and a solution of Boc₂O (4.52 g, 20.7 mmol, 1.1 equiv) in dioxane (20 ml) was added. The mixture was stirred at room temperature for 12 h. The mixture was acidified to pH 2 by addition of 1 M HCl and extracted with EtoAc. The combined organic phases were washed with brine, dried over MgSO₄ and concentrated under reduced pressure. Purification by column chromatography (SiO₂, cyclohexane:EtOAc:AcOH 90:10:0.1 to 70:30:0.1) gave 47 as colorless crystalline solid (2.8 g, 11.4 mmol, 60 %).

¹H NMR (400 MHz, CDCl₃) δ = 3.30 – 3.14 (m, 2H), 2.81 (s, 3H), 1.79 – 1.69 (m, 2H), 1.43 (s, 9H), 1.21 (s, 6H).

¹³C NMR (101 MHz, CDCl₃) δ = 183.51, 155.77, 79.60, 45.58, 40.82, 37.89, 34.21, 28.54, 25.14.

IR (thin film, cm⁻¹): ν = 3159 (m), 2975 (s), 2933 (m), 1698 (s), 1479 (m), 1395 (m), 1164 (s).

MS (ESI): calculated for C₁₂H₂₂N₁O₄ [M-H]: 244.1554, found: 244.1557.
7.3.2. Optimization of $\alpha$-ketoacid protection reaction using acid labile protecting groups

Screening of protecting groups for acid labile protection of $\alpha$-ketoacids

In Table S2 a list of evaluated protecting groups is shown.\(^{13}\) For the screening reactions, Fmoc-leucine $\alpha$-ketoacid 7 (0.2 mmol, 1 equiv) was dissolved in dry toluene (2 ml), protecting group 26 – 29 (0.6 mmol, 3 equiv) and TMSOTf (0.04 mmol, 0.2 equiv) were added. The reaction was stirred at room temperature and analyzed by LC-MS. Isolated products (ca. 5–10 mg) were dissolved in 97:3 TFA:H$_2$O (0.5 ml) and incubated for 2 h. Volatiles were removed under a stream of nitrogen and the crude analyzed by LC-MS.

Table S2: Evaluation of protecting groups for $\alpha$-ketoacids.

<table>
<thead>
<tr>
<th>Protecting group</th>
<th>product</th>
<th>protection yield</th>
<th>deprotection</th>
<th>epimerization</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>30</td>
<td>20 %</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>25</td>
<td>31</td>
<td>&lt; 5 %</td>
<td>yes</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>32</td>
<td>&lt; 5 %</td>
<td>no</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>33</td>
<td>&lt; 5 %</td>
<td>yes</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>34</td>
<td>29 %</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>29</td>
<td>35</td>
<td>28 % [a]</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

\[^{[a]}\] Initial, unoptimized yield. \[^{[b]}\] Yield after optimization. PMP = p-methoxyphenyl.

Protecting group 29 gave the best yield in the initial screening, was removable under deprotection conditions and was therefore selected for further optimization. All other protecting groups were not further investigated.

Reaction optimization

Fmoc-leucine α-ketoacid 7 was dissolved in dry solvent, catalyst and protecting group 29 were added and the reaction stirred at indicated temperature. Reactions were monitored by LC-MS.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Equivalents of 29</th>
<th>Catalyst [loading]</th>
<th>Solvent</th>
<th>Temperature</th>
<th>yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>TMSOTf [20 mol %]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>28 %</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>TMSOTf [20 mol %]</td>
<td>CH₂Cl₂</td>
<td>-78 °C to 0 °C</td>
<td>23 %</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>TMSOTf [20 mol%]</td>
<td>THF</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Tf₂O [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>26 %</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>Ms₂O [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>SnCl₂ [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>BF₃•THF [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>TBAF [20 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>Ti(OiPr)₄ [100 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>n.r.</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>31 %</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>50 °C</td>
<td>30 %</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>-20 °C</td>
<td>9 %</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>Tf₂O [10 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>18 %</td>
</tr>
<tr>
<td>14</td>
<td>1.5 [addition over 9h]</td>
<td>TMSOTf [10 mol%]</td>
<td>Toluene</td>
<td>r.t.</td>
<td>66 %</td>
</tr>
</tbody>
</table>

CH₂Cl₂ as solvent was found to perform comparable to toluene, whereas no product was obtained in THF (entries 1-3). Tf₂O gave similar yields as TMSOTf (entries 1 and 4), whereas all other catalysts investigated gave no product (entries 4-9). The yields were slightly higher if the catalyst loading was reduced to 10 % (entry 10). Increasing the reaction temperature to 50 °C had no effect on the yield, whereas lowering the temperature to – 20 °C lowered the yield dramatically (entries 11 and 12). Attempts to increase the concentration of the reaction resulted in a significant decrease in yield (entry 13). We found it to be crucial to add the protecting group 29 very slowly (over 9 to 12 hours) to the reaction. Under these conditions, a smaller excess of 29 could be used.

[a] 0.25 M substrate concentration.
Screening of deprotection conditions for protected Fmoc-leucine α-ketoacid 7

Model peptides immobilized on 2-chlorotriyl resin were treated with mixtures of TFA, indicated scavengers and H₂O (95:2.5:2.5) for 2 h at room temperature and analyzed by HPLC and LC-MS.

![Chemical Structure](image)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Scavenger</th>
<th>Model Peptide</th>
<th>Product yield by HPLC</th>
<th>Major side Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TIPS</td>
<td>A</td>
<td>0 %</td>
<td>Reduction of Ketone</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thioacetal formation[^a]</td>
</tr>
<tr>
<td>2</td>
<td>Ethanolthiol</td>
<td>B</td>
<td>&lt; 5 %</td>
<td>Incomplete Trt-deprotection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thioacetal formation</td>
</tr>
<tr>
<td>3</td>
<td>Thioanisole</td>
<td>B</td>
<td>&lt; 5 %</td>
<td>Incomplete Trt-deprotection</td>
</tr>
<tr>
<td>4</td>
<td>2,6-dimethylbenzenethiol</td>
<td>B</td>
<td>40 %</td>
<td>Incomplete Trt-deprotection</td>
</tr>
<tr>
<td>5</td>
<td>Cyclohexanethiol</td>
<td>B</td>
<td>&lt; 5%</td>
<td>Incomplete Trt-deprotection</td>
</tr>
<tr>
<td>6</td>
<td>4-methylbenzenethiol</td>
<td>B</td>
<td>80 %</td>
<td>Unidentified side product</td>
</tr>
<tr>
<td>7</td>
<td>1,2-ethanedithiol</td>
<td>B</td>
<td>&gt; 80 %</td>
<td>&lt; 5 % Thioacetal formation</td>
</tr>
</tbody>
</table>

[^a]: Proposed thioacetal side product based on LC-MS analysis:

1,2-ethanethiol was found to be the best of the evaluated scavengers. Further experiments indicated that it can be replaced by the non-malodorous 2,2’-(ethylene-dioxy)diethanethiol (DODT) without any effect on the reaction. Similarly, the presence of H₂O in the cleavage cocktail does not have a large impact on the reaction and can be omitted if desired.

Epimerization assay for protected Fmoc-Leu-α-ketoacid 35

To establish that the stereocenter of the protected Fmoc-Leu-α-ketoacid 35 was not epimerized during the protection or deprotection steps, a SFC assay was developed. During the studies it was found that the enantiomers of the Fmoc-Leu-α-ketoacid 7 are not very well separated by SFC under various conditions. Derivatization by KAHA ligation gave amides 37, which were found to be readily separable by SFC (Daicel Chiralpak OJ-H, 4.6 x 250 mm,
gradient: 5% $^3$PrOH in CO$_2$ to 50% $^3$PrOH in CO$_2$ over 10 min; flow: 3.0 mL/min.). Authentic samples of each enantiomer of the amide 37 were prepared by standard peptide coupling reactions of commercially available 2-(4-fluorophenyl)ethan-1-amine and Fmoc-(S)-Leu-OH or Fmoc-(R)-Leu-OH, respectively.

SFC traces of amide 37 obtained from deprotection and ligation of protected α-ketoacid 35 (bottom trace) and authentic samples of (S)-37 and (R)-37 (middle and top trace).

$(9H$-fluoren-9-yl)methyl $(S)$-1-((4-fluorophenethyl)amino)-4-methyl-1-oxopentan-2-yl carbamate 37:

$^1$H NMR (400 MHz, DMSO-d$_6$) $\delta$ = 7.92 (d, $J$=5.4, 1H), 7.90 – 7.86 (m, 2H), 7.76 – 7.70 (m, 2H), 7.45 – 7.37 (m, 3H), 7.36 – 7.25 (m, 2H), 7.25 – 7.17 (m, 2H), 7.10 – 7.00 (m, 2H), 4.39 – 4.26 (m, 1H), 4.27 – 4.16 (m, 2H), 3.95 (ddd, $J$=9.9, 8.3, 5.2, 1H), 3.32 – 3.14 (m, 2H), 2.69 (t, $J$=7.2, 2H), 1.57 – 1.46 (m, 1H), 1.45 – 1.25 (m, 2H), 0.85 (d, $J$=6.5, 3H), 0.81 (d, $J$=6.5, 3H).

$^{13}$C NMR (101 MHz, DMSO-d$_6$) $\delta$ = 172.13, 161.98, 159.58, 155.83, 143.90, 143.75, 140.70, 140.69, 135.49, 135.46, 130.48, 130.40, 127.60, 127.00, 125.32, 125.28, 120.08, 120.06, 114.92, 114.72, 65.49, 53.09, 46.69, 40.85, 40.00, 38.89, 34.08, 24.14, 22.92, 21.51.

$[\alpha]_D^{26}$ = -21.4° (S-37, c = 1.0, CH$_2$Cl$_2$).

IR (thin film, cm$^{-1}$): $\nu$ = 2955 (m), 1690 (w), 1653 (m), 1537 (s), 1509 (s), 1449 (w), 1234 (s).

MS (ESI, [M+H]$^+$): calculated for C$_{29}$H$_{32}$F$_1$N$_2$O$_3$ [M+H]$^+$: 475.2391, found: 475.2386.
7.3.3. Synthesis of protected Fmoc-Leu α-ketoacid 40

To a solution of Fmoc-(S)-leucine α-ketoacid 7 (9.70 g, 25.4 mmol, 1.00 equiv) in dry toluene (500 ml) was added TMSOTf (0.565 g, 2.54 mmol, 0.100 equiv). Over the course of 9 h, 29 (13.5 g, 38.1 mmol, 1.50 equiv) was added to the reaction at room temperature. After complete addition, the reaction was stirred for another 3 h; H₂O (ca. 300 ml) was added and the phases separated, the organic phase was washed with brine, dried over Na₂SO₄ and adsorbed on SiO₂. After purification by column chromatography (SiO₂, hexane:EtOAc:MeOH:AcOH 70:30:2:1) and drying under high vacuum the product 35 (mixture of non-separable diastereomers, 9.60 g, 16.7 mmol, 66%) was obtained as a white powder. Note: Not all signals of the diastereomers are resolved in the NMR data.

**1H NMR** (400 MHz, CDCl₃) δ = 7.77 – 7.70 (m, 2H), 7.57 (t, J=7.8, 2H), 7.43 – 7.20 (m, 4H), 7.13 (dd, J=8.9, 2.7, 2H), 6.82 (dd, J=8.8, 8.1, 2H), 5.05 (t, J=10.4, 1H), 4.63 – 4.48 (m, 1H), 4.47 – 4.34 (m, 2H), 4.27 – 4.14 (m, 2H), 3.78 (s, 3H), 3.77 – 3.55 (m, 2H), 1.79 – 1.51 (m, 3H), 1.04 – 0.83 (m, 9H), 0.70 – 0.56 (m, 3H).

**13C NMR** (101 MHz, CDCl₃) δ = 171.03, 170.77, 159.32, 157.10, 156.97, 144.23, 144.12, 143.77, 143.75, 141.45, 141.41, 129.46, 129.23, 128.70, 128.68, 127.76, 127.72, 127.21, 127.17, 125.39, 125.32, 125.25, 120.02, 113.24, 113.22, 101.18, 100.93, 83.19, 82.60, 74.78, 74.65, 67.30, 67.21, 55.39, 54.93, 47.33, 37.69, 33.99, 33.96, 24.64, 24.61, 24.04, 21.90, 21.61, 21.48, 18.60, 18.48.

**IR** (thin film, cm⁻¹): ν = 3436 (w), 3067(w), 2957 (m), 1725 (s), 1513 (s), 1450 (m), 1249 (s).

**MS** (ESI): calculated for C₃₄H₄₅N₁O₇ [M+H]⁺: 574.2799, found: 574.2794.
4-((5-(allyloxy)-5-oxopentyl)oxy)benzyl 2-((S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylate (protected Fmoc-Leu α-ketoacid linker allyl ester) 39

Protected Fmoc-leucine α-ketoacid 35 (6.80 g, 11.9 mmol, 1.0 equiv), allyl 5-(4-(chloromethyl)phenoxy)pentanoate 38 (4.70 g, 16.6 mmol, 1.4 equiv) and tetrabutylammonium sulfate (as 50 % aqueous solution, 0.690 g, 1.19 mmol, 0.100 equiv) were stirred at 50 °C in a biphasic mixture of toluene (20 ml) and aqueous NaHCO₃ solution (20 ml, 2.00 g NaHCO₃, 23.8 mmol, 2.00 equiv) for 14 h, after which LC-MS indicated complete conversion. The mixture was extracted with EtOAc (3 x 50 ml) and the combined organic phases were washed with H₂O and brine. The solvent was removed under reduced pressure and the residue purified by column chromatography (SiO₂, hexane:EtOAc 4:1) to give 39 (inseparable mixture of diastereomers, 7.03 g, 8.60 mmol, 72 %) as white foam. Note: Not all signals of the diastereomers are resolved in the NMR data.

**1H NMR** (400 MHz, CDCl₃) δ = 7.80 – 7.73 (m, 2H), 7.66 – 7.55 (m, 2H), 7.44 – 7.24 (m, 6H), 7.15 – 7.06 (m, 2H), 6.89 – 6.74 (m, 4H), 5.98 – 5.86 (m, 1H), 5.35 – 4.88 (m, 5H), 4.61 – 4.56 (m, 2H), 4.50 – 4.28 (m, 3H), 4.25 – 4.17 (m, 2H), 3.95 – 3.88 (m, 2H), 3.81 – 3.75 (m, 3H), 3.71 – 3.53 (m, 2H), 2.45 – 2.38 (m, 2H), 1.84 – 1.76 (m, 4H), 1.72 – 1.32 (m, 3H), 0.95 – 0.83 (m, 9H), 0.60 (d, J=9.4, 3H).

**13C NMR** (101 MHz, CDCl₃) δ = 173.15, 169.20, 169.16, 159.29, 159.27, 159.23, 156.38, 156.31, 144.32, 144.25, 144.09, 144.03, 141.43, 141.40, 141.38, 132.36, 130.89, 130.73, 129.70, 129.45, 128.77, 128.73, 128.66, 127.78, 127.72, 127.69, 127.66, 127.19, 127.16, 127.12, 125.39, 125.33, 125.29, 120.02, 120.00, 118.37, 114.62, 114.53, 113.17, 101.43, 101.32, 82.93, 82.32, 74.67, 74.58, 67.50, 67.30, 67.23, 66.98, 66.92, 65.20, 55.38, 55.37, 54.91, 54.86, 47.44, 47.41, 38.06, 33.96, 33.95, 33.91, 28.74, 24.59, 24.52, 24.04, 21.94, 21.73, 21.58, 21.41, 18.64, 18.52.

**IR** (thin film, cm⁻¹): ν = 2952 (w), 1735 (s), 1513 (s), 1247 (w), 1173 (s).

**MS** (MALDI): calculated for C₄₉H₅₇N₁Na₁O₁₀ [M+Na⁺]: 842.3875, found: 842.3875.

5-(4-(((2-((S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carbonyl)oxy)methyl)phenoxy)pentanoic acid (protected Fmoc-Leu α-ketoacid linker free acid) 40

Allyl ester 39 (2.8 g, 3.4 mmol, 1.0 equiv) and Pd(PPh₃)₄ (99 mg, 85 µmol, 0.025 equiv) were dissolved in dry, degassed CH₂Cl₂ (35 ml) and cooled to 0 °C. Morpholine (0.62 g, 7.2 mmol, 2.1 equiv) was added and the mixture
stirred for 30 min at 0 °C. After completion of the reaction as indicated by TLC and LC-MS, the mixture was diluted with CH2Cl2, washed with an aqueous solution of citric acid (10 % w/v) and brine, dried over MgSO4 and concentrated under reduced pressure. Purification by column chromatography (SiO2, hexane:MTBE:AcOH 80:20:0.1 to 50:50:0.1) gave 40 (inseparable mixture of diastereomers, 2.4 g, 3.1 mmol, 90 %) as white foam. Note: Not all signals of the diastereomers are resolved in the NMR data.

1H NMR (400 MHz, CDCl3) δ = 7.79 – 7.71 (m, 2H), 7.63 – 7.55 (m, 2H), 7.42 – 7.34 (m, 2H), 7.32 – 7.23 (m, 4H), 7.15 – 7.05 (m, 2H), 6.91 – 6.74 (m, 4H), 5.91 (s, very broad, 1H), 5.35 – 4.89 (m, 2H), 4.50 – 4.26 (m, 3H), 4.24 – 4.13 (m, 2H), 3.94 – 3.84 (m, 2H), 3.82 – 3.76 (m, 3H), 3.73 – 3.51 (m, 2H), 2.47 – 2.37 (m, 2H), 1.86 – 1.73 (m, 4H), 1.71 – 1.58 (m, 1H), 1.55 – 1.22 (m, 2H), 0.94 – 0.83 (m, 9H), 0.65 – 0.55 (m, 3H).

13C NMR (101 MHz, CDCl3) δ = 178.28, 169.28, 169.21, 159.29, 159.23, 159.19, 156.45, 156.39, 144.28, 144.22, 144.08, 144.02, 141.43, 141.40, 141.39, 130.90, 130.85, 130.75, 129.68, 129.44, 128.74, 128.66, 127.79, 127.74, 127.71, 127.66, 127.21, 127.17, 127.13, 125.39, 125.34, 125.29, 120.03, 120.01, 114.64, 114.54, 113.18, 101.44, 101.32, 82.95, 82.35, 74.69, 74.59, 67.45, 67.35, 67.26, 67.03, 66.98, 55.39, 54.92, 54.86, 47.42, 47.39, 38.05, 33.96, 33.92, 33.50, 28.63, 24.58, 24.52, 24.05, 21.95, 21.58, 21.50, 21.41, 18.63, 18.51.

IR (thin film, cm⁻¹): ν = 3425 (s), 2956 (w), 1709 (m), 1640 (m), 1513 (m), 1248 (m), 1174 (m).

MS (ESI): calculated for C46H54N1O10 [M+H]⁺: 780.3742, found: 780.3735.
7.3.4. Synthesis of orthogonally protected Fmoc-leucine α-ketoacid 49

To a solution of Fmoc-((S)-leucine α-ketoacid 7 (7.72 g, 15.0 mmol, 1.00 equiv) in dry toluene (600 ml) was added TMSOTf (0.334 g, 1.50 mmol, 0.100 equiv). Over the course of 12 h, 43 (8.57 g, 22.5 mmol, 1.50 equiv) was added to the reaction at room temperature. After complete addition, the reaction was stirred for another 2 h; H₂O (ca. 400 ml) was added and the phases separated, the organic phase was washed with brine, dried over Na₂SO₄ and adsorbed on SiO₂. After purification by column chromatography (SiO₂, hexane:EtOAc:MeOH:AcOH 70:30:2:1) and drying under high vacuum the product 44 (mixture of non-separable diastereomers, 4.60 g, 7.67 mmol, 51 %) was obtained as a white powder. Note: Not all signals of the diastereomers are resolved in the NMR data.

**1H NMR** (400 MHz, CDCl₃) δ = 7.77 – 7.69 (m, 2H), 7.57 (t, J=8.4, 2H), 7.41 – 7.30 (m, 2H), 7.30 – 7.20 (m, 2H), 7.18 – 7.09 (m, 2H), 6.91 – 6.79 (m, 2H), 6.13 – 5.96 (m, 1H), 5.48 – 5.36 (m, 1H), 5.32 – 5.22 (m, 1H), 4.56 – 4.47 (m, 3H), 4.46 – 4.33 (m, 2H), 4.29 – 4.13 (m, 2H), 3.88 – 3.46 (m, 2H), 1.74 – 1.48 (m, 3H), 1.06 – 0.79 (m, 9H), 0.67 – 0.58 (m, 3H).

**13C NMR** (101 MHz, CDCl₃) δ = 170.92, 170.71, 158.35, 156.87, 156.79, 144.23, 143.77, 141.41, 133.41, 133.39, 129.90, 128.69, 127.72, 127.68, 127.16, 127.13, 125.40, 125.27, 125.26, 119.99, 117.86, 117.82, 117.74, 114.04, 114.02, 83.19, 82.50, 74.82, 74.62, 68.95, 67.17, 54.94, 54.90, 47.35, 37.76, 33.99, 33.95, 24.65, 24.61, 24.09, 21.92, 21.65, 21.54, 18.62, 18.49.
IR (thin film, cm\(^{-1}\)): \(\tilde{\nu} = 2957\) (s), 2870 (w), 1725 (s), 1612 (m), 1511 (s), 1246 (s), 1175 (m).

MS (ESI): calculated for C\(_{36}\)H\(_{40}\)N\(_{1}\)O\(_{7}\) [M-H]: 598.2810, found: 598.2823.

2-(((S)-1-(((9H-fluoren-9-yl)methoxy) carbonyl)amino)-3-methylbutyl)-4-(4-hydroxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylic acid (Fmoc-Leu (phenol-protected) \(\alpha\)-ketoacid) 45

Orthogonally protected Fmoc-Leu-\(\alpha\)-ketoacid 44 (2.50 g, 4.16 mmol, 1.00 equiv) was dissolved in dry, degassed CH\(_2\)Cl\(_2\) (8.0 ml) and cooled to 0 °C. Pd(PPh\(_3\))\(_4\) (0.120 g, 0.104 mmol, 0.0250 equiv) and PhSiH\(_3\) (1.35 g, 12.5 mmol, 3.00 equiv) were added and the mixture stirred for 30 min at 0 °C. After completion of the reaction as indicated by TLC and LC-MS, the mixture was diluted with CH\(_2\)Cl\(_2\) and excess PhSiH\(_3\) was quenched by careful addition of an aqueous solution of citric acid (10 % w/v). The organic phase was washed with brine, dried over MgSO\(_4\) and concentrated under reduced pressure. Purification by column chromatography (SiO\(_2\), cyclohexane:EtOAc 90/10 to cyclohexane:EtOAc:MeOH:AcOH 50/40/3/1) gave 45 (inseparable mixture of diastereomers, 2.17 g, 3.88 mmol, 93 %) as white foam. Note: Not all signals of the diastereomers are resolved in the NMR data.

\(^1\)H NMR (400 MHz, CDCl\(_3\) \(\delta = 7.97\) (s, 1H), 7.79 – 7.70 (m, 2H), 7.64 – 7.48 (m, 2H), 7.42 – 7.31 (m, 2H), 7.30 – 7.16 (m, 2H), 7.06 – 6.98 (m, 1H), 6.75 – 6.65 (m, 2H), 6.45 – 6.38 (m, 1H), 5.14 (dd, \(J = 14.9, 10.6, 1H\)), 4.58 – 4.27 (m, 3H), 4.25 – 3.95 (m, 2H), 3.85 – 3.44 (m, 2H), 1.73 – 1.52 (m, 2H), 1.44 – 1.23 (m, 1H), 0.98 – 0.76 (m, 9H), 0.57 (m, 3H).

\(^{13}\)C NMR (101 MHz, CDCl\(_3\) \(\delta = 171.52, 170.04, 158.35, 157.21, 155.27, 154.64, 144.08, 143.67, 143.63, 143.59, 141.47, 141.43, 141.41, 129.09, 128.83, 128.52, 128.26, 127.99, 127.93, 127.79, 127.75, 127.35, 127.21, 127.16, 125.31, 125.22, 125.02, 124.72, 120.18, 120.16, 120.05, 120.02, 114.73, 114.61, 101.53, 100.99, 83.29, 83.17, 74.67, 74.49, 67.50, 67.35, 54.95, 54.69, 47.27, 47.18, 37.63, 37.08, 33.90, 33.85, 24.62, 24.00, 23.71, 21.83, 21.67, 21.55, 21.22, 18.40, 18.37.

IR (thin film, cm\(^{-1}\)): \(\tilde{\nu} = 3393\) (m), 2958 (s), 2871 (m), 1701 (s), 1615 (m), 1518 (s), 1224 (s), 1169 (m), 735 (s).

MS (ESI): calculated for C\(_{33}\)H\(_{38}\)N\(_{1}\)O\(_{7}\) [M+H]\(^+\): 560.2643, found: 560.2642.
4-((5-(allyloxy)-5-oxopentyl)oxy)benzyl 2-((S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-hydroxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylate (Fmoc-Leu (phenol-protected) α-ketoacid with protected linker) 46

45 (3.00 g, 5.36 mmol, 1.00 equiv) was dissolved in toluene (25 ml) and allyl 5-(4-(chloromethyl)phenoxy)pentanoate 38 (2.27 g, 8.04 mmol, 1.50 equiv), NaHCO₃ (0.675 g, 8.04 mmol, 1.5 equiv), tetrabutylammonium sulfate (as 50 % aqueous solution, 0.620 ml, 0.536 mmol, 0.100 equiv) and water (25 ml) were added. The biphasic mixture was vigorously stirred and heated to 55 °C for 72 h, after which the conversion (70 % as judged by LC-MS) did not increase further. The mixture was cooled to room temperature, diluted with toluene (25 ml), and acidified with citric acid (10 % w/v) to pH 2-3. The phases were separated and the organic phase was washed with water, brine, dried over MgSO₄ and concentrated under reduced pressure. Purification by column chromatography (SiO₂, cyclohexane:EtOAc 9:1 to 7:3) gave 46 (inseparable mixture of diastereomers, 2.80 g, 3.47 mmol, 65 %) as white foam.

Note: Not all signals of the diastereomers are resolved in the NMR data.

¹H NMR (400 MHz, CDCl₃) δ = 7.81 – 7.66 (m, 2H), 7.64 – 7.52 (m, 2H), 7.43 – 7.20 (m, 6H), 7.03 – 6.88 (m, 2H), 6.84 – 6.71 (m, 4H), 6.20 (s, 2H), 5.98 – 5.86 (m, 1H), 5.41 – 5.18 (m, 2H), 5.15 – 4.91 (m, 2H), 4.59 (dt, J=5.8, 1.4, 2H), 4.45 – 4.28 (m, 3H), 4.26 – 4.09 (m, 2H), 3.96 – 3.86 (m, 2H), 3.73 – 3.48 (m, 2H), 2.42 (td, J=6.1, 2.6, 2H), 1.88 – 1.75 (m, 4H), 1.70 – 1.58 (m, 1H), 1.55 – 1.24 (m, 2H), 0.95 – 0.69 (m, 9H), 0.60 – 0.52 (m, 3H).

¹³C NMR (101 MHz, CDCl₃) δ = 173.54, 173.46, 169.37, 169.27, 159.26, 159.22, 156.64, 156.49, 155.68, 155.61, 144.19, 144.12, 143.95, 143.93, 141.41, 141.39, 141.37, 141.36, 132.23, 132.20, 130.96, 130.88, 130.82, 130.76, 129.27, 129.08, 128.93, 128.86, 128.82, 128.78, 128.83, 127.82, 127.75, 127.74, 127.71, 127.58, 127.24, 127.20, 127.17, 127.13, 125.34, 125.31, 125.28, 125.22, 120.02, 120.00, 118.51, 118.47, 114.64, 114.54, 114.43, 101.41, 101.30, 83.04, 82.45, 74.69, 74.61, 67.52, 67.49, 67.48, 67.36, 67.22, 67.08, 67.02, 65.36, 65.32, 54.94, 54.85, 47.36, 47.34, 37.98, 37.91, 33.98, 33.93, 33.88, 33.85, 28.69, 24.55, 24.50, 24.47, 24.03, 24.00, 21.92, 21.89, 21.73, 21.72, 21.55, 21.36, 18.56, 18.45.

IR (thin film, cm⁻¹): ν = 3365 (m), 2956 (s), 2871 (m), 1735 (s), 1515 (s), 1247 (s), 1170 (m), 738 (m).

4-((5-(allyloxy)-5-oxopentyl)oxy)benzyl 2-(((S)-1-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-((4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethylbutanoyl)oxy)phenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylate (Fmoc-Leu (phenolester-protected) α-ketoacid with protected linker) 48

4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethyl butanoic acid 47 (0.285 g, 1.16 mmol, 1.25 equiv) and COMU (0.495 g, 1.16 mmol, 1.25 equiv) were dissolved in CH$_2$Cl$_2$ (1.25 ml) and DIPEA (0.299 g, 0.394 ml, 2.31 mmol, 2.50 equiv) was added. The red solution was mixed for 3 min and added to a solution of free phenol 46 (0.745 g, 0.924 mmol, 1.00 equiv) in CH$_2$Cl$_2$ (0.750 ml). DMAP (11.3 mg, 92.4 µmol, 0.100 equiv) was added and the reaction was stirred for 1.5 h. The reaction was diluted with CH$_2$Cl$_2$ (ca. 15 ml) and an aqueous solution of citric acid (10 % m/v) was added. The phases were separated and the organic phase was washed with brine, dried over MgSO$_4$ and concentrated under reduced pressure. Purification by column chromatography (SiO$_2$, cyclohexane:EtOAc 9:1 to 8:2) gave 48 (inseparable mixture of diastereomers, 0.380 g, 0.368 mmol, 40 %) as white foam. Note: Not all signals of the diastereomers are resolved in the NMR data.

$^{1}$H NMR (400 MHz, CDCl$_3$) $\delta = 7.75$ (dd, $J = 5.8$, 2.5, 2H), 7.65 – 7.55 (m, 2H), 7.42 – 7.34 (m, 2H), 7.33 – 7.23 (m, 4H), 7.19 (dd, $J = 8.4$, 5.6, 2H), 7.01 (dd, $J = 11.3$, 8.2, 2H), 6.80 (dd, $J = 8.6$, 3.2, 2H), 6.02 – 5.83 (m, 1H), 5.36 – 5.20 (m, 2H), 5.19 – 4.84 (m, 2H), 4.58 (dd, $J = 5.7$, 1.3, 2H), 4.56 – 4.41 (m, 1H), 4.41 – 4.28 (m, 2H), 4.25 – 4.14 (m, 2H), 3.95 – 3.83 (m, 2H), 3.72 – 3.52 (m, 2H), 3.38 – 3.23 (m, 2H), 2.86 (s, 3H), 2.45 – 2.37 (m, 2H), 1.98 – 1.87 (m, 2H), 1.84 – 1.74 (m, 4H), 1.70 – 1.59 (m, 1H), 1.48 – 1.42 (m, 1H), 1.37 (s, 6H), 0.95 – 0.80 (m, 9H), 0.64 – 0.58 (m, 3H).

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta = 173.17$, 169.05, 159.03, 159.30, 159.27, 156.37, 156.35, 156.83, 150.56, 150.49, 144.25, 144.15, 144.06, 143.99, 141.42, 141.38, 134.96, 134.72, 132.33, 130.96, 130.91, 130.87, 130.78, 128.66, 128.53, 128.49, 127.79, 127.73, 127.71, 127.65, 127.55, 127.48, 127.18, 127.15, 127.12, 125.58, 125.34, 125.31, 125.26, 120.80, 120.01, 119.99, 118.38, 114.62, 114.53, 101.37, 101.25, 82.68, 82.11, 79.85, 74.58, 74.48, 67.48, 67.40, 67.35, 67.00, 66.95, 65.20, 54.87, 54.83, 47.37, 41.36, 38.05, 34.35, 33.95, 33.83, 33.80, 33.77, 28.73, 28.60, 25.36, 25.32, 24.57, 24.50, 24.03, 24.01, 21.86, 21.72, 21.55, 21.40, 18.52, 18.40.

IR (thin film, cm$^{-1}$): $\nu = 2958$ (m), 1740 (s), 1693 (s), 1512 (s), 1247 (s), 1166 (s), 737 (m).

MS (ESI): calculated for C$_{66}$H$_{76}$Na$_{1}$O$_{13}$ [M+Na]$^+$: 1055.5240, found: 1055.5234.
5-(4-(((2-((S)-1-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethylbutanoyl)oxy)phenyl)-5,5-dimethyl-1,3-dioxane-2-carbonyl)oxy)methyl)phenoxy)pentanoic acid (Fmoc-Leu (phenolester-protected) α-ketoacid with linker) 49

48 (0.370 g, 0.358 mmol, 1.00 equiv) was dissolved in dry, degassed CH₂Cl₂ (2.1 ml) and cooled to 0 °C. Pd(PPh₃)₄ (20.7 mg, 17.9 µmol, 0.0500 equiv) and morpholine (65.5 mg, 0.752 mmol, 2.10 equiv) were added and the mixture stirred for 45 min at 0 °C. After completion of the reaction as indicated by TLC and LC-MS, the mixture was diluted with CH₂Cl₂ and an aqueous solution of citric acid (10 % m/v) was added. The phases were separated and the organic phase was washed with brine, dried over MgSO₄ and concentrated under reduced pressure. Purification by column chromatography (SiO₂, CH₂Cl₂:MeOH 100:0 to CH₂Cl₂:MeOH 95:5) gave 49 (inseparable mixture of diastereomers, 0.270 g, 0.272 mmol, 76 %) as white foam. Note: Not all signals of the diastereomers are resolved in the NMR data.

¹H NMR (400 MHz, CDCl₃) δ = 7.79 – 7.72 (m, 2H), 7.63 – 7.55 (m, 2H), 7.41 – 7.33 (m, 2H), 7.31 – 7.23 (m, 4H), 7.19 (t, J=8.0, 2H), 7.01 (dd, J=11.0, 8.2, 2H), 6.79 (dd, J=8.8, 2.6, 2H), 5.21 – 4.88 (m, 2H), 4.57 – 4.39 (m, 1H), 4.43 – 4.27 (m, 2H), 4.25 – 4.11 (m, 2H), 3.94 – 3.86 (m, 2H), 3.73 – 3.49 (m, 2H), 3.37 – 3.25 (m, 2H), 2.86 (d, J=3.8, 3H), 2.45 – 2.36 (m, 2H), 1.95 – 1.86 (m, 2H), 1.83 – 1.75 (m, 4H), 1.70 – 1.59 (m, 1H), 1.48 – 1.42 (m, 1H), 1.39 – 1.33 (m, 6H), 0.93 – 0.82 (m, 9H), 0.62 (d, J=9.9, 3H).

¹³C NMR (101 MHz, CDCl₃) δ = 178.22, 178.13, 169.10, 169.03, 159.25, 159.23, 156.33, 156.29, 155.76, 155.74, 150.55, 150.48, 144.25, 144.14, 144.06, 144.00, 143.98, 141.39, 141.36, 141.35, 135.22, 135.15, 135.09, 134.93, 134.67, 132.30, 132.20, 132.17, 130.88, 130.83, 130.74, 130.62, 128.71, 128.65, 128.59, 128.51, 128.47, 128.22, 128.17, 128.12, 127.71, 127.68, 127.55, 127.16, 127.13, 127.11, 127.06, 125.33, 125.30, 125.25, 120.79, 119.98, 119.96, 114.62, 114.52, 101.38, 101.25, 82.68, 82.09, 79.68, 74.57, 74.46, 67.45, 67.44, 67.38, 67.29, 66.97, 66.92, 54.84, 54.79, 47.36, 45.60, 41.33, 38.05, 37.31, 34.31, 33.81, 33.77, 33.62, 29.81, 28.61, 28.59, 25.34, 25.29, 24.56, 24.49, 24.45, 24.02, 23.85, 21.85, 21.84, 21.56, 21.51, 21.40, 21.28, 21.11, 18.50, 18.39.

IR (thin film, cm⁻¹): ν = 3332 (w), 2959 (s), 2871 (m), 1741 (s), 1512 (s), 1427 (s), 1165 (s), 734 (s).

7.3.5. Procedure for loading of protected Fmoc-Xaa α-ketoacid onto resin

As a representative example, the immobilization of 40 on Rink-amide polystyrene resin is described; similar results were obtained in the immobilization of 49. Aminomethyl polystyrene resin was found to work equally well. Rink-amide polystyrene resin (free amine form, 1.43 g, loading according to manufacturer: 0.52 mmol/g, 0.75 mmol, 1.5 equiv) was placed in a plastic syringe equipped with a frit, preswelled for 15 min with DMF and the solvent discharged. In a separate 20 ml vial, 40 (390 mg, 0.500 mmol, 1.00 equiv) and COMU (214 mg 0.500 mmol, 1.00 equiv) were dissolved in 10 ml DMF and NMM (101 mg, 1.00 mmol, 2.00 equiv) was added. The yellow solution was immediately added to the resin and shaken for 4 h at room temperature. The solvent was discharged, the resin washed with DMF (3 x) and excess amine groups on the solid support capped by treatment of the resin with a mixture of AcOH/NMM/DMF 1:1:4 (2 x 10 min). After more washings (3 x DMF, 3 x CH₂Cl₂) the resin 41 was dried under a forced flow of nitrogen and stored at 4 °C. The loading was determined to be 0.47 mmol/g (90 % based on 40) by UV(λ = 304 nm) quantification of dibenzofulvene group released after treating with 2% DBU in DMF.
7.3.6. Model studies with orthogonally protected Fmoc-leucine α-ketoacid 49

Fmoc-Gln-Leu-(phenolester-protected) α-ketoacid dipeptide 51 resembles the first two amino acid residues of H-Opr-[SUMO3 (31–75)]-(phenol-protected)-α-ketoacid 54 and was synthesized from 49 as described for 54 on page 203.

Stability of orthogonally protected α-ketoacids under cleavage conditions

Fmoc-Gln-Leu-(phenolester-protected) α-ketoacid resin (0.12 g, ca. 42 µmol) was cleaved following method B and the crude peptide analyzed by analytical HPLC. The orthogonal protection group for the α-ketoacid moiety was found to be stable to cleavage conditions; no free α-ketoacid was observed. Two diastereomers of 51 were observed and termed 51-A for the earlier eluting peak (t_R = 23.8 min) and 51-B (t_R = 24.2 min). Small amounts of (phenol-protected)-α-ketoacids 52-A (t_R = 25.8 min) and 52-B (t_R = 26.2 min) were detected in the crude product. Control experiments showed that 52-A and 52-B arose from in-situ deprotection on the heated column and not from the cleavage.

Analytical HPLC of crude Fmoc-Gln-Leu-(phenolester-protected) α-ketoacid 51. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 20 to 90 % CH_3CN in 20 min, monitored at 301 nm.

The crude residue was purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), room temperature; 40 to 90 % CH_3CN in 20 min, flow rate 40 ml/min) to give 51-A (19.6 mg, 24.0 µmol, 57 %) and 52-B (10.3 mg, 12.6 µmol, 30 %) as colorless solids. Note: The relative stereoconfiguration of the two isomers 51-A and 51-B could not be assigned based on the NMR spectra.

51-A:

^1^H NMR (600 MHz, d_6-DMSO) δ = 8.53 (q, J=5.9, 2H), 7.89 (d, J=7.6, 0.9, 2H), 7.73 (ddd, J=22.4, 7.4, 1.1, 2H), 7.62 (d, J=10.0, 1H), 7.48 (d, J=8.4, 1H), 7.44 – 7.37 (m, 3H), 7.40 – 7.37 (m,
2H), 7.33 (tdd, J = 7.4, 4.8, 1.2, 2H), 7.26 – 7.21 (m, 1H), 7.16 – 7.08 (m, 2H), 6.82 – 6.74 (m, 1H), 4.62 (s, 1H), 4.30 (ddt, J = 12.4, 10.1, 2.6, 1H), 4.28 – 4.24 (m, 1H), 4.23 – 4.17 (m, 2H), 4.11 (td, J = 8.8, 4.5, 1H), 3.60 (d, J = 11.5, 1H), 3.54 (d, J = 11.4, 1H), 2.99 (dtd, J = 13.2, 10.4, 2.8, 1H), 2.61 (t, J = 5.4, 3H), 2.12 (t, J = 8.0, 2H), 1.98 – 1.92 (m, 2H), 1.92 – 1.86 (m, 1H), 1.81 – 1.70 (m, 1H), 1.61 – 1.54 (m, 1H), 1.55 – 1.50 (m, 1H), 1.41 (ddt, J = 13.2, 10.4, 2.8, 1H), 1.31 (s, 5H), 0.85 (d, J = 6.5, 3H), 0.81 (d, J = 6.3, 3H), 0.78 (s, 3H), 0.63 (s, 2H).

13C NMR (151 MHz, d6-DMSO) δ = 174.90, 174.03, 171.33, 170.01, 155.84, 149.76, 144.00, 143.68, 140.71, 140.68, 135.36, 128.42, 127.65, 127.09, 127.07, 125.40, 125.26, 120.80, 120.11, 100.70, 81.30, 73.26, 65.75, 54.31, 51.34, 46.68, 44.88, 40.73, 37.75, 35.18, 33.09, 32.54, 31.58, 28.09, 24.63, 24.62, 23.89, 23.78, 21.35, 21.31, 18.09.

LC-MS (ESI): calculated for C45H59N4O10 [M+H]⁺: 815.4, found: 815.0.

51-B:

1H NMR (600 MHz, d6-DMSO) δ = 8.51 – 8.39 (m, 2H), 7.88 (dq, J = 7.6, 1.0, 2H), 7.68 (ddt, J = 7.4, 1.8, 0.9, 2H), 7.49 (d, J = 8.8, 1H), 7.43 – 7.41 (m, 2H), 7.41 – 7.39 (m, 1H), 7.35 – 7.29 (m, 2H), 7.28 (d, J = 8.6, 2H), 7.24 (s, 2H), 7.10 (d, J = 8.6, 2H), 6.79 – 6.72 (m, 1H), 4.54 (s, 1H), 4.30 (dddt, J = 11.7, 10.1, 2.6, 1H), 4.27 – 4.23 (m, 1H), 4.21 – 4.18 (m, 1H), 4.18 – 4.14 (m, 1H), 4.10 (td, J = 9.2, 4.5, 1H), 3.68 (d, J = 11.5, 1H), 3.52 (d, J = 11.5, 1H), 2.99 (ddt, J = 9.6, 7.3, 4.8, 2H), 2.61 (t, J = 5.5, 3H), 2.18 – 2.04 (m, 2H), 1.97 – 1.91 (m, 2H), 1.92 – 1.85 (m, 1H), 1.78 – 1.69 (m, 1H), 1.59 – 1.53 (m, 1H), 1.53 – 1.48 (m, 1H), 1.32 – 1.31 (m, 6H), 1.30 – 1.26 (m, 1H), 0.85 (d, J = 6.5, 3H), 0.79 (d, J = 6.4, 3H), 0.71 (s, 3H), 0.60 (s, 3H).

13C NMR (151 MHz, d6-DMSO) δ = 174.92, 173.81, 171.49, 170.02, 158.36, 158.13, 157.89, 155.80, 149.69, 143.86, 143.69, 140.70, 140.68, 135.32, 128.43, 127.65, 127.07, 125.24, 125.22, 120.69, 120.12, 100.78, 81.04, 73.25, 65.71, 54.71, 51.41, 46.61, 44.89, 40.73, 37.73, 35.20, 33.00, 32.55, 31.65, 28.24, 24.64, 23.89, 23.74, 21.31, 21.21, 18.08.

LC-MS (ESI): calculated for C45H59N4O10 [M+H]⁺: 815.4, found: 815.0.

Hydrolysis of phenolester 51

Crude Fmoc-Gln-Leu-(phenolester-protected) α-ketoacid 51 (ca. 3 mg) was dissolved in DMSO (50 µl) and pH 8 phosphate buffer (0.2 M) was added and the reaction progress monitored by analytical HPLC.
HPLC monitoring of hydrolysis of 51. Crude 51 (bottom trace) and reaction after 15 min (top trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 20 to 90 % CH₃CN in 20 min, monitored at 301 nm.

The crude reaction mixture was directly purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), room temperature; 40 to 90 % CH₃CN in 20 min, flow rate 40 ml/min) to give 52-A (0.8 mg, 1.1 µmol, ca. 30 %) and 52-B (0.6 mg, 0.8 µmol, ca. 24 %) as colorless solids.

The identity of both materials was confirmed by LC-MS analysis: calculated for C₃₈H₄₅N₃O₉ [M+H]+: 687.3, found: 687.8 (52-A), 687.8 (52-B).

**Acidic deprotection of 52**

52-A or 52-B (ca. 0.2 mg) were dissolved in TFA (0.3 ml) and incubated at room temperature. After 1 h, volatiles were removed with a stream of nitrogen and the residue analyzed by analytical HPLC. Both 52-A and 52-B were cleanly deprotected to give 53. The identity of 53 was confirmed by LC-MS analysis: calculated for C₂₇H₃₂N₃O₇ [M+H]+: 510.6, found: 510.4.

Note: The deprotection was unusually slow, presumably due to the absence of scavengers such as DODT. In similar reactions in the presence of scavengers, the reaction reaches complete conversion within 2 h (see page 217).
CHAPTER 7: EXPERIMENTAL

HPLC monitoring of the deprotection of 52-A (bottom two traces) and 52-B (upper two traces). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 20 to 90 % CH₃CN in 20 min, monitored at 301 nm.

**Stability assay of (phenol-protected) α-ketoacids under ligation conditions**

At the respective synthesis step during the synthesis of H-Opr-[SUMO3 (31–75)]-(phenol-protected)-α-ketoacid 54, a small amount of resin with Fmoc-[SUMO3 (62–75)]-orthogonally protected α-ketoacid was removed from the synthesizer and cleaved following Method B. The crude peptide was incubated in pH 8 buffer for 10 min and the mixture directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C 25 to 55 % CH₃CN in 30 min), separating the two diastereomers of Fmoc-[SUMO3 (62–75)]-(phenol-protected) α-ketoacid, S9-A and S9-B.

S9-A (ca. 0.2 mg) and S9-B (ca. 0.2 mg) were weighed into a glass vial insert and dissolved in 7:3 DMSO:H₂O (ca. 10 µl, ca. 20 mM) with 0.1 M oxalic acid and heated to 60 °C for 18 h. Analysis by HPLC revealed that the mixture remained unchanged and no deprotection occurred. The mixture was diluted with 97.5:2.5 TFA:DODT (90 µl) and incubated at room temperature for 75 min. Volatiles were removed under reduced pressure and the residue redissolved in 1:1 CH₃CN/H₂O with 0.1 % TFA (ca. 20 µl) and analyzed by analytical HPLC, indicating complete and clean deprotection to give Fmoc-[SUMO3 (62–75)]-α-ketoacid S10.
HPLC monitoring of the incubation of S9-A and S9-B (lower two traces) under ligation conditions (7:3 DMSO:H₂O with 0.1 M oxalic acid, 60 °C). The same reaction mixture after 75 min under deprotection conditions (top trace) Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 20 to 90 % CH₃CN in 20 min, monitored at 301 nm.

S9-A:

S9-B:
**LC-MS** (ESI): calculated for C₈₉H₁₂₄N₁₇O₃₁ [M+H]⁺: 1926.9, found: 1927.5; [M+2H]²⁺: 963.9, found: 963.9.

S10:
**LC-MS** (ESI): calculated for C₇₈H₁₁₀N₁₇O₂₉ [M+H]⁺: 1748.8, found: 1749.2; [M+2H]²⁺: 874.9, found: 874.8.
Synthesis of 51 using enantioenriched protecting group 43

Orthogonally protected Fmoc-Leu-α-ketoacid monomer 49 was synthesized as before (page 172) starting from enantioenriched 43 (60 % ee, see page 161) and used to synthesize Fmoc-Gln-Leu (phenol-protected) α-ketoacid dipeptide 51 (see page 178). The obtained dipeptides 51 gave almost identical HPLC traces regardless if they were synthesized from enantioenriched and racemic 43. Using the enantioenriched protecting group 43 showed no benefit. This result strongly suggests epimerization of the benzylic stereocenter during the synthesis, presumably during the acid-catalyzed protection of Fmoc-leucine α-ketoacid 7 with 43.

Analytical HPLC of crude 51 synthesized from racemic 43 (bottom trace) and enantioenriched 43 (top trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), 20 to 90 % CH3CN in 20 min, monitored at 301 nm.
CHAPTER 7: EXPERIMENTAL

7.4. Experimental for “CHAPTER 4: Chemical synthesis of SUMO3 proteins by KAHA ligation”: Protein synthesis and characterization

7.4.1. Synthesis of relevant small molecule starting materials

N²-(N²-(((9H-fluoren-9-yI)methoxy)carbonyl)-N⁵-trityl-L-glutaminyl)-N²-methyl-N⁵-trityl-L-glutamine (Fmoc-Gln(Trt)NMeGln(Trt)-OH) 81

Fmoc-Gln(Trt)NMeGln(Trt)-OH 81 dipeptide building block was synthesized on solid support following an adapted procedure from Chiu et al.\textsuperscript{14} Fmoc-Gln(Trt)-OH (0.977 g, 1.60 mmol, 1.00 equiv) was dissolved in CH₂Cl₂ (10 ml) and DIPEA (0.827 g, 1.11 ml, 6.40 mmol, 4.00 equiv) was added. The clear solution was added to preswelled 2-chlorotrilty polystyrene resin (1.0 g, 1.6 mmol, 1.0 equiv, substitution capacity of 1.6 mmol/g). The resin suspension was agitated for 4 h. The solvent was removed by filtration and the resin washed with CH₂Cl₂ (5 x 10 ml) and DMF (5 x 10 ml). The Fmoc group was removed as described in the General Procedure by treatment with 20 % piperidine/DMF.

2-nitrobenzenesulfonyl chloride (0.709 g, 3.20 mmol, 4.00 equiv) was dissolved in 2:1 THF:CH₂Cl₂ (6.4 ml, 0.50 M) at 0 °C and DIPEA (0.620 g, 0.835 ml, 4.80 mmol, 6 equiv) was added. The solution was added to half of the resin (0.80 mmol in theory) and the suspension agitated for 4 h. The solvent was removed by filtration and the resin washed with CH₂Cl₂ (5 x 10 ml) and DMF (5 x 10 ml).

To half of the resin (0.40 mmol in theory) was added a solution of PPh₃ (0.525 g, 2.00 mmol, 5.00 equiv) in THF (4.0 ml), followed by a solution of MeOH (0.128 g, 0.163 ml, 4.00 mmol, 10.0 equiv) in THF (2.0 ml). The resin was mixed and DIAD (0.405 g, 0.196 ml, 2.00 mmol, 5.00 equiv) was added in portions. The resin was agitated for 8 h, the solvent was removed by filtration and the resin washed with THF (5 x 10 ml), DMF (5 x 10 ml) and CH₂Cl₂ (5 x 10 ml). A small amount of resin was cleaved in a microcleavage (20 % HFIP in CH₂Cl₂, 1 h) and the crude S11 analyzed by LC-MS and NMR, indicating selective methylation on the sulfonamide nitrogen.

The resin was treated with a solution of sodium thiophenolate in DMF (1 M) (2 x 1 h). Analysis by microcleavage and LC-MS indicated only ca. 50 % deprotection to free amine; however, the conversion was not improved by repeated treatments with sodium thiophenolate in DMF (1 M) and it was decided to continue.

Fmoc-Gln(Trt)-OH (0.977 g, 1.60 mmol, 4.00 equiv) and COMU (0.629 g, 1.50 mmol 3.75 equiv) were dissolved in DMF (6.4 ml, 0.25 M) and NMM (0.326 g, 0.355 ml, 3.20 mmol, 8 equiv) was added. The yellow solution was added to the resin and the suspension was agitated for 6 h at room temperature. The resin was washed with DMF (5 x 10 ml) and CH₂Cl₂ (5 x 10 ml) and volatiles removed under a stream of nitrogen. The resin was cleaved with 20 % HFIP in CH₂Cl₂ (10 ml) for 1 h, the resin removed by filtration and the filtrate concentrated under reduced pressure. The crude residue was purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), 80 to 95 % CH₃CN with 0.1 % TFA in 15 min followed by a wash with 95 % CH₃CN for 20 min, flow rate 40 ml/min). Product containing fractions were pooled, diluted 1 X with water and lyophilized to yield Fmoc-Gln(Trt)NMeGln(Trt)-OH 81 (0.125 g, 0.125 mmol, 31 %) as white solid. NMR analysis of the final product 81 indicated a d.r. > 9:1, demonstrating that less than 10 % epimerization occurred during the synthesis.

\[\text{N}^2\text{-methyl-}\text{N}^2\text{-}(2\text{-nitro-4-sulfophenyl})\text{-N}^5\text{-trityl-}\text{L-glutamine}\]

**S11**

\[\text{^1H NMR} (600 \text{ MHz, d}_6\text{-DMSO}) \delta = 8.71 \text{ (s, 1H)}, 8.05 \text{ (dd, J}=7.9, 1.4, 1H), 7.93 \text{ (dd, J}=7.9, 1.3, 1H), 7.87 – 7.83 \text{ (m, 1H)}, 7.81 \text{ (td, J}=7.7, 1.4, 1H), 7.30 – 7.24 \text{ (m, 6H)}, 7.22 – 7.15 \text{ (m, 9H)}, 4.44 \text{ (dd, J}=10.9, 4.6, 1H), 2.88 \text{ (s, 3H)}, 2.41 – 2.26 \text{ (m, 2H)}, 2.15 – 2.03 \text{ (m, 1H)}, 1.81 \text{ (dddd, J}=14.4, 10.9, 8.9, 5.7, 1H).]

\[\text{^13C NMR} (151 \text{ MHz, d}_6\text{-DMSO}) \delta = 171.07, 170.81, 147.49, 144.89, 134.32, 132.04, 131.10, 130.32, 128.56, 127.43, 126.28, 123.89, 69.27, 59.10, 32.45, 30.33, 24.18.\]

**LC-MS**: calculated for \(\text{C}_{31}\text{H}_{30}\text{N}_3\text{O}_7\text{S} [\text{M+H}]^+ = 588.7\), found 588.1.

\[\text{N}^2\text{-}\text{N}^2\text{-}((9\text{-fluoren}-9\text{-yl})\text{methoxy})\text{carbonyl})\text{-N}^2\text{-trityl-}\text{L-glutaminyl}-\text{N}^2\text{-methyl-}\text{N}^2\text{-trityl-}\text{L-glutamine} 81\]

\[\text{^1H NMR} (600 \text{ MHz, d}_6\text{-DMSO}) \delta = 8.67 \text{ (s, 1H)}, 8.46 \text{ (s, 1H)}, 7.91 – 7.84 \text{ (m, 2H)}, 7.68 – 7.64 \text{ (m, 1H)}, 7.72 – 7.58 \text{ (m, 2H)}, 7.42 – 7.37 \text{ (m, 2H)}, 7.32 – 7.27 \text{ (m, 2H)}, 7.28 – 7.05 \text{ (m, 3OH)}, 4.93 – 4.88 \text{ (m, 1H)}, 4.45 – 4.37 \text{ (m, 1H)}, 4.18 \text{ (t, J}=7.1, 1H), 4.26 – 4.06 \text{ (m, 2H)}, 2.77 \text{ (s, 3H)}, 2.59 – 2.51 \text{ (m, 1H)}, 2.37 – 2.30 \text{ (m, 1H)}, 2.29 – 2.22 \text{ (m, 1H)}, 2.17 – 2.08 \text{ (m, 1H)}, 2.06 – 1.96 \text{ (m, 1H)}, 1.91 – 1.83 \text{ (m, 1H)}, 1.81 – 1.73 \text{ (m, 1H)}, 1.68 – 1.60 \text{ (m, 1H)}.\]
**13C NMR** (151 MHz, d$_6$-DMSO) δ = 172.76, 172.25, 171.44, 171.16, 156.19, 144.88, 144.87, 143.77, 140.69, 140.68, 128.53, 128.49, 128.47, 127.63, 127.59, 127.44, 127.38, 127.06, 126.30, 126.29, 126.19, 125.38, 125.35, 120.08, 120.04, 69.26, 69.14, 65.72, 55.78, 50.61, 46.63, 32.32, 31.79, 30.93, 26.59, 23.62.

**IR** (thin film, cm$^{-1}$): ν = 3308 (w), 3056 (m), 1716 (s), 1652 (s), 1507 (s), 1490 (s), 1264 (m), 739 (s), 700 (s).

**MS** (ESI): calculated for C$_{64}$H$_{59}$N$_4$O$_7$ [M+H]$^+$: 995.4378, found: 995.4370.

**(9H-fluoren-9-yl)methyl (2-(2-(2-hydroxyethoxy)ethoxy)ethyl)carbamate (Fmoc-PEG-OH)** 79

**(9H-fluoren-9-yl)methyl (2-(2-(2-hydroxyethoxy)ethoxy)ethyl)carbamate (Fmoc-PEG-OH)** 79 was synthesized following a modified procedure of Hawthrone et al.**15** 2-(2-(2-chloroethoxy)ethoxy)ethan-1-ol (3.37 g, 2.91 ml, 20.0 mmol), potassium phthalimide (4.44 g, 2.40 mmol, 1.20 equiv) and potassium iodide (3.32 g, 2.00 mmol, 1.00 equiv) were dissolved in DMF (40 ml, 0.50 M) and heated to 80 ºC for 36 h. The mixture was cooled to room temperature, the formed precipitate removed by filtration and DMF removed under reduced pressure. The yellowish residue was redissolved in CH$_2$Cl$_2$, washed with water, brine, dried over MgSO$_4$ and was concentrated to give crude 2-(2-(2-(2-hydroxyethoxy)ethoxy)ethyl)isoindoline-1,3-dione (5.34 g, 19.1 mmol, 96 %) as a slightly yellowish liquid which was used without further purification.

Crude 2-(2-(2-(2-hydroxyethoxy)ethoxy)ethyl)isoindoline-1,3-dione (0.279 g, 1.00 mmol, 1.00 equiv) was dissolved in CH$_2$Cl$_2$ (5.0 ml, 0.20 M) and methyl hydrazine (55.3 mg, 63.0 µl, 1.20 mmol, 1.20 equiv) was added. The mixture was stirred at room temperature for 12 h. The mixture was diluted with CH$_2$Cl$_2$ (ca. 5 ml) and the formed precipitate was removed by filtration. The mixture was concentrated under reduced pressure and the residue dissolved in DMF (3.0 ml). Fmoc N-hydroxysuccinimide ester (0.439 g, 1.30 mmol, 1.30 equiv) was added and the reaction mixture stirred at room temperature for 1 h. The mixture was diluted with CH$_2$Cl$_2$, washed with an aqueous solution of citric acid (10% w/v), H$_2$O and brine and dried over anhydrous Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The

---

crude was purified by column chromatography (SiO₂, hexane:EtOAc 9:1 to 0:1) to give 79 as sticky white solid (0.115 g, 0.310 mmol, 31 %) with identical spectroscopic properties as reported.

(9H-fluoren-9-yl)methyl (2-(2-(2-hydroxyethoxy)ethoxy)ethyl)carbamate 79

**¹H NMR** (400 MHz, CDCl₃) δ = 7.81 – 7.73 (m, 2H), 7.69 – 7.57 (m, 2H), 7.49 – 7.36 (m, 2H), 7.36 – 7.28 (m, 2H), 5.56 – 5.42 (m, 1H), 4.48 – 4.40 (m, 2H), 4.27 – 4.18 (m, 1H), 3.74 – 3.68 (m, 2H), 3.66 – 3.61 (m, 4H), 3.61 – 3.59 (m, 2H), 3.57 – 3.54 (m, 2H), 3.43 – 3.36 (m, 2H), 3.14 (s, 1H).

2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetic acid S12

2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetic acid S12 was synthesized following a modified procedure from Uekama et al.³⁻¹⁶ 3-(diethylamino)phenol (4.13 g, 25.0 mmol, 1 equiv), diethyl acetasuccinate (5.68 g, 2.63 mmol, 1.05 equiv) and dry ZnCl₂ (4.43 g, 3.25 mmol, 1.30 equiv) were suspended together with 3 Å molecular sieves (ca. 2 g) in EtOH (25 ml) and heated to reflux for 14 h. The reaction was allowed to cool to room temperature, filtered and poured into 0.01 M HCl. The aqueous phase was extracted with EtOAc, the phases separated and the organic phase dried over MgSO₄ and concentrated. The crude residue was directly dissolved in 6 M HCl (3 ml) and heated to 65 °C for 12 h. The reaction was cooled to room temperature, diluted with water (100 ml) and placed at 4 °C overnight. The formed crystals were filtered off and dried under high vacuum to give S12 (0.260 g, 0.899 mmol, 4 % over two steps) as red crystals.

**¹H NMR** (300 MHz, CDCl₃) δ = 7.43 (d, J=9.1, 1H), 6.61 (dd, J=9.1, 2.6, 1H), 6.50 (d, J=2.6, 1H), 3.71 (s, 2H), 3.41 (q, J=7.1, 4H), 2.37 (s, 3H), 1.20 (t, J=7.1, 6H).

**LC-MS**: calculated for C₁₆H₂₀NO₄ [M+H]⁺ = 290.1, found 289.8.

---

(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-((2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetamido)propanoic acid (Fmoc-Dap(Cou)-OH) 80

\[ \text{(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-((2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetamido)propanoic acid (Fmoc-Dap(Cou)-OH) 80} \]

2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetic acid \textbf{S12} (86.8 mg, 0.300 mmol, 1.00 equiv) and COMU (0.126 g, 0.294 mmol, 0.980 equiv) were dissolved in NMP (1.2 ml, 0.25 M) and NMM (60.7 mg, 66.0 µl, 0.600 mmol, 2.00 equiv) was added. The mixture was incubated 1 min at room temperature and Fmoc-(S)-2,3-diaminopropionic acid (97.9 mg, 0.300 mmol, 1.00 equiv) was added and the reaction mixture stirred at room temperature for 30 min. The mixture was diluted with CH\textsubscript{2}Cl\textsubscript{2}, washed with an aqueous solution of citric acid (10% w/v), H\textsubscript{2}O and brine, dried over anhydrous Na\textsubscript{2}SO\textsubscript{4} and concentrated. The residue was purified by preparative HPLC (Shiseido Proteonavi C4 column (50 x 250 mm), 30 to 95 % CH\textsubscript{3}CN with 0.1 % TFA in 30 min, flow rate 40 ml/min). Product containing fractions were pooled, diluted 1 X with water and lyophilized to yield Fmoc-Dap(Cou)-OH \textbf{80} (0.120 g, 0.200 mmol, 67 %) as yellowish solid.

\textbf{1H NMR} (400 MHz, CDCl\textsubscript{3}) \( \delta = 8.36 \text{ (s, 1H)}, 7.72 \text{ (dq, } J=7.6, 1.3, 2H), 7.61 – 7.51 \text{ (m, 2H), 7.51 – 7.41 \text{ (m, 1H), 7.41 – 7.32 \text{ (m, 2H), 7.32 – 7.21 \text{ (m, 3H), 6.97 – 6.90 \text{ (m, 1H), 6.84 (d, } J=2.5, 1H), 6.27 (d, } J=7.4, 1H), 4.49 – 4.38 \text{ (m, 1H), 4.21 (dd, } J=7.6, 3.3, 2H), 4.11 \text{ (t, } J=7.3, 1H), 3.84 – 3.73 \text{ (m, 1H), 3.71 – 3.63 \text{ (m, 1H), 3.61 (s, 2H), 3.39 (q, } J=7.6, 4H), 2.39 (s, 3H), 1.09 \text{ (t, } J=7.1, 6H).}

\textbf{13C NMR} (101 MHz, CDCl\textsubscript{3}) \( \delta = 172.77, 163.01, 156.68, 153.89, 151.07, 145.73, 143.83, 143.79, 141.32, 127.91, 127.30, 127.27, 126.89, 125.39, 120.05, 116.85, 116.31, 114.99, 114.00, 113.85, 103.40, 67.67, 54.55, 48.94, 47.04, 41.80, 35.12, 15.48, 11.48.

\textbf{IR} (thin film, cm\textsuperscript{-1}): \( \tilde{\nu} = 3326 \text{ (w), 2925 (m), 2359 (s), 1716 (s), 1616 (s), 1522 (s), 1355 (m), 1081 (w).}

\textbf{MS} (ESI): calculated for C\textsubscript{34}H\textsubscript{36}N\textsubscript{3}O\textsubscript{7} [M+H]\textsuperscript{+}: 598.2548, found: 598.2546.
(S)-5-(N-(2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-2-carboxyethyl)sulfamoyl)-2-(6-(diethylamino)-3-(diethyliminio)-3H-xanthen-9-yl)benzenesulfonate (Fmoc-Dap(Rho)-OH) 89

5-(chlorosulfonyl)-2-(6-(diethylamino)-3-(diethyliminio)-3H-xanthen-9-yl)benzenesulfonate (1.15 g, 2.00 mmol, 1.00 equiv) was dissolved in dry CH$_2$Cl$_2$ (5 ml) and added to a solution of Fmoc-(S)-2,3-diaminopropionic acid (0.685 g, 2.10 mmol, 1.05 equiv) in NMP (5 ml). To the dark red solution, DIPEA (0.284 g, 0.383 ml, 2.20 mmol, 1.10 equiv) was added dropwise at 0 ºC and the mixture was stirred and allowed to warm to room temperature. After 5 h, the reaction mixture was diluted with CH$_2$Cl$_2$, washed with an aqueous solution of citric acid (10% w/v), H$_2$O and brine, dried over anhydrous Na$_2$SO$_4$ and concentrated. The residue was purified by preparative HPLC (Shiseido Proteanavi C4 column (50 x 250 mm), 40 to 85 % CH$_3$CN with 0.1 % TFA in 25 min, flow rate 40 ml/min). Product containing fractions were pooled, diluted 1 X with water and lyophilized to yield Fmoc-Dap(Rho)-OH (0.394 g, 0.454 mmol, 23 %) as dark magenta solid.

$^1$H NMR (600 MHz, d$_6$-DMSO) $\delta = 8.42$ (d, J=1.9, 1H), 8.16 (t, J=6.2, 1H), 7.93 (dd, J=7.9, 2.0, 1H), 7.84 (ddt, J=13.2, 7.6, 0.9, 2H), 7.68 (dd, J=20.0, 7.6, 2H), 7.60 (d, J=8.5, 1H), 7.39 (dd, J=14.5, 7.5, 1.1, 2H), 7.36 – 7.27 (m, 3H), 7.05 – 6.95 (m, 4H), 6.94 – 6.90 (m, 2H), 4.28 – 4.17 (m, 3H), 4.12 (td, J=8.3, 4.5, 1H), 3.68 – 3.54 (m, 8H), 3.32 (ddd, J=13.2, 5.8, 4.5, 1H), 3.23 (ddd, J=13.3, 8.2, 6.6, 1H), 1.26 – 1.12 (m, 12H).

$^{13}$C NMR (151 MHz, d$_6$-DMSO) $\delta = 171.44$, 157.47, 157.08, 157.04, 155.88, 154.98, 148.01, 143.84, 143.70, 141.51, 140.62, 140.61, 133.14, 132.81, 130.65, 127.61, 127.60, 127.03, 127.01, 126.38, 125.65, 125.33, 125.32, 120.04, 120.00, 113.55, 113.50, 113.45, 113.43, 95.35, 66.05, 53.96, 46.44, 45.23, 43.41, 12.44.

IR (thin film, cm$^{-1}$): $\tilde{v} = 2975$ (w), 2931 (m), 1716 (m), 1650 (m), 1590 (s), 1417 (m), 1338 (m), 1181 (s).

MS (ESI): calculated for C$_{45}$H$_{45}$N$_{4}$O$_{10}$S$_{2}$ [M+H]$^+$: 867.2728, found: 867.2725.

---

7.4.2. Protein sequence of SUMO3 (2–92)

SUMO3 (1–103) (Native sequence with propeptide):
MSEEKPKEGV\textsubscript{10} KTENDHINLK\textsubscript{20} VAGQDGSVVQ\textsubscript{30} FKIKRHTPLS\textsubscript{40} KLMKAYCERQ\textsubscript{50} GLSMRQIRFR\textsubscript{60} FDGQPINETD\textsubscript{70} TPAQLEMEDE\textsubscript{80} DTIDVFQQQT\textsubscript{90} GGVPESSLAG\textsubscript{100} HSF

From Uniprot (P55854, SUMO3\_Human).

SUMO3 (2–92) (This work):
-SEEKPKEGV\textsubscript{10} KTENDHINLK\textsubscript{20} VAGQDGSVVQ\textsubscript{30} FKIKRHTPLS\textsubscript{40} KLMKAYCERQ\textsubscript{50} GLSMRQIRFR\textsubscript{60} FDGQPINETD\textsubscript{70} TPAQLEMEDE\textsubscript{80} DTIDVFQQQT\textsubscript{90} GG-------- ---

7.4.3. Synthesis of SUMO3 protein

Synthetic strategy

H-[SUMO3 (2-92)]-OH:
-SEEKPKEGV\textsubscript{10} KTENDHINLK\textsubscript{20} VAGQDGSVV - T\textsuperscript{30} FKIKRHTPLS\textsubscript{40} KLMKAYCERQ\textsubscript{50} GL - T\textsuperscript{5}MRQIRFR\textsubscript{60} FDGQPINETD\textsubscript{70} TPAQLEMEDE\textsubscript{80} DTIDVFQQQT\textsubscript{90} GG
7.4.3.1. Synthesis of protein segments

**Synthesis of H-[SUMO3 (2–29)]-α-ketoacid 55**

H-[SUMO3 (2–29)]-α-ketoacid 55 was prepared on a cyanosulfurylside (SY) polystyrene resin (1.00 g, 0.34 mmol/g) following previously reported procedures. Briefly, Val29 was coupled following the general procedure using COMU. The coupling was repeated once. The resin was washed thoroughly with DMF, sodium diethyldithiocarbamate (0.5 M in DMF), DMF, CH₂Cl₂ and dried under a stream of nitrogen. Manual Fmoc SPPS was performed using the procedure described in the general methods section up to Ser2. The N-terminal Fmoc group was removed, the resin was washed with DMF and CH₂Cl₂ and dried under a stream of nitrogen. Dried sulfurylside resin was cleaved following Method A. The crude H-[SUMO3 (2–30)]-cyanosulfurylside was directly oxidized without prior purification. H-[SUMO3 (2-30)]-cyanosulfurylside (120 mg, 37.3 µmol) was dissolved in 1:1 CH₃CN:H₂O (1.4 mL, 26 mM peptide) and Oxone (360 mg, Acros organics, 4.5% active oxygen, ~14 equiv) was added and allowed to react for 3.5 min. The reaction was quenched by addition of dimethyl sulfide (0.4 mL) and volatiles were evaporated under reduced pressure. The crude reaction mixture with salts was centrifuged and the supernatant was filtered through 0.22 µm filter to remove residual salts. The filtrate was purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, heated to 60 °C, 15 to 30 % CH₃CN in 20 min). The pure product fractions were collected and lyophilized to obtain H-[SUMO3 (2–30)]-α-ketoacid 55 (37.3 mg, 12.3 µmol, ca. 7 % yield for peptide synthesis, oxidation and purification steps). Analytical HPLC and MALDI FTMS were used to confirm the purity and exact mass of 55. m/z calculated for C₁₂₈H₂₁₂H₃₇O₄₅ [M+H]⁺: 3035.5280; measured 3035.5303.

---

Analytical HPLC trace of purified SUMO3 (2-29) 55. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 30 % CH₃CN in 20 min.

---

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of H-[SUMO3 (2–29)]-α-ketoacid 55.

**Synthesis of Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42**

Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42 was prepared on protected leucine α-ketoacid resin 41 at a 0.4 mmol scale (1.2 g resin with a substitution capacity of 0.34 mmol/g). The segment was elongated up to Phe31 with automated synthesis as described in the general methods. After Fmoc-deprotection under standard conditions, Fmoc-Opr-OH was coupled as described in the general procedure. The dried resin was cleaved according to Method B and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, heated to 60 °C, 27 to 42 % CH3CN in 20 min). The pure product fractions were pooled and lyophilized to obtain 87.3 mg of pure Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42 (29.1 µmol, ca. 7 % yield for peptide synthesis, cleavage and purification steps). Analytical HPLC and MALDI FTMS were used to confirm the purity and exact mass of 42. m/z calculated for C139H215N36O34S2 [M+H]+: 2996.5637; measured 2996.5645.

Analytical HPLC trace of purified Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH3CN in 20 min.
Maldi FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42.

**Synthesis of H-Opr-[SUMO3 (54-92)]-OH 56**

H-Opr-[SUMO3 (54-92)]-OH 56 was synthesized on 0.4 mmol scale on Wang-polystyrene resin (0.93 g with a substitution capacity of 0.43 mmol/g). The segment was elongated up to Met54 with automated synthesis and Boc-Opr-OH coupled as described in the general methods. The overall synthetic efficiency was found to improve by introducing two dipeptide building blocks which were coupled manually: Asp70-Thr71 was coupled as pseudoproline dipeptide (Fmoc-Asp(OtBu)-Thr(ΨMe,Mepro)-OH) and Asp62-Gly63 as Dmb-backbone protected dipeptide (Fmoc-Asp(OtBu)-N(Dmb)-Gly-OH). The resin was cleaved according to Method A and the crude peptide was purified by preparative HPLC (Shiseido Capcell Pak C4 column, 50 x 250 mm, 25 to 60 % CH3CN in 40 min and Shiseido Capcell Pak MGII C18 column, 20 x 250 mm, heated to 60 °C, 27 to 33 % CH3CN in 20 min). The pure product fractions were pooled and lyophilized to obtain 67.8 mg of pure H-Opr-[SUMO3 (54-92)]-OH 56 (14.6 μmol, ca. 4 % yield for peptide synthesis, cleavage and purification steps). Analytical HPLC and MALDI FTMS were used to confirm the purity and exact mass of 56.

m/z calculated for C_{179}H_{304}N_{56}O_{71}Na_{2} [M+Na]^+: 4677.1233; measured 4677.1246.

Analytical HPLC trace of purified H-Opr-[SUMO3 (54-92)]-OH 56. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH3CN in 20 min.
MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of H-Opr-[SUMO3 (54-92)]-OH 56.

7.4.3.2. Sequential KAHA ligations for the synthesis of H-[SUMO3 (2–92)]-OH 59

KAHA ligation for the preparation of H-Opr-[SUMO3 (31–92)]-OH 58

**a) Ligation:** Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42 (15.4 mg, 5.14 μmol, 1.50 equiv) and H-Opr-[SUMO3 (54–92)]-OH 56 (16.0 mg, 3.44 μmol, 1.00 equiv) were weighed into a glass vial and dissolved in 7:3 NMP:H2O (170 μL, 20 mM) with 0.1 M oxalic acid. The clear solution was heated 60 °C. Reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column, 4.6 x 250 mm, heated to 60 °C, 15 to 60 % CH3CN in 20 min). After completion of the ligation (13 h), the reaction mixture was diluted to 1 mL with 1:1 CH3CN/H2O with 0.1 % TFA and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C 27 to 42 % CH3CN in 30 min). The fractions containing the ligated product were pooled and lyophilized to give pure Fmoc-Opr-[SUMO3 (31–92)]-OH depsi-57 (7.2 mg, 0.95 μmol, 28 % yield for ligation and purification steps). Analytical HPLC and MALDI FTMS were used to confirm the purity and identity of
**Fmoc-Opr-[SUMO3 (31–92)]-OH depsi-57.** m/z calculated for $C_{335}H_{518}N_{92}O_{103}S_4 [M+H]^+$ 7606.7079; measured 7606.7293.

HPLC monitoring of the KAHA ligation of Fmoc-Opr-[SUMO3 (31–52)]-α-ketoacid 42 and H-Opr-[SUMO3 (54–92)]-OH 56. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 60 % CH3CN in 20 min.

Analytical HPLC trace of purified Fmoc-Opr-[SUMO3 (31–92)]-OH depsi-57. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 60 % CH3CN in 20 min.

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of Fmoc-Opr-[SUMO3 (31–92)]-OH depsi-57.

**b) One-pot Fmoc-deprotection and O to N acyl shift:** Fmoc-Opr-[SUMO3 (31–92)]-OH depsi-57 (7.3 mg, 0.95 μmol) was dissolved in DMSO (1.0 ml) containing 5 % (v/v) HNEt$_2$ and incubated at room temperature for 5 min. Ice cold rearrangement buffer (1.0 ml, pH 9.5, 0.2 M NaHCO$_3$/Na$_2$CO$_3$ buffer, 10 mM TCEP) was added and the mixture incubated for 2 h at room temperature. The reaction was quenched by addition of TFA (100 μl) and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 27 to 37 % CH3CN in 20 min, flow rate 10 ml/min). The fractions containing the product were pooled and lyophilized to give pure H-Opr-[SUMO3 (31–92)]-OH 58 (3.4 mg, 0.46 μmol, 47% yield for O to N acyl shift and purification steps). Analytical HPLC and MALDI FTICR MS...
were used to confirm the purity and identity of 58. m/z calculated for C_{320}H_{508}N_{92}O_{101}S_{4} [M+H]^+: 7384.6399; measured 7384.7833.

HPLC monitoring of the Fmoc removal (middle trace) and O to N acyl shift of Fmoc-Opr-[SUMO3 (31–92)]-OH depsi-57 to H-Opr-[SUMO3 (31–92)]-OH 58 (upper trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 60 % CH_{3}CN in 20 min.

Analytical HPLC trace of purified H-Opr-[SUMO3 (31–92)]-OH 58. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 60 % CH_{3}CN in 20 min.

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of H-Opr-[SUMO3 (31–92)]-OH 58.
KAHA ligation for the preparation of H-[SUMO3 (2–92)]-OH 59

a) Ligation: H-Opr-[SUMO3 (31–92)]-OH 58 (4.6 mg, 0.62 μmol, 1.0 equiv) and H-[SUMO3 (2–29)]-OH α-ketoacid 55 (2.8 mg, 0.94 μmol, 1.5 equiv) were weighed into a glass vial insert and dissolved in 7:3 NMP:H₂O (42 μL, 15 mM) with 0.1 M oxalic acid and allowed to react at 60 °C. The progress of the ligation was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C, 15 to 35 % CH₃CN in 30 min). After completion of the ligation (9 h), the reaction mixture was diluted to 1 mL with 1:1 CH₃CN/H₂O + 0.1 % TFA and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 20 to 40 % CH₃CN in 30 min). The fractions containing the ligated product were pooled and lyophilized to give pure H-[SUMO3 (2–92)]-OH depsi-59 (1.5 mg, 0.15 mmol, 23 % yield for ligation and purification steps). Analytical HPLC and ESI QTOF MS were used to confirm the purity and identity of H-[SUMO3 (2–92)]-OH depsi-59. m/z calculated for C₄₄₇H₇₁₉N₁₂₉O₁₄₇₃S₄ [M+H]⁺: 10374.1635; measured 10380.1867 [A+6].

HPLC monitoring of the KAHA ligation of H-[SUMO3 (2–29)]-OH α-ketoacid 55 and H-Opr-[SUMO3 (31–92)]-OH 58. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 35 % CH₃CN in 30 min.
CHAPTER 7: EXPERIMENTAL

Analytical HPLC trace of purified H-[SUMO3 (2–92)]-OH depsi-59. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 35 % CH₃CN in 30 min.

b) **O to N acyl shift:** H-[SUMO3 (2–92)]-OH depsi-59 (1.5 mg, 0.15 µmol) was dissolved in rearrangement buffer (150 µl, pH 9.5, 0.2 M NaHCO₃/Na₂CO₃, 6 M GdmCl, 10 mM TCEP) and the reaction was allowed to proceed for 3 h at room temperature. Analysis by analytical HPLC showed complete conversion to H-[SUMO3 (2–92)]-OH 59, however the difference in retention time of H-[SUMO3 (2–92)]-OH depsi-59 and H-[SUMO3 (2–92)]-OH 59 is small (< 0.1 min) under various HPLC conditions examined. The mixture was acidified to pH 1–2 with CH₃CN/H₂O/TFA 45:45:10 and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 20 to 40% CH₃CN in 30 min). The fractions containing the product were pooled and lyophilized to give pure H-[SUMO3 (2–92)]-OH 59 (1.3 mg, 0.13 µmol, 87% yield for O to N acyl shift and purification steps). Analytical HPLC and ESI-QTOF MS were used to confirm the purity and identity of H-[SUMO3 (2–92)]-OH 59. m/z calculated for C₄₄₇H₇₁₉N₁₂₅O₁₄₇S₄ [M+H]⁺: 10374.1635; measured 10379.2050 [A+5].
HPLC monitoring of O to N acyl shift of H-[SUMO3 (2–92)]-OH depsi-59 to H-[SUMO3 (2–92)]-OH 59 (left) and expansion of the same traces (right). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 35 % CH3CN in 30 min.

Analytical HPLC trace of purified H-[SUMO3 (2–92)]-OH 59. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 35 % CH3CN in 30 min.

ESI-QTOF MS: Measured (top) and deconvoluted (bottom) MS of purified H-[SUMO3 (2–92)]-OH 59.

MS/MS analysis of H-[SUMO3 (2–92)]-OH 59

MALDI TOF-TOF MS/MS sequencing: To confirm the primary sequence of synthetic H-[SUMO3 (2–92)]-OH 59, a top-down MS/MS sequencing was performed in addition to trypsin digestion and MSMS analysis using 2,5–dihydroxyacetophenone (2,5–DHAP) as the matrix. A sequence cover of 81%, including the ligation sites (Val29–Hse30 and Leu52–Hse53), was obtained confirming the primary sequence of the synthetic H-[SUMO3 (2–92)]-OH 59.
CHAPTER 7: EXPERIMENTAL

MALDI MSMS sequence coverage after tryptic digest of synthetic H-[SUMO3 (2–92)]-OH 59.

Circular dichroism spectra of H-[SUMO3 (2–92)]-OH 59
A CD spectrum of synthetic H-[SUMO3 (2–92)]-OH 59 (50 μM protein concentration, 25 °C) was recorded in PBS buffer at pH 7.4. The CD spectra of 59 compares well with a spectrum reported in the literature or recombinant SUMO3 protein suggesting that the synthetic SUMO3 adopts a fold similar to natural SUMO3.19

7.4.4. First generation of modified SUMO3 protein

Synthetic strategy

H-[SUMO3 (2–92)]-OH:

\[-\text{SEEKPKEGV}_{10}\ \text{KTENDHINLK}_{20}\ \text{VAGQDGSV}_V - \text{T}_{30}\ \text{FKIKRHTPLS}_{40}\ \text{KLMKAYCERQ}_{50}\]

GLSMRQIRFR_{40}\ \text{FDGQPINETD}_{70}\ \text{TPAQ}_L - \text{T}_{80}\ \text{MEDE}_{80}\ \text{DTIDVFQQQT}_{90}\ \text{GG}\]

7.4.4.1. Synthesis of protein segments

Synthesis of Biotin-[SUMO3 (2–29)]-α-ketoacid 67

Biotin-[SUMO3 (2–29)]-α-ketoacid 67 was prepared on protected valine α-ketoacid resin\(^{20}\) at a 0.51 mmol scale (1.50 g resin with a substitution capacity of 0.34 mmol/g). The segment was elongated up to Ser2 with automated synthesis as described in the general methods using double couplings at every step. Unlike in the synthesis of H-[SUMO3 (2–29)]-α-ketoacid 55, it was crucial to introduce Asp25-Gly26 as Dmb-protected dipeptide to avoid aspartimide formation.

For the following coupling of biotin, only an part of the resin was used (0.41 g, loading ~ 0.10 mmol/g, 41 µmol). (D)-Biotin (60.0 g, 0.243 mmol, 5.93 equiv) and HATU (91.0 mg, Kindly provided by Dr. F. Thuaud. See Thuaud, F.; Rohrbacher, F., Zwicky, A.; Bode, J. W.: Manuscript in preparation.)
0.239 mmol, 0.985 equiv) were dissolved in NMP (1.0 ml, 0.24 M) and NMM (55.2 mg, 60.0 µl, 0.541 mmol, 2.23 equiv) was added. The slightly yellow solution was incubated for 1 min and added to H₂N-Ser²–Val²⁹ α-ketoacid resin (0.41 g, loading ~ 0.10 mmol/g, 41 µmol, 1.0 equiv). The resin suspension was agitated for 2 h, the solvent removed and the resin was washed with DMF (5 x 10 ml) and CH₂Cl₂ (5 x 10 ml). The resin was cleaved following Method B and purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 60 °C; 20 to 50 % CH₃CN in 30 min, flow rate 40 ml/min) to give Biotin-[SUMO3 (2–29)]-α-ketoacid 67 (30.8 mg, 12.6 µmol, 23 %). Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of 67. m/z calculated for C₁₃₈H₂₂₆N₉₀O₅₀S₁ [M+H]⁺: 3261.60; measured 3262.59 [A+1].

Analytical HPLC trace of purified Biotin-[SUMO3 (2–29)]-α-ketoacid 67. Shiseido MGII C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 35 % CH₃CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of Biotin-[SUMO3 (2–29)]-α-ketoacid 67.
Synthesis of H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54

H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 was prepared on orthogonally protected leucine α-ketoacid resin 50 at a 0.1 mmol scale (0.29 g resin with a substitution capacity of 0.34 mmol/g). The segment was elongated up to Phe31 with automated synthesis as described in the general methods using double couplings at every step. To avoid aspartimide formation, it was crucial to introduce Asp62-Gly63 as Dmb-protected dipeptide. Boc-Opr-OH was coupled manually according to the General Procedure. The resin was cleaved according to method B. The crude peptide was dissolved in DMSO (3.0 ml) and pH 9.0 buffer (1 M NH₃/(NH₄)₂CO₃, 3.0 ml) was added and the clear solution incubated for 10 min at room temperature to reveal the free phenol on the orthogonally protected α-ketoacid. The mixture was acidified by addition of 1:4:4 TFA:H₂O:CH₃CN (1.0 ml). Purification by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 30 to 65 % CH₃CN in 35 min, flow rate 40 ml/min) gave semi-pure H-Opr-[SUMO3 31-75]-(phenol-protected)-α-ketoacid 54 (67.9 mg, 12.0 µmol, ca. 12 % yield for SPPS and purification) which was used without further purification. Analytical HPLC and MALDI FTICR MS confirmed the identity of 54. m/z calculated for C₂₅₆H₃₉₉N₇₂O₇₀S₃ [M+H]⁺: 5697.9032; measured 5697.8975.

Analytical HPLC trace of for analytical purposes repurified H-Opr-[SUMO3 31-75]-(phenol-protected)-α-ketoacid 54. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 55 % CH₃CN in 20 min.
Synthesis of H-Opr-[SUMO3 (77–92)]-OH 71

H-Opr-[SUMO3 (77–92)]-OH 71 was prepared on Wang polystyrene resin at a 0.10 mmol scale (0.31 g resin with a substitution capacity of 0.32 mmol/g). The segment was elongated up to Met77 with automated synthesis as described in the general methods. Boc-Opr-OH was coupled manually according to the general procedure. The resin was cleaved following method A. Purification by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 20 to 45 % CH₃CN in 25 min, flow rate 40 ml/min and Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 20 to 35 % CH₃CN in 20 min, flow rate 10 ml/min) gave pure H-Opr-[SUMO3 (77–92)]-OH 71 (47.3 mg, 23.9 µmol, ca. 24 % yield for SPPS and purification). Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of 71. m/z calculated for C₇₈H₁₁₈N₂₀Na₁O₃₄S₁ [M+Na]⁺: 1933.7732; measured 1933.7764.

Analytical HPLC trace of purified H-Opr-[SUMO3 (77–92)]-OH 71. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH₃CN in 20 min.
Synthesis of H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 72

H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 72 was prepared on Wang polystyrene resin at a 0.10 mmol scale (0.31 g resin with a substitution capacity of 0.32 mmol/g). The segment was elongated up to Met77 with automated synthesis as described in the general methods. Boc-Opr-OH was coupled manually according to the general procedure. The resin was cleaved following method A. Purification by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 20 to 50 % CH3CN in 30 min, flow rate 40 ml/min and Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 20 to 35 % CH3CN in 20 min, flow rate 10 ml/min) gave pure H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 72 (42.1 mg, 22.3 µmol, ca. 22 % yield for SPPS and purification). Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of 72. m/z calculated for C78H118N19O3S1 [M+H]+: 11880.7855; measured 1880.7862.

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of H-Opr-[SUMO3 (77–92)]-OH 71.

Analytical HPLC trace of purified H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 72. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH3CN in 20 min.
7.4.4.2. Sequential KAHA ligations for the synthesis of modified SUMO3 proteins 60 and 61

KAHA ligation for the preparation of Biotin-[SUMO3 (2–75)]-α-ketoacid 69

\[
\text{FKIKRHTPLS}_{40} \text{ KLMKAYCERQ}_{50} \text{ GLSMRQIRFR}_{60} \text{ FDGQPINETD}_{70} \text{ TPAQ}
\]

\[
\text{SEEKPKEGV}_{10} \text{ KTENDHINLK}_{20} \text{ VAGQDGSV}
\]

**a) Ligation:** Biotin-[SUMO3 (2–29)]-α-ketoacid 67 (31.8 mg, 9.75 µmol, 1.00 equiv) and H-Opr-[SUMO3 (31–75)]-(phenol-protected)-α-ketoacid 54 (55.1 mg, 9.75 µmol, 1.00 equiv) were weighed into a glass vial and dissolved in 7:3 NMP:H2O (485 µl, 20 mM) with 0.1 M oxalic acid. The clear solution was heated to 60 °C in a sonicator. Reaction progress was monitored by analytical HPLC (Shiseido MGII C18 column 4.6 x 250 mm, heated to 60 °C; 20 to 60 % CH3CN in 20 min, flow rate 1 ml/min).

**b) One-pot α-ketoacid deprotection:** After completion of the ligation (7 h), the crude reaction mixture was diluted with 6 mL TFA containing 2.5 % DODT and incubated at room temperature for 2 h. Volatiles were removed with a stream of nitrogen and the crude ligation product was triturated with Et2O (15 ml), centrifuged (2500 x g, 4 min), the supernatant removed by decantation and the crude ligation product dissolved in DMSO (3 ml) and 1:1 H2O:CH3CN with 0.1 % TFA (1 ml) and purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 60 °C; 15 to 55 % CH3CN in 50 min, flow rate 40 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Biotin-[SUMO3 (2–75)]-α-ketoacid depsi-69 (19.0 mg, 2.18 µmol, 23 % yield for ligation, α-ketoacid deprotection and purification steps). Analytical HPLC and ESI FTICR MS
confirmed the purity and identity of \textit{depsi-69}. m/z calculated for C$_{378}$H$_{613}$N$_{111}$O$_{116}$S$_{4}$ [M+H]$^+$: 8692.44; measured 8697.51 [A+5].

HPLC monitoring of the KAHA ligation (bottom and middle trace) of Biotin-[SUMO3 (2–29)]-\(\alpha\)-ketoacid \textit{67} and H-Opr-[SUMO3 (31–75)]-(phenol-protected)-\(\alpha\)-ketoacid \textit{54} and the one-pot deprotection of the orthogonally protected \(\alpha\)-ketoacid (upper trace). Shiseido MGII C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 60 % CH$_3$CN in 20 min.

Analytical HPLC trace of purified Biotin-[SUMO3 (2–75)]-\(\alpha\)-ketoacid \textit{depsi-69}. Shiseido MGII C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 60 % CH$_3$CN in 20 min.
ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of Biotin-[SUMO3 (2–75)]-α-ketoacid depsi-69.

**KAHA ligation for the preparation of Biotin-[SUMO3 (2–92)]-OH 60**

**a) Ligation:** Biotin-[SUMO3 (2–75)]-α-ketoacid depsi-69 (6.1 mg, 0.70 µmol, 1.0 equiv) and H-Opr-[SUMO3 (77–92)]-OH 71 (2.7 mg, 1.4 µmol, 2.0 equiv) were dissolved in 8:2 NMP:H2O (35 µl, 20 mM) with 0.1 M oxalic acid and heated to 60 °C for 8 h.

**b) Global O to N acyl shift:** The reaction mixtures were diluted with rearrangement buffer (2.0 ml, pH 9.5, 0.2 M NaHCO3/Na2CO3, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 12 h. The reaction mixtures were purified by preparative HPLC (Phenomenex Jupiter C4 column (30 x 250 mm), heated to 60 °C; 25 to 50 % CH3CN 25 min, flow rate 20 ml/min). The fractions containing the ligated product were pooled and
lyophilized to give pure Biotin-[SUMO3 (2–92)]-OH 60 (2.6 mg, 0.24 µmol, 35 %). The purity and identity of 60 was confirmed by analytical HPLC and MALDI TOF MS. m/z calculated for C_{455}H_{731}N_{131}O_{148}S_{5} [M+H]^+: 10559.2; measured 10559.7.

HPLC monitoring of the KAHA ligation of SUMO3 (2–75)-α-ketoacid depsi-69 and SUMO3 (76–92) oxaproline 71 (bottom and middle trace) and one pot global O to N acyl shift (top trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH₃CN in 20 min.

MALDI TOF MS of Biotin-[SUMO3 (2–92)]-OH 60.
**KAHA ligation for the preparation of Biotin-[SUMO3 (2–92) Gln89Pro]-OH 61**

**a) Ligation:** Biotin-[SUMO3 (2–75)]-%-ketoacid depsi-69 (8.7 mg, 1.0 µmol, 1.0 equiv) and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 72 (3.7 mg, 2.0 µmol, 2.0 equiv) were dissolved in 8:2 NMP:H₂O (50 µl, 20 mM) with 0.1 M oxalic acid and heated to 60 ºC for 8 h.

**b) Global O to N acyl shift:** The reaction mixtures were diluted with rearrangement buffer (2.0 ml, pH 9.5, 0.1 M Tris, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 12 h. The reaction mixtures were purified by preparative HPLC (Phenomenex Jupiter C4 column (30 x 250 mm), heated to 60 ºC; 25 to 50 % CH₃CN in 30 min, flow rate 20 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Biotin-[SUMO3 (2–92) Gln89Pro]-OH 61 (3.4 µg, 0.32 µmol, 32 %). The purity and identity of 61 was confirmed by analytical HPLC and ESI FTICR MS. m/z calculated for C₄₅₅H₇₃₀N₁₃₀O₁₄₇S₅ [M+H]+: 10528.23; measured 10534.275 [A+6].
Analytical HPLC trace of purified Biotin-[SUMO3 (2–92) Gln89Pro]-OH \textit{61}. Phenomenex Jupiter C4 (4.6 x 250 mm) heated to 60 °C, 20 to 50 % CH$_3$CN in 30 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) ESI FTICR MS of Biotin-[SUMO3 (2–92) Gln89Pro]-OH \textit{61}.
7.4.5. Second generation of modified SUMO3 protein

Synthetic strategy
H-[SUMO3 (2–92)]-OH:
-SEEKPKEGV\textsubscript{10} KTENDHKILK\textsubscript{20} VAGQDGSV - T\textsuperscript{30} FKIKRHTPLS\textsubscript{40} KLMKAYCEROQ\textsubscript{50} GLSMRQFR\textsubscript{60} FDQQPNETD\textsubscript{70} TPAQL - T\textsuperscript{80} MEDE DTIDVFAQQT\textsubscript{90} GG

7.4.5.1. Synthesis of protein segments

Synthesis of Biotin-Dap(Cou)-[SUMO3 (2–29)]-α-ketoacid 68

The same Ser2-Val29 protected α-ketoacid resin as in 67 (page 201) was used. Coumarin-labeled amino acid 80 (115 mg, 0.192 mmol, 2.26 equiv) and HATU (72.0 mg, 0.189 mmol, 2.23 equiv) were dissolved in DMF (2.0 ml, 96 mM) and NMM (41.4 mg, 45 µl, 0.409 mmol, 4.81 equiv) was added. The reaction was mixed for 1 min and added to preswelled Fmoc-deprotected Ser2-Val29 α-ketoacid resin (0.850 g, loading ~ 0.1 mmol/g, 85 µmol, 1.0 equiv). The resin suspension was agitated for 2 h, the solvent removed and the resin was washed with DMF (5 x 10 ml) and CH\textsubscript{2}Cl\textsubscript{2} (5 x 10 ml). After removal of residual solvents by a stream
of nitrogen, the dry resin was stored at 4°C. Microcleavage followed by HPLC analysis indicated complete coupling of the Coumarin-labeled amino acid 80.

Fmoc-PEG-OH 79 (0.200 g, 0.538 mmol, 6.33 equiv) and 4-nitrophenyl chloroformate (0.114 g, 0.565 mmol, 6.65 equiv) were dissolved in 1:1 DMF/CH₂Cl₂ (2.0 ml, 0.27 M) and DIPEA (0.139 g, 0.183 ml, 1.08 mmol, 12.7 equiv) and DMAP (6.1 mg, 50 µmol, 4.2 equiv) were added. The reaction was stirred at room temperature for 5 h and then added to Fmoc-deprotected Dap(Cou)-Ser2–Val29 α-ketoacid resin (0.850 g, loading ~ 0.10 mmol/g, 85 µmol, 1.0 equiv). The resin suspension was agitated for 3 h, the solvent removed and the resin was washed with DMF (5 x 10 ml) and CH₂Cl₂ (5 x 10 ml). The coupling was repeated once. After removal of residual solvents by a stream of nitrogen, the dry resin was stored at 4°C. Microcleavage followed by HPLC analysis indicated complete coupling of Fmoc-PEG-OH 79.

(D)-Biotin (0.122 g, 0.500 mmol, 5.88 equiv) and HATU (0.186 g, 0.490 mmol, 5.76 equiv) were dissolved in NMP (2.0 ml, 0.25 M) and NMM (0.102 g, 0.111 ml, 1.00 mmol, 11.7 equiv) was added. The slightly yellow solution was incubated for 1 min and added to Fmoc-deprotected PEG-Dap(Cou)-Ser2–Val29 α-ketoacid resin (0.850 g, loading ~ 0.1 mmol/g, 85 µmol, 1.0 equiv). The resin suspension was agitated for 2 h, the solvent removed and the resin was washed with DMF (5 x 10 ml) and CH₂Cl₂ (5 x 10 ml).

The resin was cleaved according to method B. The crude peptide was purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (50 x 250 mm), heated to 60 °C, 20 to 50 % CH₃CN with 0.1 % TFA in 30 min, flow rate 40 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Biotin-Dap(Rho)-[SUMO3 (2–29)]-α-ketoacid 68 (67.5 mg, 17.8 µmol, ca. 21 % for SPPS and purification). Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of 68. m/z calculated for C₁₆₄H₂₆₂N₄₃O₅₈S₁ [M+H]⁺: 3793.8589; measured 3793.8697.

Analytical HPLC trace of purified Biotin-Dap(Rho)-[SUMO3 (2–29)]-α-ketoacid 68. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 50 % CH₃CN in 20 min.
CHAPTER 7: EXPERIMENTAL

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of Biotin-Dap(Rho)-[SUMO3 (2–29)]-α-ketoacid 68.

Synthesis of H-Opr[SUMO3 (77–91) Gly91Ava]-OH 73

H-Opr[SUMO3 (77–91) Gly91Ava]-OH 73 was prepared on 2-chlorotrityl polystyrene resin resin at a 0.10 mmol scale (0.29 g resin with a substitution capacity of 0.35 mmol/g). For the first coupling, Fmoc-5-aminovaleric acid was used to introduce the Gly91Ava mutation. The segment was elongated up to Met77 with automated synthesis as described in the general methods. Boc-Opr-OH was coupled manually according to the general procedure. The resin was cleaved following method A. Purification by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 20 to 50 % CH₃CN with 0.1 % TFA in 30 min, flow rate 40 ml/min and Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 20 to 40 % CH₃CN with 0.1 % TFA in 20 min, flow rate 15 ml/min) gave pure H-Opr[SUMO3 (77–91) Gly91Ava]-OH 73 (40.0 mg, 21.1 µmol, ca. 21 % yield for SPPS and purification). Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of 73. m/z calculated for C₇₉H₁₂₂N₁₉O₃₀S₁ [M+H]⁺: 1896.8168; measured 1896.8165.

Analytical HPLC trace of purified H-Opr[SUMO3 (77–91) Gly91Ava]-OH 73. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH₃CN in 20 min.
CHAPTER 7: EXPERIMENTAL

Synthesis of H-Opr[SUMO3 (77–91) Gln88NMetGln89]-OH 74

H-Opr[SUMO3 (77–91) Gln88NMetGln89]-OH 74 was prepared on 2-chlorotrityl polystyrene resin resin at a 0.10 mmol scale (0.29 g resin with a substitution capacity of 0.35 mmol/g). The segment was elongated up to T90 with automated synthesis and the terminal Fmoc deprotected.

Fmoc-Gln(Trt)NMetGln(Trt)-OH 89 (0.125 g, 0.126 mmol, 1.26 equiv) and HATU (48.0 mg, 0.126 mmol, 1.26 equiv) were dissolved in 1.0 ml DMF and NMM (23.9 mg, 26.0 µl, 0.234 mmol, 2.34 equiv) was added. The slightly yellow solution was incubated for 1 min and added to the resin. The resin was agitated for 3 h, the solvent removed and the resin was washed with DMF (5 x 5 ml) and CH2Cl2 (5 x 5 ml). The segment was elongated up to Met with automated synthesis as described in the general methods and Boc-Opr-OH was coupled manually according to the general procedure. The resin was cleaved following method A. Purification by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C, 20 to 50 % CH3CN with 0.1 % TFA in 30 min, flow rate 40 ml/min and Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 20 to 37.5 % CH3CN with 0.1 % TFA in 20 min, flow rate 15 ml/min) gave two peaks of the desired peptide. The earlier elutiong peak was termed 74-A and the later eluting peak 74-B. Pure H-Opr[SUMO3 (77–91) Gln88NMetGln89]-OH 74-A (11.3 mg, 5.9 µmol, ca. 6 % yield for SPPS and purification) and H-Opr[SUMO3 (77–91) Gln88NMetGln89]-OH 74-B (10.6 mg, 5.5 µmol, ca. 5 % yield for SPPS and purification) were obtained. Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of 74-A and 74-B, respectively. m/z calculated for C79H120N20Na2O34S [M+Na]+: 1947.7889; measured 1947.7942 (74-A); C79H119N20Na2O34S [M+2Na]+: 1969.7708; 1969.7724 (74-B).
Analytical HPLC trace of purified H-Opr[SUMO3 (77–91) Gln88NMeGln89]-OH **74-A**. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH$_3$CN in 20 min.

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of H-Opr[SUMO3 (77–91) Gln88NMeGln89]-OH **74-A**.

Analytical HPLC trace of purified H-Opr[SUMO3 (77–91) Gln88NMeGln89]-OH **74-B**. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH$_3$CN in 20 min.

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of H-Opr[SUMO3 (77–91) Gln88NMeGln89]-OH **74-B**.
7.4.5.2. Sequential KAHA ligations for the synthesis of modified SUMO3 proteins 62–65-A/B

KAHA ligation for the preparation of Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70

**a) Ligation:** Biotin-PEG-Dap(Cou)-[SUMO3 (2–29)]-α-ketoacid 68 (24.0 mg, 6.32 μmol, 1.11 equiv) and H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 (32.2 mg, 5.69 μmol, 1.00 equiv) were weighed into a glass vial and dissolved in 7:3 NMP:H₂O (150 μl, 38 mM) with 0.1 M oxalic acid. The clear solution was heated to 60 °C. Reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm), heated to 60 °C; 25 to 45 % CH₃CN with 0.1 % TFA in 20 min, flow rate 1 ml/min).

**b) One-pot α-ketoacid deprotection:** After completion of the ligation (14 h), the crude reaction mixture was diluted with 4 ml TFA containing 2.5 % DODT and incubated at room temperature for 2 h. TFA was removed with a stream of nitrogen and the crude ligation product was triturated with Et₂O (15 ml), centrifuged (2500 x g, 4 min), the supernatant was removed by decantation and the trituration/washing step repeated once. The crude ligation product was dissolved in DMSO (3 ml) and 1:1 H₂O:CH₃CN with 0.1 % TFA (1 ml) and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 25 to 40 % CH₃CN with 0.1 % TFA in 20 min, flow rate 10 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70 (20.0 mg, 2.17 μmol, 38 % yield for ligation, α-ketoacid deprotection and purification steps). Analytical HPLC and ESI FTICR MS confirmed the purity and identity of depsi-70. m/z calculated for C₄₀₄H₆₄₉N₁₁₅O₁₂₅S₄ [M+H]^+: 9224.70; measured 9228.70 [A+4].
HPLC monitoring of the KAHA ligation (bottom and middle trace) of Biotin-PEG-Dap(Cou)-[SUMO3 (2–29)]-α-ketoacid 68 and H-Opr-[SUMO3 (31–75)]-(phenol-protected) α-ketoacid 54 and the one-pot deprotection of the masked α-ketoacid (upper trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₃CN in 20 min.

Analytical HPLC trace of purified Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-67. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₃CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70.
KAHA ligation for the preparation of C-terminal variants of SUMO3 (62–65)

**General procedure:** In a parallel fashion, Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid *depsi-70* and C-terminal segments *71–74-A/B* were weighed into glass vial inserts and dissolved in 8:2 NMP:H₂O with 0.1 M oxalic acid. The clear solutions were heated to 60 °C for 8 h. The reaction mixtures were diluted to 200 µl with rearrangement buffer (pH 9.5, 0.2 M NaHCO₃/Na₂CO₃, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 12 h. The reaction mixtures were diluted to 1.0 ml with 1:1 H₂O:CH₃CN + 0.1 % and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 25 to 40 % CH₃CN with 0.1 % TFA in 20 min, flow rate 15 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure SUMO3 proteins 62–65.
Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH 62

Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid *depsi*-70 (0.8 mg, 0.09 µmol, 1.0 equiv) and H-Opr-[SUMO3 (77–92)]-OH 71 (2.0 mg, 1.1 µmol, 12 equiv) were dissolved in 8:2 NMP:H2O (15 µl, 6 mM) with 0.1 M oxalic acid and reacted and purified as described in the general procedure. The purity and identity of Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH 62 (0.3 mg, 0.03 µmol, ca. 33 %) was confirmed by analytical HPLC and ESI FTICR MS. m/z calculated for C481H768N135O156S5 [M+H]+: 11091.5; measured 11096.5 [A+4].

Analytical HPLC trace of purified SUMO3 (2-92) Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH 62. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH3CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH 62.
Biotin-PEG-Dap(Cou)-[SUMO3 (2–92) Gln89Pro]-OH 63

Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70 (2.5 mg, 0.27 µmol, 1.0 equiv) and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 72 (3.6 mg, 1.9 µmol, 7.0 equiv) were dissolved in 8:2 NMP:H₂O (25 µl, 11 mM) with 0.1 M oxalic acid and reacted and purified as described in the general procedure. The purity and identity of Biotin-PEG-Dap(Cou)-[SUMO3 (2–92) Gln89Pro]-OH 63 (1.1 mg, 0.10 µmol, ca. 37 %) was confirmed by analytical HPLC and ESI FTICR MS. m/z calculated for C₄₈₁H₇₆₇N₁₃₄O₁₅₅S₅ [M+H]⁺: 11060.5; measured 11066.5 [A+6].

Analytical HPLC trace of purified Biotin-PEG-Dap(Cou)-[SUMO3 (2–92) Gln89Pro]-OH 63. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₃CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of Biotin-PEG-Dap(Cou)-[SUMO3 (2–92) Gln89Pro]-OH 63.
Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gly91Ava]-OH 64

Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid *depsi*-70 (0.8 mg, 0.08 µmol, 1.0 equiv) and H-Opr[SUMO3 (77–91) Gly91Ava]-OH 73 (2.4 mg, 1.3 µmol, 16 equiv) were dissolved in 8:2 NMP:H₂O (15 µl, 6 mM) with 0.1 M oxalic acid and reacted and purified as described in the general procedure. The purity and identity of Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gly91Ava]-OH 64 (0.2 mg, 0.02 µmol, ca. 22 %) was confirmed by analytical HPLC and ESI FTICR MS. m/z calculated for C₄₈₂H₇₇₀N₁₃₄O₁₅₅S₅ [M+H]⁺: 11076.5; measured 11081.5 [A+5].

Analytical HPLC trace of purified Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gly91Ava]-OH 64. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₃CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of 64.
Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88\_NMe\_Gln89]-OH 65-A

Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-\(\alpha\)-ketoacid \textit{depsi-70} (0.7 mg, 0.07 \(\mu\)mol, 1.0 equiv) and H-Opr[SUMO3 (77–91) Gln88\_NMe\_Gln89]-OH \textit{74-A} (1.9 mg, 1.0 \(\mu\)mol, 14 equiv) were dissolved in 8:2 NMP:H\(_2\)O (15 \(\mu\)l, 5 mM) with 0.1 M oxalic acid and reacted and purified as described in the general procedure. The purity and identity of Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88\_NMe\_Gln89]-OH \textit{65-A} (0.2 mg, 0.02 \(\mu\)mol, ca. 25 \%) was confirmed by analytical HPLC and ESI FTICR MS. \(m/z\) calculated for \(C_{482}H_{769}N_{135}O_{156}S_5\) [M+H]\(^+\): 11105.5; measured 11111.5 [A+6].

Analytical HPLC trace of purified Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88\_NMe\_Gln89]-OH \textit{65-A}. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 \% CH\(_3\)CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of SUMO3 [Q88\_NMe\_Q89] 65-A.
Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88NMeGln89]-OH 65-B

Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70 (0.8 mg, 0.09 µmol, 1.0 equiv) and H-Opr[SUMO3 (77–91) Gln88NMeGln89]-OH 74-B (2.0 mg, 1.0 µmol, 11 equiv) were dissolved in 8:2 NMP:H2O (15 µl, 6 mM) with 0.1 M oxalic acid and reacted and purified as described in the general procedure. The purity and identity of of Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88NMeGln89]-OH 65-B (0.3 mg, 0.03 µmol, ca. 31 %) was confirmed by analytical HPLC and ESI FTICR MS. m/z calculated for C482H769N135O156S5 [M+H]+: 11105.5; measured 11111.5 [A+6].

Analytical HPLC trace of purified Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88NMeGln89]-OH 65-B. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH3CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88NMeGln89]-OH 65-B.
7.4.6. Third generation of modified SUMO3 protein

Synthetic strategy

SUMO3 (2-92):

-SEEKPKEGV\textsubscript{10} KTENDHINLK\textsubscript{20} VAGQDGSV – T\textsuperscript{30} FIKKRHTPLS\textsubscript{40} KLMKAYCERQ\textsubscript{50} GLSMRQIRFR\textsubscript{60} FDGQPINETD\textsubscript{70} TPAQL – T\textsuperscript{80} MEDE FIDVFQQ\textsubscript{90} GG

7.4.6.1. Synthesis of protein segments

Synthesis of Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]-\textalpha-ketoacid 75

Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]-\textalpha-ketoacid 75 was prepared from the same protected valine \textalpha-ketoacid resin (which was elongated up to Ser2) used in the synthesis of 68 (page 212).

Rhodamine-labeled amino acid 89 (20.0 mg, 23.1 \textmu mol, 1.93 equiv) and COMU (9.9 mg, 23 \textmu mol, 1.92 equiv) were dissolved in DMF (0.50 ml, 46 mM) and NMM (9.4 mg, 11 \textmu l, 92 \textmu mol, 4.0 equiv) was added. The reaction was mixed for 1 min and added to preswelled
and Fmoc-deprotected Ser2-V29 α-ketoacid resin (0.120 g, loading ~ 0.01 mmol/g, 12 µmol, 1.0 equiv). The resin suspension was agitated for 2 h, the solvent removed and the resin was washed with DMF (5 x 5 ml) and CH₂Cl₂ (5 x 5 ml). After removal of residual solvents by a stream of nitrogen, the dry resin was stored at 4°C. Microcleavage followed by HPLC analysis indicated complete coupling of the Rhodamine-labeled amino acid 89.

Fmoc-PEG-OH 79 (0.186 g, 0.500 mmol, 41.7 equiv) and 4-nitrophenyl chloroformate (0.106 g, 0.525 mmol, 43.7 equiv) were dissolved in 1:1 DMF/CH₂Cl₂ (2.0 ml, 0.25 M) and DIPEA (71.1 mg, 95.8 µl, 0.550 mmol, 45.9 equiv) and DMAP (6.1 mg, 50 µmol, 4.2 equiv) were added. The reaction was stirred at room temperature for 5 h and then added to H₂N-Dap(Rho)-Ser2–Val29 α-ketoacid resin (0.120 g, loading ~ 0.1 mmol/g, 12 µmol, 1.0 equiv). The resin suspension was agitated for 3 h, the solvent removed and the resin was washed with DMF (5 x 5 ml) and CH₂Cl₂ (5 x 5 ml). After removal of residual solvents by a stream of nitrogen, the dry resin was stored at 4°C. Microcleavage followed by HPLC analysis indicated complete coupling of the Fmoc-PEG-OH 79.

(D)-Biotin (48.8 mg, 0.200 mmol, 16.7 equiv) and HATU (75.9 mg, 0.200 mmol, 16.7 equiv) were dissolved in NMP (0.75 ml, 0.27 M) and NMM (81.5 mg, 99.3 µl, 0.800 mmol, 33.4 equiv) was added. The slightly yellow solution was incubated for 1 min and added to H₂N-PEG-Dap(Rho)-Ser2–Val29 α-ketoacid resin (0.120 g, loading ~ 0.01 mmol/g, 12 µmol, 1.0 equiv). The resin suspension was agitated for 2 h, the solvent removed and the resin was washed with DMF (5 x 5 ml) and CH₂Cl₂ (5 x 5 ml).

The resin was cleaved according to method B. The crude peptide was purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 25 to 45 % CH₃CN with 0.1 % TFA in 20 min, flow rate 15 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]- α-ketoacid 75 (11.3 mg, 2.78 µmol, ca. 23 % for SPPS and purification). Analytical HPLC and ESI QTOF MS confirmed the purity and identity of 75. m/z calculated for C₁₇₅H₂₇₂N₄₄O₆₁S₃ [M+H]+: 4062.87; measured 4063.90 [A+1].

![Analytical HPLC trace of purified Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]- α-ketoacid 75. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 15 to 45 % CH₃CN in 20 min.](image-url)
CHAPTER 7: EXPERIMENTAL

**Synthesis of Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76**

Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76 was prepared on protected leucine α-ketoacid resin 41 at a 0.25 mmol scale (0.74 g resin with a substitution capacity of 0.34 mmol/g). The segment was elongated up to Phe31 with automated synthesis as described in the general methods using double couplings at every step. To avoid aspartimide formation, it was crucial to introduce D62–Gly63 as Dmb-protected dipeptide. Fmoc-Opr-OH was coupled manually according to the general procedure. The resin was cleaved following method B. Purification by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C, 25 to 45 % CH₃CN with 0.1 % TFA in 30 min, flow rate 40 ml/min) gave Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76 (71.3 mg, 12.5 μmol, ca. 2.5 % yield for SPPS and purification). Analytical HPLC and MALDI FTICR MS confirmed the purity and identity of Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76. m/z calculated for C₂₅₆H₃₉₉N₇₂O₇₀S₃ [M+H]+: 5697.9032; measured 5697.8975.

![Analytical HPLC trace of purified Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH₃CN in 20 min.](image)
7.4.6.2. Sequential KAHA ligations for the synthesis of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66

KAHA ligation for the preparation of H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78

a) Ligation: Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76 (24.4 mg, 4.28 μmol, 1.00 equiv) and H-Opr-[SUMO3 (77–92 Gln89Pro)]-OH 72 (9.70 mg, 5.14 μmol, 1.20 equiv) were weighed into a glass vial and dissolved in 8:2 NMP:H2O (150 μl, 28 mM) with 0.1 M oxalic acid. The clear solution was heated to 60 °C. Reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column 4.6 x 250 mm, heated to 60 °C; 25 to 45% CH3CN with 0.1% TFA in 20 min, flow rate 1 ml/min). After completion of the ligation (8 h), the reaction mixture was diluted to 3 mL with 1:1 CH3CN/H2O and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 30 to 50 % CH3CN with 0.1 % TFA in 20 min, flow rate 15 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH.

b) one-pot Fmoc-deprotection and O to N acyl shift: 5:95 Et2NH/DMSO, 5 min followed by 0.2 M Na2CO3 buffer, 6 M GdmCl, 10 mM TCEP, pH 9.5, 1 h, 49 % yield (two steps)

---

MALDI FTICR MS: Measured (top) and calculated (bottom) isotopic pattern of Fmoc-Opr-[SUMO3 (31–75)]-α-ketoacid 76.
depsi-77 (18.5 mg, 2.45 μmol, 57 % yield for ligation and purification steps). Analytical HPLC and ESI QTOF MS confirmed the purity and identity of Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH depsi-77. m/z calculated for C_{333}H_{515}N_{91}O_{101}S_4 [M+H]^+: 7533.69; measured 7536.72 [A+3].

HPLC monitoring of the KAHA ligation of Fmoc-Opr-[SUMO3 (31–75)]- α-ketoacid 76 and H-Opr-[SUMO3 (77–92 Gln89Pro)-OH 72. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₂CN in 20 min.

Analytical HPLC trace of purified Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH depsi-77. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₂CN in 20 min.

b) One-pot Fmoc-deprotection and O to N acyl shift: Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH depsi-77 (18.5 mg, 2.45 μmol) was dissolved in DMSO (2.50 ml, 0.98 mM) containing 5 % (v/v) HNEt₂ and incubated at room temperature for 5 min. Ice cold rearrangement buffer (2.50 ml, pH 9.5, 0.2 M NaHCO₃/Na₂CO₃ buffer, 6 M GdmCl, 10 mM TCEP) was added and the mixture incubated for 1 h at room temperature. The reaction was

![](image)
quenched by addition of TFA (100 µl) and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 25 to 45 % CH₃CN with 0.1 % TFA in 20 min, flow rate 15 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78 (8.8 mg, 1.2 µmol, 49 % yield for ligation and purification steps). Analytical HPLC and ESI QTOF MS confirmed the purity and identity of H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78. m/z calculated for C₃₁₈H₅₀₅N₉₁O₉₉S₄ [M+H]+: 7311.62; measured 7314.64 [A+3].

HPLC monitoring of the Fmoc removal (middle trace) and O to N acyl shift of Fmoc-Opr-[SUMO3 (31–92) Gln89Pro]-OH depsi-77 to H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78 (upper trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₃CN in 20 min.

Analytical HPLC trace of purified H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH₃CN in 20 min.

ESI QTOF MS: Measured (top) and deconvoluted (bottom) MS of H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78.
KAHA ligation for the preparation of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66

a) Ligation: H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78 (8.5 mg, 1.2 µmol, 1.0 equiv) and Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]-α-ketoacid 75 (5.3 mg, 1.3 µmol, 1.1 equiv) were weighed into a glass vial insert and dissolved in 8:2 NMP:H₂O (40 µl, 30 mM) with 0.1 M oxalic acid. The clear solution was heated to 60 °C. Reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column 4.6 x 250 mm, heated to 60 °C, 25 to 45 % CH₃CN with 0.1 % TFA in 20 min, flow rate 1 ml/min).

b) Global O to N acyl shift: After completion of the ligation (7 h), the reaction mixture was diluted to 1 ml with rearrangement buffer (pH 9.5, 0.2 M NaHCO₃/Na₂CO₃, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 12 h. The reaction mixture was diluted with 1:1 H₂O:CH₃CN containing 0.1 % TFA (3 ml) and purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 25 to 55 % CH₃CN with 0.1% TFA in 30 min, flow rate 8 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 (3.0 mg, 0.26 µmol, 23 % yield for ligation and purification steps). Analytical HPLC and ESI-QTOF MS confirmed the purity and identity of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66. m/z calculated for C₄₉₂H₇₇₇N₁₃₅O₁₅₈S₇ [M+H]^+: 11329.50; measured 11335.53 [A+6].
HPLC monitoring of the KAHA ligation of Biotin-PEG-Dap(Rho)-[SUMO3 (2–30)]-\(\alpha\)-ketoacid 75 and H-Opr-[SUMO3 (31–92) Gln89Pro]-OH 78 (bottom and middle trace) and one pot global O to N acyl shift (top trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH\(_3\)CN in 20 min.

Analytical HPLC trace of purified Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 25 to 45 % CH\(_3\)CN in 20 min.

ESI-QTOF MS: Measured (top) and deconvoluted (bottom) MS of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66.

Circular dichroism spectra of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66
A CD spectrum of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 (25 μM protein concentration, 25 °C) was recorded in PBS buffer at pH 7.4. The spectrum compares well to the unlabeled H-[SUMO3 (2–92)]-OH 59 (see page 200), indicating that the introduced labels and mutations do not affect the protein folding.
CD spectrum of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66, 25 μM protein concentration in PBS buffer, pH 7.4, 25 ºC.

**UV-VIS and fluorescence spectrum of 66**

UV-VIS spectrum of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66, 10 μM in PBS + 1 mM DTT, pH 7.4, 25 ºC, $\lambda_{\text{max}} \text{(vis)} = 571$ nm.

Fluorescence spectrum of Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66, 10 μM in PBS + 1 mM DTT, pH 7.4, 25 ºC, $\lambda_{\text{ex}} = 571$ nm, $\lambda_{\text{em, max}} = 587$ nm.
7.5. Experimental for “CHAPTER 4: Chemical synthesis of SUMO3 proteins by KAHA ligation”: Biochemical experiments and SUMOylation assays with synthetic SUMO proteins

7.5.1. Materials and General Methods

a) Materials

**In vitro SUMOylation reactions:** For *in vitro* SUMOylation reactions, a commercially available SUMOylation kit from Enzo Life Sciences (BML-UW8955) was used. Components that were used from the kit included: SUMO E1 (SUMO activating enzyme E1 (human), (recombinant), BML-UW9330), SUMO E2 (SUMO conjugating enzyme E2, Ubc9 (human), (recombinant) (untagged), BML-UW9320), recombinant SUMO 3 protein (SUMO-3 (human), (recombinant) (His-tag), BML-KW9195), SUMOylation buffer (BML-KW9890, vide infra), RanGAP1 (model substrate protein for SUMOylation, RanGAP1 fragment (418–587), (human), (recombinant) (GST-tag), BML-UW9755), Mg-ATP solution (BML-KW9805) and SUMO2/3 antibody (BML-PW9465).

For a number of experiments (e.g. large scale SUMOylation experiments), certain components of the kit were either bought separately from the same supplier (Ubc9 (human), (recombinant) (untagged), Enzo Life Sciences BML-UW9320-0100) or replaced by in house made solutions that were found to perform equally well. 10X SUMOylation buffer (500 mM Tris-HCl, pH 7.6, 500 mM MgCl₂, 10 mM DTT) was prepared following a reported procedure.²¹ 20 X Mg-ATP (50 mM Tris, pH 7.5, 0.10 M ATP, 0.10 M MgCl₂) was prepared by dissolving Mg-ATP (Sigma Aldrich A9187, 10 mg, 20 µmol) in Tris buffer (150 µl, 50 mM, pH 7.5) and adjusting the pH to 7.5 by addition of NaOH solution (2.0 M). A solution of MgCl₂ (1.0 M, 20 µl) was added and the mixture brought to a total volume of 200 µl by addition of Tris buffer (50 mM, pH 7.5).

**SENP protease assays:** In deSUMOylation assays, SENP2 enzyme (SENP2 (catalytic domain) (human), (recombinant) (GST-tag), Enzo Life Sciences BML-UW9765) was used.

**SDS-PAGE:** Proteins were mixed with loading buffer (2X, Sigma Aldrich S3401 or – if in-gel fluorescence detection was used – a in-house made buffer: 100 mM Tris-HCl, pH 6.8, 4% w/v SDS, 20 % glycerol, 200 mM DTT), heated to 95 ºC for 5 min, loaded onto the gels (typically 10 or 20 µl per lane) and separated by SDS-PAGE using a Mini-PROTEAN® system (Biorad) and Mini-PROTEAN® TGX Stain-Free™ Precast Gels (Biorad, 7.5 % Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 10 well, 50 µl, 4568024; 8–16 % Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 10 well, 50 µl, 4568104; 4–20 % Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 10 well, 50 µl, 4568094) according to the

instructions provided by the supplier. Proteins were visualized by staining with Coomassie Brilliant Blue, using a silver staining protocol or were directly visualized by in-gel fluorescence measurements using a Biorad Molecular Imager FX System (equipped with a 532 nm laser for excitation and a 555 nm lowpass filter). For better visibility, in-gel fluorescence and Western blot images were digitally processed in a manner treating the whole pictures equally (enhancement of contrast, adjustment of brightness and grey curves) using commercially software (Adobe Photoshop). As molecular weight marker, Precision Plus Protein™ Dual Xtra Prestained Protein Standards (Biorad 1610377) were used.

**Western blots:** For immunoblots, proteins were transferred onto PVDF or nitrocellulose membranes using precut blotting transfer packs (Biorad Trans-Blot® Turbo™ Mini PVDF Transfer Packs, 1704156; Trans-Blot® Turbo™ Mini Nitrocellulose Transfer Packs, 1704158) and a Biorad Trans-Blot® Turbo™ Transfer System following the protocol of the manufacturer. As primary antibodies were used: SUMO-2/3 (human) (NT) polyclonal antibody (rabbit, Enzo Life Sciences BML-PW9465) and Ran GAP1 polyclonal antibody (rabbit, Santa Cruz Biotechnology sc-25630). As secondary antibody, a goat anti-rabbit IgG polyclonal antibody, conjugated to HRP (Enzo Life Sciences ADI-SAB-300-J) was was used. Biotin-labeled proteins were detected using a streptavidin-HRP conjugate (Pierce 21130). Protein bands on Western blots were visualized by chemical staining using a commercial kit (Opti-4CN Substrate Kit, Biorad 1708235). As molecular weight marker, Precision Plus Protein™ WesternC™ Blotting Standards (Biorad 1610385) in combination with a StrepTactin-HRP Conjugate (for detection of the marker) were used.

**Cell culture:** Dulbecco’s Modified Eagle’s Medium - high glucose (with 4500 mg/L glucose, L-glutamine, sodium pyruvate, and sodium bicarbonate, Sigma Aldrich D6429), supplemented with 10 % FBS (Gibco 10270) and 1 % Pen Strep (Sigma Aldrich P4458) was used to grow HEK293 and HeLa cells at 37 °C and 5 % CO₂ in standard cell culture flasks. Cells were regularly passed at ca. 90 % confluency (depending on the seeding amount ca. every 3–4 days) and media was changed whenever necessary (typically once between two passages). To mobilize adherent cells, trypsin (0.5 g/l porcine trypsin, 0.2 g/l EDTA, in PBS, from 10X stock, Sigma Aldrich T4174) was used. For cell lysis, a cell lysis buffer (50 mM HEPES, pH 7.4, 150 mM KCl, 1 mM EDTA, 1% Nonidet P-40 (Nonidet® P40 (Substitute) solution 10 % peroxide-free, AppliChem A2239), 0.1 % CHAPS), described by Salvesen et

---

al.\textsuperscript{24} was used. Just before use, a protease inhibitor cocktail (1 % v/v, unless otherwise indicated; Sigma Aldrich P8340, containing AEBSF, Aprotinin, Bestatin, E-64, Leupeptin, Pepstatin A) was added to the buffer.

**Biotin-tag affinity purification:** SUMOylated proteins from *in vitro* SUMOylation reactions using probe \textsuperscript{66} were purified using Pierce™ Monomeric Avidin Agarose (ThermoFisher Scientific 20228), in-house packed in Mini Bio-Spin® Chromatography Columns (Biorad 7326207) following a modified protocol based on the instructions provided by the manufacturer.

**b) General Procedures**

**In vitro SUMOylation reactions:** Small scale reactions (for analysis by SDS-PAGE and Western blot) using a commercially available SUMOylation kit from Enzo Life Sciences were performed following the procedure provided by the manufacturer. If synthetic SUMO proteins were used in experiments, they were added as 20 X stock solution in SUMOylation buffer (50 µM protein concentration, ~ 0.5 mg/ml, stored at –20 ºC). If crude cell lysate was used as SUMOylation substrate instead of RanGAP, it was freshly prepared prior to the experiment and directly added to the reaction mixture (total protein concentration in lysate: ~ 35 mg/ml). Other kit components were stored in aliquots at –80 ºC and thawed before the experiments. Homogeneity of the solution was ensured by complete thawing and mixing by pipet. In a typical experiment, the following components were added to a micro centrifuge tube:

\[
\begin{align*}
MQ-H2O & \quad 13 \mu l \\
10 X SUMOylation buffer & \quad 2 \mu l \\
20 X Mg-ATP & \quad 1 \mu l \\
20 X SUMO E1 & \quad 1 \mu l \\
20 X SUMO E2 & \quad 1 \mu l \\
20 X RanGAP1 (or cell lysate) & \quad 1 \mu l \\
SUMO3 (from kit or synthetic) & \quad 1 \mu l
\end{align*}
\]

The solution was mixed by pipet and incubated at 37 ºC for 1 h (unless otherwise indicated). For analysis by SDS-PAGE, the reaction was quenched by addition of 2 X loading buffer, heated to 95 ºC for 5 min and analyzed by SDS-PAGE, in-gel fluorescence or Western blot. Large scale SUMOylation reactions of cell lysates (for pull down experiments and MS analysis) used a modified protocol and are described in full detail at individual experiments.

**Western blots:** Proteins were transferred from SDS gels onto PVDF or nitrocellulose membranes using a Biorad Trans-Blot® Turbo™ Transfer System following the protocol of the manufacturer. To block non-specific binding, the membranes were incubated with PBS-T (PBS buffer with 0.02 % Tween-20) with 1 % BSA for 60 min and washed thoroughly with PBS-T (3 x 10 min). The membrane was then treated with the primary antibody (anti SUMO2/3 or anti RanGAP) at a dilution of 1:2000 in PBS-T with 1 % BSA (5 µl antibody in 10 ml) and incubated overnight at 4 ºC on a rocking platform. The membrane was thoroughly washed with PBS-T (3 x 10 min), followed by incubation with the secondary antibody (Goat Anti rabbit-HRP conjugate) at a dilution of 1:3000 in PBS-T with 1 % BSA together with Precision protein™ StrepTactin-HRP conjugate for detection of the marker (3.3 µl secondary antibody and 2 µl protein standard in 10 ml) at 25 ºC. Biotin-labeled proteins were detected using a streptavidin-HRP conjugate at a dilution of 1:2000 in PBS-T with 1 % BSA (5 µl streptavidin-HRP conjugate in 10 ml) instead of antibodies. After 60 min, the membrane was thoroughly washed with PBS-T (6 x 10 min) and dried. For colorimetric detection of the proteins on the blot, the dried membrane was incubated with opti-4CN substrate solution (15 mL) until the desired level of signal was obtained.

**Cell culture:** Cell culture was performed following standard protocols. Briefly, a typical procedure used for growing HEK293 in a T25 cell culture flask is described. For larger T75 flasks, all quantities/volumes were tripled.

Starting the cell culture from frozen stocks: HEK293 cells (frozen and stored at – 80 ºC; ca. 1 mio cells) were thawed and transferred into 3 ml full DMEM medium in a centrifuge tube. The cells were pelleted by centrifugation (5 min at 125 x g), the medium carefully removed with an aspirator and the cells resuspended in 3 ml fresh DMEM medium. The suspension was added to a T25 cell culture flask and the cells evenly distributed by carefully tilting the flask. The cells were incubated at 37 ºC and 5 % CO₂ until reaching ca. 90 % confluency; the medium was changed whenever necessary (typically once between every passage).

Cell passage: The medium was removed from the flask with an aspirator and the adherent cells were washed once with PBS (3 ml). Trypsin (1 ml) was added to the cells and incubated for 3–5 min at 37 ºC until the cells detached from the flask. The cell suspension was immediately diluted with full DMEM medium (2 ml), transferred to a centrifuge tube and the cells pelleted by centrifugation (5 min at 125 x g). The supernatant was removed and the cells resuspended in full DMEM medium (3 ml) and the number of cells estimated using a Neubauer chamber. Aliquots (containing ca. 300000 cells for a T25 flask) of the cell suspension were transferred into the desired number of new flasks and filled up with full DMEM medium to a total volume of 3 ml. The cells evenly distributed by carefully tilting the flask and incubated at 37 ºC and 5 % CO₂ until the next passage. After ca. 20 passages, a new culture from frozen stocks was started.
Cell lysis: Cells that were grown in a T25 flask and reached 90 % confluency (ca. 9 mio cells) were trypsinated and pelleted in a micro centrifuge tube as described above. Ice cold lysis buffer (75 µl, supplemented with 1 % protease inhibitor cocktail) was added to the cell pellet which was resuspended using a pipett and incubated for 5 min at 4 ºC. The cell debris was precipitated by centrifugation (8000 x g, 10 min, 4 ºC) and the supernatant transferred into a new micro centrifuge tube. The cell lysate was kept on ice until it was used in SUMOylation experiments.

**Biotin-tag affinity purification:** Typically, a large scale SUMOylation reaction (total volume 750 µl or 1.5 ml) was subjected to Biotin-tag affinity purification. Following the instructions provided by the manufacturer, 400 µl avidin agarose (as 50 % aqueous slurry) were pipetted into an empty Mini Bio-Spin column and the liquid drained. To block non-specific binding sites, the beads were washed with PBS (800 µl), elution buffer (600 µl, 2 mM D-biotin in PBS), regeneration buffer (1.2 ml, 0.1 M glycine, pH 2.8) and PBS (800 µl). The crude, unquenched SUMOylation reaction (750 µl) was directly added to the beads and incubated on a rocking platform at room temperature for 1 h. The liquid was drained and the beads washed with PBS (10 x 600 µl PBS); all fractions were collected in separate 2.2 ml centrifuge tubes. The proteins were eluted with elution buffer (3 x 400 µl, each aliquot incubated ca. 1 min on the beads). For SDS-PAGE analysis, the elution fractions were pooled and concentrated using molecular weight cutoff spin filters (Pierce™ Protein Concentrators PES, 3K MWCO, 0.5 mL, 88512). For proteomic analysis, the elution fractions were kept at -20 ºC until MS analysis. The beads were washed with regeneration buffer (3 x 400 µl), PBS (800 µl) and stored wet with PBS containing 0.05% NaN₃ until next usage.
7.5.2. SUMOylation of RanGAP1 with H-[SUMO3 (2–92)]-OH

RanGAP1 was SUMOylated using recombinant SUMO3 (as positive control, provided with the SUMOylation kit) or synthetic H-[SUMO3 (2–92)]-OH as described in the General Methods. As negative control, the SUMOylation reaction was performed without the addition of Mg-ATP solution. The reactions were analyzed by SDS PAGE (8–16 % pre-cast gels) and Western blot as described above.

Results: The Western blot clearly showed bands for the RanGAP1 conjugate similar to the positive control using recombinant SUMO3. As expected for the negative controls that lacked Mg-ATP in the enzymatic reaction, no bands for either RanGAP1 or RanGAP1-recombinant SUMO3 were observed. These results demonstrate that the synthetic H-[SUMO3 (2–92)]-OH is readily recognized and processed by the SUMOylation machinery. This further shows that the synthetic H-[SUMO3 (2–92)]-OH (with the two homoserine mutations Gln30Hse and S53T) can perform the full biological function similar to the natural SUMO3.
7.5.3. SUMOylation studies of RanGAP1 using Biot-[SUMO3 (2–92)]-OH 60 and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61

SUMOylation of RanGAP1 with Biot-[SUMO3 (2–92)]-OH 60 and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61

RanGAP1 was SUMOylated as described in the General Methods using recombinant SUMO3 (as positive control, provided with the SUMOylation kit) or Biot-[SUMO3 (2–92)]-OH 60 or H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 proteins. The reactions were analyzed by SDS PAGE (8–16 % pre-cast gels) and Western blot (anti SUMO 2/3) as described above.

Results: Analysis by Coomassie staining indicated that the concentrations and absolute amounts of proteins used in a typical SUMOylation reaction are close to the detection limit by
Coomassie staining, as band corresponding to RanGAP1 and the SUMO3-RanGAP1 conjugates are barely visible. Impurities in recombinant SUMO3 and in the SUMOylation buffer (from the kit) are marked by an asterisk.

The Western blot clearly showed bands for the RanGAP1–Biotin-[SUMO3 (2–92)]-OH 60 and RanGAP1–H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 conjugates similar to the positive control using recombinant SUMO3 and the previously used unlabeled H-[SUMO3 (2–92)]-OH 59. These results demonstrate that the synthetic Biotin-[SUMO3 (2–92)]-OH 60 and H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 proteins are readily recognized and processed by the SUMOylation machinery in comparable efficiency as the recombinant SUMO3. This further indicates that the N-terminal biotin label, the homoserine mutations (Gln30Hse and Glu76Hse) and the Gln89Pro mutation in 61 do not affect the biological function.

DeSUMOylation assays of RanGAP1–Biotin-[SUMO3 (2–92)]-OH 60 and RanGAP1–H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 conjugates

To generate the corresponding conjugates, RanGAP1 was SUMOylated as before using biotin labeled synthetic Biotin-[SUMO3 (2–92)]-OH 60 or H-Opr-[SUMO3 (77–92) Gln89Pro]-OH 61 proteins. The two crude SUMOylation reactions were split into four equal aliquots (4 x 5 µl) and SENP2 enzyme (from 5 µM stock solution) was added to a final SENP2 concentration of 1 µM, 500 nM and 50 nM (2 µl, 1 µl or 1 µl of a 1:10 dilution of SENP stock solution, respectively). To one aliquot, no SENP2 was added (control). The reactions were filled up to a total volume of 10 µl with SENP reaction buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.5 mM DTT) and incubated at 37 ºC for 1 h. The reactions were quenched by addition of loading buffer (2X, 10 µl) and the samples heated to 95 ºC for 5 min, followed by analysis using SDS PAGE (8–16 % pre-cast gels) and Western blot (anti RanGAP1).
Results: For Biotin-[SUMO3 (2–92)]-OH, the Western blot analysis indicated a complete SUMOylation of the RanGAP1 substrate after the SUMOylation reaction. Addition of SENP2 enzyme lead to quantitative deSUMOylation after 1 h at 1 µM and 500 nM SENP2 concentration as indicated by complete disappearance of the band corresponding to the RanGAP1–Biotin-[SUMO3 (2–92)]-OH conjugate and reappearance of the band corresponding to unconjugated RanGAP1 protein. Even at the lowest tested SENP2 concentration (50 nM), significant deSUMOylation was observed.

In contrast to this, the SUMOylation of RanGAP1 using H-Opr-[SUMO3 (77–92) Gln89Pro]-OH proceeded with incomplete conversion as indicated by remaining unconjugated RanGAP1 substrate. Addition of SENP2 enzyme lead to minimal deSUMOylation only at the highest tested SENP2 concentrations (1 µM and 500 nM), as the majority of RanGAP1–H-Opr-[SUMO3 (77–92) Gln89Pro]-OH conjugate was still present after 1 h. At the lowest SENP2 concentration (50 nM), no change was observed in comparison to the control. These results demonstrate that H-Opr-[SUMO3 (77–92) Gln89Pro]-OH is deconjugated by SENP2 only at non-physiologically high enzyme concentrations and to a small extent.

7.5.4. SUMOylation studies of RanGAP1 with Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH proteins

RanGAP1 was SUMOylated as described in the General Methods using Biotin-PEG-Dap(Cou)-[SUMO3 (2–92)]-OH proteins, which differ in their C-terminal region. After the SUMOylation reaction, the reaction mixtures were split in half; one batch was directly analyzed by SDS PAGE (8–16 % pre-cast gels) and Western blot (anti SUMO2/3); the other batch was used in a deSUMOylation assay as described above (final concentration of SENP2: 1 µM).

Western blot of SUMOylation reactions (left panel, anti SUMO2/3) and deSUMOylation assay (right panel, anti RanGAP1) of SUMO3 C-terminal variants. In the anti RanGAP1 Western blot, the SUMO3 variants are labeled by the Biotin-targeting StrepTactin-HRP conjugate used for detection of the marker.
Results: Western blot analysis (anti SUMO2/3) of the SUMOylation reactions showed bands corresponding to RanGAP1-SUMO3 conjugates for all tested SUMO3 variants, indicating that they all were recognized by the SUMOylation machinery and conjugated to RanGAP1. However, especially Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gly91Ava]-OH 64 and Biotin-PEG-Dap(Cou)-[SUMO3 (2–91) Gln88NMeGln89]-OH 65-A showed much less intense bands for the RanGAP1-SUMO3 conjugate, indicating that these two SUMO3 variants are not as well recognized and processed by the SUMOylation machinery as the other SUMO3 variants 62, 63 and 65-B.

In the deSUMOylation assay, Western blot analysis (anti RanGAP) showed no band corresponding to unconjugated RanGAP1 protein in case of Biotin-PEG-Dap(Cou)-[SUMO3 (2–92) Gln89Pro]-OH 63, indicating that a) the SUMOylation of this substrate was quantitative and b) no deconjugation occurred. These results underline, that the Gln89Pro mutant is only to a very small extent susceptible to deconjugation by SENP2. For all other SUMO3 mutants the results are not conclusive, as it cannot be distinguished if the observed band of RanGAP1 arose from incomplete conjugation during the SUMOylation reaction or from deconjugation during the deSUMOylation assay.

Control experiments indicated that the bands observed close to the SUMO variants are low abundant (as judged by SDS PAGE and Coomassie staining), inseparable impurities from the synthesis (Biotin-PEG-Dap(Cou)-[SUMO3 (2–75)]-α-ketoacid depsi-70) or formed during the heating cycle of the sample preparation for SDS PAGE. None of these impurities was conjugated to RanGAP1 or had any other observable effect on the reactions.
7.5.5. SUMOylation studies of proteins from crude HEK293 cell lysates with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66

Three HEK293 cell lysates were prepared using lysis buffers with different protease inhibitor (PI) mix concentrations (10 %, 1 %, no protease inhibitor) and SUMOylated as described in the General Methods using the Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66. In the positive control, RanGAP1 was used as SUMOylation substrate. The reactions were analyzed by SDS PAGE (8–16 % pre-cast gel) and in-gel fluorescence imaging.

In-gel fluorescence image of SDS PAGE of SUMOylation reactions of HEK293 cell lysates (containing different amounts of protease inhibitor mix (PI)) using Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66.

Results: An intense band corresponding to Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 and a multitude of bands corresponding to cellular protein–Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH conjugates were observed, showing that in-gel fluorescence is a suitable method for the detection of 66 and its protein conjugates. Importantly, the HEK293 cell lysate alone showed no detectable bands, demonstrating that there was no endogenous fluorescence background and all observed bands came from Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 or its protein conjugates.

The observation of a band corresponding to the RanGAP1–Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 conjugate showed that 66 is recognized and processed by the SUMOylation machinery.

For all three SUMOylation reactions of HEK293 cell lysate, a multitude of bands with high molecular weight were observed, indicating that 66 was conjugated to multiple cellular proteins. The observed band patterns and intensity distributions were very similar in all three samples showing that the amount of protease inhibitor present in the lysis buffer does not affect the SUMOylation machinery.
These results demonstrate that proteins from crude cell lysates can be SUMOylated and the conjugates readily detected using in-gel fluorescence. Unlike Western blots, this provides a straightforward method to distinguish naturally SUMOylated proteins from proteins labeled with 66.

**Optimization of the SUMOylation reaction: Requirement for E1 and E2 enzyme**

HEK293 cell lysates or RanGAP1 were SUMOylated with 66 as described in the General Methods (positive control) or in a similar manner with the exception that in these samples, either E1 (SAE1/SAE2) or E2 (UBC9) enzymes were not added to the SUMOylation reactions. The reactions were analyzed by SDS PAGE (7.5 % pre-cast gel) and in-gel fluorescence imaging.

Results: The SUMOylation experiment of RanGAP1 showed a band corresponding to the RanGAP1–Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 conjugate only if both E1 and E2 enzyme are added to the reaction (lane 4). If one of the enzymes is omitted, no conjugated product is observed (lanes 5 and 6).

In contrast to this, the SUMOylation experiment of HEK293 cell lysate where no E1 enzyme was added showed almost identical band pattern and intensity compared to the positive control (lane 3 vs. 1). When no E2 enzyme was added, only very weak bands corresponding to conjugates were observed (lane 2). These results indicate that the endogenous E1 enzyme in the cell lysate retains its activity to a high degree, making the addition of exogenous E1 enzyme in *in vitro* SUMOylation assays unnecessary. The addition of E2 enzyme is necessary to achieve a high degree of SUMOylation in the *in vitro* assay, indicating a low activity of the endogenous E2 enzyme.
SUMOylation of HEK293 and HeLa cell lysates with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66: influence of external stimuli

a) Addition of E3 ligases

HEK293 and HeLa cell lysates were prepared as described in the general methods and SUMOylated with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 in presence or absence of SUMO E3 ligases PIAS1 (Enzo Life Sciences BML-UW9960, final concentration ca. 0.1 µM) or RanBP2 (Enzo Life Sciences BML-UW9455, final concentration ca. 0.1 µM). SUMOylation reactions were analyzed by SDS PAGE (8–16 % pre-cast gels) and in-gel fluorescence.

Results: Qualitatively, the addition of PIAS1 had no obvious effect on protein SUMOylation (e.g. lanes 1 vs. 2). The addition of RanBP2 – which itself is being SUMOylated – enhanced the general protein SUMOylation significantly (e.g. lanes 1 vs. 3).

b) Heat shock experiments

Flasks with 90 % confluent cell cultures (HEK293 and HeLa cells) were incubated at 43 °C (heat shock) or 37 °C (control) for 1 h, followed by cell lysis and SUMOylation (+/- E2 enzyme) as described in the General Methods. SUMOylation reactions were analyzed by SDS PAGE (8–16 % pre-cast gels) and in-gel fluorescence.
Results: Qualitatively, cell lysates derived from heat shocked cells showed slightly more intense bands corresponding to SUMOylated proteins (e.g. lane 4 vs. 3) compared to lysates from cells grown under regular conditions.

**SUMOylation of HEK293 and HeLa cell lysates with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 and affinity purification**

HEK293 and HeLa cell lysates were prepared as described in the general methods. For the SUMOylation reaction, the following components were added to a 2.2 ml centrifuge tube (total volume: 750 µl):

<table>
<thead>
<tr>
<th>Volume added [µl]</th>
<th>Final concentration [µM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ-H2O</td>
<td>518.2</td>
</tr>
<tr>
<td>10 X SUMOylation buffer</td>
<td>75.0</td>
</tr>
<tr>
<td>20 X Mg-ATP</td>
<td>37.5</td>
</tr>
<tr>
<td>20 X SUMO E2</td>
<td>6.8</td>
</tr>
<tr>
<td>Cell lysate</td>
<td>75.0</td>
</tr>
<tr>
<td>Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66</td>
<td>37.5</td>
</tr>
</tbody>
</table>

The SUMOylation reactions were incubated at 37 °C, followed by affinity purification as described in the General Methods. The fractions from the washings with regeneration buffer were neutralized (120 µl 1 M Tris-HCl, pH 9.5 per fraction) and pooled with the elution buffer fractions. The combined fractions were concentrated to a volume of ca. 25 µl using a spin filter as described above. Aliquots from the crude SUMOylation reactions, from the first elution and washing buffer fractions and from the concentrated elution fraction were analyzed by SDS PAGE (8–16 % pre-cast gels), Coomassie staining and in-gel fluorescence.

Coomassie stained (left panel) and in-gel fluorescence image (right panel) of SUMOylation reactions using HEK293 or HeLa cell lysates and Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66
Results: Analysis by Coomassie staining and in-gel fluorescence shows qualitatively that the performed affinity purification workflow provides significant enrichment and purification of the SUMOylated cellular proteins (lanes 7 and 8 vs. crude reactions in lanes 1 and 2). The different cell lysates (lanes 7 and 8) show no major difference in band pattern or intensity, suggesting that similar proteins are being SUMOylated.

**SUMOylation of HEK293 cell lysates with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 and proteomic analysis of SUMOylated proteins**

Cell lysates were prepared from three different cultures of HEK293 cells as described in the General Methods. The cell lysates were independently subjected to SUMOylation with Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66; the negative control samples were subjected to the same conditions with the exception that no Biotin-PEG-Dap(Rho)-[SUMO3 (2–92) Gln89Pro]-OH 66 was added. For one set of SUMOylation reaction and negative control was added to a 2.2 ml centrifuge tube (total volume: 1500 µl) and reacted as described above:

<table>
<thead>
<tr>
<th>SUMOylation reaction</th>
<th>Negative control</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ-H2O</td>
<td>-</td>
</tr>
<tr>
<td>10 X SUMOylation buffer</td>
<td>-</td>
</tr>
<tr>
<td>20 X Mg-ATP</td>
<td>75.0</td>
</tr>
<tr>
<td>20 X SUMO E2</td>
<td>13.5</td>
</tr>
<tr>
<td>Cell lysate</td>
<td>150.0</td>
</tr>
<tr>
<td><strong>66</strong></td>
<td>75.0</td>
</tr>
</tbody>
</table>

After the SUMOylation reaction, the SUMOylated proteins were purified by biotin affinity purification as described in the General Methods. The three elution fractions of each sample were pooled, frozen at -20ºC and transferred to the Wollscheid Lab at ETH Zurich for tryptic digest followed by LC-MS/MS analysis. 127 proteins could be identified. A comparison with five publications showed that out of the 127 identified proteins, at least 37 proteins are known to be SUMOylated. For a comprehensive list, see Appendix.


7.6. Experimental for “CHAPTER 5: Chemical synthesis of irisin protein”

Protein sequence of irisin

Irisin is an extracellular part (32–143) of the Fibronectin type III domain-containing protein 5 (FNDC5), from which it is cleaved and released.

**FNDC5 (1–212):**

```
MHPGSPSAWP10  PRARAALRLW20  LGCCVFALVQ30  ADSPSAPNV40  TVRHLKANSA50 
VVSWDVLDEE60  VVIGFAISQQ70  KKDVRMLRFI80  QEVNTTTRSC90  ALWDLEEDTE100 
YIVHVQAIS110  QGQSPASEPV120  LFKTPREAEK130  MASKNKDEV140  TMKEMGRNQQL150 
RTGEVLIIVV160  VLFMWAGVIA170  LFCRQYDIIK180  DNEPNNNEK190  TKSASETSTP200 
EHQGGGLRS210  KI212
```

From Uniprot (Q8NAU1-1, FNDC5_HUMAN)

**Irisin (1–112) (=FNDC5 (32–143), this work):**

```
DSPSAPVNVT10  VRHLKANSA20  VVVSWDVLDEE30  VVIGFAISQQ40  KKDVRMLRFI50 
EVNTTTRSCA60  LWLDEEDTE70  IVHVQAISI80  QGQSPASEPV90  FKTTPREAEKM100 
ASKNKDEVTM110  KE112
```

---


Synthetic strategy

Irisin (1–112)

DSPSAPVNT₁₀ VRHLKANSAV₂₀ VSWDVL – T₆DEV₃₀ VIGFAISQK₄₀ KDVRMLRFIQ₅₀
EVNTTTRSCA₆₀ LWDL – T₆EDTEY₇₀ IVHVQAISIQ₈₀ GQSPASEPVL₉₀ FKTPREAEMK₁₀₀
ASKNDEVTM₁₁₀ KE₁₁₂

7.6.1. Synthesis of segments

Synthesis of H-[Irisin (1–26)]-α-ketoacid 82

H-[Irisin (1–26)]-α-ketoacid 82 was prepared on protected leucine α-ketoacid resin 41 at a 0.20 mmol scale (0.59 g resin with a substitution capacity of 0.34 mmol/g). The segment was elongated up to Asp1 with automated synthesis as described in the general methods using double couplings at every step. After the final Fmoc-deprotection, ca. a quarter of the resin (0.63 g) was cleaved according to Method B and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C, 30 to 60 % CH₃CN with 0.1 % TFA in 30 min, flow rate 40 ml/min). Product containing fractions were pooled and lyophilized to give pure H-[Irisin (1–26)]-α-ketoacid 82 (19.5 mg, 7.00 µmol, ca. 14 % yield for SPPS and purification). Analytical HPLC and MALDI FTICR MS confirmed the
purity and identity of 82. m/z calculated for \( \text{C}_{123}\text{H}_{197}\text{N}_{35}\text{Na}_{4}\text{O}_{39}^{+} \): 2811.4400; measured 2811.4379.

![Analytical HPLC trace of purified H-[Irisin (1–26)]-\( \alpha \)-ketoacid 82. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH\(_3\)CN in 20 min.](image)

MALDI FTICR MS: Measured (top) and calculated isotopic pattern (bottom) of H-[Irisin (1–26)]-\( \alpha \)-ketoacid 82.

**Synthesis of labeled H-Dap(Rho)-[Irisin (1–26)]-\( \alpha \)-ketoacid 83**

Protected leucine \( \alpha \)-ketoacid resin (0.30 g, 24 \( \mu \)mol) used in the synthesis of H-[Irisin (1–26)]-\( \alpha \)-ketoacid 82 (elongated up to Asp1) was Fmoc deprotected. Rhodamine labeled amino acid 89 (0.102 mg, 0.118 mmol, 4.92 equiv) and COMU (49.3 mg, 0.115 mmol, 4.80 equiv) were dissolved in NMP (0.75 ml, 0.16 M) and NMM (24.0 mg, 26.0 \( \mu \)mol, 0.234 mmol, 9.80) was added. The mixture was incubated for 1 min and added to the resin. The resin was agitated for 2.5 h, washed with DMF (5 x 5 ml) and CH\(_2\)Cl\(_2\) (5 x 5 ml) and the N-terminal Fmoc deprotected. The dried resin was cleaved following Method B and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C, 30 to 60 % CH\(_3\)CN with 0.1 % TFA in 30 min, flow rate 40 ml/min). Product
containing fractions were pooled and lyophilized to give pure H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83 (10.1 mg, 2.96 µmol, ca. 12 % yield for SPPS and purification). Analytical HPLC and ESI QTOF MS confirmed the purity and identity of H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83. m/z calculated for C_{153}H_{231}N_{39}O_{46}S_{2} [M+H]^+: 3416.64; measured 3416.66.

Analytical HPLC trace of purified H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH₃CN in 20 min.

ESI QTOF MS: Measured (top) and deconvoluted (bottom) MS of H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83.

**Synthesis of 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84**

Segment 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 was prepared on protected leucine α-ketoacid resin 41 at a 0.5 mmol scale (1.56 g resin with a substitution capacity of 0.34 mmol/g). The segment was elongated up to Asp28 with automated synthesis as described in the general methods using double couplings at every step. To 1.6 g of the resin (0.10 mmol at a loading of 0.063 mmol/g) was coupled 1-NPEOC-Opr-OH using HATU as described in the general methods. The dried resin was cleaved according to Method B and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C, 30 to 60 % CH₃CN with 0.1 % TFA in 30 min, flow rate 40 ml/min and Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 30 to 60 % CH₃CN with 0.1 % TFA in 30 min, flow rate 10 ml/min). The fractions containing the
product were pooled and lyophilized to give pure 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 (32.6 mg, 7.04 μmol, ca. 7% for SPPS and purification). Analytical HPLC and ESI QTOF MS confirmed the purity and identity of 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84. m/z calculated for C_{204}H_{321}N_{55}O_{64}S_{2} [M+H]^+: 4629.30; measured 4629.29.

Analytical HPLC trace of purified 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 40 to 60 % CH_{3}CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (middle) MS and calculated isotopic pattern (bottom) of 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84.

**Synthesis of H-Opr-[Irisin (66–112)]-OH 85**

H-Opr-[Irisin (66–112)]-OH 85 was synthesized on 2-chlorotrityl polystyrene resin at a 0.40 mmol scale (1.14 g resin with a substitution capacity of 0.35 mmol/g). The segment was elongated up to Glu64 with automated synthesis using double couplings at every step and Boc-Opr-OH was coupled with COMU as described in the general methods. The resin was cleaved according to Method A and the crude peptide purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C, 30 to 60 % CH_{3}CN with 0.1 % TFA in 30 min, flow rate 40 ml/min and Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 20 to 50 % CH_{3}CN with 0.1 % TFA in 30 min, flow rate 8 ml/min). The fractions containing the product were pooled and lyophilized to give pure H-Opr-[Irisin (66–112)]-OH 85 (136.0 mg, 25.2 μmol, ca. 6% for SPPS and purification). Analytical HPLC and ESI QTOF MS confirmed the purity and identity of H-Opr-[Irisin (66–
CHAPTER 7: EXPERIMENTAL

112)-OH 85. m/z calculated for C_{233}H_{376}N_{62}O_{80}S_{2} [M+H]^+: 5386.670; measured 5389.700 [A+3].

Analytical HPLC trace of purified H-Opr-[Irisin (66–112)]-OH 85. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 40 % CH_{3}CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of H-Opr-[Irisin (66–112)]-OH 85.

7.6.2. Sequential KAHA ligations for the synthesis of irisin proteins 87 and 88

KAHA ligation for the preparation of H-Opr-[Irisin (28-112)]-OH deps-86

a) Ligation: 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 (30.0 mg, 6.48 μmol, 1.00 equiv) and H-Opr-[Irisin (66–112)]-OH 85 (41.9 mg, 7.77 μmol, 1.20 equiv) were weighed into a glass vial and dissolved in a mixture of 9:1 NMP:H_{2}O (0.40 ml, 16 mM) with 0.1 M oxalic acid, forming instantly a gel. The mixture was heated to 60 °C and reaction progress was
monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column 4.6 x 250 mm, heated to 60 °C, 20 to 80 % CH₃CN with 0.1 % TFA in 20 min, flow rate 1 ml/min).

**b) One-pot photodeprotection of oxaproline**

The crude ligation mixture was first diluted with 9:1 NMP:H₂O (1.0 ml) with 0.1 M oxalic acid to solubilize the protein mixture and further diluted to 8 ml with 1:1 H₂O: CH₃CN + 0.1 % TFA. The clear solution was irradiated with a regular laboratory bench top UV lamp at a wavelength of 365 nm for 30 min. Longer irradiation times were found not to improve the conversion. The crude mixture was purified by preparative HPLC (Shiseido Capcell Pak UG80 C18 column (50 x 250 mm), heated to 65 °C; 30 to 70 % CH₃CN with 0.1 % TFA in 40 min, flow rate 40 ml/min) and the product containing fractions pooled and lyophilized to give pure H-Opr-[Irisin (28-112)]-OH depsi-86 (11.5 mg, 1.18 µmol, 18 % yield for ligation, deprotection and purification). Analytical HPLC and ESI FTICR MS confirmed the purity and identity of H-Opr-[Irisin (28-112)]-OH depsi-86. m/z calculated for C₄₂₇H₆₉₀N₁₁₆O₁₃₈S₄ [M+H]⁺: 9779.95; measured 9785.00 [A+6].

![HPLC monitoring of the KAHA ligation (bottom and middle trace) of 1-NPEOC-Opr-[Irisin (28–64)]-α-ketoacid 84 and H-Opr-[Irisin (66–112)]-OH 85 and the one-pot fotodeprotection of the oxaproline (upper trace). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 80 % CH₃CN in 20 min.](image)

![Analytical HPLC trace of purified H-Opr-[Irisin (28-112)]-OH depsi-86. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 20 to 80 % CH₃CN in 20 min.](image)
KAHA ligation for the preparation of H-[Irisin (1–112)]-OH 87

a) Ligation: H-[Irisin (1–26)]-α-ketoacid 82 (2.2 mg, 0.77 µmol, 1.2 equiv) and H-Opr-[Irisin (28-112)]-OH depsi-86 (6.3 mg, 6.4 µmol, 1.0 equiv) were weighed into a glass vial insert and a mixture of 9:1 NMP:H₂O (60 µl, 11 mM) with 0.1 M oxalic acid was added. The mixture forms instantly a gel with some insoluble material left. The mixture was heated to 60 °C for 18 h and reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column 4.6 x 250 mm, heated to 60 °C, 30 to 50 % CH₃CN with 0.1 % TFA in 20 min, flow rate 1 ml/min).

b) Global O to N acyl shift: The crude ligation mixture was first diluted with 9:1 NMP:H₂O (0.12 ml) with 0.1 M oxalic acid to solubilize the protein mixture and further diluted to 1.0 ml with rearrangement buffer (pH 9.5, 0.2 M NaHCO₃/Na₂CO₃ buffer, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 3 h. The mixture was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C, 32.5 to 47.5 % CH₃CN with 0.1 % TFA in 20 min, flow rate 10 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure H-[Irisin (1–112)]-OH 87 (1.4 mg, 0.11 µmol, 17 %). Analytical HPLC and ESI FTICR MS confirmed the purity and
identity of 87. m/z calculated for C_{546}H_{887}N_{151}O_{175}S_{4} [M+H]^+: 12523.40; measured 12530.46 [A+7].

HPLC monitoring of the KAHA ligation (bottom and middle trace) of H-[Irisin (1–26)]-α-ketoacid 82 and H-Opr-[Irisin (28–112)]-OH depsii-86 and the one-pot global O to N acyl shift (upper trace, 4 x exaggerated to compensate for high dilution). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH_{3}CN in 20 min.

Analytical HPLC trace of purified H-[Irisin (1–112)]-OH 87. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH_{3}CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of H-[Irisin (1–112)]-OH 87.

Circular dichroism spectra of H-[Irisin (1–112)]-OH 87.
A CD spectrum of H-[Irisin (1–112)]-OH 87 (10 μM protein concentration, 25 °C) was recorded in 20 % TFE/10 mM phosphate buffer pH 6.0.
CD spectrum of H-[Irisin (1–112)]-OH 87, 10 µM protein concentration in 20 % TFE/10 mM phosphate buffer pH 6.0 at 25 ºC.

KAHA ligation for the preparation of H-Dap(Rho)-[Irisin (1–112)]-OH 88

**a) Ligation:** H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83 (2.8 mg, 0.81 µmol, 1.2 equiv) and H-Opr-[Irisin (28-112)]-OH depsi-86 (6.6 mg, 6.8 µmol, 1.0 equiv) were weighed into a glass vial insert and a mixture of 9:1 NMP:H2O (60 µl, 11 mM) with 0.1 M oxalic acid was added. The mixture forms instantly a gel with some insoluble material left. The mixture was heated to 60 ºC for 18 h and reaction progress was monitored by analytical HPLC (Shiseido Capcell Pak UG80 C18 column 4.6 x 250 mm, heated to 60 ºC, 30 to 50 % CH3CN with 0.1 % TFA in 20 min, flow rate 1 ml/min). It was found that both starting materials (83 and 86) and H-Dap(Rho)-[Irisin (1–112)]-OH depsi-88 elute very closely from the HPLC, making assignment of the peaks impossible.
b) Global O to N acyl shift: The crude ligation mixture was first diluted with 9:1 NMP:H$_2$O (0.12 ml) with 0.1 M oxalic acid to solubilize the protein mixture and further diluted to 1.0 ml with rearrangement buffer (pH 9.5, 0.2 M NaHCO$_3$/Na$_2$CO$_3$ buffer, 6 M GdmCl, 10 mM TCEP) and incubated at room temperature for 3 h. The mixture was directly purified by preparative HPLC (Shiseido Capcell Pak MGII C18 column (20 x 250 mm), heated to 60 °C; 32.5 to 47.5 % CH$_3$CN with 0.1 % TFA in 20 min, flow rate 10 ml/min). The fractions containing the ligated product were pooled and lyophilized to give pure H-Dap(Rho)-[Irisin (1–112)]-OH 88 (1.3 mg, 0.10 µmol, 15 %). Analytical HPLC and ESI FTICR MS confirmed the purity and identity of 88. m/z calculated for C$_{579}$H$_{921}$N$_{155}$O$_{182}$S$_6$ [M+H]$^+$: 13149.59; measured 13157.70 [A+8].

HPLC monitoring of the KAHA ligation (bottom and middle trace) of H-Dap(Rho)-[Irisin (1–26)]-α-ketoacid 83 and H-Opr-[Irisin (28-112)]-OH depsi-86 and the one-pot global O to N acyl shift (upper trace, 2 x exaggerated to compensate for high dilution). Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH$_3$CN in 20 min.

Analytical HPLC trace of purified H-Dap(Rho)-[Irisin (1–112)]-OH 88. Shiseido Capcell Pak UG80 C18 column (4.6 x 250 mm) heated to 60 °C, 30 to 50 % CH$_3$CN in 20 min.

ESI FTICR MS: Measured (top) and deconvoluted (bottom) MS of labeled H-Dap(Rho)-[Irisin (1–112)]-OH 88.
Circular dichroism spectra of H-Dap(Rho)-[Irisin (1–112)]-OH 88

A CD spectrum of H-Dap(Rho)-[Irisin (1–112)]-OH 88 (10 μM protein concentration, 25 °C) was recorded in 20 % TFE/10 mM phosphate buffer pH 6.0. The shape of the spectrum is very similar to the one of H-[Irisin (1–112)]-OH 87, suggesting that the labeling does not alter the protein folding.

CD spectrum of H-Dap(Rho)-[Irisin (1–112)]-OH 88, 10 μM protein concentration in 20 % TFE/10 mM phosphate buffer pH 6.0 at 25 °C.
CHAPTER 8: APPENDIX

NMR SPECTRA

PROTEOMICS RESULTS
CHAPTER 8: Appendix

8.1. NMR spectra

Fmoc-Ala-Leu-α-ketoacid 4
H-Opr-(4F)Phe-OH 5
Fmoc-(S)-Leu cyanosulfurylide S1
Fmoc-(S)-Leu-α-ketoacid 7
H-Trp-Thr-Leu-α-ketoacid 20
di-tert-butyl isoxazolidine-2,5-dicarboxylate S2
tert-butyl (S)-3-(hydroxymethyl)isoxazolidine-2-carboxylate S3
tert-butyl isoxazolidine-2-carboxylate S4
tert-butyl 1,2-oxazinane-2-carboxylate S5
Fmoc-Ala-Leu-(depsi)-Hse-(4-F)Phe-OH  

**depsi-6:**

![Chemical structure](image)
Fmoc-Ala-Leu-Hse-(4-F)Phe-OH amide-6:

[Chemical structure images]
Fmoc-Ala-(S)-Leu-(S)-Hse-(4-F)Phe-OH 6-A
Fmoc-Ala-(R)-Leu-(S)-Hse-(4-F)Phe-OH 6-B
Fmoc-Ala-(S)-Leu-(R)-Hse-(4-F)Phe-OH 6-C
Fmoc-Leu-(depsi)-Hse-(4F)-Phe-OH 8
Succinyl-(depsi)-Hse-(4F)-Phe-OH 9
Propionyl-(depsi)-Hse-(4F)-Phe-OH 10
Benzoyl-(*depsi*)-Hse-(4F)-Phe-OH 11
4-amino-2-(2-(4-hydroxyphenyl)acetoxy)butanoic acid *depsi*-13
(S)-3-amino-4-hydroxybutyl (((9H-fluoren-9-yl)methoxy)carbonyl)-L-leucine *depsi*-15
(9H-fluoren-9-yl)methyl ((S)-1-(((S)-1,4-dihydroxybutan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate amide-15
3-aminopropyl 2-phenylacetate \textit{depsi-17}
N-(3-hydroxypropyl)-2-phenylacetamide \textit{amide-17}
4-aminobutyl 2-phenylacetate *depsi*-19
N-(4-hydroxybutyl)-2-phenylacetamide amide-19
1-(4-(allyloxy)phenyl)-2,2-dimethylpropane-1,3-diol S7
4-(4-methoxyphenyl)-2,2,5,5,8,8-hexamethyl-3,7-dioxa-2,8-disilanonane 29
4-(4-(allyloxy)phenyl)-2,2,5,5,8,8-hexamethyl-3,7-dioxo-2,8-disilanonane 43
allyl 5-(4-(hydroxymethyl)phenoxy)pentanoate S8
allyl 5-(4-(chloromethyl)phenoxy)pentanoate 38
4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethylbutanoic acid 47
(9H-fluoren-9-yl)methyl (S)-(1-((4-fluorophenethyl)amino)-4-methyl-1-oxopentan-2-yl) carbamate 37
2-((S)-1-(((9H-fluoren-9-yl)methoxy) carbonyl) amino)-3-methylbutyl)-4-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylic acid 35
4-((5-(allyloxy)-5-oxopentyl)oxy)benzyl 2-(((S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methyl butyl)-4-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylate 39
5-(4-(((2-(S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carbonyl)oxy)methyl)phenoxy)pentanoic acid 40
2-((S)-1-(((9H-fluoren-9-yl)methoxy) carbonyl)amino)-3-methylbutyl)-4-(4-(allyloxy)phenyl)-
5,5-dimethyl-1,3-dioxane-2-carboxylic acid 44
2-((S)-1-(((9H-fluoren-9-yl)methoxy) carbonyl)amino)-3-methylbutyl)-4-(4-hydroxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylic acid 45
4-((5-(allyloxy)-5-oxopentyl)oxy)benzyl 2-((S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-hydroxyphenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylate 46
4-((5-(allyloxy)-5-oxopentyl)oxy)benzyl 2-(((S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methyl butyl)-4-(4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethylbutanoyl)oxy)phenyl)-5,5-dimethyl-1,3-dioxane-2-carboxylate 48
5-(4-(((2-(S)-1-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-methylbutyl)-4-(4-((tert-butoxycarbonyl)(methyl)amino)-2,2-dimethylbutanoyl)oxy)phenyl)-5,5-dimethyl-1,3-dioxane-2-carbonyl)oxy)methyl)phenoxy)pentanoic acid 49
Fmoc-Gln-Leu-(phenolester-protected) \( \alpha \)-ketoacid 51-A
Fmoc-Gln-Leu-(phenolester-protected) α-ketoacid 51-B
$N^2$-methyl-$N^2$-(2-nitro-4-sulfophenyl)-$N^5$-trityl-$L$-glutamine S11
$N^2-(N^2-(((9H\text{-fluoren-9-yl})\text{methoxy})\text{carbonyl})-N^2\text{-trityl-}L\text{-glutaminyl})-N^2\text{-methyl-N^2}\text{-trityl-L-glutamine} \ 81
(9H-fluoren-9-yl)methyl (2-(2-(2-hydroxyethoxy)ethoxy)ethyl)carbamate 79

2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetic acid S13
S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-(2-(7-(diethylamino)-4-methyl-2-oxo-2H-chromen-3-yl)acetamido)propanoic acid 80
(S)-5-((N-(2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-2-carboxyethyl)sulfamoyl)-2-(6-(diethylamino)-3-(diethyliminio)-3H-xanthen-9-yl)benzenesulfonate 89
### CHAPTER 8: APPENDIX

#### 8.2. Proteomics results

Proteins enriched in the SUMOylation reaction compared to the negative control

<table>
<thead>
<tr>
<th>Entry</th>
<th>Entry name</th>
<th>Known to be SUMOylated</th>
<th>Protein names</th>
</tr>
</thead>
<tbody>
<tr>
<td>E9PAV3</td>
<td>NACAM_HUMAN</td>
<td>no</td>
<td>Nascent polypeptide-associated complex subunit alpha, muscle-specific form (Alpha-NAC, muscle-specific form) (skNAC)</td>
</tr>
<tr>
<td>O00299</td>
<td>CLIC1_HUMAN</td>
<td>no</td>
<td>Chloride intracellular channel protein 1 (Chloride channel ABP) (Nuclear chloride ion channel 27) (NCC27) (Regulatory nuclear chloride ion channel protein) (hRNCC)</td>
</tr>
<tr>
<td>O14980</td>
<td>XPO1_HUMAN</td>
<td>yes</td>
<td>Exportin-1 (Exp1) (Chromosome region maintenance 1 protein homolog)</td>
</tr>
<tr>
<td>O60361</td>
<td>NDK8_HUMAN</td>
<td>yes</td>
<td>Putative nucleoside diphosphate kinase (NDK) (NDP kinase) (EC 2.7.4.6)</td>
</tr>
<tr>
<td>O75083</td>
<td>WDR1_HUMAN</td>
<td>no</td>
<td>WD repeat-containing protein 1 (Actin-interacting protein 1) (AIP1) (NORI-1)</td>
</tr>
<tr>
<td>O75347</td>
<td>TBCA_HUMAN</td>
<td>no</td>
<td>Tubulin-specific chaperone A (TCP1-chaperonin cofactor A) (Tubulin-folding cofactor A) (CFA)</td>
</tr>
<tr>
<td>O75369</td>
<td>FLNB_HUMAN</td>
<td>no</td>
<td>Filamin-B (FLN-B) (ABP-278) (ABP-280 homolog) (Actin-binding-like protein) (Beta-filamin) (Filamin homolog 1) (Fh1) (Filamin-3) (Thyroid autoantigen) (Truncated actin-binding protein) (Truncated ABP)</td>
</tr>
<tr>
<td>O75937</td>
<td>DNJC8_HUMAN</td>
<td>yes</td>
<td>DnaJ homolog subfamily C member 8 (Splicing protein sp3f1)</td>
</tr>
<tr>
<td>O76003</td>
<td>GLRX3_HUMAN</td>
<td>yes</td>
<td>Glutaredoxin-3 (PKC-interacting cousin of thioredoxin) (PICOT) (PKC-theta-interacting protein) (PKCq-interacting protein) (Thioredoxin-like protein 2)</td>
</tr>
<tr>
<td>O95817</td>
<td>BAG3_HUMAN</td>
<td>yes</td>
<td>BAG family molecular chaperone regulator 3 (BAG-3) (Bcl-2-associated athanogene 3) (Bcl-2-binding protein Bis) (Docking protein CAIR-1)</td>
</tr>
<tr>
<td>O95861</td>
<td>BPNT1_HUMAN</td>
<td>no</td>
<td>3‘(2‘),5‘-bisphosphate nucleotidase 1 (EC 3.1.3.7) (Bisphosphate 3’-nucleotidase 1) (PAP-inositol 1,4-phosphatase) (PPIP)</td>
</tr>
<tr>
<td>P00338</td>
<td>LDHA_HUMAN</td>
<td>no</td>
<td>L-lactate dehydrogenase A chain (LDH-A) (EC 1.1.1.27) (Cell proliferation-inducing gene 19 protein) (LDH muscle subunit) (LDH-M) (Renal carcinoma antigen NY-REN-59)</td>
</tr>
<tr>
<td>P00441</td>
<td>SODC_HUMAN</td>
<td>no</td>
<td>Superoxide dismutase [Cu-Zn] (EC 1.15.1.1) (Superoxide dismutase 1) (hSod1)</td>
</tr>
<tr>
<td>P00492</td>
<td>HPRT_HUMAN</td>
<td>yes</td>
<td>Hypoxanthine-guanine phosphoribosyltransferase (HGPRT) (HGPRTase) (EC 2.4.2.8)</td>
</tr>
<tr>
<td>P00558</td>
<td>PGK1_HUMAN</td>
<td>no</td>
<td>Phosphoglycerate kinase 1 (EC 2.7.2.3) (Cell migration-inducing gene 10 protein) (Primer recognition protein 2) (PRP 2)</td>
</tr>
<tr>
<td>P00918</td>
<td>CAH2_HUMAN</td>
<td>no</td>
<td>Carbonic anhydrase 2 (EC 4.2.1.1) (Carbonate dehydratase II) (Carbonic anhydrase C) (CAC) (Carbonic anhydrase II) (CA-II)</td>
</tr>
<tr>
<td>P04075</td>
<td>ALDOA_HUMAN</td>
<td>yes</td>
<td>Fructose-bisphosphate aldolase A (EC 4.1.2.13) (Lung cancer antigen NY-LU-1) (Muscle-type aldolase)</td>
</tr>
<tr>
<td>P04406</td>
<td>G3P_HUMAN</td>
<td>no</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (EC 1.2.1.12) (Peptidyl-cysteine S-nitrosylase GAPDH) (EC 2.6.99.-)</td>
</tr>
<tr>
<td>P05455</td>
<td>LA_HUMAN</td>
<td>yes</td>
<td>Lupus La protein (La autoantigen) (La ribonucleoprotein) (La protein) (Sjogren syndrome type B antigen) (SS-B)</td>
</tr>
<tr>
<td>P06733</td>
<td>ENOA_HUMAN</td>
<td>yes</td>
<td>Alpha-enolase (EC 4.2.1.1) (2-phospho-D-glycerate hydratase) (C-myc promoter-binding protein) (Enolase 1) (MBP-1) (MPB-1) (Non-neural enolase) (NNE) (Phosphopyruvate hydratase) (Plasminogen-binding protein)</td>
</tr>
<tr>
<td>P06744</td>
<td>G6PI_HUMAN</td>
<td>no</td>
<td>Glucose-6-phosphate isomerase (GPI) (EC 5.3.1.9) (Autocrine motility factor) (AMF) (Neuroleukin) (NLK) (Phosphoglucone isomerase) (PGI) (Phosphohexose isomerase) (PHI) (Sperm antigen 36) (SA-36)</td>
</tr>
<tr>
<td>P06753</td>
<td>TPM3_HUMAN</td>
<td>no</td>
<td>Tropomyosin alpha-3 chain (Gamma-tropomyosin) (Tropomyosin-3) (Tropomyosin-5) (hTM5)</td>
</tr>
<tr>
<td>P07195</td>
<td>LDHB_HUMAN</td>
<td>yes</td>
<td>L-lactate dehydrogenase B chain (LDH-B) (EC 1.1.1.27) (LDH heart subunit) (LDH-H) (Renal carcinoma antigen NY-REN-46)</td>
</tr>
<tr>
<td>P07237</td>
<td>PDIA1_HUMAN</td>
<td>yes</td>
<td>Protein disulfide-isomerase (PDIs) (EC 5.3.4.1) (Cellular thyroid hormone-binding protein) (Prolyl 4-hydroxylase subunit beta) (p55)</td>
</tr>
</tbody>
</table>
Annexin A2 (Annexin II) (Annexin-2) (Calpain I heavy chain) (Calpain-1 heavy chain) (Chromobindin-8) (Lipocortin II) (Placental anticoagulant protein IV) (PAP-IV) (Protein I) (p36) (Profilin-1 (Epididymis tissue protein Li 184a) (Profilin I) (p38) (Annexin A6 (67 kDa calelectrin) (Annexin VI) (Annexin-6) (Calphobindin-II) (CPB-II) (Chromobindin-20) (Lipocortin VI) (Protein III) (p68) (p70) (Annexin A5 (Anchorin CII) (Annexin V) (Annexin-5) (Calphobindin I) (CBP-I) (Endonexin II) (Lipocortin V) (Placental anticoagulant protein 4) (PP4) (Placental anticoagulant-alpha) (Vascular anticoagulant) (VAC-alpha) (Glutathione S-transferase P (EC 2.5.1.18) (GST class-pi) (GSTP1-1) (High mobility group protein B1 (High mobility group protein 1) (HMG-1) (Ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCH-L1) (EC 3.4.19.12) (EC 3.4.19.12) (Neuron cytoplasmic protein 9.5) (PGP 9.5) (PGP9.5) (ubiquitin thioesterase L1) (Leukotriene A-4 hydrolase (LTA-4 hydrolase) (LTA-4 hydrolase) (Leukotriene A(4) hydrolase) (Inosine-5'-monophosphate dehydrogenase 2 (IMP dehydrogenase 2) (IMPD 2) (IMPDH 2) (IMPDH-2) (Creatine kinase B-type (EC 2.7.3.2) (B-CK) (Creatine kinase B chain) (Elongation factor 2 (EF-2) (Protein disulfide-isomerase A4 (EC 5.3.4.1) (Endoplasmic reticulum resident protein 70) (ER protein 70) (ERp70) (Endoplasmic reticulum resident protein 72) (ER protein 72) (ERp72) (Translational-control tumor protein (TCTP) (Fortilin) (Histamine-releasing factor) (HRF) (p23) (Plastin-3 (T-plastin) (Aldose reductase (AR) (EC 1.1.1.21) (Aldo-keto reductase family 1 member B1) (Eukaryotic peptide chain release factor GTP-binding subunit (ERF3A (Eukaryotic peptide chain release factor subunit 3a) (ERF3a) (G1 to S phase transition protein 1 homolog) (Nucleoside diphosphate kinase A (NDK A) (NDP kinase A) (NDP kinase II) (EC 2.7.4.6) (Granzyme A-activated DNase) (GAAD) (Metastasis inhibition factor nm23) (NM23-H1) (Tumor metastatic process-associated protein) (Carbonyl reductase [NADPH] 1 (EC 1.1.1.184) (15-hydroxyprostaglandin dehydrogenase [NADP(+)] (EC 1.1.1.197) (NADPH-dependent carbonyl reductase 1) (Prostaglandin 9-ketoreductase) (Prostaglandin-E(2) 9-reductase) (EC 1.1.1.189) (Short chain dehydrogenase/reductase family 21C member 1) (ATP-dependent 6-phosphofructokinase, liver type (ATP-PFK) (PFK-L) (EC 2.7.1.11) (6-phosphofructokinase type B) (Phosphofructo-1-kinase isozyme B) (PFK-B) (Phosphohexokinase) (Phosphoglycerate mutase 1 (EC 3.1.3.13) (EC 1.5.2.11) (EC 5.4.2.4) (BPG-dependent PGAM 1) (Phosphoglycerate mutase isozyme B) (PGAM-B) (Transcription factor BTF3 (Nascent polypeptide-associated complex subunit beta) (NAC-beta) (RNA polymerase B transcription factor 3) (Inosine-5'-monophosphate dehydrogenase 1 (IMP dehydrogenase 2) (IMPD 1) (IMPDH 1) (IMPDH-I)) (Trifunctional purine biosynthetic protein adenosine-3 [Includes: Phosphoribosylamine--glycine ligase (EC 6.3.4.13) (Glycinamide
<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Gene Symbol</th>
<th>Expression Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P22234</td>
<td>PUR6_HUMAN</td>
<td>no</td>
<td>Multifunctional protein ADE2 [Includes: Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6) (SAICAR synthetase); Phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) (AIR carboxylase) (AIRC)]</td>
</tr>
<tr>
<td>P22314</td>
<td>UBA1_HUMAN</td>
<td>yes</td>
<td>Ubiquitin-like modifier-activating enzyme 1 (Protein A1S9) (ubiquitin-activating enzyme E1)</td>
</tr>
<tr>
<td>P23264</td>
<td>PP1B_HUMAN</td>
<td>no</td>
<td>Peptidyl-prolyl cis-trans isomerase B (PPlase B) (EC 5.2.1.8) (CYP-S1) (Cyclophilin B) (Rotamase B) (S-cyclophilin) (SCYLP)</td>
</tr>
<tr>
<td>P23526</td>
<td>SAHH_HUMAN</td>
<td>no</td>
<td>Threonine-tRNA ligase, cytoplasmic (EC 6.1.1.3) (Threonyl-tRNA synthetase) (ThrRS)</td>
</tr>
<tr>
<td>P23528</td>
<td>COF1_HUMAN</td>
<td>yes</td>
<td>Cofilin-1 (18 kDa phosphoprotein) (p18) (Cofilin, non-muscle isoform)</td>
</tr>
<tr>
<td>P25205</td>
<td>MCM3_HUMAN</td>
<td>yes</td>
<td>DNA replication licensing factor MCM3 (EC 3.6.4.12) (DNA polymerase alpha holoenzyme-associated protein P1) (P1-MCM3) (RLF subunit beta) (p102)</td>
</tr>
<tr>
<td>P26583</td>
<td>HMGB2_HUMAN</td>
<td>yes</td>
<td>High mobility group protein B2 (High mobility group protein 2) (HMG-2)</td>
</tr>
<tr>
<td>P26639</td>
<td>SYTC_HUMAN</td>
<td>no</td>
<td>14-3-3 protein theta (14-3-3 protein T-cell) (14-3-3 protein tau) (Protein HS1)</td>
</tr>
<tr>
<td>P27348</td>
<td>1433T_HUMAN</td>
<td>no</td>
<td>Threonine-tRNA ligase, cytoplasmic (EC 6.1.1.3) (Threonyl-tRNA synthetase) (ThrRS)</td>
</tr>
<tr>
<td>P29401</td>
<td>TKT_HUMAN</td>
<td>no</td>
<td>Transketolase (TK) (EC 2.2.1.1)</td>
</tr>
<tr>
<td>P29692</td>
<td>EF1D_HUMAN</td>
<td>no</td>
<td>Elongation factor 1-delta (EF-1-delta) (Antigen NY-CO-4)</td>
</tr>
<tr>
<td>P30041</td>
<td>PRDX6_HUMAN</td>
<td>yes</td>
<td>Peroxiredoxin-6 (EC 1.11.1.15) (1-Cys peroxiredoxin) (1-Cys PRX) (24 kDa protein) (Acidic calcium-independent phospholipase A2) (aiPLA2) (EC 3.1.1.1) (Antioxidant protein 2) (Liver 2D page spot 40) (Non-selenium glutathione peroxidase) (NSGPx) (EC 1.11.1.9) (Red blood cells page spot 12)</td>
</tr>
<tr>
<td>P30048</td>
<td>PRDX3_HUMAN</td>
<td>no</td>
<td>Thioredoxin-dependent peroxide reductase, mitochondrial (EC 1.11.1.15) (Antioxidant protein 1) (AOP-1) (HBC189) (Peroxiredoxin III) (Prx-III) (Peroxiredoxin-3) (Protein MER5 homolog)</td>
</tr>
<tr>
<td>P30085</td>
<td>KCY_HUMAN</td>
<td>no</td>
<td>UMP-CMP kinase (EC 2.7.4.14) (Deoxyctydylate kinase) (CK) (dCMP kinase) (Nucleoside-diphosphate kinase) (EC 2.7.4.6) (Uridine monophosphate/cytidine monophosphate kinase) (UMP/CMP kinase) (UMP/CMPK)</td>
</tr>
<tr>
<td>P30086</td>
<td>PEBP1_HUMAN</td>
<td>no</td>
<td>Phosphatidylethanolamine-binding protein 1 (PEBP-1) (HCNPPp) (Neuropolypeptide h3) (Prostatic-binding protein) (Raf kinase inhibitor protein) (RKKIP) (Cleaved into: Hippocampal cholergic neurostimulating peptide (HCNP))</td>
</tr>
<tr>
<td>P30101</td>
<td>PDIA3_HUMAN</td>
<td>no</td>
<td>Protein disulfide-isomerase A3 (EC 5.3.4.1) (58 kDa glucose-regulated protein) (58 kDa microsomal protein) (p58) (Disulfide isomerase ER-60) (Endoplasmic reticulum resident protein 57) (ER protein 57) (ERp57) (Endoplasmic reticulum resident protein 60) (ER protein 60) (ERp60)</td>
</tr>
<tr>
<td>P31939</td>
<td>PUR9_HUMAN</td>
<td>no</td>
<td>Bifunctional purine biosynthesis protein PURH (Cleaved into: Bifunctional purine biosynthesis protein PURH, N-terminally processed) [Includes: Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3) (5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase) (AICAR transformylase); IMP cyclohydrase (EC 3.5.4.10) (ATIC) (IMP synthase) (Inosinicase)]</td>
</tr>
<tr>
<td>P33316</td>
<td>DUT_HUMAN</td>
<td>no</td>
<td>Deoxyuridine 5-triphosphate nucleotidohydrolase, mitochondrial (dUTPase) (EC 3.6.1.23) (dUTP pyrophosphatase)</td>
</tr>
<tr>
<td>P34932</td>
<td>HSP74_HUMAN</td>
<td>no</td>
<td>Heat shock 70 kDa protein 4 (HSP70RY) (Heat shock 70-related protein APG-2)</td>
</tr>
<tr>
<td>P37802</td>
<td>TAGL2_HUMAN</td>
<td>yes</td>
<td>Transgelin-2 (Epididymis tissue protein L1 7e) (SM22-alpha homolog)</td>
</tr>
<tr>
<td>P38646</td>
<td>GRP75_HUMAN</td>
<td>no</td>
<td>Stress-70 protein, mitochondrial (75 kDa glucose-regulated protein)</td>
</tr>
<tr>
<td>Accession</td>
<td>Human Name</td>
<td>Active/inactive</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>P39687</td>
<td>AN32A_HUMAN</td>
<td>no</td>
<td>Protein (GRP-75) (Heat shock 70 kDa protein 9) (Mortalin) (MOT) (Peptide-binding protein 74) (PBPA74) (Acidic leucine-rich nuclear phosphoprotein 32 family member A) (Acidic nuclear phosphoprotein pp32) (leucine-rich acidic nuclear protein) (LANP) (Mapmodulin) (Potent heat-stable protein phosphatase 2A inhibitor 1P2P2A) (Putative HLA-DR-associated protein I) (PHAPI)</td>
</tr>
<tr>
<td>P40925</td>
<td>MDHC_HUMAN</td>
<td>no</td>
<td>Malate dehydrogenase, cytoplasmic (EC 1.1.1.37) (Cytosolic malate dehydrogenase) (Diiodophenylpyruvate reductase) (EC 1.1.1.96)</td>
</tr>
<tr>
<td>P40926</td>
<td>MDHM_HUMAN</td>
<td>no</td>
<td>Malate dehydrogenase, mitochondrial (EC 1.1.1.37)</td>
</tr>
<tr>
<td>P46060</td>
<td>RAGP1_HUMAN</td>
<td>yes</td>
<td>Ran GTPase-activating protein 1 (RanGAP1)</td>
</tr>
<tr>
<td>P49321</td>
<td>NASP_HUMAN</td>
<td>yes</td>
<td>Nuclear autoantigenic sperm protein (NASP)</td>
</tr>
<tr>
<td>P49366</td>
<td>DHYS_HUMAN</td>
<td>no</td>
<td>Deoxyhypusine synthase (DHS) (EC 2.5.1.46)</td>
</tr>
<tr>
<td>P49588</td>
<td>SYAC_HUMAN</td>
<td>no</td>
<td>Alanine--tRNA ligase, cytoplasmic (EC 6.1.1.7) (Alanyl-tRNA synthetase) (AlaRS) (Renal carcinoma antigen NY-REN-42) (Histidine triad nucleotide-binding protein 1) (Protein kinase C inhibitor 1) (Protein kinase C-interacting protein 1) (PKCI-1)</td>
</tr>
<tr>
<td>P49773</td>
<td>HINT1_HUMAN</td>
<td>no</td>
<td>Hepatoma-derived growth factor (HDGF) (High mobility group protein 1-like 2) (HMG-1L2)</td>
</tr>
<tr>
<td>P51858</td>
<td>HDGF_HUMAN</td>
<td>yes</td>
<td>Rho GDP-dissociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha)</td>
</tr>
<tr>
<td>P52565</td>
<td>GDIR1_HUMAN</td>
<td>yes</td>
<td>ATP-citrate synthase (EC 2.3.3.8) (ATP-citrate (pro-S)-lyase) (ACL) (Citrate cleavage enzyme)</td>
</tr>
<tr>
<td>P54090</td>
<td>PEPD1_HUMAN</td>
<td>no</td>
<td>Tyrosine--tRNA ligase, cytoplasmic (EC 6.1.1.1) (Tyrosyl-tRNA synthetase) (TyrRS) [Cleaved into: Tyrosine--tRNA ligase, cytoplasmic, N-terminally processed]</td>
</tr>
<tr>
<td>P54577</td>
<td>SYYC_HUMAN</td>
<td>no</td>
<td>Adenylate kinase 2, mitochondrial (AK 2) (EC 2.7.4.3) (ATP-AMP transphosphorylase 2) (ATP:AMP phosphotransferase) (Adenylate monophosphate kinase) [Cleaved into: Adenylate kinase 2, mitochondrial, N-terminally processed]</td>
</tr>
<tr>
<td>P54819</td>
<td>KAD2_HUMAN</td>
<td>no</td>
<td>Transitional endoplasmic reticulum ATPase (TER ATPase) (EC 3.6.4.6) (15S Mg(2+)-ATPase p97 subunit) (Valosin-containing protein) (VCAP)</td>
</tr>
<tr>
<td>P55072</td>
<td>TERA_HUMAN</td>
<td>yes</td>
<td>Triosephosphate isomerase (TIM) (EC 5.3.1.1) (Triosephosphate isomerase)</td>
</tr>
<tr>
<td>P60174</td>
<td>TPIS_HUMAN</td>
<td>yes</td>
<td>Actin, cytoplasmic 1 (Beta-actin) [Cleaved into: Actin, cytoplasmic 1, N-terminally processed]</td>
</tr>
<tr>
<td>P60709</td>
<td>ACTB_HUMAN</td>
<td>yes</td>
<td>Ubiquitin-conjugating enzyme E2 K (EC 2.3.2.23) (E2 ubiquitin-conjugating enzyme K) (Huntingtin-interacting protein 2) (HIT-2) (ubiquitin carrier protein) (ubiquitin-conjugating enzyme E2-25 KDa) (ubiquitin-conjugating enzyme E2(25K)) (ubiquitin-conjugating enzyme E2-25K) (ubiquitin-protein ligase) (Peptidyl-prolyl cis-trans isomerase A (PPIase A) (EC 5.2.1.8) (Cyclophilin A) (Cyclophilin A-binding protein) (Rotamase A) [Cleaved into: Peptidyl-prolyl cis-trans isomerase A, N-terminally processed]</td>
</tr>
<tr>
<td>P60842</td>
<td>IF4A1_HUMAN</td>
<td>no</td>
<td>Growth factor receptor-bound protein 2 (Adapter protein GRB2) (Protein Ash) (SH2/SH3 adapter GRB2) (14-3-3 protein zeta/delta) (Protein kinase C inhibitor protein 1) (KCI-1)</td>
</tr>
<tr>
<td>P61086</td>
<td>UBE2K_HUMAN</td>
<td>yes</td>
<td>Guanine nucleotide-binding protein subunit beta-2-like 1 (Cell proliferation-inducing gene 21 protein) (Guanine nucleotide-binding protein subunit beta-like protein 12.3) (Human lung cancer oncogene 7 protein) (HLC-7) (Receptor for activated C kinase) (Receptor of activated protein kinase C 1) (RACK1) [Cleaved into: Guanine nucleotide-binding protein subunit beta-2-like 1, N-terminally processed]</td>
</tr>
<tr>
<td>P62937</td>
<td>PPIA_HUMAN</td>
<td>yes</td>
<td>SUMO-conjugating enzyme UBC9 (EC 6.3.2.1) (SUMO-protein ligase) (ubiquitin carrier protein 9) (ubiquitin carrier protein I)</td>
</tr>
<tr>
<td>P62993</td>
<td>GRB2_HUMAN</td>
<td>no</td>
<td>Growth factor receptor-bound protein 2 (Adapter protein GRB2)</td>
</tr>
<tr>
<td>P63104</td>
<td>1433Z_HUMAN</td>
<td>no</td>
<td>Guanine nucleotide-binding protein subunit beta-2-like 1 (Cell proliferation-inducing gene 21 protein) (Guanine nucleotide-binding protein subunit beta-like protein 12.3) (Human lung cancer oncogene 7 protein) (HLC-7) (Receptor for activated C kinase) (Receptor of activated protein kinase C 1) (RACK1) [Cleaved into: Guanine nucleotide-binding protein subunit beta-2-like 1, N-terminally processed]</td>
</tr>
<tr>
<td>P63241</td>
<td>IF5A1_HUMAN</td>
<td>yes</td>
<td>SUMO-conjugating enzyme UBC9 (EC 6.3.2.1) (SUMO-protein ligase) (ubiquitin carrier protein 9) (ubiquitin carrier protein I)</td>
</tr>
</tbody>
</table>
CHAPTER 8: APPENDIX

315

N-Box which binds to Ub-conjugating enzyme E2 I (ubiquitin-protein ligase I) (p18)

Platelet-activating factor acetylhydrolase IB subunit beta (EC 3.1.1.47) (PAF acetylhydrolase 30 kDa subunit) (PAF-AH 30 kDa subunit) (PAF-AH subunit beta) (PAFAH subunit beta)

Peptidyl-prolyl cis-trans isomerase FKBP3 (FKBP3) (EC 5.2.1.8) (25 kDa FK506-binding protein) (25 kDa FKBP) (FKBP-25) (FK506-binding protein 3) (FKBP-3) (Immunophilin FKBP25) (Rapamycin-selective 25 kDa immunophilin) (Rotamase)

Fatty acid-binding protein, epidermal (Epidermal-type fatty acid-binding protein) (E-FABP) (Fatty acid-binding protein 5) (Psoriasis-associated fatty acid-binding protein homolog) (PA-FABP)

Peptidyl-prolyl cis-trans isomerase FKBP4 (FKBP4) (EC 5.2.1.8) (51 kDa FK506-binding protein) (FKBP51) (52 kDa FKBP) (52 kDa FK506-binding protein) (52 kDa FKBP) (59 kDa immunophilin) (p59) (FK506-binding protein 4) (FKBP4) (FKBP59) (HSP-binding immunophilin) (HBI) (Immunophilin FKBP52) (Rotamase) [Cleaved into: Peptidyl-prolyl cis-trans isomerase FKBP4, N-terminally processed]

Lactoylglutathione lyase (EC 4.4.1.5) (Aldoketomutase) (Glyoxalase I) (Glx I) (Ketone-aldehyde mutase) (Methylglyoxalase) (S-D-lactoylglutathione methylglyoxal lyase)

Kinesin light chain 1 (KLC 1)

Peroxiredoxin-4 (EC 1.11.1.15) (Antioxidant enzyme AOE372) (AOE37-2) (Peroxiredoxin IV) (Prx-IV) (Thioredoxin peroxidase AO372) (Thioredoxin-dependent peroxide reductase A0372)

Microtubule-associated protein RP/EB family member 1 (APC-binding protein EB1) (End-binding protein 1) (EB1)

Fascin (55 kDa actin-bundling protein) (Singer-like protein) (p55)


Anamorsin (Cytokine-induced apoptosis inhibitor 1) (Fe-S cluster assembly protein DRE2 homolog)


Microtubule-associated protein RP/EB family member 1 (APC-binding protein EB1) (End-binding protein 1) (EB1)

Fascin (55 kDa actin-bundling protein) (Singer-like protein) (p55)


Anamorsin (Cytokine-induced apoptosis inhibitor 1) (Fe-S cluster assembly protein DRE2 homolog)

Programmed cell death 6-interacting protein (PDCD6-interacting protein) (ALG-2-interacting protein 1) (ALG-2-interacting protein X) (Hsp90 co-chaperone) (Progesterone receptor complex p23) (Telomerase-binding protein p23)

Microtubule-associated protein RP/EB family member 1 (APC-binding protein EB1) (End-binding protein 1) (EB1)

Fascin (55 kDa actin-bundling protein) (Singer-like protein) (p55)


Anamorsin (Cytokine-induced apoptosis inhibitor 1) (Fe-S cluster assembly protein DRE2 homolog)

Programmed cell death 6-interacting protein (PDCD6-interacting protein) (ALG-2-interacting protein 1) (ALG-2-interacting protein X) (Hsp90 co-chaperone) (Progesterone receptor complex p23) (Telomerase-binding protein p23)
<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9P258</td>
<td>RCC2_HUMAN (RCC1-like protein TD-60) (Telophase disk protein of 60 kDa)</td>
</tr>
<tr>
<td>Q9UBE0</td>
<td>SAE1_HUMAN (SUMO-activating enzyme subunit 1 (ubiquitin-like 1-activating enzyme E1A) [Cleaved into: SUMO-activating enzyme subunit 1, N-terminally processed])</td>
</tr>
<tr>
<td>Q9UBT2</td>
<td>SAE2_HUMAN (SUMO-activating enzyme subunit 2 (EC 6.3.2.-) (Anthracycline-associated resistance ARX) (ubiquitin-like 1-activating enzyme E1B) (ubiquitin-like modifier-activating enzyme 2))</td>
</tr>
<tr>
<td>Q9UHD1</td>
<td>CHRD1_HUMAN (Cysteine and histidine-rich domain-containing protein 1 (CHORD domain-containing protein 1) (CHORD-containing protein 1) (CHP-1) (Protein morgana))</td>
</tr>
<tr>
<td>Q9UK76</td>
<td>HN1_HUMAN (Hematological and neurological expressed 1 protein (Androgen-regulated protein 2) [Cleaved into: Hematological and neurological expressed 1 protein, N-terminally processed])</td>
</tr>
<tr>
<td>Q9UKY7</td>
<td>CDV3_HUMAN (Protein CDV3 homolog)</td>
</tr>
<tr>
<td>Q9Y230</td>
<td>RUVB2_HUMAN (RuvB-like 2 (EC 3.6.4.12) (48 kDa TATA box-binding protein-interacting protein) (51 kDa erythrocyte cytosolic protein) (ECP-51) (INO80 complex subunit J) (Repressing pontin 52) (Reptin 52) (TIP49b) (TIP60-associated protein 54-beta) (TAP54-beta))</td>
</tr>
<tr>
<td>Q9Y265</td>
<td>RUVB1_HUMAN (RuvB-like 1 (EC 3.6.4.12) (49 kDa TATA box-binding protein-interacting protein) (49 kDa TBP-interacting protein) (54 kDa erythrocyte cytosolic protein) (ECP-54) (INO80 complex subunit H) (Nuclear matrix protein 238) (NMP 238) (Pontin 52) (TIP49a) (TIP60-associated protein 54-alpha) (TAP54-alpha))</td>
</tr>
<tr>
<td>Q9Y617</td>
<td>SERC_HUMAN (Phosphoserine aminotransferase (EC 2.6.1.52) (Phosphohydroxythreonine aminotransferase) (PSAT))</td>
</tr>
</tbody>
</table>