Chemical Synthesis of Betatrophin and the Integral Membrane Protein IFITM3 by KAHA Ligation

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. Sc. ETH ZURICH)

presented by

Thibault Jean Roger Harmand
Master of Science in Chemistry, University of Lyon
Born on 15.03.1989
Citizen of France

Accepted on the recommendation of
Prof. Dr Jeffrey W. Bode, examiner
Prof. Dr. Donald Hilvert, co-examiner

2017
Abstract

Proteins are involved in and perform a myriad of biological functions. Understanding their structures, functions and interactions with small molecules or other proteins is essential for modern day science. In recent years, proteins as therapeutics have become more and more relevant in the pharmaceutical industry, because of their low toxicity, low risk of side effects as well as their high specificity.

Hence, obtaining pure and homogenous samples of proteins has become an important field of research. Although recombinant production allows access to a large number of proteins, it does present limitations: incorporation of unnatural amino acids and the highly sought after posttranslational modifications (PTMs) – that are often critical for the biological activity of the protein of interest – are difficult, if not impossible, to implement. Chemical protein synthesis appears to be a good alternative to the recombinant approach, as it can easily bypass the mentioned limitations. It can deliver homogenous protein samples with almost any kind of amino acid component, without biological contaminants.

In 2006, the Bode group developed a method called the α-ketoacid–hydroxylamine ligation (KAHA ligation) which allows the chemical synthesis of proteins by ligating two unprotected peptide segments. The KAHA ligation proceeds via the decarboxylative condensation of an α-ketoacid with a hydroxylamine to form an amide bond under acidic organic buffering conditions at slightly elevated temperature. Since its discovery, the KAHA ligation has been subject to constant development, allowing the Bode group to synthesize proteins like SUMO2, SUMO3, irisin, AS48, NP4, betatrophin, interleukin-2 proteins and IFITM3.

This dissertation describes, in the first chapter, the synthetic improvements made for two main building blocks: the cyanosulfur ylide linker and the (S)-N-Boc-5-oxaproline, used for the preparation of α-ketoacid peptide segments and hydroxylamine peptide segments respectively. The new procedures enable the production of these monomers on a multigram scale, allowing our group to scale up the synthesis of synthetic proteins.

The second chapter describes the usage of the KAHA ligation for the preparation of multimilligram quantities of the protein hormone betatrophin. This protein had attracted attention after primary results claimed it had direct effects on the proliferation of β-cells, giving rise to hope for a new approach to treat diabetes. This synthesis represents the first report of a synthetic protein assembled via five peptide segments by KAHA ligation.

Integral membrane proteins are, because of their hydrophobicity, challenging to obtain both by recombinant overexpression and chemical synthesis. In a third chapter, the KAHA
ligation was used to synthesize one of these challenging molecules: the antiviral transmembrane protein IFITM3. The unique features of the KAHA ligation – acidic organic solvents and the formation of a depsi-peptide at the ligation site – appeared to be ideally suited for the synthesis of such hydrophobic proteins, and we could successfully synthesize the desired protein in a multimilligram quantity. A phosphorylated, a palmitoylated and a fluorescent variant of IFITM3 could also be synthesized using the KAHA ligation. We showed that the synthetic IFITM3 was incorporating itself into a lipid membrane and displaying an alpha helix fingerprint, as predicted by the calculated 3D structure. Preliminary results using lipid vesicles and fluorescently labeled viruses showed that our synthetic material displays antiviral activity and more experiments are currently being conducted.
Résumé

Les protéines sont impliquées et performent un large nombre de processus biologiques. La compréhension de leurs structures, de leurs fonctions ou encore de leurs interactions avec d'autres molécules ou d'autres protéines représente une partie importante de la recherche scientifique actuelle. De plus, depuis quelques années, les protéines comme nouveaux médicament est un secteur de l’industrie pharmaceutique en pleine expansion grâce à leur faible toxicité, le faible risque d’effets secondaires ainsi que leur grande spécificité.

Pour toutes ces raisons, obtenir des protéines de manière pure et homogène est devenu un champ de recherche important. Bien que l’obtention de protéine par les techniques recombinantes classiques ont permis de produire et d’isoler nombre d’entre elles, ces approches présentent quelques limitations : incorporer dans la chaîne d’acides aminés des acides aminés non naturels ou incorporer des modifications post-traductionnelles est très compliqué voir souvent impossible. La synthèse chimique de protéine apparaît de plus en plus comme une bonne alternative puisqu’il elle permet de contourner ces problèmes et peut générer des échantillons de protéine de manière pure, sans bio contamination et pouvant comporter n’importe quelle sorte d’acide aminé ou autre molécule.

En 2006, le groupe du Prof. Bode a développé une réaction chimique entre un α-cétoacide et une hydroxylamine, appelée la KAHA ligation (α-ketoacid-hydroxylamine ligation) qui permet la synthèse de protéines en attachant ensemble deux segments de protéines. Cette réaction procède par une condensation du α-cétoacide et de l’hydroxyle amine pour former une liaison amide en condition acide, dans des solvants organiques à des températures légèrement élevées. Depuis sa découverte, la KAHA ligation est en constante évolution et de nombreux monomères ont été développés permettant au Bode groupe de synthétiser un grand nombre de protéines : SUMO2, SUMO3, irisin, AS48, NP4, betatrophin, interleukin-2 et IFITM3.

Cette dissertation détaille dans un premier chapitre les améliorations synthétiques de deux principaux monomères, le cyanosulfur ylide linker et le (S)-N-Boc-5-oxaproline, utilisés pour la préparation de peptide α-cétoacid et de peptide hydroxylamine respectivement. Ces nouvelles procédures ont permis la production de ces molécules en multi grammes et ce qui a permis à notre group d’augmenter les quantités de protéines synthétisées dans nos laboratoires.

Dans un second chapitre, la KAHA ligation a été utilisée pour préparer plusieurs milligrammes de la protéine betatrophin. Cette protéine a attiré l’attention après que de premiers résultats aient présenté cette protéine comme ayant une influence avec la
prolifération de cellules β, laissant penser qu’une nouvelle approche pour traiter le diabète était possible. Cette synthèse est le premier exemple de l’utilisation de la KAHA ligation pour la synthèse de protéine en combinant cinq segments de protéine.

Les protéines membranaires, de par leur nature hydrophobique, sont compliquées à obtenir que ce soit par recombinaison ou par synthèse chimique. Dans un troisième chapitre, la KAHA ligation a été utilisé pour synthétiser l’une de ces molécule complexe: la protéine membranaire antiviral IFITM3. Les caractéristiques uniques de la KAHA ligation – l’utilisation de solvant organique à pH acide ainsi que la formation de depsi-peptide au site de ligation – rend cette ligation particulièrement adaptée à la synthèse de protéines membranaires et nous avons pu synthétiser plus de 10 milligrammes de cette protéine. Toujours en utilisant la KAHA ligation nous avons pu synthétiser différentes variantes de cette protéine : l’une avec une tyrosine phosphorylée, l’une imitant la palmitylation des trois cystéines et une dernière fluorescente. Nous avons également montré que notre IFITM3 synthétique s’incorporait naturellement dans la membrane de vésicules lipidiques et que le signal obtenu par CD spectroscopie – montrant la signature d’alpha hélix – correspondait aux prédictions de sa structure 3D.

Nous avons aussi obtenu des résultats préliminaires utilisant des virus fluorescents et des vésicules lipidiques imitant la membrane de cellules, montrant que notre IFITM3 possédait une activité antivirale.
Acknowledgments

First and foremost, I would like to thank Prof. Jeffrey Bode, who has given me the opportunity to join his group. He has been a great mentor and has always supported me during my PhD. Jeff let me a lot of freedom in my work, letting me try anything I had in mind, and this greatly contributed to make my PhD an amazing experience. I also knew that his door was always open and that I could always come to him if I had problems. I also want to thank him for giving me the opportunity to go to Nagoya where I spent three great months in a truly wonderful country. For all of this, I will always be grateful.

I further want to thank Prof. Donald Hilvert for accepting the co-examination of this thesis. His valuable comments and many corrections have resulted in a major improvement of this manuscript.

I would like to thank the two senior scientists of our group: Dr. Vijay Pattabiraman who showed me the basics of peptide and protein chemistry and Dr. Shunsuke Oishi, who welcomed me in Japan, making it really easy to integrate and work in good conditions.

A big thank you to Mario Kessinger for making any kind of administrative problems I encountered during my PhD a formality, and for the couple of French words we exchanged every morning, merci Monsieur.

Before going into more details, I would like to thank all the former and current members of the Bode group for the stimulating environment. It was a great experience to be surrounded by so many brilliant and motivated people.

I had the chance to start at a similar time with Dr. Claudia Murar, Dr. Benedikt Wanner, Dr. Thomas Wucherpfennig, Dr. Imants Kreituss and Florian Rohrbacher. We spent a lot of time together, some may say “maybe too much”, but that was really great to have them with me almost all along. I have such great memories with every one of them and I think that it will be hard to find a group of people like that ever again. We created a really great dynamic and atmosphere and I am really proud to have friends like them, even Florian. A big thank you to all of you!

I also I would like to thank all my “E floor” colleague-friends; the old ones: Dr. Ivano Pusterla, Dr. Thomas Wucherpfennig, Dr. Claudia Murar, Florian Rohrbacher, Dr. Satomi Shimura, Gabor Boross, Simon Baldauf, Andre Zwicky “Zwückly” and the new ones: Mamiko (Mamikoko) Ninomiya and Jakob Farnung. We had an amazing time together and I will miss you.

A special and important thank you to the three ladies that help me correcting my thesis and my SNSF earlyfellowship application: Dr. Claudia Murar, Dr. Satomi Nayamura and
Haewon Song. I do not know if I would have got the grant without them (probably not), thank you a lot!

Thanks also to the member of the Kon-Tiki “fan club”: Dr. Imants Kreituss and Dr. Gabriel Schäfer, for the great evening we had together. I hope we will be able to have some Züri shots together in a close future.

Two persons deserve a particular mention: Dr. Takahiro Hayashi and Dr. Adrien Joliton. I guess I do not need to go into details, I just wanted to say thanks a lot for the fun we had, and for having, together, brought a special “touch” to the PhD apéros in the LOC!

I would like to thank Valeria, for being by my side during almost all my PhD time, and for making, more than once, all the dark clouds go away.

Finally, I would like to thank my Family. Papa, Maman, merci pour m'avoir toujours soutenu, encouragé et supporté et pas seulement durant ma thèse. Je sais que j'ai pu et je peux toujours compter sur vous. Merci pour avoir fait le déplacement pour ma soutenance, cela m'a fait énormément plaisir. Vous êtes de vrais modèles pour moi et je suis chanceux et fier d'avoir des parents comme vous. Merci pour tout.

Merci aussi à Pierre pour avoir fait le déplacement et m'avoir aidé à organiser l'apéro, mais aussi pour être un super petit frère.
List of Publications and Presentation


Oral Presentation


Poster Presentation

# Table of Contents

## CHAPTER 1. Introduction

1.1. Introductory remarks ........................................ 2  
1.2. Peptide Chemistry ........................................... 4  
1.3. Condensation of protected peptide segments ............. 8  
1.4. Protein synthesis by chemical ligations ................. 9  
   1.4.1. Native Chemical Ligation ............................... 10  
   1.4.2. Serine Threonine Ligation ............................. 11  
1.5. α-Ketoacid-hydroxylamine ligation (KAHA ligation) .... 12  
   1.5.1. α-ketoacid and hydroxyl amine for KAHA ligation 13  
   1.5.1.1. Peptide-α-ketoacid .................................. 13  
   1.5.1.2. Peptide-hydroxyl amine ............................ 15  
   1.5.2. Protein synthesis by KAHA ligation .................. 17  
   1.5.3. KAHA ligation mechanism ............................. 18  
1.6. Conclusions ................................................. 20  
1.7. References .................................................. 21

## CHAPTER 2. Synthesis and optimization of the key building blocks – Cyanosulfur ylide, Boc- and Fmoc-Oxaproline – for KAHA ligation

2.1. General Introduction ........................................... 32  
2.2. Cyanosulfur ylide linker for the preparation of α-ketoacid peptide segments 32  
   2.2.1. Introduction ............................................... 32  
   2.2.2. Synthesis of the cyanosulfur ylide linker 2.1 .......... 34  
   2.2.3. Alkylation of the cyanosulfur ylide linker on resin 36  
   2.2.4. Fmoc SPPS, resin cleavage and oxone oxidation ... 37  
   2.2.5. Conclusions .............................................. 38  
2.3. New route for the synthesis of (S)-N-Boc-5-Oxaproline and (S)-N-Fmoc-5-Oxaproline 38  
   2.3.1. Introduction ............................................... 38  
   2.3.2. Enantioselective organomediated route for the synthesis of (S)-N-Boc-5-Oxaproline 40  
   2.3.3. Synthesis of (S)-N-Fmoc-5-Oxaproline from (S)-N-Boc-5-Oxaproline 42  
   2.3.4. Conclusions and outlook .............................. 43
CHAPTER 3. Chemical synthesis of the protein betatrophin  46

3.1. Introduction  46
3.1.1. The betatrophin protein as a new inducer for β-cell proliferation  47
3.2. Chemical synthesis of betatrophin  47
3.2.1. Retrosynthesis, disconnection and assembling strategies towards betatrophin  47
3.2.2. Segments synthesis towards betatrophin  50
3.2.3. Assembly of the peptide segments by KAHA ligation towards betatrophin  51
3.3. The fall of betatrophin  54
3.4. Conclusion and outlook  55
3.5 References  56

CHAPTER 4. Chemical synthesis of the antiviral integral membrane protein IFITM3  60

4.1. Introduction  60
4.1.1. Generalities about membrane proteins  60
4.1.2. A specific class of integral membrane protein: the transmembrane proteins  62
4.1.3. Production and purification of integral membrane proteins, a challenging topic  63
4.2. IFITM3: a transmembrane antiviral protein  64
4.2.1. Introduction to IFITM3  64
4.2.2. Mode of action of IFITM3  64
4.2.3. Posttranslational modifications involved in IFITM3 bioactivity  66
4.2.4. IFITM3 can adopt different topology in lipid membranes  69
4.3. Chemical synthesis of IFITM3 by KAHA ligation  70
4.3.1. Generalities  70
4.3.2. Retrosynthesis, disconnection and segment synthesis towards IFITM3  71
4.3.2.1. First strategy: NCL/KAHA ligation  71
4.3.2.2. Second strategy: KAHA ligation  78
4.3.2.3. Synthesis of phosphorylated Tyr20 IFITM3  83
4.3.2.4. Synthesis of cysteine palmitoylated IFITM3  85
4.3.2.5. Incorporation in lipid membranes  88
4.3.2.6. Synthesis of fluorescent IFITM3  89
4.4. Conclusions and outlook 90
4.5. References 92

CHAPTER 5. Summary and outlook 102

5.1. Optimization and new synthetic route for the synthesis of SY linker, Boc-Opr and Fmoc-Opr 102
5.2. Synthesis of the betatrophin protein 102
5.3. Synthesis of the antiviral membrane protein via KAHA ligation 103
5.4. Combining KAHA ligation and NCL in one pot for the synthesis of proteins, application to U18 ctenitoxin-Co1b 104
5.5. Synthesis of the Call1 protein by combining KAHA ligation and NCL 105
5.6. References 107

CHAPTER 6. Experimental part 110

6.1. General Methods 110
6.2. Experimental for “Chapter 2: Synthesis and Optimization of the Key Building Blocks – Cyanosulfur Ylide, Boc- and Fmoc Oxaproline – For KAHA Ligation” 114
6.2.1. Cyanosulfur ylide linker synthesis 114
6.2.2. Typical procedure for the loading and alkylation of the linker 2.1 and attachment of the first amino acid 116
6.2.3. Synthesis of Boc and Fmoc Oxaproline 117
6.3. Experimental for “Chapter 3: Chemical Synthesis of the Betatrophin Protein” 123
6.3.1. Synthesis of the peptide segments 123
6.3.2. KAHA ligation for the synthesis of betatrophin 127
6.3.3. CD spectra 131
6.4. Experimental for “Chapter 4: Chemical Synthesis of the Antiviral Integral Membrane protein IFITM3” 133
6.4.1. Synthesis of the peptide segments 133
6.4.2. Synthesis of IFITM3 via the KAHA-NCL strategy 142
6.4.2.1. KAHA ligations with thioester moiety 142
6.4.2.2. Native Chemical Ligation as the final ligation for the synthesis of IFITM3 144
6.4.3. Synthesis of IFITM3-ArgTAG by KAHA ligation 149
6.4.3.1. KAHA ligations with the peptide segments 149
6.4.3.2. CD spectra measurement of IFITM3-ArgTAG

6.4.4. Synthesis of IFITM3 by KAHA ligation
   6.4.4.1. KAHA ligation with peptide segments
   6.4.4.2. SDS-PAGE analysis of IFITM3-ArgTAG and IFITM3 and MS/MS analysis of IFITM3

6.4.4.3. CD spectra measurement of IFITM3

6.4.5. Synthesis of phosphorylated-IFITM3

6.4.6. Synthesis of fluorescent-IFITM3

6.4.7. Synthesis of palmitoylated IFITM3
   6.4.7.1. Synthesis of the different alkylating reagents
   6.4.7.2. IFITM3 cysteine alkylation using palmitoylhydrazide
   6.4.7.3. IFITM3 cysteine alkylation using dehydroalanine and thiopalmitic acid
   6.4.7.4. IFITM3 cysteine alkylation using 2-chloro-N-undecylacetamide
   6.4.7.5. IFITM3 Cysteine alkylation with hexyldecyl-dithiopyridine

6.4.8. Vesicles assay and fluorescence microscopy
   6.4.8.1. Preparation of the vesicles
   6.4.8.2. Incorporation of IFITM3 in lipid vesicles
   6.4.8.3. Fluorescence microscopy of Fluoro-IFITM3 in the membrane of vesicles

CHAPTER 7. Appendix: NMR spectra
## Symbols and abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<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>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>Cl-</td>
<td>6-Chloro-1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOBt</td>
<td>(1-cyano-2-ethoxy-2-oxo-ethylidenaminoxy)dimethylaminomorpholinocarbenium hexafluorophosphate</td>
</tr>
<tr>
<td>COMU</td>
<td>1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate</td>
</tr>
<tr>
<td>d</td>
<td>chemical shift</td>
</tr>
<tr>
<td>d. r.</td>
<td>diastereomeric ration</td>
</tr>
<tr>
<td>DIC</td>
<td>N,N'-Diisopropylcarbodiimide</td>
</tr>
<tr>
<td>DIPEA</td>
<td>N,N-diisopropylethylamine</td>
</tr>
<tr>
<td>DEA</td>
<td>N,N-diethylamine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DODT</td>
<td>3,6-dioxo-1,8-octanediethiol</td>
</tr>
<tr>
<td>E</td>
<td><em>entgegen</em> (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-fluorenylethoxycarbonyl</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HATU</td>
<td>1-[bis(dimethylamino)methylene]-5-chlorobenzotriazolium 3-oxid hexafluorophosphate</td>
</tr>
<tr>
<td>HCCA</td>
<td>α-cyano-4-hydroxycinnamic acid</td>
</tr>
<tr>
<td>HCTU</td>
<td>1-[bis(dimethylamino)methylene]-5-chlorobenzotriazolium 3-oxid hexafluorophosphate</td>
</tr>
<tr>
<td>HOBt</td>
<td>Hydroxybenzotriazole</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>Hse</td>
<td>(S)-homoserine</td>
</tr>
<tr>
<td>J</td>
<td>coupling constant</td>
</tr>
<tr>
<td>KAHA</td>
<td>α-ketoacid hydroxylamine (ligation)</td>
</tr>
<tr>
<td>m</td>
<td>multiplet (NMR), milli</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Term/Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------</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>MPA</td>
<td>3-mercaptopropionic acid</td>
</tr>
<tr>
<td>MPAA</td>
<td>4-mercaptophenylacetic acid</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>NCL</td>
<td>native chemical ligation</td>
</tr>
<tr>
<td>Nff</td>
<td>nonafluorobutanesulfonyl fluoride</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>ppm</td>
<td>parts per million</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>rt</td>
<td>room temperature</td>
</tr>
<tr>
<td>s</td>
<td>singlet</td>
</tr>
<tr>
<td>S</td>
<td>sinister (configuration)</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>TBS</td>
<td>tert-butylidiphenylsilyl</td>
</tr>
<tr>
<td>tert</td>
<td>tertiary</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoacetic acid</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TIPS</td>
<td>triisopropylsilane</td>
</tr>
<tr>
<td>TOF</td>
<td>time of flight</td>
</tr>
<tr>
<td>TLC</td>
<td>thin-layer chromatography</td>
</tr>
<tr>
<td>TMS</td>
<td>trimethylsilyl</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>Z</td>
<td>zusammen (olefin geometry)</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION
CHAPTER 1. Introduction

1.1 Introductory remarks

Proteins are involved and perform a myriad of remarkable functions in living organism. They catalyze a large majority of reactions that occur in cells, they make up the antibodies in the immune system, they convert chemical energy into mechanical work, they transport small molecules and ions inside the organism, they maintain the structure of the cell and they are involved in cell-cell adhesion.

![Amino Acid Structures](image)

**Figure 1.1.** List of the 20 amino acids encoded by the universal genetic code with their three- and one-letter codes.

Although these biomolecules serve many different functions, their composition is merely the result of the different combination of 21 amino acid residues (Figure 1.1) linked to each other via an amide (or peptide) bond linkage. All these amino acids, except glycine, have side chains, some of them bearing functional groups such as carboxylic acids, primary amides, alcohols, amines, urea, imidazole, thiol and selenol. All the amino acids bearing a side chain are chiral at the \( \alpha \)-carbon atom and natural proteins contain exclusively \((S)\)-\( \alpha \)-amino acids, with the exception of cysteine and selenocysteine that have the R configuration. The length of these polymers can vary from only few amino acids residues (which are referred to as oligopeptides) to large proteins containing thousands of them.

The different functions and properties that proteins display are dictated by their 3-dimensional structures. These structures provide specific layout of the amino acid functional
groups that are responsible for the reactions and interactions that a protein has with itself and with other molecules. There are four different types of protein structure (Figure 1.2).\(^1\) A linear chain of amino acids in an oligopeptide is called the primary structure. These oligopeptide chains can fold, generating defined secondary structures, the main types being the \(\alpha\)-helix and the \(\beta\)-strand or \(\beta\)-sheet. The amino acid chain can also fold into a compact tertiary structure containing various different secondary structure elements. When a protein is composed of more than one amino acid chain, each chain having its own tertiary structure, the complex is referred to as a quaternary structure.

![Figure 1.2. The four different levels of protein structure.](image)

The combination of these 20 amino acids in nature gives rise to around 20 000 different proteins in human. Studying, analyzing and understanding the role of all these biomolecules is already a labor-intensive task. Yet another level of complexity is added: several amino acid residues can be modified after translation by the ribosome. These modifications, named posttranslational modifications (PTMs), are crucial for the bioactivity of proteins.\(^2\) PTMs are highly diverse and can range from a simple oxidation of the amino acid side chain (e.g. oxidation of the sulfur atom,\(^3\) nitration of an aromatic ring\(^4\)) to the modification of the side chain functional group with small molecules (methylation,\(^5\) palmitoylation,\(^6\) glucosylation,\(^7\) phosphorylation\(^8\)) to even more complex modifications such as the conjugation of other proteins (ubiquitination,\(^9\) sumoylation,\(^10\) etc.) (Figure 1.3).

![Figure 1.3. Selected examples of posttranslationally modified amino acids](image)
In recent years, protein therapeutics have become a significant segment of the pharmaceutical industry, because of their low toxicity and low risk of side effects as well as their high specificity. They are already widely used to treat various life-threatening diseases such as cancer, diabetes and multiple sclerosis. Predictions show a possible market size of nearly $248.7 billion by 2020, when it was only 174.7 billion in 2015.\textsuperscript{11}

For both studying proteins as well as their use as potential drugs, access to pure and homogeneous proteins is essential. For a long time, the only way to obtain them was by isolating proteins from biological samples. Over the last 50 years, biology has made impressive advances and the recombinant production of a wide range of proteins in different organisms is now a routine procedure. However, there are still limitations.

First, the incorporation of non-canonical amino acids is challenging, thus creating variants and mutants of proteins in order to change their bioactivity and properties is complicated. Although different methods to overcome these problems have been proposed and are still in development, such as the amber codon suppression technology\textsuperscript{12} or the pyrrolysine translational machinery,\textsuperscript{13} it is still complicated to incorporate multiple or even one unnatural amino acids into a recombinant protein.

Second, the incorporation of the desired PTMs in a site specific and homogenous fashion is technically difficult and impossible to control.\textsuperscript{14}

To overcome these limitations, the chemical synthesis of proteins has made important progress over the last twenty years and offers an attractive option to obtain pure small- to mid-sized proteins for research and pharmaceutical applications.\textsuperscript{15} The incorporation of non-canonical amino acids as well as PTMs in this way is readily achieved.

\subsection*{1.2 Peptide Chemistry}

Two main problems need to be resolved in order for chemists to generate polypeptides using simple amino acid building blocks without Nature’s powerful machinery. The first problem is that the condensative coupling reaction between a carboxylic acid and an amine to form an amide bond is thermodynamically disfavored and the energy of activation is high.\textsuperscript{16} The second problem is that twelve of the twenty one amino acids bear reactive functional groups on their side chains and the C or N terminus of the growing peptide is, as well, a reactive functional group. The synthesis of a controlled polypeptide would be impossible without the use of suitable protecting groups (Figure 1.4).
To overcome the first problem, the carboxylic acid moiety must be activated. To overcome the second problem, protecting groups must be introduced for each reactive functional group present on the side chain and termini.

In 1903, Emil Fischer reported a coupling reaction between two glycine residues, where one carboxylic acid was activated as an acyl chloride moiety. However such derivatives are highly reactive and side reactions and epimerization at the α-carbon are common. Over the years new reagents have been developed in order to perform similar reactions under faster and milder conditions. A wide variety of such reagents are now known, displaying different physical properties such as solubility, reactivity and safety.

The activation of the carboxylic acid moiety proceeds by reacting it with a so-called coupling reagent in order to generate an activated ester, which in turn reacts with the desired amine. The most common coupling reagents can be divided into three different classes: carbodiimides, phosphonium-based reagents and uronium-based reagents (Figure 1.5).
The problem of the reactive groups of amino acids and growing peptide have been overcome using two main sets of protecting groups. Both of these strategies make use of two orthogonal protecting groups, a temporary one that protects the N-terminal amine and one that is permanently on the side chain functional groups and the C-terminus during the whole synthesis. They can all be removed once the synthesis is complete. These two commonly employed protecting group strategies are called the Boc-and Fmoc-chemistry, named after the temporary protecting group on the N terminus.

In the Boc strategy, the N-terminal Boc protecting group can be removed using trifluoroacetic 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 Fmoc protecting group can be removed using a secondary amine – like piperidine – and the side chain functional groups are masked with TFA labile protecting groups (Figure 1.6). Although Boc and Fmoc strategies are compatible with SPPS, most standard syntheses are nowadays preferably performed using Fmoc SPPS to avoid the safety risks arising from the usage of pure HF in the Boc strategy.

While the coupling of a few of amino acids in solution is straightforward, the repetitive coupling steps, the deprotection of the N-terminal amine and the purification of intermediates after each coupling step makes the synthesis fastidious, or even impossible, for the assembly of longer peptides. R.B. Merrifield proposed a different approach to peptide synthesis by developing a method where the peptides are assembled on a solid phase. This had a

Figure 1.6. Fmoc Chemistry protecting groups commonly used for amino acid side chains.
revolutionary impact on the synthesis and development of peptides, and Merrifield was awarded the Nobel Prize for chemistry in 1984. Solid phase peptide synthesis (SPPS) presents important advantages over peptide synthesis in solution, by using an excess of activated amino acids that can be easily removed by washing. Coupling reactions can therefore be carried out more rapidly and nearly to completion (Figure 1.7).

![Figure 1.7. General representation of Fmoc-solid phase peptide synthesis and frequently utilized linkers.](image)

Since this development, the field of peptide science has seen numerous modifications and improvements. As mentioned before, numerous coupling reagents have been developed, but the resins have also evolved. Although the original Merrifield resin: cross-linked poly(styrene-divinylbenzene) is still in use, more polar resins such as cross-linked poly(dimethylacrylamide) resins,$^{24}$ combinations of soft polyamides with rigid, highly permeable matrices, or highly cross-linked polystyrene, often to give better results.$^{25}$

Solvation behavior is a critical parameter during SPPS and resins made by copolymerization of ethylene oxide and polystyrene$^{26}$ or by grafting PEG chains onto polystyrene beads show better swelling properties.$^{27}$ Nowadays, commercially available resins are modified with different linkers, enabling attachment of the protected C-terminal amino acid
residue by the formation of ester or amide bonds, allowing the synthesis of peptide acids and peptide amides, respectively.\textsuperscript{28}

Theoretically, it would be possible to assemble a peptide chain of any length by SPPS. However, in practice a peptide chain length of more than 40-50 amino acids is difficult to obtain,\textsuperscript{29} The continuous accumulation of side products – truncated and capped peptides – makes the process no longer advantageous. Nonetheless, there exist some examples of proteins that were made entirely by SPPS: ribonuclease A (124 residues)\textsuperscript{30}, human immunodeficiency virus Tat protein (85 residues)\textsuperscript{31}, HIV protease (99 residues)\textsuperscript{32} and the ubiquitin protein (76 residues).\textsuperscript{33} However most eukaryotic proteins are significantly larger, 200-300 amino acid residues and can by no means be produced by SPPS.\textsuperscript{34}

1.3 Condensation of protected peptide segments

A solution to overcome the problem mentioned above exists. A protein target longer than 50 amino acids can be divided into smaller fragments, which are possible to synthesize by SPPS. Once cleaved from the solid support, these fragments can be assembled in solution. The assembly reaction is carried out using coupling reagents while the fragments are still side chain protected (Figure 1.8). The conditions used to cleave the growing peptide from the resin also results in the deprotection of the side chain protecting groups, making the synthesis of side chain protected peptides with free C-terminal carboxylic acid and a free N-terminal amine group complicated.\textsuperscript{35} However the development of the mild acid labile 2-chloro trityl linker allowed cleavage from the resin while keeping the protecting groups on the side chain intact.\textsuperscript{36}

![Figure 1.8. Coupling of protected peptides in solution.](image-url)
Several proteins have been made this way: pleiotrophin (136 residues)\textsuperscript{37} and a precursor protein to the green fluorescent protein (GFP) (238 residues).\textsuperscript{38} However the protecting groups still present on the protein give a strong hydrophobic character, making both purification and characterization challenging. Solubility presents another issue: fully protected peptides exhibit poor solubility in classic organic solvents. The reaction consequently has to be done under diluted conditions, thus slowing down the reaction rate and promoting epimerization of the activated ester.\textsuperscript{39}

These drawbacks encouraged the development of an alternative strategy for the synthesis of proteins: chemical ligation.

### 1.4 Protein synthesis by chemical ligations

A chemoselective reaction brings together and ligates two highly functionalized, unprotected molecules, selectively. Chemical ligation reactions achieve their selectivity by using unique chemical functional groups at the ligation site, which react together in a highly chemoselective fashion (Figure 1.9).

In order for a coupling reaction to be considered as a chemical ligation, some criteria need to be fulfilled.

![Figure 1.9. General scheme of a chemical ligation reaction, affording an amide bond at the ligation site.](image)

**1. Chemoselectivity.** The reactive groups involved have to exclusively react with each other and must be able to do so with many other functional groups present.

**2. Fast rate.** Chemical ligation reactions are bimolecular reactions; thus, the reaction rate depends of the concentration of both starting materials. Furthermore, the molecules used for ligation reactions are often extremely valuable and using near equimolar amount of material is often required. Moreover, the molecular weights of the reaction partners are regularly greater than 5 kDa. This constrains the concentration range at which the chemist can operate: reactions are often carried out at the highest possible concentration allowed by the solubility of the two ligation partners in the selected solvent.
3. Mild conditions. The ligation reaction has to be successfully conducted under conditions (pH, solvents, temperature) that do not alter the functional groups present or the stereochemical information of the starting materials.

These considerations led to the development of a large variety of different chemical ligation reactions. The linkage formed can be of different types. The most common ligation techniques give rise to: amides (native chemical ligation (NCL), α-ketoacid hydroxylamine ligation (KAHA ligation), serine/threonine ligation (STL), traceless Staundinger), thioethers (maleimides and derivaties, thioclick), triazoles – that have been shown to be good mimics of amide bonds – (Huisgen cycloaddition) or oximes (oxime ligation). Depending on the objective, the best-suited ligation reaction method can be chosen from a still expanding pool of techniques.

For the synthesis of a protein, the amide bond forming ligation techniques are by definition the most frequently used and relevant. Hence, they will be described in more detail here. Huisgen cycloaddition has been used for the synthesis of an analogue of the cystatin A protein, which showed comparable bioactivity to the native molecule. The traceless Staudinger reaction has been used for the synthesis of proteins. Although these two techniques have been successful in some cases, NCL, STL and KAHA ligation are the most common methods used for the synthesis of proteins.

1.4.1 Native Chemical Ligation (NCL)

Native chemical ligation is nowadays the most widely used method for the chemical preparation of proteins. Developed by Kent et al. this method ligates under mild conditions an unprotected C-terminal thioester peptide with a peptide containing an unprotected N-terminal cysteine to form the corresponding amide bond. This reaction is highly chemoselective and tolerates all canonical amino acid side chain functionalities. Inspired by the work of Wieland et al. the mechanism of the reaction starts with a first reversible trans thioesterification step followed by a S to N acyl transfer that is irreversible (Figure 1.10).

The reaction proceeds optimally in buffered aqueous condition at a pH typically between 6.5 and 8.5. These conditions present two intrinsic limitations: at acidic pH the thiol group of the cysteine becomes less nucleophilic, but at basic pH the thioester moiety suffers from instability, leading to direct hydrolysis or attack from primary amines such as lysine.
NCL remains to date the most used chemical ligation method for the assembly of proteins. It has been used for the synthesis of a large variety of proteins, from globular to membrane proteins as well as L-amino acid mirror image proteins. Another advantage of NCL is that, by using N-terminal cysteine as the ligating agent, large overexpressed proteins can be generated that would not be obtainable by synthetic chemistry. Thioester proteins can also be obtained directly by overexpression using inteins.56

1.4.2 Serine/Threonine Ligation (STL)

The serine/threonine ligation (STL) was introduced by Li et al.42 in 2010, and is based on the prior work of Tam et al. in 1994.58 The first step consists of the capture of the N-terminal amine of a serine or threonine residue by a C-terminal salicylaldehyde ester that forms a stable cyclic N,O benzylidene acetal intermediate at the ligation site (Figure 1.11). This can be readily cleaved under acidic conditions to yield a native amide bond.
Although C-terminal lysine, aspartic and glutamic acid residues are not compatible with STL, this method provides more synthetic flexibility compared to NCL as it relies on an N-terminal serine or threonine that are significantly more abundant than cysteine. However, it requires a relatively high substrate concentration and is carried out in an acetic acid/pyridine mixture which limit both the solubility of the protein segments and the biocompatibility of the reaction.

Nevertheless, STL was successfully applied in the synthesis of human erythrocyte acylphosphatase (98 residues), the semi-synthesis of peptoid-modified RNase A and several cyclic peptides.

1.5 α-Ketoacid-hydroxylamine ligation (KAHA ligation)

The Bode group has worked on the development of amide bond forming reactions and has developed a novel chemoselective ligation. The reaction involves two unprotected peptides, one bearing a C-terminal α-ketoacid and one bearing an N-terminal hydroxylamine. Together they form an amide bond at the ligation site (Figure 1.12). No additional reagents or catalysts are required and the reaction proceeds at slightly elevated temperatures in a mixture of water and organic solvent without problematic byproducts.
Figure 1.12. α-Ketoacid-hydroxylamine ligation (KAHA ligation).

Over the years, several hydroxylamines for the KAHA ligation have been screened and different methods for the generation of the α-ketoacid moiety at the C-terminus have been developed.

1.5.1 α-Ketoacids and hydroxylamines for KAHA ligation.

1.5.1.1 Peptide-α-ketoacid

The most reactive position of an α-ketoacid group is the electrophilic “keto” group. This position is sensitive to nucleophilic attack and converts into its hydrate form in the presence of water. Although ketoacids are really stable under acidic conditions they tend to easily form enolates under basic ones, resulting in epimerization of the stereocenter.

The first C-terminal α-ketoacid peptides were developed and synthesized by Wassermann et al.\textsuperscript{64} by first coupling cyanophosphorus ylides to protected peptides followed by ozonolysis and hydrolysis (Scheme 1.1). However, this method presents certain drawbacks such as long reaction times, the use of ozone gas, and sometimes epimerization of the α-stereocenter.
Our group has developed an alternative to Wassermann’s approach using a cyanosulfur ylide. Originally, the cyclic sulfur ylide was directly coupled to the C-terminal carboxylic acid via classic coupling methods and would afford the desired α-ketoacid after oxidation with oxone under aqueous acidic conditions (Scheme 1.1).  

Encouraged by this successful new approach, the group developed a solid supported cyanosulfur ylide linker that allowed us to generate C-terminal sulfur ylide peptides directly by SPPS (Figure 1.13).

Unfortunately, the oxidation step is not compatible with cysteine, methionine and tryptophan residues. To overcome this problem, another recent approach has been proposed that directly liberates the ketoacid moiety after TFA cleavage from the resin. This allowed the preparation of cysteine, methionine and tryptophan containing peptide segments and the preparation of proteins, that were not accessible with our previous method.
The group has more recently reported another approach, using a photo-labile version of the protected ketoacid that allows protein assembly from the N-terminus to the C-terminus. However, despite the fact that these recent advancements are often more practical than the original sulfur ylide linker, the sulfur ylide linker remains the only method that allows the preparation of C-terminal arginine α-ketoacids. Furthermore, the sulfur ylide is a perfectly orthogonal protecting group for ketoacids, and allows sequential KAHA ligation from the N- to the C-terminus, providing a flexible tool for the synthesis of proteins.

1.5.1.2 Peptide-hydroxyl amine

The hydroxylamine is the second partner of the KAHA ligation. Hydroxylamines are significantly less basic than regular amines, for examples, pKa of O-benzylhydroxyl amine and O-methylhydroxylamine are 4.23 and 5.80 respectively, while the N-terminal amino group of a peptide and the amino group of a lysine side chain have a pKa of 7.7 and 10.5, respectively. Because of the α-effect of the oxygen substituent, hydroxylamines are good nucleophiles.
Hence, while all the other amino groups are protonated during KAHA ligation, the hydroxylamine is not, remaining the most nucleophilic functional group of the protein segment.

Over time, numerous hydroxylamines have been synthesized and evaluated for their use in the KAHA ligation by our group (Figure 1.15), and many others are in preparation.

**Figure 1.15. Hydroxylamines used in KAHA ligation for the synthesis of proteins.**

So far, the two monomers that present the best balance between easy of synthesis, reactivity and stability are the 5-(S)-oxaproline (Opr)\(^\text{72}\) and the 4-ethoxyl-5-oxaproline (Asa).\(^\text{73}\) Interestingly, KAHA ligation with oxaproline first generates an ester linkage between the two protein segments that can be readily converted into the corresponding amide product under mild basic conditions.\(^\text{74}\) This was seen as a disadvantage at first, it later appeared that this depsi-peptide linkage dramatically increases the solubility of the ligated compound (Scheme 1.2). This unique feature played a critical role for the synthesis of the intramembrane protein IFITM3, described in Chapter 4.

After a rearrangement of the ester linkage, a homoserine residue is generated at the ligation site (Scheme 1.2). Even though this has raised concerns, it has so far never been a problem during biological evaluations.\(^\text{75-78}\)

**Scheme 1.2. Formation of the depsi-peptide and homoserine residue in KAHA ligation using oxaproline.**

Nevertheless, our group has developed a monomer that gives a canonical amino acid at the ligation site. This compound contains a four-membered ring called an oxazetidine and shows incredibly fast ligation rates giving a serine residue at the ligation site.\(^\text{79}\) Serine is among the most frequent amino acids in human proteins. However, the oxazetidine suffers from low stability and the synthetic route to access it is challenging.
1.5.2 Protein synthesis by KAHA ligation

The oxaproline monomer, despite an originally complicated and limiting synthesis – an optimized protocol enabling kilogram scale synthesis is described in Chapter 2 – has allowed our group to synthesize numerous proteins. The first one described by our group was the synthesis of the 63 amino acid long Pup protein (Scheme 1.3).\textsuperscript{72}

![Scheme 1.3. Synthesis of the Pup protein by KAHA ligation.]

Although the first version of oxaproline was Boc-protected, the Boc group limited growth of the protein from the C- to N-terminus. To get more flexibility, an Fmoc-protected oxaproline was developed, allowing sequential synthesis from the C-terminus to the N-terminus.\textsuperscript{80} This, together with the possibility of using the sulfur ylide linker as a masked

![Scheme 1.4. Convergent synthesis of the Nitrophorin 4 protein (1–184) using five segments and four KAHA ligations.]

\textsuperscript{72} AQEOTKGGGTTGDDDDGGRKAGQERKEL

\textsuperscript{80} pH 9.5 0.2 M NH4HCO3 buffer
ketoacid, has allowed our group to synthesize larger proteins like betatrophin\(^ {81} \) and nitrophorin \(4^{76} \), respectively 19.8 and 20.4 kDa.

More recently, a photolabile version of the oxaproline monomer was also developed, giving us even more flexibility regarding the assembly strategies.\(^ {68} \)

KAHA ligation with oxaproline has also been successfully applied in the synthesis of a variety of cyclic peptides\(^ {62} \) and has even allowed the chemical synthesis of the cyclic protein AS-48\(^ {83} \) for which NCL appeared to be unsuited.\(^ {84} \)

### 1.5.3 KAHA ligation mechanism

As described above, several hydroxylamines for the KAHA ligation have been screened and it appeared that these different hydroxylamines not only showed differences in stability and reactivity, but also different reaction pathways. This led us to distinguish between two types of reaction: type I, unsubstituted \(O\)-hydroxylamine; and type II, substituted \(O\)-hydroxylamines. The mechanism of type I KAHA ligations was elucidated by our group via \(^ {18}O\) isotope labeling, which showed that the oxygen atom of the generated amide was derived originally from the oxygen of the hydroxylamine (Scheme 1.5).\(^ {85} \)

![Scheme 1.5. Mechanism of the type I KAHA ligation.](image)

The reaction proceeds to form E- and Z-nitrones after dehydration, 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.

The mechanism of the type II KAHA ligation was investigated after our group discovered that the link formed in KAHA ligations with oxaproline was not an amide bond but an ester.\(^ {74} \) Heavy water was used for \(^ {18}O\) isotope labeling and, surprisingly, as opposed to type
I, the oxygen of the amide bond does not come from the oxygen of the hydroxylamine, but the oxygen of the water (Scheme 1.6).

Following these results, a mechanism was proposed. From their classic reactivity, one would expect the first step to be an attack of the oxaproline on the $\alpha$-ketoacid, forming an iminium ion. Indeed, the ligation studies with $^{18}$O showed that unreacted material did not incorporate $^{18}$O, suggesting that iminium ion formation is not only the first step but is also irreversible. The iminium ion intermediate can undergo decarboxylation and cleavage of the N,O bond – possibly in a concerted fashion – yielding the nitrilium intermediate. We identified this potential intermediate at the branch point of two paths yielding the amide and ester compound (Scheme 1.7).

In path A, water directly attacks the nitrilium generating the amide bond right away. It explains the almost complete incorporation of $^{18}$O in the oxygen of the carbonyl. In path B, the nitrilium undergoes a 6-endo-dig attack from the hydroxyl group generating the cyclic
imminium ether. Reversible nucleophilic attack of water onto the iminium leads to the elimination of the amino group, affording the ester linked compound.

Although the existence of the nitrilium intermediate 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 as well as the incorporation of $^{18}$O labeled into the molecule.

1.6. Conclusions

Obtaining proteins by chemical synthesis has dramatically improved over the last years, and can now provide access to proteins in high purity, free from biological contaminants, and with the possibility of incorporating PTMs, unnatural amino acids and various tags.

Although NCL is still the reference method for the synthesis of proteins, the KAHA ligation method developed in the Bode group is constantly improving, making it an interesting alternative to NCL. Also, because of its features, like functioning in organic and acidic solvents, KAHA ligation can be helpful when NCL does not work.

The goal of this thesis was first to optimize the synthesis of the two building blocks necessary for the KAHA ligation: the sulfur ylide linker and the Boc protected oxaproline (Chapter 2). Second, we attempted to synthesize one of the most important classes of protein, which is also the most complicated to obtain either by recombinant expression or by the current NCL based techniques, namely membrane proteins. We focused on the IFITM3 protein which displays antiviral activity against a large variety of viruses (Chapter 4). Finally, we applied the KAHA ligation for the synthesis of a 19.8 kDa protein that was believed to have an effect on β-cell proliferation: the betatrophin protein (Chapter 3).
1.7 References


11. Technology Advancements in Protein Therapeutics Spurring Market Growth, Reports BCC Research.


42. Bode, J. W.; Fox, R. M.; Baucom, K. D.: Chemoselective amide ligations by

43. Chun, R.; Glabe, C. G.; Fan, H.: Chemical synthesis of biologically active Tat trans-

44. B. Bader, K. Kuhn, D. J. Owen, H.Waldmann, A.Wittinghofer, J. Kuhlmann.: Bioorganic

45. S. Wittrock, T. Becker, H. Kunz.: Synthetic vaccines of tumor-associated glycopeptide

46. Chun, R.; Glabe, C. G.; Fan, H.: Chemical synthesis of biologically active Tat trans-


49. Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R.: Engineering chemical reactivity on cell surfaces

50. Stijn M. Agten, Rory R. Koenen, Hans Ippel, Veit Eckardt, Philipp von Hundelshausen,


CHAPTER 1. Introduction


78. Unpublished work from the Bode group.


The work described in this Chapter was performed in collaboration with Dr. Sameer Kulkarni for the synthesis of the sulfur ylide linker and with Claudia Murar for the synthesis of the oxaproline monomers.
CHAPTER 2. Synthesis and Optimization of the Key Building Blocks – Cyanosulfur Ylide, Boc- and Fmoc-Oxaproline – for KAHA Ligation

2.1 General introduction

In this chapter, the synthesis of three different building blocks used for KAHA ligation, the cyanosulfur ylide linker and the (S)-N-Boc- and (S)-N-Fmoc-5-oxaproline, are described (Figure 2.1). The cyanosulfur ylide linker was the first way to generate, on solid support, C-terminal sulfur ylide peptide segments that, upon oxone oxidation afforded the corresponding \( \alpha \)-ketoacid peptide segments required for KAHA ligation. Stable to the ligation conditions, it behaves as a masked ketoacid allowing chemical synthesis of a protein from the N-terminus to the C-terminus.

Boc- and Fmoc-oxaprolines are the other partners required in the KAHA ligation, and while the Boc-Opr readily delivers free Opr after TFA cleavage, the Fmoc-Opr allows us to synthesize proteins in the C to N direction.

With the success of the KAHA ligation for the synthesis of proteins more and more projects based on the KAHA ligation have appeared in our group, resulting in an increased demand for monomers. Furthermore, in order to make our method accessible to a broad range of research groups, the previous protocols described by our group needed optimization.

![Figure 2.1. The three different building blocks described in this Chapter.](image)

2.2 Cyanosulfur ylide linker for the preparation of \( \alpha \)-ketoacid peptide segments

2.2.1 Introduction

The cyanosulfur ylide linker was first reported by our group a few years ago\(^8\) and allowed us to prepared several proteins by KAHA ligation. This linker is easy to use and can be stored for a long time without traces of decomposition.

It is coupled on the Rink amide polystyrene resin or ChemMatrix\(^8\) resin via standard coupling protocols. These resins can be stored for months without any degradation. Further alkylation of the sulfur with halide-acetonitrile generates the reactive ylide species that can be used to couple the first amino acid of the desired sequence (Scheme 2.1). Once the first Fmoc-
CHAPTER 2. Synthesis and Optimization of the Key Building Blocks – Cyanosulfur Ylide, Boc- and Fmoc-Oxaprolines – for KAHA Ligation

Amino acid has been coupled on the resin, automated Fmoc SPPS can be used in order to synthesize the desired peptide segment. After TFA cleavage – using standard scavenger cocktails – the unprotected peptide segment sulfur ylide can be stored for months in the fridge.

To obtain the α-ketoacid moiety, the sulfur ylide peptide segment must be oxidized under aqueous conditions. Over the years, our group has found that DMDO\(^{87}\) and oxone\(^{88}\) were the best-suited oxidants for this reaction. However, for practical reasons the use of oxone is preferred. This oxidation step, even though quite mild, faces problems when some amino acids are present in the peptide segment. Sulfur containing amino acids like cysteine and methionine, as well as tryptophan residues, appear to be unstable to the oxidation step, limiting the use of the sulfur ylide linker to peptide segments lacking these three amino acids. To overcome this problem, our group has developed different building blocks that readily generate an α-ketoacid moiety directly after TFA cleavage (Chapter 1, section 1.6.1.1).\(^{67}\)
Nevertheless, the sulfur ylide linker remains a practical monomer that tolerates a large number of amino acids in alpha position, avoiding the tedious and long syntheses needed for the protected ketoacid monomers mentioned above. It is also, so far, the only method that can generate an arginine ketoacid; ArgKA has been recently used for the preparation of some proteins such as ubiquitin and IFITM3.

But more importantly, because of the stability of the SY linker to KAHA ligation, the ligated product bearing the SY linker can be formed and isolated. Thus, after KAHA ligation, oxidation can be performed on the ligated product affording the protected α-ketoacid segment ready to use in further ligations. This feature allows sequential synthesis in the N to C direction (Scheme 2).

![Scheme 2.3. The different ligation options available with C-terminal peptide segments.](image)

**2.2.2 Synthesis of the cyanosulfur ylide linker 2.1**

The previous synthetic routes used by our group faced different problems that blocked us from scaling up this building block. As part of our scale up plans, we tried to improve the limiting steps and provide an optimized protocol that could be easily and safely executed on a multigram scale. Scheme 2.4 compares the previous synthetic routes and the one we have optimized. As depicted in the scheme, the α,β-unsaturated ester 2.21 was obtained – for both
syntheses – starting from tetrahydrothiophen-3-one through a Horner–Wadsworth–Emmons reaction in 97% yield.

Scheme 2.4. a) Previous synthetic route for the synthesis of the linker 2.1 b) Optimized route to linker 2.1.

Originally, alcohol 2.23 was obtained in two-step fashion from ester 2.21. The 1,4-conjugate reduction using metal-catalyzed hydrogenation appeared to be problematic; the presence of sulfide results in catalyst poisoning. To overcome this problem, a two-step process was applied, using a tedious procedure with in situ generated nickel boride, followed by LiAlH₄ reduction, providing alcohol 2.23 in low yield (< 60% over 2 steps).

Although the yield was quite poor, the main problem was the use of the nickel boride, which was extremely difficult and could not be scaled to more than a few grams. We wanted to overcome this difficulty and found that these two steps could be combined into a single one using LiAlH₄ under refluxing conditions. The conjugated ester 2.21 was treated with LiAlH₄ in refluxing THF for 3 days followed by workup under basic conditions (10 M NaOH solution) to directly obtain the alcohol 2.23 in 82% yield. Importantly, this reaction could be executed on a 50 gram scale, which was critical to successful scale-up.
The O-alkylation step was also one of the main limitations of the earlier route and never exceeded 40–50% yield. We decided to change the base normally used, NaOH (12 M in H2O), to a stronger one (KOH 12 M in H2O), and replaced the solvent (benzene) with n-hexanes. Additionally, we had noticed that side reactions occurred when the reaction was carried out at room temperature. We found that it was possible to suppress these undesired alkylations by maintaining the internal temperature at 0 °C. Strong magnetic stirring was a key criterion for affording the tert-butyl ester 2.24 in almost 80 % yield.

The tert-butyl group was previously removed using classic TFA conditions in CH2Cl2 and triisopropylsilane as a scavenger. This method provided surprisingly poor yields when performed on a gram scale. We switched to a different approach involving an acidic resin (Dowex-50) in water. This method happened to be extremely convenient and could be executed on a more than 10 g scale to furnish the sulfur ylide linker 2.1 cleanly in 95% yield. Using this new route, we prepared more than 50 grams of linker 2.1.

2.2.3 Alkylation of the linker on resin

With the synthetic problem solved, we focused our attention on improving the cyanometylation of the solid supported tetrahydrothiophene linker (Scheme 2.5). As mentioned before the linker 2.1 can be loaded to Rink amide MBHA or ChemMatrix® resins using a standard coupling protocol with HCTU. After coupling, the resin could be capped by treatment with 20% acetic anhydride and NMM in DMF for capping the eventual remaining free amine. Once the linker 2.1 is loaded on the solid support, the tetrahydrothiophene moiety was subjected to alkylation. Although the previous method used bromo-acetonitrile or iodo-acetonitrile in the presence of silver hexafluorophosphate (AgPF6), we noticed that silver tetrafluoroborate (AgBF4) in combination of iodo-acetonitrile led to the formation of a more stable salt.

The procedure is as follows: iodo-acetonitrile (4 equiv) and AgBF4 (4 equiv) in CH3CN are combined to generate the active alkylation agent. The resulting mixture is added to the linker resin 2.5 and allowed to react in a mixture of 2:1 toluene/CH3CN at 60 °C for 8 h (Scheme 2.5).

Scheme 2.5. Alkylation conditions of resin-linker 2.5 and loading of the first Fmoc amino acid.
CHAPTER 2. Synthesis and Optimization of the Key Building Blocks – Cyanosulfur Ylide, Boc- and Fmoc-Oxaproline – for KAHA Ligation

The sulfur ylide resin is washed intensively and the first amino acid (5 equiv) is coupled using standard reagents such as HCTU or HATU. After the first coupling, in order to remove all the remaining silver ions present, the resin is washed three times with sodium diethylthiocarbamate solution (0.5 M in DMF). The loading efficiency is measured by Fmoc-deprotection of the first amino acid that can be quantified by UV spectrophotometry. Usually, >90% loading was consistently observed for various amino acids (Arg, Ile, Leu, Phe, Tyr, Glu, Gln and Val).

2.2.4 Fmoc SPPS, resin cleavage and oxone oxidation

With resin 2.6 ready, the desired peptide can be synthesized using standard coupling reagents and protocols commonly used for Fmoc SPPS. Either manual or automated peptide synthesis can be performed without any deviation from the usual protocols. Segments around 50 amino acids long can be routinely synthesized without trouble arising from the cyanosulfur ylide linker.

Cleavage of the sulfur ylide peptides from the resin can be accomplished under standard conditions using 95:2.5:2.5 TFA/H₂O/TIPS or with scavengers; thiols and electron-rich aromatics can be used for example without issue. The cyanosulfur ylide is perfectly stable and neither protonation of the carbon— even in neat TFA— nor reduction of the ketone moiety — even when a large excess of silane reducing agent is employed— was observed.

After the cleavage, cyanosulfur ylide peptides are stable and can be stored for months. Purification by HPLC proceeds under standard conditions, the sulfur ylide moiety being extremely UV active at 214, 254 and 301 nm.

To oxidize the cyanosulfur ylide to obtain the corresponding α-ketoacid, the unprotected cyanosulfur ylide is treated with oxone (2 equiv of active oxygen) in 1:1 CH₃CN/H₂O containing 0.1% TFA for 5 to 10 min (Scheme 2.6). The reaction time depends on the nature of the first amino acid as well as the concentration. The later can change depending of the solubility of the peptide under the reaction conditions. HPLC analysis showed that prolonged reaction times led to the formation of undesired carboxylic acids. The reaction is quenched with a sulfur-based compound such as dimethylsulfide (150 equiv), and the volatiles can be removed with a nitrogen flux and/or reduced pressure. The resulting crude peptide can then be redissolved in 1:1 CH₃CN/H₂O and directly purified by reverse phase HPLC.
Additional reactions to the sulfur ylide peptide are oxidized with oxone salts. Oxidation afforded the corresponding α-ketoacid peptide segment.

Additionally, on large scale a lot of oxone salts tend to precipitate, and quick centrifugation of the mixture is recommended before HPLC purification.

This oxidation protocol can also be applied directly to crude sulfur ylide peptide up to 200 mg scale, thereby saving one purification step. To date, we have been able to prepare α-ketoacid peptides with arginine, isoleucine, leucine, phenylalanine, valine, tyrosine, and glutamic acid as the C-terminal residues.

2.2.5 Conclusions

We have developed an optimized synthesis of the sulfur ylide linker for preparing C-terminal peptide α-ketoacids by rapid oxidation of the resulting cyanosulfur ylide peptides prepared by Fmoc-SPPS. This synthesis is reliable and provides easy access to multigram quantities of the linker. This preparation has allowed us to prepare a large variety of peptide α-ketoacids that we have then used for KAHA ligation for the total chemical synthesis of proteins.

2.3 New Route for the Synthesis of (S)-N-Boc-5-Oxaproline and (S)-N-Fmoc-5-Oxaproline

2.3.1 Introduction

As described in Chapter 1, the other component of the KAHA ligation is the (S)-5-oxaproline. This building block is generally prepared either as Boc- ((S)-N-Boc-5-Opr, or Boc-Opr), Fmoc- ((S)-N-Fmoc-5-Opr, or Fmoc-Opr) or, more recently, photolabile protected peptides.
version. From a synthetic point of view, the original approach to the synthesis of Boc-Opr 2.2 and Fmoc-Opr 2.3 was based on a modified procedure of Vasella et al. and utilized a [3+2] cycloaddition with ethylene in a pressurized reactor (Scheme 2.7). However, this protocol requires a chiral auxiliary that needs to be prepared over several steps from the expensive L-gulonic acid-1,4-lactone 2.26 and only affords a 6:4 diastereomeric ratio after the cycloaddition reaction. After several recrystallizations, the two diastereoisomers could be separated but with relatively poor recovery. Above all, the requirement of a pressurized reactor in this procedure is a major impediment for preparing this building block in a practical way (Scheme 2.7).

**Scheme 2.7. Original synthesis of Boc- and Fmoc-oxaproline.**

In an effort to meet our own demand for oxaproline derivatives as well as initiate the use of KAHA ligations for the chemical synthesis of peptides or proteins, our goal was to develop a scalable and practical synthesis of the oxaproline building block.

In order to overcome these problems, we established a practical, scalable and economical route to the enantiopure oxaproline based on an enantioselective organocatalyst-mediated reaction (Scheme 2.8). An outsourcing company provided 1 kg of the desired Boc-Opr after using this new synthetic protocol.

**Scheme 2.8. New route towards the preparation of (S)-(N)-Boc-5-Oxaproline.**
2.3.2 Enantioselective organomediated route for the synthesis of (S)-N-Boc-5-Oxaproline

The first step of the synthesis was the asymmetric addition of silyl protected hydroxylamines to \( \alpha,\beta \)-unsaturated aldehydes in the presence of catalytic amounts of chiral catalyst. For this step, we were inspired by the extensive work of Cordova and MacMillan in the field of conjugate addition reactions involving \( N \)-centered nucleophiles and chiral amine catalysts (Scheme 2.9).\(^{91-94}\)

![Scheme 2.9. Conjugate addition of silyl protected hydroxylamines to \( \alpha,\beta \)-unsaturated aldehydes.]

We started by screening a range of diarylprolinol-derived catalysts for the conjugate addition between the \( N \)-Boc-protected hydroxylamine \( \text{2.30} \) and ethyl trans-4-oxo-2-butenoate \( \text{2.31} \). Reaction time, concentration and solvent were screened. In the initial catalyst screen we identified the TMS- and TBS-protected diphenylprolinols \( \text{2.33} \) and \( \text{2.34} \) as catalysts of choice, which gave the desired product in enantiomeric ratios of 96:4 and 98:2, respectively (Table 2.1, entries 1 and 2).

The synthesis of these two catalysts can be readily prepared on a more than 50 gram scale from the commercially available and affordable intermediate (S)-diphenyl(pyrrolidin-2-yl)methanol. Although catalyst \( \text{2.34} \) showed better selectivity, catalyst \( \text{2.33} \) was a good compromise between an acceptable enantiomeric ratio and better yield. We therefore decided to focus our efforts on the TMS-protected prolinol \( \text{2.33} \).

Although better yields were obtained using \( \text{2.33} \), it was still not sufficient to provide decent amounts of material for our scale-up purpose. To increase productivity, we started by screening different acid additives such as benzoic acid, acetic acid and dimethylhydrogen phosphate, but no improvement was observed. We tested stronger acids such as \( p \)-toluene sulfonic acid and 2,4-dinitrobenzenesulfonic acid as promoters but these additives led to a drop in the stereo-selectivity without increasing the yield.
After several tries, we found that using 50 mol% of the organocatalyst 2.33 improved
the addition yield to 48% (Table 2.1, entry 9). To our surprise, increasing the quantity of the
TMS-protected diphenylprolinol to 75% or to an equimolar ratio did not show any further yield
improvement. In no case did the reaction go to completion.

Table 2.1. Screening of catalysts for the asymmetric reaction.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Catalyst [mol%]</th>
<th>Solvent</th>
<th>Temperature</th>
<th>Yield[a]</th>
<th>ee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.33 [20 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>35%</td>
<td>96:4</td>
</tr>
<tr>
<td>2</td>
<td>2.34 [20 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>25%</td>
<td>98:2</td>
</tr>
<tr>
<td>3</td>
<td>2.35 [20 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>4</td>
<td>2.36 [20 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>5</td>
<td>2.33 [20 mol%]</td>
<td>1,2-dichloroethane</td>
<td>4 °C to rt</td>
<td>30%</td>
<td>95:5</td>
</tr>
<tr>
<td>6</td>
<td>2.33 [20 mol%]</td>
<td>CHCl₃</td>
<td>4 °C[c]</td>
<td>30%</td>
<td>97:3</td>
</tr>
<tr>
<td>7</td>
<td>2.33 [20 mol%]</td>
<td>CHCl₃</td>
<td>-20 °C</td>
<td>20%</td>
<td>98:2</td>
</tr>
<tr>
<td>8[d]</td>
<td>2.33 [20 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>20%</td>
<td>n.d.</td>
</tr>
<tr>
<td>9</td>
<td>2.33 [50 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>48%</td>
<td>96:4</td>
</tr>
<tr>
<td>10</td>
<td>2.33 [75 mol%]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>50%</td>
<td>n.d.</td>
</tr>
<tr>
<td>11</td>
<td>2.33 [1 equiv]</td>
<td>CHCl₃</td>
<td>4 °C to rt</td>
<td>48%</td>
<td>n.d.</td>
</tr>
<tr>
<td>12</td>
<td>2.33 [20 mol%]</td>
<td>CHCl₃[e]</td>
<td>4 °C to rt</td>
<td>35%</td>
<td>96:4</td>
</tr>
</tbody>
</table>

[a] Isolated yield; [b] n.d. stands for not determined. Product could not be obtained
pure; lower enantioselectivity than with previous catalysts; [c] Time of reaction 36 h;
[d] Reaction was conducted with 1.0 equiv of 2.30 and 3.0 equiv of 2.31; [e] 0.6 M
substrate concentration. The best condition is highlighted.

Even though the yield is still suboptimal and the reaction not really catalytic, almost 50%
yield at this stage of the synthesis was acceptable for us. We then proceeded to the next
step of the synthesis of Boc-Opr.

After isolation of the β-amino aldehyde 2.32, reduction with sodium borohydride
provided alcohol 2.37 cleanly and in good yield (Scheme 2.10). We attempted to cyclize this
alcohol in one pot using nonafluorobutanesulfonyl fluoride (Nff) and DBU as base. Treatment
of alcohol with Nff (1.1 equiv) and DBU (3.2 equiv) in a mixture of (1:2) CH₂Cl₂:pentane at 0
°C for 1 h afforded partial conversion to the desired cyclic product 2.29. However due to the
high basicity of DBU, the alcohol 2.37 underwent transesterification and side product 2.38 was
formed together with the desired \( \text{2.29} \) in a 1:1 ratio. Using a milder base such as pyridine and performing the reaction at lower temperature like \(-15\) °C did not show any improvement and lowered the yield of the desired product.

**Scheme 2.10. Design of a new strategy towards the synthesis of \((S)-N\)-Boc-5-Oxaproline.**

We therefore considered a two-step approach for the preparation of cyclic ester \( \text{2.29} \). Alcohol \( \text{2.37} \) was first mesylated, affording compound \( \text{2.41} \) in complete conversion. Without any purification, the crude mesylate \( \text{2.39} \) was treated with TBAF (1.5 equiv, 1 M in THF) at low substrate concentration (0.05 M) at 4 °C for 2 h, affording the desired cyclic compound \( \text{2.29} \) as a single product in 64% yield from aldehyde \( \text{2.32} \) (over three steps) (Scheme 2.11).

Hydrolysis of the ethyl ester, as described previously, afforded \((S)-N\)-Boc-5-Opr \( \text{2.1} \) in 94% ee and good overall yield (30% over five steps).

**Scheme 2.11. Synthesis of \((S)-N\)-Boc-5-oxaproline from alcohol \( \text{2.37} \).**

The new method that we have developed overcomes all the problems we had with the previous route, allowing us to easily scale up, at a reasonable cost, the synthesis of Boc-Opr \( \text{2.2} \).

More importantly the synthetic route has been used by a chemical company to prepare 1 kg of Boc-Opr. We hope that this optimized procedure will allow us to commercialize this building block soon.

### 2.3.3 Synthesis of \((S)-N\)-Fmoc-5-Oxaproline from \((S)-N\)-Boc-5-Oxaproline

Now that we have ready access to Boc-Opr \( \text{2.2} \) we tried to see if it would be possible to directly convert Boc-Opr into protected Opr with a protecting group of our choice. Our first attempt involved deprotection of the Boc group by treatment with TFA in \( \text{CH}_2\text{Cl}_2 \) for 1 h. The
solvent and TFA were co-evaporated with toluene several times in order to remove the remaining TFA affording oxaproline 2.40 (Scheme 2.12).


Without isolation, oxaproline 2.40 was dissolved in either a mixture of acetonitrile/water or dioxane/water containing sodium carbonate and Fmoc succinimide was added. The reaction was stirred overnight at room temperature and afforded the desired Fmoc-Opr 2.3 in approximately 70% yield after purification.

2.3.4 Conclusions

We have developed a scalable, affordable and practical synthesis of the Boc- and Fmoc-oxaproline building blocks using an improved enantioselective organocatalyst-mediated reaction. The synthesis of Boc-Opr has been performed on a kilogram scale, affording 1 kg of pure Boc-Opr. The ability to produce our KAHA ligation monomers on such scale, coupled with our desire to make them commercially available soon, can be expected to encourage the use of KAHA ligation as an alternative approach for the chemical synthesis of peptides or proteins.
2.4 References


CHAPTER 3

CHEMICAL SYNTHESIS OF THE PROTEIN BETATROPHIN

The work in this Chapter was performed in equal collaboration with Claudia Murar.
CHAPTER 3. Chemical Synthesis of the Betatrophin Protein

In this chapter, the application of KAHA ligation to the chemical synthesis of betatrophin is described. Betatrophin is a protein hormone that is mainly synthesized and secreted by liver and adipose tissue. It was believed to have a direct impact on β-cell proliferation and thus be a potential target for new diabetes therapy. We executed a chemical synthesis of this protein that afforded multimilligram quantities of betatrophin.

3.1 Introduction

Diabetes is a group of metabolic diseases in which the patient presents high blood sugar levels over a long period of time. If left untreated, it can give rise to many complications such as heart disease, stroke, chronic kidney failure, foot ulcers, damage to the eyes or even death. In 2015 more than 400 million people were diagnosed with diabetes worldwide, which represents around 8.3% of the population. It is projected that, by 2040, this number will grow to 640 million.  

Although there are three types of diabetes, namely type 1, type 2 and gestational diabetes, about 90% of clinical cases are type 2 diabetes (T2D). This type of diabetes, mainly caused by obesity and lack of exercise, results in either reduced insulin secretion and/or insulin resistance, both leading to an improper blood glucose metabolism. In order to reduce the glucose level in blood resulting from insulin resistance, β-cells intensify their production of insulin until they reach a threshold. At this point, β-cells can no longer cope with the insulin demand. As nutrient excess persists, the excess of sugars and lipids negatively impacts the functioning of β-cells, lowering insulin secretion, reducing insulin gene expression and ultimately causing apoptosis.  

Although there are several drugs that either control plasma glucose levels or increase insulin production, reaching a full glycemic control similar to healthy β-cells is very difficult to achieve.  

As a result, many studies have been dedicated to identifying new and novel mechanisms to induce β-cell proliferation, which is valued as the ultimate treatment for both type 1 and type 2 diabetes. Therefore, numerous strategies and alternative cell sources have been utilized to mimic β-cell function.
3.1.1 The betatrophin protein as a new inducer for β cell proliferation

In 2013, Melton and coworkers described an inducer of β-cell proliferation that is secreted in liver and adipose tissues. This protein, already known as ANGPTL8, was renamed betatrophin. In their report, they showed that betatrophin increased β-cell replication and β-cell mass in insulin-resistant mouse model. Over-expression of betatrophin resulted in a 17-fold increase in β-cell proliferation and a threefold increase in β-cell mass. As a result, mice overexpressing betatrophin had improved glucose tolerance and lowered fasting blood glucose.\textsuperscript{106}

The identification of betatrophin as a novel inducer of β-cell proliferation has attracted remarkable attention from the scientific community as well as the media. It was welcomed as a next-generation drug for diabetes and was described as a potential alternative to insulin injections.\textsuperscript{106}

Given all the excitement about a protein that could become a potential new therapy target, our group decided to pursue a chemical synthesis of betatrophin. As mentioned in the introduction chapter, chemical synthesis of proteins allows the addition of posttranslational modifications, fluorescent tags or biotin to proteins of interest.

3.2 Chemical synthesis of betatrophin

Betatrophin consists of 177 amino acid residues, placing it in the accessible range for chemical synthesis. Surprisingly, of these 177 amino acids, none is a cysteine residue, making it ideally suited for our KAHA ligation method. Our objective was to establish an efficient milligram-scale chemical synthesis of betatrophin accessible to any researcher. This would allow other researchers to easily resynthesize the protein with desired tags or unnatural amino acids and ultimately provide significant insight into understanding not only its function and activities but also help identify its receptor.

3.2.1 Retrosynthesis, disconnection and assembling strategies towards betatrophin

The first step towards the synthesis of betatrophin was an analysis of the amino acid sequence and the structure of betatrophin. In order to bring down the size of the individual segments to around 30–40 amino acid residues (readily accessible by SPPS), it was clear that a four ligation-five segment strategy would be required. Recent examples of protein synthesis from 5 segments are not common but can be found in the literature.\textsuperscript{76,107,108} However, at the time when we decided to chemically synthesize betatrophin, the KAHA ligation had never been
used in a synthesis involving more than 3 peptide segments nor for a protein longer than 100 amino acid residues.

We challenged our approach and selected four ligation sites at position Leu60-Thr61, Leu96-Glu97, Leu133-Glu134 and Leu166-Thr167 which would introduce four mutations if the 5-oxaproline (Opr) was used for the ligation: T61T⁵, E97T⁶, E134T⁷ and T167T⁸. Methionine residues are well tolerated in KAHA ligations, but in order to avoid oxidation while handling, storage and refolding, we chose to replace the three methionine residues (Met 24, Met 100, Met 177) by norleucine residues.¹⁰⁹⁻¹¹² As described in Section 2.2.1, leucine, phenylalanine, valine and tyrosine α-ketoacids are effective ligation partners for 5-oxaproline. For betatrophin, we chose leucine as the C-terminal residue of the α-ketoacids due to its high abundance and our experience in using this residue at a ligation site.

With the different peptide segments set, we thought about ways to assemble them. To gain optimal convergence we envisaged two different strategies. Fortunately, the flexibility of the KAHA ligation to grow the peptide toward either the C or the N terminus gave us some freedom with regard to the assembling strategy. For the first strategy (Scheme 3.1), we decided to go for a convergent synthesis with a final ligation between the ligated ketoacid segment H₂N-S1S2-KA 3.6 and the oxaprole segment Opr-S3S4S5-OH 3.8.
To follow this strategy, segment 2 3.2 would have to have a masked \( \alpha \)-ketoacid at the C terminus by employing the sulfur ylide linker that would readily give the corresponding \( \alpha \)-ketoacid \( \text{H}_2\text{N-S1S2-KA 3.6} \) upon oxone oxidation. The other protein segment, \( \text{Opr-S3S4S5-OH 3.8} \), could also be assembled via KAHA ligation, starting with the ligation between the oxaproline segment \( \text{Opr-S5-OH 3.5} \) and the Fmoc protected oxaproline ketoacid segment \( \text{FmocOpr-S4-KA 3.4} \). The ligated product could be deprotected and ligated to the Fmoc protected oxaproline ketoacid segment \( \text{FmocOpr-S3-KA 3.3} \), affording after another Fmoc deprotection \( \text{Opr-S3S4S5-OH 3.8} \). The final ligation between \( \text{H}_2\text{N-S1S2-KA 3.6} \) and \( \text{Opr-S3S4S5-OH 3.8} \) would generate the betatrophin protein 3.9.

We also envisaged a second strategy with a final ligation between the ligated ketoacid segment \( \text{H}_2\text{N-S1S2S3-KA 3.10} \) and the oxaproline segment \( \text{Opr-S3S4S5-OH 3.7} \). This would allow us to have common segments between the two strategies, saving time, and yet potentially having a backup plan in case of a ligation problem.

In this strategy, \( \text{H}_2\text{N-S1-KA 3.1} \) and \( \text{Opr-S2-SY 3.2} \) would be assembled in a similar fashion to get \( \text{H}_2\text{N-S1S2-KA} \), but the elongation would continue with the ligation of \( \text{Opr-S3-SY 3.10} \) affording upon oxone oxidation step the 100 amino acid long \( \alpha \)-ketoacid segment \( \text{H}_2\text{N-} \).
S1S2S3-KA 3.11. Combining this ketoacid segment with the ligated product Opr-S4S5-OH 3.7 would afford the betatrophin protein 3.9.

### 3.2.2 Segment synthesis towards betatrophin

Over the years, our group has designed efficient protocols for the synthesis of peptides by Fmoc SPPS with the different monomers developed by the group. Following the method described in the introduction, we were able to synthesize all 6 segments on a hundred-milligram scale. The segment syntheses are outlined in Scheme 3.3; details and characterization data can be found in the experimental section (Chapter 6).

Using a common batch of leucine-protected α-ketoacid resin, H2N-S1-KA 3.1, FmocOpr-S3-KA 3.3 and FmocOpr-S4-KA 3.4 could be synthesized by automated Fmoc-SPPS. Fmoc oxaproline could then be coupled manually using HATU as coupling reagent, and TFA cleavage furnished – in all three cases – the desired protein segment without any trouble (Scheme 3.3.a). In a similar fashion, using a common batch of the sulfur ylide linker resin preloaded with a leucine residue, Opr-S2-SY 3.2 and Opr-S3-SY 3.10 were prepared via

---

**Scheme 3.3. Synthesis of betatrophin segments from:**

a) Protected leucine α-ketoacid resin

b) Leucine-sulfurylide resin

c) Wang polystyrene resin
standard automated Fmoc SPPS. Boc oxaproline was manually coupled with HATU, and TFA cleavage afforded both segments in good yield and good purity (Scheme 3.3.b).

Finally, the C-terminal segment 5 Opr-S5-OH 3.5 was readily synthesized by SPPS on Wang polystyrene resin to cleanly afford the desired peptide after resin cleavage (Scheme 3.3.c). All six peptide segments could be synthesized and purified on a scale between 70–350 mg.

3.2.3 Assembly of betatrophin by KAHA ligation of the peptide segments

With all required protein segments in hand, we proceeded to assemble betatrophin by sequential KAHA ligations. In the first ligation, H2N-S1-KA 3.1 and Opr-S2-SY 3.2 were ligated at a protein segment concentration of 20 mM (on 106 mg scale) in a mixture of 8:2 DMSO/H2O with 0.1 M oxalic acid at 60 °C (Scheme 3.4). The ligation proceeded smoothly with complete consumption of the limiting starting material in 8 h. The crude mixture containing the ligated sulfur ylide was dissolved in 1:1 CH3CN/H2O with 0.1% TFA to a final concentration of 7 mM and subjected to our standard sulfur ylide oxidation conditions to deliver ligated H2N-S1S2-KA 3.6 in an overall yield of 43% (92 mg of purified peptide). We optimized the one-pot conditions for these particular segments. However our experience suggests that the optimal

![Scheme 3.4. KAHA ligation with peptide Opr segments 3.2 and α-ketoacid 3.1. HPLC monitoring of the ligation at 60 °C. MALDI-MS spectrum of purified 3.6.](image)

concentrations and times for the sulfur ylide oxidation may differ. In order to increase the overall efficiency of the synthesis and improve the solubility of 3.6 we decided to proceed with the ester bond at the ligation site and perform a global O to N acyl shift after the final ligation.

Segment 4 FmocOpr-S4-KA 3.4 and segment 5 Opr-S5-OH 3.5 were coupled under standard KAHA conditions at 120 mg scale to furnish the Fmoc-protected depsi-peptide
intermediate in 8 h. The ligation mixture was directly diluted in a mixture of 2:2:1 NH$_4$OH/H$_2$O/DMSO (v/v/v, pH 13.5) to a final concentration of 2 mM and subjected to a one pot *depsi*-amide rearrangement and Fmoc-deprotection for 1 h (Scheme 3.5). The desired product Opr-S4S5-OH 3.7 was isolated in 52% yield (126 mg of purified peptide). The one pot KAHA ligation and Fmoc-deprotection with 5% Et$_2$NH in DMSO worked smoothly and was extensively used in our group. On the other hand, rearrangement of the *depsi*-peptide to the amide peptide requires basic aqueous buffers. In this case the ligated mixture was not soluble in sodium bicarbonate (pH 9.5) or ammonium buffers. To circumvent the solubility issue, we adjusted the ammonium hydroxide buffer with DMSO to a final pH of 13.5. These conditions were optimal for the one pot KAHA ligation, Fmoc-deprotection and rearrangement for the betatrophin segments, but depending on the solubility properties of a peptide sequence, this reaction should be first performed stepwise and carefully monitored by RP-HPLC.

With Opr-S4S5-OH 3.7 in hand, we proceeded with the ligation of segment 3 FmocOpr-S3-KA 3.3 to afford the Fmoc-protected *depsi*-peptide, which was readily deprotected and rearranged *via* the previously established protocol (Scheme 3.6). The ligation was performed at a concentration of 15 mM at 60 °C for 15 h on a 40 mg scale and we obtained the ligated product Opr-S3S4S5-OH 3.8 in 50% isolated yield (30.8 mg of purified peptide).

In parallel with this strategy, we also attempted the ligation of H$_2$N-S1S2-KA 3.6 and Opr-S3-SY 3.10 at 15 mM on a 5 mg scale; the desired product would serve us for the second strategy (see Scheme 3.2 above). The ligated product was, however, obtained in modest yield (<30 %) and we noticed that upon dissolution, the protein segments instantly formed a gel, which most likely affected the mixing efficiency and conversion of the reaction. Dilution of the
ligation mixture to 10 mM did not improve the outcome of the reaction. We assume that gel formation was due to the hydrophobicity of the N-terminal segment 1 H$_2$N-S1-KA 3.1. In addition, the poor solubility of the isolated product H$_2$N-S1S2-KA 3.6 would pose difficulties for the final ligation. In order to enable a high yielding, efficient synthesis of betatrophin protein, we therefore decided to focus on the first convergent strategy.

With H$_2$N-S1S2-KA 3.6 and Opr-S3S4S5-OH 3.8 in hand, we proceeded to the final ligation. This ligation was performed on a 10.0 mg scale and under more dilute conditions (10 mM) in order to ascertain the solubilization of both ligation partners. Additionally, with a view to favor the consumption of Opr-S3S4S5-OH 3.8 and enhance the conversion of the reaction we added 1.5 equiv of H$_2$N-S1S2-KA 3.6. After 24 h, the desired ligated depsi-peptide was diluted in a mixture of 2:2:1 NH$_2$OH/H$_2$O/DMSO (v/v/v, pH 13.5) to a final concentration of 1 mM and subjected to the final depsi-amide rearrangement to afford the synthetic protein H$_2$N-S1S2S3S4S5-OH 3.9 in 27% yield (>4 mg final protein). The purity and identity of 3.9 was confirmed by analytical RP-HPLC, ESI-HRMS and SDS-PAGE (Scheme 3.7).

The protein was folded in 10 mM acetate buffer, pH 4.0 by gradually decreasing the GnHCl concentration in the PBS buffer by dialysis (from 6 M to 2 M). The circular dichroism (CD) spectrum of betatrophin 3.9 shows two minima at 223 and 208 nm characteristic for an $\alpha$-helix conformation. Although a computational structure of ANGPTL8 has been proposed by Siddiq et al.,$^{113}$ no CD spectra of ANGPTL8 has been reported.

The chemical synthesis of the 177-mer was repeated several times and we were pleased to find reproducible yields and similar amount of recovered protein.
3.3 The fall of betatrophin

Soon after we successfully synthesized betatrophin – in less than 9 months using KAHA ligation – many researchers began to question the initial results on the effect of betatrophin on β-cell proliferation.¹¹⁴ First, different groups showed that betatrophin did not affect glucose and insulin tolerance and did not show significant changes in glucose concentration regulation.¹¹⁵ And more importantly, they showed that the increased betatrophin expression was not capable of inducing β-cell proliferation, increasing β-cell mass or improve glucose homeostasis.¹¹⁶-¹¹⁸ In 2017, the original paper of Melton and coworkers was retracted, ending the debate about the possible action of betatrophin on β-cell proliferation and β-cell mass.

In a preliminary study in collaboration with the Wolfrum group at ETH, we could never assess the bioactivity of the synthetic betatrophin protein, as there is no cell based assay known and we were not able to establish a good positive control. As little is known about betatrophin, we hope that our synthetic protein will be a useful tool for further investigations into its biological function.
3.4 Conclusions and outlook

We successfully established a facile protocol for the synthesis of betatrophin, resulting in more than 5 mg of product. Importantly, the one-pot conditions and ligation procedures were optimized to an easy and straightforward system that is simple to implement by people outside the field of chemistry. Our protocol can be applied to other targets independent of scale and we have conducted ligations with >300 mg of peptide segment. In order to get optimal results for the one pot reactions, a small-scale test is highly recommended as concentration, reaction time and depsi-amide rearrangement buffer may need to be adjusted. We hope that this robust protocol will encourage the research community to use KAHA ligation as an alternative to NCL for the chemical synthesis of proteins.

To our delight, after our data were published, several other groups showed high interest in testing our synthetic protein.
3.5 References


CHAPTER 3. Chemical Synthesis of the Betatrophin Protein


CHAPTER 4
CHEMICAL SYNTHESIS OF THE ANTIVIRAL INTEGRAL
MEMBRANE PROTEIN IFITM3
CHAPTER 4. Chemical Synthesis of the Antiviral Integral Membrane Protein IFITM3

4.1 Introduction

This Chapter describes the synthesis of the highly challenging membrane-associated protein, Interferon-Induced Transmembrane protein 3 (IFITM3), using the \( \alpha \)-ketoacid–hydroxylamine ligation. IFITM3 has been shown to play a crucial role in inhibiting the replication of a large --and still expanding-- range of viruses: influenza A, dengue, Ebola, Marburg and Zika viruses. The mechanism of action and structural topology remains unknown and is a subject of considerable interest and debate.

Although the interest in this protein is high and continues to increase, IFITM3 is currently not accessible by recombinant approaches, which limits greatly its biological and biophysical study. We have developed a chemical synthesis that allows access to this protein and its prominent isoforms, including phosphorylated and palmitoylated variants. The site-specific attachment of a fluorescent dye enables the production of protein for biophysical studies.

4.1.1 Generalities about membrane proteins

Membranes of cells are composed of a phospholipid double layer, which is decorated with a wide range of proteins and protein complexes. These proteins are responsible for a variety of functions such as transporting ions and molecules through the membrane, exposing markers to the surface of the cell, or processing signals from external and internal stimuli. These functions make such proteins really important biomolecules for understanding a large variety of biological events.

All membrane proteins can be classified into two categories: integral membrane proteins and peripheral membrane proteins. They are distinguished by their ability to be extracted from the lipid bilayer. Integral membrane proteins are embedded in the lipid bilayer and can be removed only by disrupting the membrane with detergents. Peripheral membrane proteins can be removed from the membrane without dissolving the bilayer. Most frequently, these peripheral proteins are removed by shifting the ionic strength or the pH of the aqueous solution to disrupt the ionic interactions of the peripheral protein with either phospholipid polar head groups or other membrane proteins.

- **Integral membrane proteins** (also known as *intrinsic proteins*) are proteins that have one or more domains integrated into the phospholipid bilayer. These proteins present regions rich in hydrophobic amino acid rich that can interact with the hydrophobic tails of the lipids, thus
anchoring the protein to the membrane. We can distinguish different types of integral membrane proteins (Figure 4.1).

The most common are the transmembrane proteins. These proteins span the entire phospholipid bilayer and can have one or several membrane spanning domains together with hydrophilic amino acid chains in either the extracellular or intracellular region. We find two types of these proteins depending of the nature of their membrane domain. The latter can be composed either of $\alpha$-helices (Figure 4.1.a-b) or multiple $\beta$-strands (Figure 4.1.c). Another type of membrane protein is also strongly embedded in the phospholipid layer but does not pass through the membrane (Figure 4.1.d). Their C and N termini can both be outside or inside the membrane. The last type of integral membrane protein refers to proteins that are anchored to one of the membrane sides by covalently binding to fatty acids or phospholipids. A famous example is the glycosylphosphatidylinositol-anchored proteins, commonly known as GPI-anchored proteins. The fatty acid to which the protein is attached inserts itself in the membrane, leaving the protein totally outside the bilayer.

Peripheral membrane proteins (also known as extrinsic proteins) are proteins that, instead of being embedded into the hydrophobic core of the lipid bilayer, interact with the membrane surface (Figure 4.2). There are two main pathways for peripheral membrane proteins to interact with the membrane: either by interacting with the polar phospholipid head groups via complexed cations (Figure 4.2.a), or by protein-protein interactions with integral membrane proteins (Figure 4.2.b). They can localize to either the cytosolic or the extracellular face of the plasma membrane.
### 4.1.2 A specific class of integral membrane protein: the transmembrane proteins

As briefly mentioned in the previous section, transmembrane proteins are integral membrane proteins that span the entire membrane. Two types of transmembrane proteins can be distinguished depending on the tertiary structure they adopt inside the lipid bilayer: α-helices and β-barrels.

α-Helical transmembrane proteins are the more common of the two and can be found in the membrane of both prokaryotes and eukaryotes. Voltage-gated ion channels, such as K⁺ and Cl⁻ channels, are good examples of this type of transmembrane protein (Figure 4.3.a).

The β-barrel transmembrane proteins are present in the outer membranes of Gram-negative bacteria, cell walls of Gram-positive bacteria, and outer membrane of mitochondria and chloroplasts. Good examples of this kind of transmembrane protein are the Porin proteins that act as pores through which molecules can diffuse (Figure 4.3.b).
4.1.3 Production and purification of integral membrane proteins, a challenging topic

Due to their vital roles in many essential cell functions, membrane proteins are among the most interesting and important biomolecules. However, they remain under studied. Although it is estimated that 30% of all genes are encode for integral membrane proteins, less than 1% of the structures deposited at the Protein Data Bank are membrane proteins. This lack of data mainly comes from their highly hydrophobic nature as well as their intricate subunit structure and difficulties associated with obtaining three-dimensional crystals suitable for X-ray analysis. Moreover, detergents are required for solubilizing, and more importantly, purifying membrane proteins. Yet, complexes formed by detergent-solubilized membrane proteins are often unstable and have a strong tendency to aggregate. Also, it is hard to predict whether they are in their native conformation.

These limitations arise from three main difficulties:

- A primary difficulty encountered in the study of membrane proteins is to obtain the protein of interest in a decent amount and purity. Membrane proteins are usually present at low concentration in membranes, and the protein of interest is, in most cases, not the only membrane protein present in the membrane. However, it does happen in certain cases that membrane proteins can be isolated by directly extracting them from the lipid bilayer. Some examples include the pacteriorhodopsin protein, which is the only protein present in the membrane of Halobacteria salinaria, and the protein band 3, which is the predominant membrane transporter in red blood cells. Unfortunately, in most cases, membrane proteins cannot be easily obtained in sufficient amounts directly from the corresponding membrane. In order to bypass this problem, they are often overexpressed. However, here again, a problem arises. At high concentration, these proteins aggregate in the cytoplasm, killing the cells and providing only low yields of the protein of interest. Also, posttranslational modifications that are often required for biological function cannot be obtained.

- A second difficulty is that membrane proteins are, by definition, inside the core of the phospholipid bilayer, which in even the simplest organism is a complex, heterogeneous and dynamic environment. In their native state, membrane proteins cannot be analyzed by classic biophysical methods like NMR, X-ray crystallography, circular dichroism, nano-discs or ligand-binding studies. In order to perform such analyses, the proteins need to be extracted from the membrane and the analysis has to be conducted in a detergent or an artificial lipid environment. This leads to complicated sample preparation and non-negligible background noise from the lipid environment.

- Finally, membrane proteins suffer extremely poor solubility in aqueous solutions and require the use of complex systems that satisfy their high hydrophobicity. For in vitro studies, the
reconstitution of the proteins is critical. However, it is unfortunately non-trivial in the case of membrane proteins.122

4.2 IFITM3: A transmembrane antiviral protein

4.2.1 Introduction to IFITM3

Among all the different functions that membrane proteins can perform, they may also exhibit antiviral activity. This is the case for IFITM3, a 133 amino acids long transmembrane protein that has a strong antiviral activity against a large range of viruses. IFITM3 restricts cellular infection by a variety of important viruses in humans and animals.123-135 IFITM3 is composed of two hydrophobic segments, which are expected to be embedded in the lipid bilayer, linked by a hydrophilic loop (Figure 4.4).

Numerous residues in IFITM3 contain post-translational modifications (PTMs). These PTMs include palmitoylation of cysteines 71, 72 and 105;124 methylation of lysine 88; ubiquitination of lysines 24, 83, 88 and 104;136,137 and phosphorylation of tyrosine 20.138,139 (Figure 4.4).

4.2.2 Mode of action of IFITM3

The different viruses that are inhibited by IFITM3 – influenza virus, SARS coronavirus, dengue virus and West Nile virus, among many others – present a common trait: all of them enter cells via endocytosis and utilize pH-dependent fusion mechanisms.140,141 In contrast, IFITM3 showed poor restriction when cells were infected with viruses like Sendai virus and Moloney leukemia virus, which fuse at the cell surface.123,142 Interestingly, treatment of SARS coronavirus (which is normally inhibited by IFITM3) with trypsin allowed the virus to fuse at the cell surface, resulting in virus evasion of IFITM3-mediated restriction.126
Furthermore, different experiments have shown that IFITM3 co-localizes with different markers that are found on endosomes and lysosomes.\textsuperscript{126,136,143,144} Influenza virus particles could be visualized entering cells expressing IFITM3 but were eliminated by degradation prior to their fusion/escape from IFITM3-positive endolysosomes.\textsuperscript{143,145} Work form Hang and coworkers using IFITM3 lacking ubiquitination sites have suggested that upon IFITM3 expression, endosomes and acidic lysosomes were aberrantly merged, providing a compartment for viral degradation.\textsuperscript{136}

These experimental data have suggested a model where viruses are degraded in acidic endolysosomes in cells expressing IFITM3 (Figure 4.5). Although the mechanism by which IFITM3 is able to block this fusion remains unknown, it is now quite certain that IFITM3 affects the rigidity of the endosomal membrane, blocking the fusion of the virus. The fluidity of the membrane has been known for some time to be important for efficient fusion.\textsuperscript{146}

![Figure 4.5. General mechanism of endosome fusing viral entry without and with IFITM3.](image)

To explain this phenomena, two concepts have emerged: first, researchers have proved that IFITM3-containing endolysosomes contain increased levels of cholesterol – thus affecting the fluidity of the membrane – compared to cells where IFITM3 was absent.\textsuperscript{147,148} This change in the membrane composition was explained by the ability of IFITM3 to interact with the VAPA protein and thus disrupt an interaction between VAPA and a cholesterol trafficking protein.
However, recent publications have contested this theory, showing that modulation of the cholesterol level had no significant effect on IFITM3 restriction. Overexpression of VAPA did not show a greater effect either. In a similar fashion, influenza A infectivity was tested in Niemann–Pick type C1 fibroblasts – cells that contain high levels of cholesterol in the late endosomal compartment – but showed only little difference when compared with control cells.

A second theory has been proposed where IFITM3, owing to its biophysical properties, has a direct impact on the membrane fluidity of the endosome. If this mechanism is confirmed, it would make IFITM3 the first cellular antiviral effector that limits infection by directly altering the ‘rigidity’ of cell membranes.

Despite all the research on IFITM3, a clear mechanism explaining how IFITM3 is able to block viral fusion remains to be determined.

### 4.2.3 Posttranslational modifications involved in IFITM3 bioactivity

As mentioned earlier, IFITM3 is highly posttranslationally modified, and these modifications play a crucial role for its bioactivity (Figure 4.6). In the following section these PTMs and their effect on IFITM3 are described one by one.

- **Palmitoylation of Cys 71, 72 and 105:**
  
  IFITM3 is palmitoylated on all its cysteine residues (71, 72 and 105), and palmitoylation-deficient versions of IFITM3 (where cysteine residues were converted to alanine) showed a more diffuse pattern of localization than wild type IFITM3, indicating that S-palmitoylation promotes IFITM3 clustering. Not only does this PTM change the pattern of localization, palmitoylation deficiency also decreases the anti-influenza virus activity of IFITM3. Interestingly, when a mutant IFITM3, lacking palmitoylation sites, was modified with lipids at the N-terminus (by myristoylation) or at the C-terminus (by prenylation) showed antiviral activity again. These results suggest that the palmitoylation moiety probably does not change the conformation of IFITM3, but rather helps IFITM3 to accumulate in membranes. So far, the palmitoyltransferase(s) responsible for modifying IFITM3 have not been identified.
Phosphorylation of Tyr 20:

IFITM3 has a 4 amino acids sequence (YEML) that is recognized by the endocytic machinery and is essential for IFITM3 trafficking to endolysosomes.\textsuperscript{138,139} When this motif is mutated, IFITM3 loses its antiviral activity and accumulates in the plasma membrane.\textsuperscript{138,139,152,153} When IFITM3 is phosphorylated on Tyr 20 – by the protein kinase Fyn\textsuperscript{138,139} – a loss of antiviral activity has been observed. This result suggests that phosphorylation blocks the recognition motif from interacting with the endosomal machinery, thus acting as a negative regulator of activity.

Interestingly, during the H1N1 pandemic of 2009, human cases associated with severe infection presented a variant of IFITM3 in which the first 21 amino acids – including the YEML motif – were absent.\textsuperscript{154} These severe cases could be explained, at least in part, by the loss of this YEML motif, impairing proper trafficking of IFITM3 in the endosomes.

It has also been shown that both phosphorylated-IFITM3 (obtained by overexpressing Fyn) and Tyr 20 mutant of IFITM3 were significantly less ubiquitinated than native IFITM3, suggesting that Tyr 20 might also be a critical component of a ubiquitin ligase recognition motif.\textsuperscript{139}

Studies of IFITM3 phosphorylation have revealed important details regarding IFITM3 localization and trafficking. They have provided indications that IFITM3 is localized in the plasma membrane before trafficking to endolysosomes.\textsuperscript{139,153} In order to be taken into the endocytosis pathway, the phosphorylated group needs to be removed by a
phosphatase. This hypothesis correlates with the observation that tyrosine phosphorylation of IFITM3 is difficult to detect without pretreatment of cells with protein tyrosine phosphatase inhibitors.\textsuperscript{138,139} The precise nature of the phosphatases involved in regulating Fyn and IFITM3 is unknown, but would aid in understanding the sequence of events involved in phosphorylation and dephosphorylation of IFITM3, as well as the physiological triggers for addition and removal of this PTM.

- **Ubiquitination of Lys 24, 83, 88, 104:**

  Mono and di-ubiquitinated IFITM3 have been observed by both western blotting and MS/MS analysis after trypsin digestion.\textsuperscript{136} These analyses also revealed that IFITM3 is polyubiquitinated – mainly at Lys 24 – with polyubiquitins branched both at Lys 48 and at Lys 63.\textsuperscript{136} These polyubiquitin linkages, which typically control protein turnover or trafficking, respectively,\textsuperscript{155} suggest that ubiquitination of IFITM3 has multiple effects.

  IFITM3 ubiquitination can occur at four different lysines, though Lys 24 appeared to be the most frequently modified residue.\textsuperscript{136} However, when Lys 24 was mutated to alanine, no change in antiviral activity, localization, or abundance of IFITM3 was observed. This suggests that the other lysines may functionally compensate for ubiquitination of this residue.\textsuperscript{136,138,152} When all the lysine residues were converted to alanine, IFITM3 showed an increase in antiviral activity and appeared to localize entirely in endolysosomes.\textsuperscript{136}

  The specific roles of the different ubiquitinated IFITM3 variants remain to be understood, and the ubiquitin ligase(s) responsible have yet to be identified. So far, based on available data, it appears that ubiquitination reduces IFITM3’s antiviral activity. Thus, identifying these enzymes could potentially be a target for improving IFITM3 antiviral activity.

- **N-Methylation of Lys 88:**

  In addition to being ubiquitinated, Lys 88 of IFITM3 can also be monomethylated by Set7, a lysine methyltransferase.\textsuperscript{137} Monomethylation of IFITM3 reduces its antiviral activity as several experiments have shown. Interestingly, infection with influenza virus or vesicular stomatitis virus showed a rapidly increasing level of methylated IFITM3, while IFN-\textalpha treatment decreased methylation. This may suggest that these viruses actively promote IFITM3 methylation as an immune evasion strategy.\textsuperscript{137}

  Surprisingly however, mutation of Lys 88 does not result in major changes in IFITM3 activity against the influenza virus, dengue virus or vesicular stomatitis virus,\textsuperscript{137,152} even though one would have expected an increase in antiviral activity. These surprising experimental data could be explained by analogy with the multiple lysines of IFITM3 that
are ubiquitinated, meaning that compensatory methylation may occur at another lysine in the absence of Lys 88.

Further investigations still need to be carried out to fully understand the effect that Lys 88 methylation has on IFITM3 bioactivity. The fact that Set7 can modulate IFITM3 antiviral activity expands the catalogue of IFITM3-modulating enzymes that can be targeted for antiviral therapy.

4.2.4 IFITM3 can adopt different topologies in lipid membranes

The membrane topology of IFITM3 is a subject of controversy, and evidence suggests that several conformations are possible – depending of the stress state of the cell – with the antivirally active topology being a topic of debate. Determining the topology of active IFITM3 is essential for understanding its biophysical interactions with cellular membranes and other cellular proteins, and ultimately its mechanism of antiviral action. So far, determination of its topology has been attempted by mapping of IFITM3 PTMs.

The first topology proposed was a dual transmembrane conformation where the N and C termini were both outside the cell (Figure 4.7.a). This topology has been supported by the detection of both C and N termini by antibody staining of non-permeabilized cells.\(^{123,125}\)

The second topology that has been proposed was a dual intramembrane conformation, where this time, the N and C termini face the cytosol (Figure 4.7.b). This theory was supported by analysis of the different PTMs on IFITM3. First, ubiquitination occurs in the cytoplasm since this is the location of ubiquitin ligases. Lys 24, which is strongly ubiquitinated, thus has to be located in the cytoplasm. By the same reasoning Lys 83, 88 and 104 in the loop between the two membrane domains have to be cytoplasmic oriented.\(^{136}\) Second, Tyr 20 is phosphorylated by Fyn a cytoplasmic kinase.\(^{138,139}\) Tyr 20 is also a part of the YEML motif that interacts with the endocytosis machinery and is essential for IFITM3 localization to endolysosomes. Furthermore, IFITM3 was engineered to have glycosylation sites at both termini, but these sites appeared to be unmodified, supporting the proposition that both the N- and C-termini are cytoplasmic.\(^{136}\)

These two topologies could represent an inactive and an active form, respectively. The inactive form allows staining with N-terminus-directed antibodies at the cell surface of non-permeabilized cells, whereas the active IFITM3 has a cytoplasmically oriented N-terminus, allowing the different PTMs to happen.\(^{136,140,152,153,157}\)
Figure 4.7. The different topologies that IFITM3 is believed to adopt.

A third topology was recently proposed that contests the topology of the second membrane domain (Figure 4.7.c). Data showed that IFITM3 staining at the cell surface is stronger when C-terminal epitope tags are targeted, compared to N-terminal tags, suggesting that a more significant fraction of the second hydrophobic domain fully spans the plasma membrane.\(^{153,157}\) When the C and the N termini of IFITM3 were modified with epitope tags, it appeared that the C-terminal, but not the N-terminal, tags were degraded within lysosomes, which would be consistent with their lumenal orientation.\(^{157}\) From these additional data, a type II transmembrane topology has been suggested. However, the importance of the type II transmembrane conformation in the antiviral activity of IFITM3 remains to be demonstrated, especially because the percentage of IFITM3 adopting this conformation is unknown and also because this model was developed using IFITM3 modified with C-terminal epitope tags, where the tag was more than double the length of the C-terminal domain. This could potentially modify the biophysical properties of IFITM3.\(^{153,157}\)

### 4.3 Chemical synthesis of IFITM3 by KAHA ligation

#### 4.3.1 Generalities

IFITM3 is without any doubt a key element of the immune system against viral fusion. Furthermore, although this protein has been under the spotlight for years, there is still much to learn about its bioactivity/PTMs/actual topology.

Studies in the literature highlight some limitations that directly arise from the mode of preparation of IFITM3: expression with large tags that can alter conformation, difficulties associated with IFITM3 isolation, and characterizing which posttranslational modifications are present.

In order to bring insight into the field, we decided to develop a robust chemical synthesis of IFITM3 and different variants bearing a phosphorylate group on Tyr 20, mimics of the
palmitic acid on the three cysteines, as well as a fluorescent dye. If successful, this route would also allow us to synthesize IFITM3 variants labeled with photo-affinity tag to help identify the E3 ubiquitin enzyme, and the phosphatases involved in the PTMs.

4.3.2 Retrosynthesis, disconnection and segment synthesis towards IFITM3

The first step towards the synthesis of a protein is the analysis of its amino acid sequence. IFITM3 is a 133 amino acids long protein that consists of two intra/trans-membrane domains linked by a loop and a hydrophilic N-terminal segment that is not supposed to bind to the membrane. In contemplating the synthesis of IFITM3, the second membrane region deserved particular consideration. This region, spanning residues 105–129, contains only non-polar residues, including seven leucines, six isoleucines, two valines, a phenylalanine and a tryptophan. The remainder – one alanine, one glycine, one methionine, one threonine, one proline and one asparagine – offer no polar/charged side chains to facilitate synthesis or solubilization. We anticipated that the synthesis and handling of this region would be exceptionally challenging. The other intramembrane domain (IM1) is less intimidating, but the lack of polar residues in this segment also suggested that the synthesis of this region could be problematic. The N-terminal domain, reported to be cytosolic or extracellular rather than membrane-associated, is more polar and we hoped that the properties of synthetic intermediates containing this region would be more favorable.

4.3.2.1 First strategy: NCL/KAHA Ligation

The presence of a cysteine residue (Cys 105) immediately prior to the IM2 domain argued strongly for a synthetic strategy featuring a final native chemical ligation to attach the IM2 region to a larger – and hopefully more soluble – segment containing the rest of the protein. Although the slightly basic aqueous buffers typically used for NCL are not ideal for working with membrane proteins, several notable examples of MP synthesis with NCL have emerged. Impressive examples include the synthesis of the diacylglycerol kinase by Becker,\textsuperscript{158} the NS4A Protein by Pessi,\textsuperscript{159} the MscL by Kochendoerfer,\textsuperscript{160} the channel protein Kir5.1, the HCV ion channel and the multidrug resistant transporter EmrE by Liu,\textsuperscript{161,162} and the landmark synthesis of the voltage-gated potassium channel KcsA by Muir and MacKinnon.\textsuperscript{163,164}

To construct the 103-residue peptide thioester required for this strategy, we wished to use the α-ketoacid–hydroxylamine (KAHA) ligation developed by our group as a complement to NCL for protein synthesis. The KAHA ligation proceeds under aqueous, acidic conditions that also tolerate the presence of organic solvents such as DMSO or NMP. These solvent mixtures excel at solubilizing peptides, including many hydrophobic sequences. Furthermore,
the KAHA ligation conditions are completely orthogonal to those of NCL and we hoped that C-terminal thioesters would be stable to these conditions, obviating the need to mask them for sequential ligation. This would allow us to construct peptides with a C-terminal thioester and an N-terminal hydroxylamine that could serve as a lynchpin for assembling IFITM3 from four segments using two KAHA ligations and a final NCL (Figure 4.8). Li and co-workers recently introduced a similar approach, combining their Serine/Threonine Ligation (STL) methodology and NCL to successfully synthesize interleukin 25.\(^\text{165}\)

![Diagram of IFITM3 protein](image)

**Figure 4.8. General strategy for the synthesis of IFITM3 via KAHA ligation and NCL.**

KAHA ligation requires peptides bearing C-terminal \(\alpha\)-ketoacids and we have established several routes to such peptides using either a cyanosulfur ylide linker or protected \(\alpha\)-ketoacids as described in Chapter 2. Using these linkers, the preparation of the peptide-\(\alpha\)-ketoacids proceeds smoothly on large scale without any special precautions using standard Fmoc-SPPS. One interesting aspect of the KAHA ligation with oxaproline is that it generates a *depsi*-peptide link at the ligation site as described in Chapter 1. This unexpected outcome is beneficial for the synthesis of hydrophobic proteins – in analogy to Kiso’s isopeptide approach – as it breaks up secondary structure and improves solubility.\(^\text{166}\) These esters can be readily and rapidly rearranged to the native amides upon treatment with basic buffers used for folding or NCL.
The three segments required for the synthesis of the 103-residue thioester H₂N-S1S2S3-COSR 4.6 by KAHA are shown in Figure 4.8. In this approach, Ser 50 and Met 68 would be replaced by norleucines in order to avoid sulfur oxidation, which can arise during peptide handling. Arginine α-ketoacid-segment 4.1 (H₂N-S1-KA) was synthesized via our sulfur ylide linker method; oxidation of the crude peptide provided easy access to the desired peptide α-ketoacid 1. Segment 2, containing both Fmoc-protected oxaproline and C-terminal phenylalanine ketoacid (FmocOpr-S2-KA) 4.2 could also be prepared without difficulty by following our standard protocols using a protected phenylalanine α-ketoacid monomer and Fmoc-5-oxaproline (see Chapter 6).

We wished to prepare the lynchpin segment 3 (Opr-S3-COSR), which contains both the oxaproline and thioester, by Fmoc-SPPS. To do so, we adopted the hydrazine protocol developed by Liu and co-workers. The synthesis of the desired peptide hydrazide bearing an oxaproline at the N-terminus was straightforward. Unfortunately, during oxidation of the hydrazide with NaNO₂, the free oxaproline also underwent oxidation, probably because it is not protonated at this pH, allowing it to react with the oxidant. By using a hydrazide segment bearing Fmoc protected-oxaproline, the oxidation and conversion to the thioester proceeded smoothly (Scheme 4.1).

Although we feared that deprotection of the Fmoc group in the presence of the thioester would be difficult, it proceeded surprisingly well by treating FmocOpr-S3-COSR at 0.5 mM with 1% DBU in dry DMSO for 30 seconds at rt. The resulting lynchpin segment could be purified directly by HPLC to afford the desired product Opr-S3-COSR 4.3 (Scheme 4.1).

The synthesis of segment 4 containing the IM2 domain proved to be the biggest challenge. Due to its extreme hydrophobicity, we first attempted to synthesize it by Boc SPPS,
as Boc chemistry is known to be better suited for the preparation of hydrophobic peptides.\textsuperscript{168} Unfortunately, although we could observe the desired mass, all attempts to purify the peptide obtained after HF cleavage were unsuccessful. We next attempted the synthesis of segment 4 by Fmoc SPPS, however this was even less efficient. After coupling only 10 residues, the amount of impurities was greater than the desired product (see Supporting Information).

We therefore sought for alternative methods to synthesize this hydrophobic sequence. The addition of solubilizing tags and backbone protection have been reported to improve the synthesis of hydrophobic peptides.\textsuperscript{162,169} We decided to employ an oligo-arginine tag – similar to one described by Wade and co-workers\textsuperscript{170} – that could be removed by treatment with base at room temperature. This tag provides two advantages. First, oligo-arginines have been shown to greatly increase the solubility of hydrophobic sequences, making the handling and purification of these segments easier. Second, by placing this arginine tag at the C-terminus rather than the backbone of the peptide, we hoped that it would not interfere with the 3D structure of the folded IFITM3 and, if necessary, could be kept in place to improve handling of the full protein. After coupling Fmoc–Gly–OH followed by seven Fmoc–Arg(Pbf)–OH directly onto a polystyrene Rink amide resin, we introduced methyl alcohol benzoic acid as a cleavable linker and initiated the synthesis of segment 4 from there. Using this strategy, Fmoc-SPPS of the hydrophobic segment proceeded smoothly and the cleaved peptide bearing the (Arg)\textsubscript{7} tag \textbf{4.4} could be readily purified by reverse phase HPLC on C4 columns (Scheme 4.2).

With all the segments in hand, we proceeded with the ligations (Scheme 4.3). FmocOpr-S2-KA \textbf{4.2} and Opr-S3-COSR \textbf{4.3} were subjected to KAHA ligation conditions and, despite our initial concern about the stability of the thioester during KAHA ligation, we were...
Heating the reaction mixture to 37 °C or 40 °C did not improve the solubility. Analysis of the peptides failed; even at 1 mM buffer, 0.2 M Na₂HPO₄, pH 6.8 and 4-mercaptophenylacetic acid (MPAA) as thio-additive, the key NCL between H₂N-S1S2S3-COSR 4.6 and Cys-S4-ArgTAG 4.4 we first attempted to solubilize the peptides under standard conditions for NCL: 6 M guanidinium buffer, 0.2 M Na₂HPO₄, pH 6.8 and 4-mercaptophenylacetic acid (MPAA) as thio-additive. Heating the reaction mixture to 37 °C or 40 °C did not improve the solubility. Analysis of the
reaction mixture after 2 hours showed formation of the MPAA thioester, but unfortunately the ligation never proceeded. In attempts to improve the nucleophilicity of the cysteine residue, we increased the pH to 7 and 7.2, but without any improvement. At pH 7.5, new peaks appeared and the HPLC traces began to be difficult to interpret, but the desired product was not detected.

In order to study this ligation in more detail, we screened for conditions using thioester Opr-S3-COSR 4.3 and Cys-S4-ArgTAG 4.4. Due to the hydrophobicity of the peptides, we faced the same solubility problems; precipitates formed at the bottom of the vial. Despite this solubility issue, the formation of the MPAA thioester 4.7 could again be observed (Scheme 4.5). Poor solubility during native chemical ligation of membrane proteins is a well-known phenomenon. To overcome this problem, many groups have proposed optimized strategies in addition to the use of solubilizing tags. Detergents such as dodecylphosphocholine (DPC)\textsuperscript{106,171}, sodium dodecylsulfate (SDS)\textsuperscript{162} or \(\beta\)-octylglucoside (OG)\textsuperscript{158,159} and organic co-solvent such as TFE,\textsuperscript{172} HFIP\textsuperscript{173} or even DMF,\textsuperscript{174,175} can be added directly to the ligation mixture, enhancing the solubility of the hydrophobic segments.

![Scheme 4.5. NCL between Opr-S3-COSR 4.3 and Cys-S4-ArgTAG 4.4, typical HPLC trace obtained after 2 hours and MALDI-MS spectra of the different formed products 4.7, 4.8 and 4.9.](image)

For our segments, the most effective solvent mixture appeared to be the use of 10% organic solvent such as TFE or HFIP with 8 M urea. But in the presence of organic solvent, the trans-thioesterification required higher pH than in pure aqueous buffer and the formation of the MPAA thioester 4.7 could not be observed at pH 7.5 or below. Unfortunately, at pH 7.5
or higher, two main side reactions occurred: 1) hydrolysis of the thioester product to give 4.8 and 2) cyclization of the side chain lysine to give 4.9 (Scheme 4.5). The details of the ligation conditions we screened are summarized in the Supporting Information (Page 146–148).

To confirm that the origin of the difficult ligation was the extreme hydrophobicity of segment 4, we attempted NCL between Opr-S3-COSR 4.3 and the short N-terminal cysteine test peptide 4.10 (Scheme 4.6.a). In this case, the ligation proceeded after 4 hours at 2 mM without solubility problems to afford the desired ligated product 4.11. A similar experiment was made with a short glycine thioester peptide 4.12 and the Cys-S4-ArgTAG 4.4. At 2 mM after 4 h in 6 M guanidinium buffer (pH 6.8) the MPAA thioester segment 4.13 could be clearly detected, but we also observed precipitation and no ligated product was formed (Scheme 4.6.b).

Scheme 4.6 a) NCL test with 4.3 and a short cysteine peptide 4.10. HPLC of the reaction mixture after 4 hours and MALDI-MS spectrum of the major peak, showing the desired ligated product mass 4.11. b) NCL test with 4.4 and a short thioester peptide test 4.12. MALDI-MS spectrum of the only peak formed after 4 hours, corresponding of the MPAA thioester segment 4.13.

From these experiments, we conclude that the problem stemmed from cysteine segment 4.4, most likely due to its hydrophobicity. Recently, Tam and co-workers reported that they could not synthesize the highly hydrophobic cyclic protein AS-48 by NCL for the same reason and resorted to a more diluted, enzymatic approach.84
4.3.2.2 Second Strategy: KAHA ligation

Despite the elegance of a combined KAHA/NCL strategy, we were forced to reconsider our approach. KAHA ligation proceeds well under acidic conditions and tolerates much more organic co-solvent than NCL. This allows the solubilization of even very hydrophobic peptide segments, often at 10-20 mM. Furthermore, by using 5-oxaproline as the hydroxylamine and isolating the ligation products as the initially formed ester, we can often improve the solubility and handling of the ligation products, a strategy that we recently applied to the synthesis of S100A4. We therefore decided to attempt an “all-KAHA” ligation approach to IFITM3 that would test the suitability of KAHA ligation for the synthesis of membrane proteins.

In redesigning our strategy, we took the opportunity to reconsider the ligation sites. Analysis of the IFITM3 sequence and important biological PTM sites (Figure 4.9), as well as comparison with the family of IFITM proteins, allowed us to select the most suitable sites for introducing the nonstandard Hse amino acid. We aimed to synthesize IFITM3 from three segments using two KAHA ligations. For the first ligation site, we chose again Ser50Hse, as this falls in an exposed loop region that shows considerable variability among IFITM proteins. As the second site, we selected Thr95Hse as this site was reported to be one of the few variable sites within an otherwise highly conserved region. Among the known variants, threonine, glutamic acid, aspartic acid, serine, alanine and asparagine residues have been

Figure 4.9. New strategy for the synthesis of IFITM3 by an “all KAHA ligation” approach
observed at this position. This strategy called for the use of a valine \( \alpha \)-ketoacid for the ligation, which although somewhat slower than less hindered residues has proven to be an excellent choice for KAHA ligations (Figure 4.9).\(^6\) This strategy required that we extend the synthesis of the challenging C-terminal segment by 11 residues.

Using the solubilizing (Arg)\(_7\) tag at the C-terminus, we were able to successfully prepare Opr-S3-ArgTAG 4.15 without too much difficulty on a Rink amide polystyrene resin (0.32 mmol/g loading). The synthesis of 1 gram of this resin afforded 128 mg of the purified IFITM3 Opr-S3-ArgTAG peptide 4.15 (7% overall yield) bearing an N-terminal 5-oxaproline. Central segment 4.14 required a protected Opr residue (see Supporting Information). We have previously used Fmoc-Opr for this purpose but have recently introduced oxaprolines bearing photolabile protecting groups that can be removed directly after the ligation, saving time and material.\(^6\) The central segment, bearing a C-terminal valine \( \alpha \)-ketoacid, was prepared by Fmoc-SPPS and the requisite photoprotected [(2-nitrophenyl)ethyl]carbamate-Opr (NpecOpr) was introduced as the final residue. The \( \text{H}_2\text{N-S1-KA} \) segment 4.1 remained the same as described above (see Supporting Information).

![Scheme 4.7. KAHA ligation with peptide Opr segments 4.15 and ketoacid 4.14. HPLC traces of the progression of the ligation at 60 °C. ESI-MS spectrum of purified 4.16.](image)

NpecOpr-S2-KA 4.14 and Opr-S3-ArgTAG 4.15 were joined by the first KAHA ligation. We were pleased to find that the segments were perfectly soluble in a 95:5 NMP/H\(_2\)O mixture at 20 mM. Despite the use of the hindered valine \( \alpha \)-ketoacid, the ligation was almost complete after 10 hours, at which time a colorless gel had formed. The reaction mixture was diluted to 10 mM with NMP and left at 60 °C for an additional 10 hours but no change was observed, indicating that the ligation reaction had reached maximum conversion after 10 hours. The ligation mixture was diluted to 1 mM with 1:1 CH\(_3\)CN/H\(_2\)O containing 0.1% TFA with the addition of 2,2-(Ethylenedioxy)diethanethiol (DODT), and the resulting mixture was stirred.
under handheld UV light at a wavelength of 365 nm for 45 minutes to remove the Npec group. The reaction mixture was directly purified on a C4 column heated at 60 °C to afford the desired ligated ester peptide 4.16 in an overall yield of 37 % (Scheme 4.7). The second — and final — ligation was performed between 1 equiv peptide 4.16 and 1.5 equiv H$_2$N-S1-KA 4.1 at 15 mM in a mixture of 95:5 NMP/H$_2$O. The ligation was somewhat slower than typically observed. After 20 hours, a gel had formed and the ligation did not proceed beyond 60% conversion. Despite gel formation, the desired ligated product formed cleanly. The mixture was diluted to 0.5 mM with carbonate buffer (pH 8.5) and left for 8 hours to initiate $O$ to $N$ acyl shift of the two homoserine esters.

Scheme 4.8. KAHA ligation with peptide Opr segments 4.16 and $\alpha$-ketoacid 4.1. HPLC traces of the progression of the ligation and of the final IFITM3-ArgTAG at 60 °C. ESI-MS spectra of the purified IFITM3-ArgTAG 4.17 and SDS-PAGE.

After concentration and purification by HPLC, we isolated 16.3 mg (22% yield) of the desired IFITM3-ArgTAG 4.17 (Scheme 4.8). The synthetic IFITM3 possessing a C-terminal ArgTAG was characterized by HPLC, HRMS, and SDS-PAGE analysis. For CD spectroscopy, the protein was incorporated into phosphatidylcholine (PC) lipid vesicles and the spectrum showed a strong $\alpha$-helix fingerprint, in accordance with the calculated 3D structure (Figure 4.10).
Figure 4.10. CD spectra of the reconstituted IFITM3-ArgTAG 4.17 in PC micelles. (50 mM in 1 mm quartz cell).

To release the IFITM3 protein from the C-terminal solubilizing tag, 4.17 was treated with 0.1 M NaOH for 5 minutes. The outcome of this hydrolysis was surprising. Although we observed the desired native IFITM3 by SDS-PAGE, MALDI-MS analysis and SDS-PAGE showed that the majority of the material was truncated products. The main truncated product, with a molecular weight of 6545 Da, was identified as a peptide segment from Pro 40 to Hse 94. Observation of this segment confirmed that the three segments were properly ligated, but revealed that the highly proline rich region (35–41; 4 proline residues within 7 amino acid residues) of H$_2$N-S1-KA was susceptible to hydrolysis under basic conditions and was not stable to the conditions needed for removing the ArgTAG.

An alternative approach was to remove the (Arg)$_7$ tag prior to the final ligation (Scheme 4.8). Although this seemed like a logical solution, it would require the isolation and subsequent ligation of the hydrophobic segment 4.18 lacking the solubilizing tag. Fortunately, this turned out to be less challenging than expected. After the first ligation and UV irradiation of the photolabile protecting group, the reaction mixture was diluted with a 0.25 M NaOH solution

Scheme 4.9. KAHA ligation with the Opr segment 4.15 and $\alpha$-ketoacid 4.14. Direct removal of the arginine tag with NaOH solution. HPLC traces of the progression of the ligation at 60 °C. ESI-MS spectrum of the purified 4.18.
until a concentration of 0.1 M was reached. After 5 minutes, the solution was cooled in an ice bath and slowly quenched with a solution of 50% TFA in water and directly purified by HPLC to afford the desired Opr-S2S3-OH \(4.18\) in 24% yield after 3 steps (Scheme 4.9). During the step to remove the ArgTAG, only one single peak was observed by HPLC, confirming that no \(O\to N\) shift was complete.

The final ligation – now without the solubilizing group – was performed using 95% NMP with 5% aqueous oxalic acid. Despite the hydrophobic segment, the ligation proceeded relatively well. We again observed gel formation and the conversion was less than the ligation with the more soluble segment. Following dilution of the ligation mixture to 0.5 mM with a carbonate buffer (pH 8.5) and stirring for 8 hours at 23 °C to effect the \(O\to N\) shift, 8.1 mg of unmodified IFITM3 \(4.19\) was isolated by preparative HPLC on a heated C4 column (Scheme 4.10).

![Scheme 4.10. KAHA ligation with peptide Opr segments 4.18 and ketoacid 4.1. HPLC traces of the progression of the ligation and of the purified IFITM3 at 60 °C and SDS-PAGE of IFITM3 4.19.](image)

All attempts at measuring the MS of the intact IFITM3 protein lacking the solubilizing tag, however, were unsuccessful. A similar problem was noted by Liu and coworkers for their synthesis of the multidrug resistance transporter EmrE, another membrane associated protein, and can be rationalized by the highly hydrophobic nature of the protein and paucity of ionizable sites. We attempted to analyze the protein after digestion, but the lack of accessible cleavage sites and the hydrophobic nature of the fragments precluded sequence determination by the usual methods. Finally, we succeeded by digestion of the protein under acidic conditions and analysis of the resulting fragments by MALDI-MS. This procedure gave 95.5% sequence coverage of the synthetic IFITM3 and showed no trace of the ArgTAG, confirming the synthesis of IFITM3 (S50T, T95T) (Figure 4.11).
Figure 4.11. MS results after acid digestion of IFITM3. The grey bands correspond to the mass of the different peptides found and to which part of the amino acid sequence it corresponds to. When the intensity of the MS signal was strong enough, MS/MS analysis was performed and is displayed by the red marks. The MS sequence coverage was greater than 95%.

To record the CD spectrum of synthetic IFITM3, we followed the same procedure as applied for the protein containing the ArgTAG. After reconstitution of the protein in PC vesicles, the CD spectra showed a strong fingerprint for an α-helix conformation. The two CD spectra were nearly identical, showing that the C-terminal Arg-tag did not influence the 3D structure (Figure 4.12). The mass difference between these two versions of IFITM3 was large enough to be observed by SDS-PAGE (Scheme 4.9).

Figure 4.12. CD spectra analysis of the reconstituted IFITM3 in PC micelles (green trace). Comparison with the CD spectra of the reconstituted IFITM3-ArgTAG in PC micelles (grey trace) (50 mM in 1 mm quartz cell).

4.3.2.3 Synthesis of phosphorylated IFITM3-ArgTAG

Several studies have established that posttranslational modifications of IFITM3 are essential for its anti-viral activity. The role of these numerous modifications remains a subject
of debate, and the inability to prepare and isolate homogeneous protein complicates further efforts. A key advantage of chemical protein synthesis is the ability to precisely control the presence or absence of posttranslational modifications, as well as the power to incorporate probes such as affinity tags or fluorescent dyes that facilitate biochemical and biophysical studies.\textsuperscript{177,178}

In our initial efforts to synthesize IFITM3 derivatives, we focused on the two most important posttranslational modifications: phosphorylation of Tyr 20 and palmitoylation of Cys 71, Cys 72, and Cys 105. In addition, we planned to prepare a fluorescently labeled IFITM3 for studies on its incorporation into membrane vesicles.

![Scheme 4.11](image)

Scheme 4.11. a) Synthesis of phosphorylated ketoacid segment 4.21 via oxidation of the phosphorylated segment 4.20 with Oxone. b) KAHA ligation between phosphorylated ketoacid 4.21 and Opr segment 4.16. HPLC monitoring of the ligation and the purified phosphorylated IFITM3-ArgTAG 4.22 at 60 °C. ESI-MS spectrum of 4.22.

We had not previously attempted KAHA ligations with phosphorylated peptides. Our synthetic strategy required the phosphorylated Tyr to be stable to the oxidative conditions for conversion of the Segment 1 cyanosulfur ylide 4.20 to the \( \alpha \)-ketoacid and to the acidic conditions of the KAHA ligation. For preparation of Segment 1 containing the phosphorylated-
tyrosine, we chose O-benzyl protected phosphorylated-tyrosine for Fmoc-SPPS, which affords the unprotected residue upon TFA cleavage. The segment synthesis proceeded smoothly and the phosphorylated-tyrosine was stable to oxidation of the sulfur ylide with aqueous oxone and Segment 1 α-ketoacid 4.21 was isolated without difficulty (Scheme 4.11.a). This underwent KAHA ligation with the Segment 2-3 peptide 4.16 to give phosphorylated IFITM3-ArgTAG 4.22, which could be purified by preparative HPLC (Scheme 4.11.b).

### 4.3.2.4 Synthesis of cysteine palmitoylated IFITM3-ArgTAG

Palmitoylation of cysteines – the reversible attachment of C16 or other long chain acids to form thioesters – is a common posttranslational modification often found in membrane-associated proteins. IFITM3 is reported to be palmitoylated at all three of its cysteine residues, but the extent and role of this modification as well as the type and location of the palmitoyltransferase are yet to be fully elucidated.

We first wanted to generate the native palmitoyl thioester linkage at all three cysteines. This kind of linkage cannot be introduced by modifying the Fmoc-cysteine residue and introducing them by SPPS as the thioester linkage would not survive the numerous piperidine treatments. As there is no method so far describing synthetic palmitoylation of proteins, we took inspiration from the hydrazide method of Liu et al.,167 which has been used to generate thioester peptide segments. The idea was to oxidize palmitoylhydrazide 4.24 into the corresponding palmitoyl azide intermediate 4.28 and to add it to IFITM3-ArgTAG 4.17. Here the cysteines play the same role as the thio-additive in NCL to generate the thioester moiety.

![Scheme 4.12. IFITM3-ArgTAG palmitoylation. a) Synthesis of palmitoylhydrazide 4.24 b) Acylation of IFITM3 cysteines using palmitoylhydrazide 4.24.](image)

We synthesized palmitoylhydrazide 4.24 from the corresponding palmitic acid N-hydroxysuccinimide ester 4.23 (Scheme 4.12.a). Once purified, 4.24 was subjected to oxidation with sodium nitrite in 6M GnHCl and NaH₂PO₄ buffer at pH 3. Unfortunately, and as we expected, the solubility of this molecule in aqueous buffer was an issue, and addition of SDS did not improve its solubility. We nevertheless pursued the protocol and added IFITM-
ArgTAG 4.17 – previously diluted in the same buffer – to the palmitoyl azide 4.25 solution (Scheme 4.12.b). The resulting mixture became even more turbid and was stirred overnight at pH 6. The solution was subjected to spin filtration and dialysis in order to remove the excess of reagents and then lyophilized. Unfortunately, SDS-PAGE of the resulting solution did not show any difference in mass, and only the unmodified starting protein could be seen by HPLC.

We then tried a different approach to generate the native thioester linkage in which hexadecanethioic-acid 4.26 was used as a nucleophile to attack IFITM3-ArgTAG variants containing pre-formed dihydroalanine (Dha) 4.29 (Scheme 4.13.c). Under slightly acidic condition (pH 4–5), the thioacid would be the only anionic species and could react with the Dha which is known to be a really good electrophile. Although it would give a mixture of diastereoisomers, we considered the resulting product as a potentially good mimic of native palmitoylated IFITM3.

The palmitic thioacid 4.26 was directly synthesized from palmitic acid N-hydroxysuccinimide ester 4.23 with sodium hydrosulfide hydrate (Scheme 4.13.a). To generate the dihydroalanine IFITM3 4.29 we used the method of Davis et al. using α,α′-dibromo-adipyl(bis)amide 4.28, which is readily synthesized from adipic acid 4.27 (Scheme 4.13.b). IFITM3 was then treated with 4.28 in phosphate buffer at pH 8. (Scheme 4.12.c) Unfortunately, here again, solubility was an issue, and the Dha-IFITM3-ArgTAG 4.29 could not be detected by HRMS mass.

We decided to change our approach and focused our efforts on selective alkylation of the three cysteines. Our choice was the use of chloroacetamide derivatives which are known for selectively modifying cysteine residues. We synthesized 2-chloro-N-undecylacetamide 4.32 from undecanamine 4.31 and 2-chloroacetyl chloride (Scheme 4.14.a). IFITM3 was dissolved in phosphate at pH 7.8 and 120 equivalents of 4.32 were added, and the resulting mixture stirred at 37 °C overnight. Again, we faced solubility issues in aqueous buffer, and the addition of different concentrations of SDS did not show any improvement. Starting material 4.17 was the only molecule that could be detected by SDS-PAGE and MS analysis (Scheme 4.14.b).


Since solubility was obviously a main issue in all these cases, we decided to look for methods that had been used for the lipidation of molecules. We found a method by Post et al.\textsuperscript{180} that could be run in an acetic acid buffer and organic solvent (n-propanol) mixture to generate a disulfide linkage. Disulfide linkages have been successfully used to modify proteins and can be cleaved by mild reducing agents.\textsuperscript{179,181} To do so, we synthesized hexyldecyl-dithiopyridine 4.34 from hexadecathiol 4.33 and 2,2'-dipyridyldisulfide. IFITM3-ArgTAG 4.17 was treated with an excess of hexyldecyl-dithiopyridine 4.34 in a mixture of DMSO, n-propanol GnHCl and AcOH for 24 hours at room temperature, leading to the attachment of three alkyl chains (Scheme 4.15). The reaction mixture was dialyzed for 24 hours against H$_2$O/AcOH and the resulting solution was lyophilized.

SDS-PAGE analysis showed the formation of a single product with a higher molecular weight than the unmodified IFITM3; however, it was not possible to observe the alkylated IFITM3 4.23 by HPLC because of its extreme hydrophobicity. All attempts to analyze the sample by mass spectrometry failed. When this newly formed product was treated with TCEP for 1 hour at 40 °C, however MS analysis of the product showed the mass of the unmodified IFITM3, providing strong evidence that the cysteines were modified via disulfide formation (Scheme 4.15).
We tried to perform an Ellman test in order to confirm that all the cysteines were modified. Unfortunately, because of the poor solubility of the product (a white turbid solution in different buffers) obtaining accurate and meaningful UV measurements was not possible. Although it just gives a qualitative information, when 5,5'-dithiobis(2-nitrobenzoic acid) was added to a solution of 4.35, the solution remained white. However, when 4.35 was incubated with TCEP, dialyzed and spin-filtered against water, addition of 5,5'-Dithiobis(2-nitrobenzoic acid caused the solution to turn yellow. The same result was obtained with IFITM3-ArgTAG 4.17.

This cannot guarantee at 100% that the three cysteines are alkylated, but it strongly suggests that it is the case.

4.3.2.5 Incorporation in lipid membranes

Several reports have attributed the antiviral activity of IFITM3 to its ability to block the fusion of viral membranes in the endosomes.143,149 The activity of IFITM3 is therefore dependent on its ability to localize in the cell membrane. In order to use synthetic IFITM3 and its variants to probe its mode of action, we wished to incorporate the protein into a suitable membrane mimic and chose egg phosphatidylcholine as a suitable starting point. Synthetic IFITM3 was incubated with vesicles and then passed through Sepharose CL-4B exclusion resin with an exclusion limit of 2 x 10^7 Da. If IFITM3 is inserted into membrane vesicles, it should elute with the larger vesicles; the smaller, non-incorporated proteins will be retained. Deionized water (300 µL) was used to ensure that all the vesicles passed through the resin and the filtrate was lyophilized and analyzed by SDS-PAGE. Chymotrypsinogen A was used...
CHAPTER 4. Chemical Synthesis of the Antiviral Integral Membrane Protein IFITM3

Figure 4.13. Proteins were incubated with lipid vesicles and passed through size exclusion resin. SDS-PAGE of the filtrates showed that IFITM3 4.17 and Palm-IFITM3 4.35 were incorporated into the vesicles as a negative control as shown in the work of Francis and co-workers. We were pleased to see that these experiments were conclusive and that IFITM3 was incorporated into the lipid membrane as expected. The palmitoylated version behaved as expected and also eluted with the vesicles (Figure 4.13).

4.3.2.6 Synthesis of fluorescent IFITM3-ArgTAG 4.39

For further confirmation that synthetic IFITM3 is incorporated into the membrane, we prepared a H$_2$N-S1-KA segment with an N-terminal 5(6)-carboxyfluorescein 4.38 – from the corresponding sulfur ylide segment 4.37 – by introducing it during Fmoc-SPPS. Resin cleavage, oxidation of the cyanosulfur ylide and ligation with Opr-S2S3-ArgTAG segment 4.16 afforded IFITM3 with a fluorescent dye 4.39. No side reactions or decomposition occurred with the dye during these steps (Scheme 4.16.a).

The fluorescently modified IFITM3 4.39 was incorporated in medium size vesicles, passed through the size exclusion resin and detected by fluorescent microscopy. As shown in Scheme 4.16, IFITM3 is clearly localized in the membrane of the vesicles, as expected from the previous experiment.

Even though it could arise from the sample preparation, fluorescence microscopy shows labeled IFITM3 as compact and defined “dots” inside the lipid bilayer. This could indicate that IFITM3 localizes in lipid membranes as clusters or aggregates.
CHAPTER 4. Chemical Synthesis of the Antiviral Integral Membrane Protein IFITM3

4.4 Conclusions and outlook

In this chapter, we have described a robust chemical synthesis of pure IFITM3 on multimilligram scale using the KAHA ligation. This approach provides easy access to IFITM3 containing both natural posttranslational modifications as well as other derivatives that are not possible to prepare with current recombinant approaches. The synthetic IFITM3 behaved as expected by easily incorporating itself into lipid membranes, suggesting that it may be adopting a biologically relevant fold. Furthermore, our preliminary data using synthetic lipid vesicles as mimic of the cell membrane and fluorescent-labeled virus show that our synthetic IFITM3 is indeed reducing virus fusion (Chapter 5).

Even though we are currently collecting data to confirm and quantify the antiviral properties of our synthetic IFITM3, our work represents an important step to better understand the role and mechanism of this antiviral protein. With our well-designed protocols, one could synthesize modified versions of phosphorylated Tyr 20 IFTM3 possessing, for example, a biotin moiety or a photo affinity tag in order to fish out the currently unknown phosphatase enzymes involved in the dephosphorylation process. Or a similar approach could be used to identify the E3 ubiquitinligase involved in IFITM3 ubiquitination.
In general, synthesis of pure membrane proteins is a challenging task that limits greatly our understanding of these fascinating molecules. Chemical synthesis of membrane proteins can provide access to substantial amounts of homogeneous material in decent quantities as well as providing posttranslationally modified versions of these proteins.

Although the chemical synthesis approach provides important advantages over overexpression, it still suffers from a major limitation: the solubility of the peptide and protein segments. We proved during our attempts to synthesize IFITM3 that the ability of KAHA ligation to operate under acidic conditions with organic co-solvents as well as the depsi-peptide generated at the ligation site makes it ideally suited for the preparation of such molecules.
CHAPTER 4. Chemical Synthesis of the Antiviral Integral Membrane Protein IFITM3

4.5 References


CHAPTER 4. Chemical Synthesis of the Antiviral Integral Membrane Protein IFITM3


CHAPTER 5
SUMMARY AND OUTLOOK
CHAPTER 5. Summary and outlook

The work presented in this Thesis covers the technical improvements for the synthesis of key build blocks used in the KAHA ligation as well as the synthesis of two proteins: the protein hormone betatrophin and the antiviral membrane protein IFITM3.

5.1 Optimization and new synthetic route for the synthesis of SY linker, Boc-Opr and Fmoc-Opr

The synthetic routes for three main building blocks used in the KAHA ligation: sulfur ylide linker (SY) 2.1, (S)-N-Boc-5-oxaproline (Boc-Opr) and (S)-N-Fmoc-5-oxaproline (Fmoc-Opr) were optimized and developed. The SY linker allows the preparation of C-terminal sulfur ylide peptides that can be readily converted upon oxone oxidation into the corresponding α-ketoacids. With the optimized synthesis, more than 50 grams of the SY 2.1 could be prepared, allowing a scale up of numerous α-ketoacid peptide segments in our group. With the new synthesis of the 2.2 an outsourcing company could prepare a kilogram of Boc-Opr 2.2, sustaining us for several years and making our goal to have it commercially available possible. The Boc-Opr 2.2 can be the starting point for the synthesis of the Fmoc-Opr 2.3 and has recently been used by our group for generating oxaproline with photo labile protecting groups.

5.2 Synthesis of the betatrophin protein

Convergent synthesis using KAHA ligation was used to synthesize the betatrophin protein via a five-protein segment, 4 ligation approach. This 19.8 kDa protein was the first example from our group describing the utilization of the KAHA ligation for the synthesis of a protein of this size. Although the original bioactivities that motivated us to synthesize it were biased, we demonstrated that using KAHA ligation, we could synthesize a protein of this size in multi-milligram quantities in less than 9 months.

This encouraged us in going even further and our group described an even larger protein using the KAHA ligation: Nitrophorin 4.76
5.3 **Synthesis of the antiviral membrane protein via KAHA ligation**

We also described an efficient chemical synthesis of pure IFITM3 on multi-milligram scale using the KAHA ligation. Using the same approach, we prepared different variants of IFITM3 – containing both natural posttranslational modifications as well as a fluorescent dye – that were not obtainable with current recombinant approaches. We showed that our synthetic IFITM3 behaved as expected, easily incorporating itself into lipid membranes and the strong \(\alpha\)-helix fingerprint of the CD spectra expected the calculated 3D structure.

Furthermore, in collaboration with Prof. Yamauchi, we are currently establishing a protocol using synthetic lipid vesicles as endosomal membrane mimics with the aim to evaluate the antiviral activity of our synthetic IFITM3 as well as its variants. Liposomes have already been used to show that fluorescently-labeled viruses can fuse with liposomes. The fluorescence signals increased as a result of fluorophore dilution and fluorescence dequenching can be measured.\(^{183,184}\)

For our early experiment, we used an R18 (red) labeled influenza X31 strain virus and mixed it with liposomes (DOPC/Cholesterol 4:1) and IFITM3-containing liposomes. Then the pH of the mixture was acidified to pH 5 and the fluorescent values were measured every 5 seconds for 30 minutes. After 30 min, tritonX-100 (final 0.1%) was added to allow maximum fluorescence. This value was used to calculate the liposomal membrane merging.
A graph could be obtained using the following equation:

\[ y = \frac{f(t) - f(0)}{f(max) - f(0)} \]

Where \( f(0) \) is the fluorescence read before acidification and \( f(max) \) is the fluorescence read after complete dequenching, achieved by the addition of Triton X-100 detergent to a final concentration of 0.1% (Figure 5.1).

![Figure 5.1. Virus fusion assay with lipid vesicle containing IFITM3.](image)

We were glad to see that fluorescence – and so virus fusion – was considerably reduced when IFITM3 was incorporated into the membrane of the liposomes. We are currently designing a more accurate system in order to be as close as possible to what endosomes look like in a living cell.

In a more general way, we showed that the KAHA ligation, because of its unique features – acidic conditions with organic co-solvents as well as the "depsi"-peptide generated at the ligation site – was an ideal method for the chemical preparation of membrane proteins.

5.4 Combining KAHA ligation and NCL in one pot for the synthesis of proteins, application to U18 ctenitoxin-Co1b

During protein synthesis every purification step results in a loss of precious material as well as a loss of time and solvent. Therefore, methodologies combining 2 or 3 ligations in one pot are highly sought. While synthesizing IFITM3, we found out that thioesters were stable under KAHA ligation conditions. We imagined combining these two ligation methods into a
one-pot ligation for the synthesis of proteins. We used a small peptide, U18, as a test model and combined these two ligations to synthesize U18 (Scheme 5.1).

![Image of a scheme showing the synthesis process](image)

**Scheme 5.1. KAHA ligation and NCL in one pot for the synthesis of U18.**

We were glad to see that the KAHA ligation, proceeded smoothly in the presence of the thioester moiety and that the reaction mixture, when further subjected to NCL conditions, afforded the desire protein in one step (Scheme 5.1). The use of excess starting material, which is necessary for the synthesis of some proteins, did not, as anticipated, interfere which the ligation. This encouraging result led us to initiate the synthesis of more complex and longer proteins.

### 5.5 Synthesis of the Call1 protein by combining KAHA ligation and NCL

As a part of a collaboration with our satellite laboratory in Nagoya, we were interested in applying this technique to the synthesis of the 105 amino acid long protein Call1. This protein was recently discovered and has shown interesting activity in the mechanism of pollen tube guidance in plants. So far, after trying different possible combinations, we are planning a synthesis of this protein via a three peptide segment strategy (Scheme 5.2).
Scheme 5.2. General strategy for the synthesis of Call1 via KAHA ligation and NCL.

Although the syntheses of H$_2$N-S1-KA 5.6 and Cys-S3-OH 5.8 are more complicated than we expected, using Acm protected cysteines as well as numerous pseudo-prolines appeared to be a good solution to obtain these peptide segments in good purity and quantity.

We are now finishing the synthesis of Call1 and will prepare different versions of it with tags and unnatural modifications in order to help to understand its biological role and activity better.
5.6 References


CHAPTER 6. Experimental Part

6.1. 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 (hexanes, 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), ABCR (Karlsruhe, Germany) and TCI Europe (Zwijndrecht, Belgium). 1-NPEOC-2, Boc-3 and Fmoc-Opr-OH 4 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 13 C 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 Q-TOF detector or with a Bruker solariX instrument (ESI or MALDI FTICR MS). 4-Hydroxy-α-cyanocinnamic acid was used as the matrix for MALDI analysis. MS-MS analysis was performed by the Functional Genomics Center Zurich, Switzerland. Low 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 1000cm⁻¹. Optical rotations were measured on a JASCO P-2000 instrument at
ambient temperature operating at the sodium D-line ($\lambda = 589.3$ nm) in a cell with 100 mm path length. CD-spectra were recorded on an 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. Reaction mixtures 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 SiO$_2$ 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 $\mu$L 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-H$_2$O containing 0.1 % (v/v) TFA and HPLC grade CH$_3$CN 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 $\mu$m, 120 Å pore size, 4.6 mm I.D. x 250 mm) column, on a Shiseido Capcell Pak C18 UG 80 (5 $\mu$m, 120 Å pore size, 4.6 mm I.D. x 250 mm) column or on a Shiseido MGII C18 column (5 $\mu$m, 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 $\mu$m, 100 Å pore size, 20 mm I.D. x 250 mm) column, on Shiseido Capcell Pak C4 or C18 columns (5 $\mu$m, 80 Å pore size, 50 x 250 mm) or on a Vydac C4 column (5 $\mu$m, 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 types of solvent gradient were used: The column was pre-equilibrated at the desired composition (e.g. 5% CH$_3$CN) for typically 10 min. After injection of the sample, the solvent composition was held at the same concentration of CH$_3$CN for 3 min (5 min or 7.5 min for preparative HPLC) and within the gradient run time, a linear solvent gradient was run to the final solvent composition (e.g. 50% CH$_3$CN). After the gradient run, the solvent composition was changed to 95% CH$_3$CN within 1 min and the column was flushed for 7 min. Within 1 min, the solvent composition was changed back to the original concentration of CH$_3$CN and the run ended. For the sake of simplicity, only the gradient time and the starting and end composition of the eluent are stated for the individual experiments, although all experiments included the full cycle as described above.
**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 a 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. 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) was dissolved in a minimal amount of CH$_2$Cl$_2$ 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 allowed to react for 1-2 h.

**Manual coupling of special amino acids**

Valuable non-standard monomers (e.g. protected (S)-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 shaken 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 were 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/CH₃CN + 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 shaken 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 were 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/CH₃CN + 0.1 % TFA or mixtures thereof) for RP-HPLC purification.
6.2 Experimental for “Chapter 2: Synthesis and Optimization of the Key Building Blocks – Cyanosulfur Ylide and Boc/Fmoc Oxaproline – For KAHA Ligation”

6.2.1 Cyanosulfur ylide linker synthesis

(E) and (Z)-Ethyl 2-(dihydrothiophen-3(2H)-ylidene)acetate (2.21).

In a flame dried 2 L flask, NaH, 60% dispersion in mineral oil (22.5 g, 563 mmol, 1.20 equiv) was added to dry THF (840 mL). The resulting suspension was placed into an ice bath and triethylphosphonatoacetate (126 g, 112 mL, 563 mmol, 1.20 equiv) was added dropwise. The mixture was allowed to warm to room temperature and stirred for 30 min. The mixture was cooled to -78 °C and 4,5-dihydro-3(2H)-thiophenone (47.9 g, 40.2 mL, 470 mmol, 1.00 equiv) was added dropwise. The resulting mixture was stirred at -78 °C for 3 h, allowed to warm and stirred overnight at room temperature. The reaction was filtered, the filtrate was concentrated and the solid redissolved in water and extracted with ethyl acetate (3 x 400 mL). The resulting organic layers were combined and washed with brine (2 x 150 mL), dried over Na₂SO₄ and concentrated in vacuo to afford the crude product as a dark brown oil. The crude sample was purified by column chromatography on silica gel (hexanes/EtOAc 95:5) to afford the title compound 2.21 (78.5 g, 456 mmol) as a colorless oil in 97% yield.

1H NMR (400 MHz, CDCl₃) δ = 5.91–5.87 (m, 1H), 4.17 (q, J=7.0 Hz, 2H), 4.00 (s, 1H), 3.61 (s, 1H), 3.20 (t, J=4.0 Hz, 1H), 2.97 (t, J=6.8 Hz, 1H), 2.87 (t, J=4.0 Hz, 2H), 1.29 (t, J=7.0 Hz, 3H).

13C NMR (101 MHz, CDCl₃) δ = 162.5, 113.0, 112.9, 59.7, 39.9, 38.6, 34.6, 34.4, 31.5, 29.3, 14.2.


2-(Tetrahydrothiophen-3-yl)ethanol (2.23).

In a three-neck flame-dried flask, LiAlH₄ (30.0 g, 791 mmol, 4.00 equiv) was added slowly to dry and cold THF (420 mL) in an ice bath. A solution of ethyl 2-(dihydrothiophen-3(2H)-ylidene)acetate 2.21 (34.0 g, 198 mmol, 1.00 equiv) in dry THF (420 mL) was added dropwise. The ice bath was then removed and the solution heated to reflux. TLC analysis (2:1 hexanes/EtOAc) after 3 days showed consumption of the SM and formation of a major product. The reaction was allowed to cool down, placed into an ice bath and quenched with 36.0 mL of water over 5 hours. Once finished, 24.0 mL of
sodium hydroxide (10 M) was added followed by 72.0 mL of water. The resulting mixture was then filtered and the filtrate extracted with (3 x 150 mL) of EtOAc. The organic layers were combined, dried over Na$_2$SO$_4$ and concentrated to afford the crude compound as a yellow oil. The crude material was purified by column chromatography on silica gel (gradient of 30–50% EtOAc in hexanes) to afford the title compound 2.23 (21.4 g, 162 mmol) as a colorless oil in 82% yield.

$^1$H NMR (500 MHz, CDCl$_3$) $\delta = 3.72$ (t, J=6.75 Hz, 2H), 2.97 (dd, J=6.5 Hz, 4 Hz, 1H), 2.88–2.83 (m, 2H), 2.52 (m, 1H), 2.32–2.25 (m, 1H), 2.21–2.15 (m, 1H), 1.77–1.66 (m, 2H), 1.64–1.57 (m, 1H) ppm.

$^{13}$C NMR (125 MHz, CDCl$_3$) $\delta = 61.2, 41.2, 36.7, 36.6, 36.2, 30.6$ ppm.


tert-Butyl 2-(2-(tetrahydrothiophen-3-yl)ethoxy)acetate (2.24).

Prior to the experiment, hexanes (150 mL) and KOH (12 M, freshly prepared, 150 mL) were cooled in an ice bath. Then 2-(tetrahydrothiophen-3-yl)ethanol 2.23 (10.0 g, 75.8 mmol, 1.00 equiv) was dissolved in the cold hexanes in a 1 L flask placed in an ice bath, and vigorously stirred. Cold KOH, the phase transfer catalyst, (((n-Bu)$_4$N)$_2$)SO$_4$ (7.58 mmol, 0.100 equiv) and finally tert-butyl 2-bromoacetate (22.2 g, 16.8 mL, 114 mmol, 1.50 equiv) were added. The reaction was stirred vigorously at 0 °C. After 1 h, TLC analysis (8:2 hexanes/EtOAc) showed consumption of the starting material and formation of a major product. 500 mL of NH$_4$Cl were added and the aqueous phase was extracted with EtOAc (4 x 75.0 mL). The organic layers were combined, dried over Na$_2$SO$_4$, and concentrated to afford the crude compound as a colorless oil. The crude material was purified by column chromatography on silica gel (5% EtOAc in hexanes) to afford compound 2.24 (14.7 g, 59.7 mmol) as a colorless oil in 79% yield.

$^1$H NMR (500 MHz, CDCl$_3$) $\delta = 3.94$ (s, 2H), 3.56 (t, J=7.5 Hz, 2H), 2.97 (dd, J=6.5 Hz, 4.0 Hz, 1H), 2.87–2.84 (m, 2H), 2.50 (t, J=9.2 Hz, 1H), 2.35–2.22 (m, 1H), 2.20–2.14 (m, 1H), 1.80–1.69 (m, 2H), 1.64–1.55 (m, 1H), 1.47 (s, 9H) ppm.

$^{13}$C NMR (125 MHz, CDCl$_3$) $\delta = 169.6, 81.5, 70.5, 68.8, 41.6, 36.6, 33.2, 30.6, 28.1$ ppm.

HR-MS (ESI) calcd. for C$_{12}$H$_{22}$O$_3$S [M+H]$^+$: 246.1290, measured: 246.1302.
2-(2-(Tetrahydrothiophen-3-yl)ethoxy)acetic acid 2.1.

The tert-butyl ester 2.24 (20.0 g, 81.3 mmol) was dissolved in 300 mL of water, and Dowex-50WX2 (10.0 g, previously washed with water) was added. The resulting suspension was stirred at reflux. After 2.5 h, TLC analysis (8:2 hexanes/EtOAc) showed consumption of the starting material and formation of a major product. The reaction was allowed to cool and filtered in order to remove the Dowex resin. The resulting solution was basified with NaHCO₃ (pH>10) and extracted with Et₂O (2 x 70.0 mL). KHSO₄ (2 M solution) was added to the aqueous phase until it reached acidic pH (pH<2) and the resulting solution was extracted with CH₂Cl₂ (4 x 100 mL). The organic layers were combined, dried over Na₂SO₄, and concentrated to afford compound 2.1 (14.7 g, 77.4 mmol) as a pale yellow oil in 95% yield. The linker 2.1 was used without further purification.

1H NMR (400 MHz, CDCl₃) δ = 4.13 (s, 2H), 3.62 (t, J=6.3 Hz, 2H), 2.96 (dd, J=6.5 Hz, 4.0 Hz, 1H), 2.87–2.84 (m, 2H), 2.50 (t, J=9.5 Hz, 1H), 2.31–2.28 (m, 1H), 2.18–2.15 (m, 1H), 1.80–1.74 (m, 2H), 1.62–1.58 (m, 1H) ppm.

13C NMR (125 MHz, CDCl₃) δ = 175.1, 70.9, 67.8, 41.3, 36.6, 33.1, 30.6 ppm.


6.2.2 Typical procedure for the loading and alkylation of the linker 2.1 and attachment of the first amino acid

The tetrahydrothiophene linker 2.1 (285 mg, 1.50 mmol, 3.00 equiv) was preactivated with HCTU (621 mg, 1.50 mmol, 3.00 equiv) and NMM (330 mL, 3.00 mmol, 6.00 equiv) in DMF (10.0 mL) for 5 min. The reaction mixture was transferred to the Rink amide resin (1.56 g, 0.499 mmol, 1 equiv) and the resulting suspension was agitated for 2.5 h. The resin was drained and washed thoroughly with DMF and CH₂Cl₂ (3 times each) and dried. In an oven-dried 20 mL vial covered with aluminum foil, AgBF₄ (487 mg, 2.50 mmol, 4.00 equiv) was dissolved in anhydrous CH₂CN (1.50 mL). ICH₂CN (0.180 mL, 2.50 mmol, 4.00 equiv) was added to the reaction mixture and allowed to react at room temperature for 30 min. The mixture was diluted with anhydrous CH₂CN (2.00 mL) and anhydrous toluene (2.00 mL) and added to a dark coated 40.0 mL vial containing the resin-linker previously swelled in anhydrous toluene (4.50 mL). The vial was purged with N₂ and reacted at 60 °C for 8 hours. The resin was dried and washed with CH₂CN, DMF, CH₂Cl₂, and finally DMF (3 times each).
Fmoc amino acid (2.50 mmol, 5.00 equiv) was preactivated with HCTU (2.35 mmol, 4.70 equiv) and NMM (5.00 mmol, 10.0 equiv) in DMF (10.0 mL) for 5 min. The resulting mixture was transferred to the Rink amide resin-cyanosulfurylide and allowed to react at room temperature for 4 h. The resin was dried and washed with DMF, a solution of sodium diethyldithiocarbamate (0.500 M solution in DMF), DMF, CH$_2$Cl$_2$ (3 times each) and dried. The level of attachment of the Fmoc-amino acid on cyanosulfurylide resin was determined by UV ($\lambda = 304$ nm) by quantification of dibenzofulvene group released after treating with 2% DBU in DMF. We typically obtained loadings ranging from 0.25 to 0.30 mmol/g.

6.2.3 Synthesis of Boc and Fmoc Opr

(S)-(−)-α,α-Diphenyl-2-pyrrolidinemethanol trimethylsilyl ether 2.34.

![Chemical Structure Image]

In a 1-L round bottom flask (S)-(−)-α,α-diphenyl-2-pyrrolidinylmethanol (25.0 g, 98.8 mmol, 1.0 equiv) was dissolved in THF (220 mL). To the resulting solution, imidazole (20.0 g, 294 mmmol, 3.0 equiv) was added in one portion. After complete dilution, the reaction mixture was cooled to 4 °C and trimethylchlorosilane (31.3 mL, 247 mmol, 2.50 equiv) was added dropwise over 20 min. The reaction was stirred for 15 h at room temperature under N$_2$. MTBE (150 mL) was added to the reaction and stirred for an additional 15 min. The mixture was filtered Celite and MTBE (3 x 50 mL) was used to wash the precipitate. The organic layer was separated and washed with H$_2$O (3 x 150 mL), saturated aqueous NaCl (2 x 250 mL), and dried over MgSO$_4$. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 10 h, affording the desired compound 2.34 (29.9 g, 91.9 mmol, 93% yield) as a pale yellow oil

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.54 – 7.46 (m, 2H), 7.42 – 7.36 (m, 2H), 7.35 – 7.20 (m, 6H), 4.07 (t, $J = 7.4$ Hz, 1H), 2.98 – 2.75 (m, 2H), 1.84 – 1.72 (m, 1H), 1.68 – 1.55 (m, 3H), 1.48 – 1.37 (m, 1H), -0.06 (s, 9H) ppm.

$^{13}$C NMR (101 MHz, CDCl$_3$) δ 146.83, 145.78, 128.44, 127.61, 127.57, 127.53, 126.90, 126.73, 83.17, 65.42, 47.16, 27.51, 25.06, 2.20 ppm.

HR-MS (ESI) calcd. for C$_{20}$H$_{28}$NOSi [M+H]$^+$ 326.1935, measured 326.1935.
**tert-Butyl (tert-butyldimethylsilyl)oxycarbamate 2.30.**

In a 2-L round-bottomed flask, N-Boc hydroxylamine (44.0 g, 330 mmol, 1.0 equiv) was dissolved in CH$_2$Cl$_2$ (1.10 L). The solution mixture was cooled to 4 °C in an ice-water bath and Et$_3$N (49.0 mL, 363 mmol, 1.10 equiv) was added dropwise over 15 min. tert-Butyldimethylsilyl chloride (49.7 g, 330 mmol, 1.0 equiv) was dissolved in CH$_2$Cl$_2$ (150 mL) and added dropwise over 60 min at 4 °C (final substrate concentration of 0.25 M). The solution was stirred for 16 h at room temperature under N$_2$.

After completion, H$_2$O (250 mL) was added to the reaction mixture and the organic layer was separated, washed with saturated aqueous NaCl (250 mL), dried over MgSO$_4$ and filtered. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 10 h, affording the desired compound 2.30 (79.0 g, 319 mmol, 96.7% yield) as a white solid.

**1H NMR** (400 MHz, CDCl$_3$) δ 6.67 (s, 1H), 1.48 (s, 9H), 0.95 (s, 9H), 0.162 (s, 6H) ppm.

**13C NMR** (101 MHz, CDCl$_3$) δ 158.5, 82.2, 28.8, 26.5, 18.7, -5.1 ppm.

**HR-MS (ESI)** calcd. for C$_{11}$H$_{25}$NO$_3$Si [M+Na]$^+$ 270.1496, measured 270.1497.

**HR-MS (ESI)** calcd. for C$_{11}$H$_{25}$NO$_3$Si [M+Na]$^+$ 270.1496, measured 270.1497.

(S)-Ethyl 2-((tert-butoxycarbonyl)((tert-butyldimethylsilyl)oxy)amino)-4-oxobutanoate 2.32.

In a 1-L, three-necked, round-bottomed flask equipped with a 250-mL addition funnel, a thermometer fitted with a glass adaptor, and a rubber septum through which nitrogen atmosphere was ensured, (S)-(−)-α,α-diphenyl-2-pyrrolidinemethanol trimethylsilyl ether 2.34 (19.0 g, 58.6 mmol, 0.50 equiv) was dissolved in chloroform (120 mL). The flask was cooled to 4 °C using an ice-water bath. (2E)-4-Oxo-2-butenoate (14.1 mL, 117 mmol, 1.0 equiv) was added dropwise over 15 min, maintaining the internal temperature at 4 °C. tert-Butyl (tert-butyldimethylsilyl)oxycarbamate 3.30 (34.8 g, 141 mmol, 1.20 equiv) was dissolved in chloroform (60 mL), and added dropwise over 1 h. The internal temperature was maintained at 4 °C through the entire course of addition (the final concentration of the substrate was 0.6 M). The addition funnel and the thermometer were removed and the flask was equipped with a glass stopper and a rubber septum. The reaction was stirred for 16 h under N$_2$ at room temperature. After completion, the reaction mixture was transferred to a 1L round-bottomed flask and concentrated by rotary evaporation and dried on the vacuum pump for 5 h. The crude compound was purified by column chromatography (gradient of 5–15% EtOAc in hexanes). The desired fractions were collected and concentrated by rotary evaporation and further dried.
for 5 h under vacuum pump affording the desired compound $\textbf{3.32}$ (21.2 g, 95% ee, 56.1 mmol, 48% yield) as a pale yellow oil

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 9.83 (t, $J = 1.1$ Hz, 1H), 4.91 (dd, $J = 8.1, 5.3$ Hz, 1H), 4.32 – 4.09 (m, 2H), 3.24 (ddd, $J = 17.5, 8.1, 1.3$ Hz, 1H), 2.80 (ddd, $J = 17.6, 5.3, 0.9$ Hz, 1H), 1.49 (s, 9H), 1.36 – 1.19 (m, 3H), 0.93 (s, 9H), 0.17 (d, $J = 7.0$ Hz, 6H) ppm.

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 198.70, 169.05, 157.90, 82.56, 61.65, 60.52, 42.76, 28.17, 28.14, 25.88, 25.86, 17.94, 14.16, -4.71, -4.79 ppm.

HR-MS (ESI) calcd. for C$_{17}$H$_{33}$NO$_6$Si [M+Na]$^+$ 398.1969, measured 398.1968.

Ethyl $N$-(tert-butoxycarbonyl)-$N$-((tert-butylidimethylsilyl)oxy)-L-homoserinate

In a 500-mL, three-necked, round-bottomed flask equipped with a plastic stopper, a low-temperature thermometer and a rubber septum through which nitrogen atmosphere was ensured, aldehyde $\textbf{2.32}$ (15.0 g, 40.0 mmol, 1.0 equiv, 95% ee) was dissolved in MeOH (200 mL, concentration of substrate is 0.20 M) and the reaction mixture cooled to −20 ºC using a CH$_3$CN-dry ice bath. Sodium borohydride (3.02 g, 79.8 mmol, 2.0 equiv) was added in ten portions (~300 mg every three min) and the internal temperature was maintained at −20 ºC. After complete addition, the reaction was stirred for 45 min in the CH$_3$CN-dry ice bath at -20 ºC, after which the reaction was allowed to warm up to 0 ºC. After completion, a mixture of ice-water (170 mL) was added to the solution with vigorous stirring and the solution was stirred for 10 min at 0 ºC. To this mixture EtOAc (900 mL) and H$_2$O (50 mL) were added and the aqueous solution was extracted with EtOAc (3 x 150 mL). The combined organic layers were washed with saturated aqueous NH$_4$Cl (100 mL), saturated aqueous NaCl (100 mL), dried over Na$_2$SO$_4$ and filtered. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 5 h affording the desired product $\textbf{2.37}$ as a pale yellow oil (14.8 g, 39.2 mmol, 98.2%), which was used without further purification.

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 4.47 (dd, $J = 8.8, 5.8$ Hz, 1H), 4.22 (dd, $J = 7.1, 5.1$ Hz, 2H), 3.77 (t, $J = 5.7$ Hz, 2H), 2.50–2.00 (m, 2H), 1.50 (s, 9H), 1.29 (d, $J = 7.2$ Hz, 3H), 0.95 (s, 9H), 0.21 (d, $J = 3.1$ Hz, 6H) ppm.

$^{13}$C NMR (101 MHz, CDCl$_3$): $\delta$ 172.77, 157.40, 83.09, 65.67, 60.86, 60.39, 50.88, 28.20, 28.13, 26.92, 25.90, 24.95, 21.05, 17.98, 14.20, -4.87, -4.90 ppm.

HR-MS (ESI) calcd. for C$_{17}$H$_{36}$NO$_6$Si [M+H]$^+$ 378.2306, measured 378.2310.
**Ethyl (N-(tert-butoxycarbonyl)-N-((tert-butyldimethylsilyl)oxy)-O-(methylsulfonyl))-L-homoserinate 2.39.**

In a 500-mL, three-necked, round-bottomed flask equipped with an addition funnel, a thermometer fitted with a glass adaptor, and a rubber septum through which \( \text{N}_2 \) atmosphere was ensured, the alcohol 2.37 (14.8 g, 39.2 mmol, 1.0 equiv) was dissolved with \( \text{CH}_2\text{Cl}_2 \) (175 mL). The flask was cooled to 4 °C using an ice-water bath and \( \text{Et}_3\text{N} \) (16.6 mL, 119 mmol, 3.0 equiv) was added to the flask dropwise over 15 min maintaining the internal temperature at 3-4 °C. Methanesulfonyl chloride (7.40 mL, 75.7 mmol, 1.93 equiv) was added dropwise over 7 min and the internal temperature was maintained at 4 °C through the entire course of addition (final concentration of substrate 0.2 M). The addition funnel was removed and the flask was equipped with a glass stopper.

The reaction mixture was stirred for 15 min at 4 °C after which the ice-water bath was removed and the reaction was stirred at room temperature for 1 h. After completion, saturated aqueous \( \text{NH}_4\text{Cl} \) (150 mL) was added to the reaction solution and the aqueous layer was separated and extracted with \( \text{CH}_2\text{Cl}_2 \) (4 x 150 mL). The combined organic layers were washed with saturated aqueous \( \text{NH}_4\text{Cl} \) (200 mL), saturated aqueous \( \text{NaHCO}_3 \) (200 mL), saturated aqueous \( \text{NaCl} \) (200 mL), dried over \( \text{Na}_2\text{SO}_4 \) and filtered. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 6 h affording a dark-orange oil 2.39 that was used in the next step with no further purification (17.6 g, 38.6 mmol, 98.1% crude yield).

**\(^1\text{H} \text{NMR} \) (400 MHz, CDCl\(_3\)) \( \delta \)** 4.53 (dd, \( J = 8.5, 6.1 \text{ Hz}, 1\text{H} \)), 4.46 – 4.29 (m, 2H), 4.21 (dd, \( J = 12.8, 7.1 \text{ Hz}, 1\text{H} \)), 3.02 (s, 3H), 2.55 – 2.34 (m, 1H), 2.35 – 2.15 (m, 1H), 1.49 (s, 9H), 1.28 (t, \( J = 7.1 \text{ Hz}, 3\text{H} \)), 0.92 (s, 9H), 0.26 – 0.11 (m, 6H) ppm.

**\(^{13}\text{C} \text{NMR} \) (101 MHz, CDCl\(_3\)) \( \delta \)** 177.7, 162.3, 87.9, 70.5, 65.78, 33.0, 30.8, 29.8, 22.8, 0.03 ppm.

**2-(tert-Butyl) 3-ethyl (S)-isoxazolidine-2,3-dicarboxylate 2.29**

In a 1-L, three-necked, round-bottomed flask equipped with an addition funnel, a thermometer fitted with a glass adaptor, and a rubber septum through which nitrogen atmosphere was ensured, 2.39 (17.6 g, 38.6 mmol, 1.0 equiv) was dissolved in THF (750 mL). The flask was cooled to 4 °C using an ice-water bath and tetrabutylammonium fluoride (1 M in THF, 58.0 mL, 58.0 mmol, 1.50 equiv) was added to the flask dropwise over 90 min maintaining the internal temperature at 3-4 °C. The reaction mixture was stirred for 1 hour at 4 °C. After completion, saturated aqueous \( \text{NaHCO}_3 \) (120 mL) was added to the reaction and stirred for 10 min. The aqueous solution was diluted with \( \text{Et}_2\text{O} \) (300
mL) and the aqueous layer was separated and washed with Et₂O (3 x 150 mL). The combined organic layers were washed with saturated aqueous NaHCO₃ (150 mL), saturated aqueous NaCl (150 mL), dried over Na₂SO₄ and filtered. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 10 h to a yellow oil. Column chromatography (20% EtOAc in hexanes) furnished 2.29 (6.34 g, 25.8 mmol, 65.0% over 3 steps) as a clear yellow oil.

³¹H NMR (400 MHz, CDCl₃) δ 4.66 (ddd, J = 4.8 Hz, 9.0 Hz, 9.4 Hz, 9.0 Hz), 4.21 (q, J = 7.1 Hz, 2H), 4.15–4.08 (m, 1H), 3.81 (ddd, J = 8.0 Hz, 8.0 Hz, 8.0 Hz, 1H), 2.66–2.55 (m, 1H), 2.49–2.38 (m, 1H), 1.48 (s, 9H), 1.28 (t, J = 7.1 Hz, 3H) ppm.

¹³C NMR (101 MHz, CDCl₃): δ 170.6, 155.8, 82.6, 68.5, 61.8, 59.6, 33.2, 28.3, 14.4 ppm.

HRMS (ESI) calcd. for C₁₁H₂₀NO₅ [M+H]⁺ 246.1341, found 246.1331.

(S)-N-Boc-5-oxaproline 2.2.

In a 500-mL, three-necked, round-bottomed flask equipped with an addition funnel, a thermometer fitted with a glass adaptor, and a nitrogen inlet, cyclic compound 2.29 (6.34 g, 25.8 mmol, 1.0 equiv) was dissolved in THF (95 mL). The flask was cooled to 4 ºC using an ice-water bath and a chilled solution of aqueous 1 M LiOH (103 mL, 103 mmol, 4.0 equiv) was added to the flask dropwise over 45 min maintaining the internal temperature at 3-4 ºC. The reaction mixture was stirred for 1.5 h at room temperature. After completion, CHCl₃ (250 mL) was added and the solution was acidified to pH 3 with an aqueous solution of 2 M KHSO₄ (100 mL). The aqueous layer was separated and washed with CHCl₃ (5 x 100 mL), and the combined organic layers were dried over Na₂SO₄ and filtered. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 10 h affording a clear yellow oil 2.2 that solidifies upon standing in a refrigerator (5.39 g, 24.8 mmol, 96.0% yield, 62.4% overall yield, 95% ee).

³¹H NMR (400 MHz, CDCl₃) δ 7.43 (s, 1H), 4.75–4.70 (m, 1H), 4.73 (dd, J = 5.2 Hz, 9.3 Hz, 1H), 3.85 (ddd, J = 8.0 Hz, 8.0 Hz, 8.0 Hz, 1H), 2.72–2.50 (m, 2H), 1.50 (s, 9H) ppm.

¹³C NMR (101 MHz, CDCl₃) δ 174.5, 156.0, 83.2, 68.4, 59.4, 32.7, 28.0 ppm.

HRMS (ESI) calcd. for C₉H₁₅NO₅ [M+Na]⁺ 240.0841, found 240.048.
(S)-N-Fmoc-5-oxaproleine 2.3.

In a 250 mL round-bottomed flask (S)-N-Boc-5-oxaproleine 2.2 (5.0 g, 23 mmol, 1 equiv) was dissolved in 2:1 CH₂Cl₂/TFA (50 mL). The resulting mixture was stirred at rt for 1 h under nitrogen atmosphere. After completion, the solvents were co-evaporated by rotary evaporation with toluene (3 x 50 mL). The resulting solid was dissolved in 3:1 1,4-dioxane/H₂O (100 mL) containing NaHCO₃ (7.7 g, 92 mmols, 4.0 equiv) and Fmoc-OSu (11.6 g, 34.5 mmol, 1.5 equiv) was added in one portion to the reaction mixture and stirred overnight at rt.

After completion the reaction was acidified using 6M HCl until pH 2 was reached. The aqueous layer was separated and extracted with EtOAc (4 x 100 mL), and the combined organic layers were dried over Na₂SO₄ and filtered. The filtrate was concentrated by rotary evaporation and dried on the vacuum pump for 10 h to a yellow oil. Column chromatography (gradient of 0–5% CH₃OH in CH₂Cl₂) furnished 2.3 (6.9 g, 29.1 mmol, 79% yield) as a white solid.

¹H NMR (400 MHz, CDCl₃) δ 7.81–7.78 (d, 2H), 7.68–7.62 (m, 2H), 7.48–7.40 (m, 2H), 7.37–7.32 (m, 2H), 4.73–4.68 (dd, 1H), 4.68–4.60 (dd, 1H), 4.53–4.48 (dd, 1H), 4.34–4.27 (dd, 1H), 4.25–4.17 (m, 1H), 3.85–3.76 (dd, 1H), 2.67–2.56 (m, 2H) ppm.

¹³C NMR (101 MHz, CDCl₃) δ 157.3, 143.3, 143.2, 141.4, 141.3, 128.0, 127.3, 127.2, 125.2, 125.1, 120.0, 69.2, 68.8, 59.8, 46.9, 32.6 ppm.

HR-MS (ESI) calcd. for C₁₉H₁₇NO₅ [M+Na]⁺ 362.1004, measured 362.1017.
6.3 Experimental for “Chapter 3: Chemical Synthesis of the Betatrophin Protein”

6.3.1 Synthesis of the peptide segments.

Synthesis of leucine α-ketoacid segment 1 (H$_2$N-S1-KA) (3.1).

H$_2$N-S1-KA 3.1 was synthesized on Rink amide polystyrene resin (0.36 mmol/g scale: 1.7 g, 0.6 mmol) by first loading protected Leu-KA using HATU and NMM in DMF. We obtained a final loading of 0.30 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS, the N-terminal Fmoc was removed and the resin was washed with DMF and CH$_2$Cl$_2$. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA:DODT:H$_2$O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et$_2$O; centrifugation and drying afforded the crude peptide.

The crude peptide was purified by HPLC using a Shiseido pak C18 column (50 x 250 mm) with a gradient of 15% CH$_3$CN for 7.5 min and 15% CH$_3$CN to 90% CH$_3$CN over 45 min at rt with a flow rate of 40 mL/min to obtain 183 mg of H$_2$N-S1-KA 3.1 in an overall yield of 8.5%

![H$_2$N-S1-KA 3.1](image)

Analytical RP-HPLC traces with a gradient 5% CH$_3$CN for 5 min and 5% CH$_3$CN to 95% CH$_3$CN over 20 min at rt with flow rate of 1 mL/min a) RP-HPLC trace of the crude H$_2$N-S1-KA 3.1. b) RP-HPLC trace of the purified H$_2$N-S1-KA 3.1. c) MALDI-FTICR MS of purified H$_2$N-S1-KA 3.1, calcd for C$_{186}$H$_{300}$N$_{53}$O$_{56}$ [M+H]$^+$: 4230.7805, measured 4230.2211.

Synthesis of oxaproline segment 2 leucine cyanosulfur ylide (Opr-S2-SY) (3.2).

Opr-S2-SY was synthesized by loading the cyanosulfur ylide liker on a Rink amide polystyrene resin (loading 0.36 mmol/g, scale: 1.43 g, 0.50 mmol). Fmoc-Leu was couple to the cyanosulfur ylide resin using HATU and NMM in DMF for 4 h followed by several washings with DMF, sodium diethylthiocarbamate, DMF
and CH₂Cl₂. We obtained a final loading of 0.33 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS the N-terminal Fmoc was removed and Boc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/TIPS/H₂O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et₂O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 10% CH₃CN for 7.5 min and 10% CH₃CN to 90% CH₃CN over 45 min at rt with a flow rate of 40 ml/min to obtain 396 mg of Opr-S2-SY 3.2 in an overall yield of 19%.

Analytical RP-HPLC traces with a gradient of 5% CH₃CN for 5 min and 5% to 95% CH₃CN over 20 min at rt with flow rate 1 mL/min a) RP-HPLC trace of the crude Opr-F2-SY 3.2. b) RP-HPLC trace of the purified Opr-F2-SY 3.2. c) MALDI-FTICR MS of purified Opr-S2-SY 3.2, calcd C₁₇₈H₃₀₂N₇O₅₅S [M+H]⁺: 4152.7775, measured 4152.2339.

Synthesis of Fmoc protected oxaproline segment 3 leucine α-ketoacid (FmocOpr-S3-KA) (3.3).

FmocOpr-S3-KA 3.3 was synthesized on a Rink amide polystyrene resin (0.36 mmol/g scale: 1.2 g, 0.43 mmol) by first loading protected Leu-KA using HATU and NMM in DMF. We obtained a final loading of 0.30 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS the N-terminal Fmoc was removed and Fmoc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/DODT/H₂O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et₂O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC
using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 35% CH$_3$CN for 7.5 min and 35% to 80% CH$_3$CN over 45 min at rt with a flow rate of 40 mL/min to obtain 277 mg of FmocOpr-S3-KA 3.3, in an overall yield of 18%.

Analytical RP-HPLC traces with a gradient of 35% CH$_3$CN for 3 min and 35% to 80% CH$_3$CN over 30 min at rt with flow rate 1 mL/min a) RP-HPLC trace of the crude segment 3 FmocOpr-S3-KA 3.3. b) RP-HPLC trace of the purified FmocOpr-S3-KA 3.3. c) MALDI-FTICR MS of purified FmocOpr-S3-KA 3.3, calcd C$_{191}$H$_{307}$N$_{50}$O$_{65}$[M+H]$^+$: 4343.8415, measured 4343.2424.

Synthesis Fmoc protected oxaproline segment 4 leucine $\alpha$-ketoacid (FmocOpr-S4-KA) (3.4).

FmocOpr-S3-KA 3.4 was synthesized on a Rink amide polystyrene resin (0.36 mmol/g scale: 1.7 g, 0.6 mmol) by first loading protected Leu-KA using HATU and NMM in DMF. We obtained a final loading of 0.30 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS the N-terminal Fmoc was removed and Fmoc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH$_2$Cl$_2$. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/DODT/H$_2$O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et$_2$O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 30% CH$_3$CN for 7.5 min and 30% to 80% CH$_3$CN over 25 min at rt with a flow rate of 40 mL/min to obtain 229 mg of FmocOpr-S4-KA 3.4, in an overall yield of 12%.
Synthesis of Opr segment 5 (Opr-S5-OH) (3.5).

Opr-S5-OH 3.5 was synthesized by loading Fmoc-Ala on a Wang polystyrene resin loading 1.08 mmol/g, scale: 0.5 g, 0.50 mmol) using the symmetrical anhydride method with DIC. We obtained a final loading of 0.25 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection. After automated Fmoc SPPS the N-terminal Fmoc was removed and Boc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/DODT/H₂O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et₂O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 20% CH₃CN for 7.5 min and 20% to 75% CH₃CN over 25 min at rt with a flow rate of 40 mL/min to obtain 269 mg of Opr-S5-OH 3.5, in an overall yield of 14%.
6.3.2 KAHA ligation for the synthesis of betatrophin

Synthesis of H$_2$N-S1S2-KA (3.6).

H$_2$N-S1-KA 3.1 (140 mg, 33.2 µmol, 1.3 equiv) and Opr-S2-SY 3.2 (106 mg, 25.5 µmol, 1.0 equiv) were mixed together and dissolved in 1.28 mL of 8:2 NMP/H$_2$O and 0.1 M oxalic acid (final concentration of 20 mM). The resulting mixture was mixed by vortex until complete solubilization and magnetically stirred at 60 ºC for 8 h.

After 8 h, the ligation mixture was diluted with 1:1 CH$_3$CN/H$_2$O + 0.1% TFA (5.12 mL) and oxone (1.71 g, 0.13 mmol, 5.0 equiv) was added to the solution. The resulting mixture was shaken at rt for 5–10 min and DMS (467 µl, 650 mmol, 50 equiv) was added to quench the reaction. The excess DMS was evaporated and 1:1 CH$_3$CN/H$_2$O + 0.1% TFA (6 mL) added to the solution. After centrifugation, the supernatant was filtered through a microporous membrane filter and purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 30% CH$_3$CN for 7.5 min and 30% to 75% CH$_3$CN over 40 min at rt with a flow rate of 40 mL/min to obtain 103 mg of H$_2$N-S1S2-KA 3.6, in an overall yield of 45%.
Analytical RP-HPLC traces with a gradient of 30% CH₃CN for 3 min and 30% to 75% CH₃CN over 22 min at rt with flow rate 1 mL/min. a) Monitoring of the ligation between H₂N-S1-KA 3.1 and Opr-S2-SY 3.2 and Oxone oxidation of H₂N-S1S2-SY. b) RP-HPLC trace of the purified H₂N-S1S2-KA 3.6. c) MALDI-FTICR of purified Opr-S4-OH 3.5, calcd C₁₃₆H₁₉₇N₁₀₆O₁₁1 [M+H]⁺: 8156.2565, measured 8155.3976.

Synthesis of Opr-S4S5-OH (3.7).

FmocOpr-S4-KA 3.4 (193 mg, 47.0 μmol, 1.50 equiv) and Opr-S5-OH 3.5 (120 mg, 31.0 μmol, 1.0 equiv) were mixed together and dissolved in 1.25 mL of 8:2 NMP/H₂O and 0.1 M oxalic acid (final concentration of 20 mM). The resulting mixture was mixed by vortex until complete solubilization and magnetically stirred at 60 °C for 8 h.
After 8 h, to the reaction mixture was added 2:2:1 NH₃OH/H₂O/DMSO pH 13 (15.8 mL) at rt for 30 minutes, and the ligation mixture was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 25% CH₃CN for 7.5 min and 25% to 85% CH₃CN over 23 min at rt with a flow rate of 40 mL/min to obtain 127 mg of Opr-S4S5-OH 3.7 in an overall yield of 53%.

Analytical RP-HPLC traces with a gradient of 20% CH₃CN for 3 min and 20% to 95% CH₃CN over 22 min at rt with flow rate 1 mL/min. a) Monitoring of the ligation between FmocOpr-S4-KA 3.4 and Opr-S5-OH 3.5 followed by Fmoc deprotection and O-N shift in a one pot fashion. b) RP-HPLC trace of the purified Opr-S4S5-OH 3.7. c) MALDI-FTICR of purified Opr-S4-OH 3.5, calcd C₃₉₉H₆₄₂N₁₁₁O₉₀ [M+H]+: 7614.7675, measured 7613.1773.

Synthesis of Opr-S3S4S5-OH (3.8).

FmocOpr-S3-KA 3.3 (34 mg, 7.8 µmol, 1.50 equiv) and Opr-S4S5-OH 3.7 (40 mg, 5.2 µmol, 1.0 equiv) were mixed together and dissolved and dissolved in 350 µL of 8:2 NMP/H₂O and
0.1 M oxalic acid (final concentration of 20 mM). The resulting mixture was mixed by vortex until complete solubilization and placed in a heated shaker at 60 °C for 15 h.

After 15 h, 2:2:1 NH₃OH/H₂O/DMSO (2.7 mL, pH 13) was added to the reaction mixture at rt for 1 h, and the ligation mixture was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 30% CH₃CN for 7.5 min and 30% to 80% CH₃CN over 25 min at rt with a flow rate of 40 mL/min to obtain 29–32 mg of Opr-S3S4S5-OH 3.8 in an overall yield of 47–52%.

Analytical RP-HPLC traces with a gradient of 30% CH₃CN for 3 min and 30% to 80% CH₃CN over 22 min at rt with flow rate 1 mL/min. a) Monitoring of the ligation between FmocOpr-S3-KA 3.3 and Opr-S4S5-OH 3.6 followed by Fmoc deprotection and O-N shift in a one pot fashion. b) RP-HPLC trace of the purified Opr-S3S4S5-OH 3.8. c) MALDI-FTICR of purified Opr-S3S4S5-OH 3.8, calcd C₅₁₄H₈₄O₁₉₁N₁₈₁O₁₅₁ [M+H]+: 11690.3420, measured 11691.3495.

Synthesis of betatrophin (3.9).
H$_2$N-S1S2-KA 3.6 (10.5 mg, 1.29 µmol, 1.50 equiv) and Opr-S3S4S5-OH 3.8 (10 mg, 0.86 µmol, 1.0 equiv) were mixed together and dissolved in 200 µL of 8:2 NMP/H$_2$O and 0.1 M oxalic acid (final concentration of 10 mM). The resulting mixture was mixed by vortex until complete solubilization and placed in a heated shaker at 60 ºC for 24h.

After 24 h, NaHCO$_3$ buffer pH 9.5 (800 µL) was added to the reaction mixture at rt for 2 h, and the ligation mixture was purified by HPLC using a Shiseido Capcell pak C18 column (20 X 250 mm) with a gradient of 30% CH$_3$CN for 5.0 min and 30% to 80% CH$_3$CN over 40 min at rt with a flow rate of 10 mL/min to obtain 5.6 mg of betatrophin 3.9 in an overall yield of 33%.

Analytical RP-HPLC traces with a gradient of 30% CH$_3$CN for 3 min and 30% to 80% CH$_3$CN over 30 min at rt with flow rate 1 mL/min. a) Monitoring of the ligation between H$_2$N-S1S2-KA 3.6 and Opr-S3S4S5-OH 3.8 followed by O-N shift in a one pot fashion. b) RP-HPLC trace of the purified betatrophin 3.9. c) MALDI-FTICR of purified betatrophin 3.9, calcd C$_{869}$H$_{1426}$N$_{269}$O$_{260}$ [M+H]$^+$: 19802.5895, measured 19801.6021

6.3.3 CD Spectra

In collaboration with Prof. Michael Ploug (University of Copenhagen), we measured CD spectra at pH 4, 5, 6 and 7. For all these measurements the procedure was the same:

The lyophilized betatrophin 3.9 was dissolved in 6M GdnHCl pH (4,5, 6, or 7) and dialyzed against 10 mM acetate buffer containing 4M GdnHCl (pH 4,5, 6, or 7) for 5 hours at 4 ºC. After 5 hours, the buffer was changed to 10 mM acetate buffer containing 2M GdnHCl (pH 4,5, 6, or 7) and dialyzed again for 5 hours at 4 ºC. The exchange buffer was changed one last time to 10 mM acetate buffer pH (4,5, 6, or 7) without GdnHCl and dialyzed for 10 hours at 4 ºC.

Although the protein was perfectly soluble at pH 4 and 5 resulting in CD spectra showing a strong α-helix fingerprint, it was poorly soluble at pH 6 and 7. The CD spectra recorded at pH
7, despite showing two minima at 208 and 223 nm, cannot guarantee the nature of the folded state of betatrophi n 3.9.

CD spectra of betatrophin 3.9. a) CD spectra at pH 4, green line measure at 4 °C and blue line measure at rt. b) CD spectra at pH 7.
6.4 Experimental for “Chapter 4: Chemical Synthesis of the Antiviral Integral Membrane Protein IFITM3”

6.4.1 Synthesis of the peptide segments.

Synthesis of arginine α-ketoacid segment 1 (H$_2$N-S1-ArgKA) (4.1).

H$_2$N-S1-KA 4.1 was synthesized by loading the cyanosulfur ylide linker on a Rink amide polystyrene resin (loading 0.36 mmol/g, scale: 1.0 g, 0.36 mmol) FmocArg was coupled to the cyanosulfur ylide resin using HATU and NMM in DMF for 4 h, followed by several washings with DMF, sodium diethyldithiocarbamate, DMF and CH$_2$Cl$_2$. We obtained a final loading of 0.33 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS, the N-terminal Fmoc was removed, and the resin washed with DMF and CH$_2$Cl$_2$. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/TIPS/H$_2$O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et$_2$O; centrifugation and drying afforded the crude peptide.

The crude peptide (100 mg) was dissolved 1:1 H$_2$O/CH$_3$CN + 0.1% TFA (4 mL) and after complete dissolution, oxone (588 mg, 0.96 mmol, 50 equiv) was added. The reaction mixture was gently agitated for 5 min and the reaction was quenched with 250 µL of DMS. The excess DMS was evaporated using a N$_2$ flow and the resulting mixture was re-dissolved, centrifuged and purified by preparative HPLC using a Shiseido Capcell pak C18 column (20 x 250 mm) with a gradient of 10% CH$_3$CN for 5 min and 10% to 65% CH$_3$CN over 30 min at rt with a flow rate of 10 mL/min to obtain 34 mg of H$_2$N-S1-KA 4.1 in an overall yield of 35%.

Analytical RP-HPLC traces with a gradient of 5% CH$_3$CN for 3 min and 5% to 98% CH$_3$CN over 20 min at rt with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude H$_2$N-S1-KA 4.1. b) RP-HPLC trace of the purified H$_2$N-S1-KA 4.1. c) MALDI-TOF MS of purified H$_2$N-S1-KA 4.1, calcd for C$_{238}$H$_{363}$N$_{66}$O$_{72}$ [M+H]$^+$: 5300.8273, measured 5300.6241.
Synthesis of Fmoc protected oxaproline segment 2 phenylalanine α-ketoacid (FmocOpr-S2-KA) (4.2).

FmocOpr-S2-KA 4.2 was synthesized on a Rink amide polystyrene resin (0.36 mmol/g scale: 1.0 g, 0.36 mmol) by first loading protected Leu-KA using HATU and NMM in DMF. We obtained a final loading of 0.25 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection. After automated Fmoc SPPS the N-terminal Fmoc was removed, and Fmoc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA:DODT:H₂O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et₂O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 25% CH₃CN for 7.5 min and 25% to 85% CH₃CN over 35 min at rt with a flow rate of 40 mL/min to obtain 175 mg of FmocOpr-S2-KA 4.2, in an overall yield of 30%.

Analytical RP-HPLC traces with a gradient of 20% CH₃CN for 3 min and 20% to 96% CH₃CN over 30 min at rt with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude FmocOpr-S2-KA 4.2. b) RP-HPLC trace of the purified FmocOpr-S2-KA 4.2. c) MALDI-TOF MS of purified FmocOpr-S2-KA 4.2, calcd for C₁₁₄H₁₅₁N₂₂O₃₂ [M+H]⁺: 2341.5464, measured 2341.6612

Synthesis of Fmoc protected oxaproline segment 3 lysine hydrazide (FmocOpr-S3-NHNH₂) SI.4.3.

FmocOpr-S3-NHNH₂ was synthesized on chlorotrityl resin (loading 0.45 mmol/g scale: 0.85 g, 0.38 mmol) following the peptide hydrazide method described by Liu and coworkers.¹⁷³ After automated Fmoc SPPS, the N-terminal Fmoc was removed and Fmoc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA:TIPS:H₂O,
filtered, and the filtrate was evaporated under vacuum. The peptide segment was precipitated with Et<sub>2</sub>O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 x 250 mm) with a gradient of 25% CH<sub>3</sub>CN for 7.5 min and 25% CH<sub>3</sub>CN to 95% CH<sub>3</sub>CN over 35 min at rt with a flow rate of 40 mL/min to obtain 130 mg of FmocOpr-S3-NHNNH<sub>2</sub> **SI.4.3** in an overall yield of 10%.

The purified FmocOpr-S3-NHNNH<sub>2</sub> **SI.4.3** (80 mg, 19.3 mmols) was dissolved using 0.2 M phosphate solution containing 6 M GdnHCl (pH 3.0-3.1) and cooled to −15 °C using a salt/ice bath. A solution of 0.5 M NaNO<sub>2</sub> was added dropwise to the solution and stirred at −15 °C for 15 min. After 15 min, 3.5 mL of a solution of 0.25 M 3-mercaptopropionic acid (MPA) in 0.2 M phosphate solution containing 6 M GdnHCl (pH 8) was added and the reaction mixture was allowed to warm to rt. Once at rt, the pH was adjusted to 6 using a 0.25 M NaOH solution and the solution stirred for 2 h at rt. After 2 h, the resulting mixture was diluted with H<sub>2</sub>O and 50 mg of TCEP was added. The resulting mixture was directly purified by HPLC using a Shiseido Capcell pak C18 column (50 x 250 mm) with a gradient of 25% CH<sub>3</sub>CN for 7.5 min and 25% CH<sub>3</sub>CN to 95% CH<sub>3</sub>CN over 35 min at rt with a flow rate of 40 mL/min to obtain 43 mg of FmocOpr-S3-COSR **4.3** in 53% yield.

**Fmoc deprotection:** The lyophilized powder was dissolved in 8.0 mL of dry DMSO and 100 µL of DBU was added. After 30 seconds, the reaction was quenched with a pre-cooled
TFA/H₂O (1:1) solution and purified by HPLC using Shiseido Capcell pak C18 column (20 x 250 mm) with a gradient of 25 to 95% CH₃CN (with 0.1% TFA) in 35 min to obtain 18 mg of the desired Opr-S3-COSR 4.3 in 45% yield.

Analytical RP-HPLC traces with a gradient of 5% CH₃CN for 3 min and 5% CH₃CN to 95% CH₃CN over 20 min at rt with a flow rate of 1 mL/min. a) RP-HPLC trace and MALDI-TOF MS of purified FmocOpr-S3-COSR, calcd for C₁₉₀H₂₈₇N₄₈O₅₅S₃ [M+H]⁺: 4219.7948, measured 4220.0014. b) RP-HPLC trace and MALDI-TOF MS of purified Opr-S3-COSR 4.3, calcd for C₁₇₅H₂₇₇N₄₈O₅₃S₃ [M+H]⁺: 3997.5561, measured 3997.7231

Synthesis of the cysteine segment 4 (Cys-S4-OH) (SI.4.4).

For the first attempt, Cys-S4-OH was purchased – prepared by Boc SPPS, already cleaved and ready for purification – from a commercial peptide supplier. The segment is highly hydrophobic and had to be dissolved in a mixture of 8:1:1 HFIP:CH₃CN:H₂O before being directly injected into the HPLC.

Analytical HPLC using Vydac C4 column (4.6 x 250 mm) 80 to 95 % CH₃CN in 20 min at 60 °C displayed numerous broad peaks. After 2 purifications, the trace was still not exploitable and the recovery after every purification step was dramatically low. From 150 mg of crude sample we obtained 2.5 mg of a peptide mixture after 2 purification steps.
Analytical HPLC trace with a gradient of 80% CH$_3$CN to 95% CH$_3$CN over 20 min at 60 $^\circ$C with a flow rate of 1 mL/min. a) HPLC trace of the crude peptide Cys-S4-OH purchased from CS-Bio. b) HPLC trace of the Cys-S4-OH after 2 purifications.

The synthesis of Cys-S4-OH by Fmoc SPPS was attempted on an HMPB-chemMatrix® resin (loading 0.35 mmol/g, scale: 1.0 g, 0.35 mmol). After elongation to residue Val$_{123}$, the resin was washed with DMF, CH$_2$Cl$_2$, dried and a micro-cleavage using 95:2.5:2.5 TFA:TIPS:H$_2$O was performed. After filtration, the filtrate was evaporated under vacuum and the peptide segment was precipitated with Et$_2$O; centrifugation and drying afforded the crude peptide. Analytical HPLC using a Shiseido Capcell pak C18 column (4.6 x 250 mm) with a gradient of 30% CH$_3$CN for 3 min and 35% CH$_3$CN to 95% CH$_3$CN (with 0.1% TFA) in 30 min showed a large number of impurities for a 10 residue peptide, making this strategy unsustainable.

Synthesis of the cysteine segment 4 arginine tag (Cys-S4-ArgTAG) (4.4).

Cys-S4-ArgTAG 4.4 was synthesized on a Rink Amide polystyrene resin (loading 0.32 mmol/g, scale: 1.0 g, 0.32 mmol) by coupling a first Fmoc-Gly residue followed by seven FmocArg(Pbf) residues using HCTU, NMM in DMF. After N-terminal Fmoc deprotection, 4-(hydroxymethyl)benzoic acid was manually coupled using HATU and NMM in DMF for 4 h and Fmoc-Gly-OH was attached to the resin via a double coupling using the symmetrical anhydride method with DIC in CH$_2$Cl$_2$/DMF for 2 h. We obtained a final loading of 0.30 mmol/g, determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS the N-terminal Fmoc was removed, and the resin was washed with DMF and CH$_2$Cl$_2$. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/TIPS/H$_2$O, filtered and the filtrate evaporated under vacuum. The peptide segment was
Chapter 6. Experimental

Precipitated with Et₂O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 40% CH₃CN for 7.5 min and 40% to 98% CH₃CN over 35 min at 60 °C with a flow rate of 40 mL/min to obtain 178 mg of Cys-S3-ArgTAG 4.4 in an overall yield of 11%.

Analytical RP-HPLC traces with a gradient of 35% CH₃CN for 3 min and 35% to 98% CH₃CN over 30 min at 60 °C with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude Cys-S4-ArgTAG 4.4. b) RP-HPLC trace of the purified Cys-S4-ArgTAG 4.4. c) MALDI-TOF MS of purified Cys-S4-ArgTAG 4.4, calcd for C₂₁₂H₃₅₇N₆₂O₅₃S [M+H]+: 4494.5570, measured 4494.2199.

Synthesis of the photo-oxaproline segment 2 valine α-ketoacid (NpecOpr-S2-ValKA) (4.14).

NpecOpr-S2-KA 4.14 was synthesized on a Rink amide polystyrene resin (0.36 mmol/g scale: 1.7 g, 0.6 mmol) by first loading protected Val-KA using HATU and NMM in DMF. We obtained a final loading of 0.30 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection. After automated Fmoc SPPS, the N-terminal Fmoc was removed, and NpecOpr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/DODT/H₂O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et₂O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 20% CH₃CN for 7.5 min and 20% to 85% CH₃CN over 35 min at 60 °C with a flow rate of 40 mL/min to obtain 180 mg of NpecOpr-S2-KA 4.14, in an overall yield of 14%.
Analytical RP-HPLC traces with a gradient of 20% CH$_3$CN for 3 min and 20% to 98% CH$_3$CN over 25 min at 60 °C with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude Npec-S2-KA 4.14. b) RP-HPLC trace of the purified Npec-S2-KA 4.14. c) MALDI-TOF MS of purified Npec-S2-KA 4.14, calcd for C$_{237}$H$_{353}$N$_{88}$O$_{66}$S$_2$ [M+H]$^+$: 5166.8180, measured 5166.5499.

Synthesis of the Oxaproline segment 3 arginine tag (Opr-S3-ArgTAG) (4.15).

Opr-S3-ArgTAG 4.15 was synthesized on a Rink Amide polystyrene resin (0.32 mmol/g scale: 1.7 g, 0.6 mmol) by coupling a first Fmoc-Gly residue, followed by seven Fmoc(Pbf) residues, using HCTU and NMM in DMF. After N-terminal Fmoc deprotection, 4-(hydroxymethyl)benzoic acid was manually coupled using HATU and NMM in DMF for 4 h and Fmoc-Gly-OH was attached to the resin via a double coupling using the symmetrical anhydride method with DIC in CH$_2$Cl$_2$/DMF for 2 h. We obtained a final loading of 0.30 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS, the N-terminal Fmoc was removed, and Boc-Opr was manually coupled using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH$_2$Cl$_2$. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/TIPS/H$_2$O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et$_2$O; centrifugation and drying afforded the crude peptide. The crude peptide was purified by HPLC using a Shiseido Capcell pak C18 column (50 X 250 mm) with a gradient of 40% CH$_3$CN for 7.5 min and 40% to 98% CH$_3$CN over 35 min at 60 °C with a flow rate of 40 mL/min to obtain 128 mg of Opr-S3-ArgTAG 4.15, in an overall yield of 7%.

Analytical RP-HPLC traces with a gradient of 35% CH$_3$CN for 3 min and 35% to 95% CH$_3$CN over 30 min at 60 °C with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude Opr-S3-ArgTAG 4.15. b) RP-HPLC trace of the purified Opr-S3-ArgTAG 4.15. c) MALDI-TOF MS of purified Opr-S3-ArgTAG 4.15, calcd for C$_{257}$H$_{428}$N$_{72}$O$_{66}$S [M+H]$^+$: 5442.6639, measured 5542.3129.
Synthesis of phosphorylated arginine α-ketoacid segment 1 (Phosphorylated-S1-KA) (4.21).

H$_2$N-S1-KA 4.1 was synthesized by loading the cyanosulfur ylide linker on a Rink amide polystyrene resin (loading 0.36 mmol/g, scale: 1.0 g, 0.36 mmol). FmocArg was coupled to the cyanosulfur ylide resin using HATU and NMM in DMF for 4 h, followed by several washings with DMF, sodium diethyldithiocarbamate, DMF and CH$_2$Cl$_2$. We obtained a final loading of 0.33 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS up to Glu 21 and N-terminal Fmoc deprotection, Fmoc-Tyr(PO(OBzl)OH)-OH was manually coupled using HATU and NMM in DMF for 4 h. The resin was again subjected to automated Fmoc SPPS until completion of the segment, followed by washing with DMF and CH$_2$Cl$_2$. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/TIPS/H$_2$O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et$_2$O; centrifugation and drying afforded the crude peptide.

The crude peptide (100 mg) was dissolved 1:1 H$_2$O/CH$_3$CN + 0.1% TFA (4 mL) and after complete dissolution, oxone (597 mg, 0.97 mmol, 50 equiv) was added. The reaction mixture was gently agitated for 5 min and the reaction was quenched with 250 µL of DMS. The excess of DMS was evaporated using a N$_2$ flow and the resulting mixture was re-dissolved, centrifuged and purified by preparative HPLC using a Shiseido Capcell pak C18 column (20 x 250 mm) with a gradient of 10% CH$_3$CN for 5 min and 10% to 65% CH$_3$CN over 30 min at rt with a flow rate of 10 mL/min to obtain 24 mg of phosphorylated-S1-KA 4.21 in an overall yield of 25%.

Analytical RP-HPLC traces with a gradient of 5% CH$_3$CN for 3 min and 5% to 98% CH$_3$CN over 20 min at 60 ºC with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude phosphorylated-S1-KA 4.21. b) RP-HPLC trace of the purified phosphorylated-S1-KA 4.21. c) MALDI-TOF MS of purified phosphorylated-S1-KA 4.21, calcd for C$_{238}$H$_{386}$N$_{66}$O$_{75}$P [M+H]$^+$: 5380.8072, measured 5380.6671.
Synthesis of fluorescent arginine α-ketoacid segment 1 (Fluorescent-S1-KA) (4.38).

Fluorescent-S1-KA 4.38 was synthesized by loading the cyanosulfur ylide linker on a Rink amide polystyrene resin (loading 0.36 mmol/g, scale: 1.0 g, 0.36 mmol). FmocArg was coupled to the cyanosulfur ylide resin using HATU and NMM in DMF for 4 h followed by several washings with DMF, sodium diethyldithiocarbamate, DMF and CH₂Cl₂. We obtained a final loading of 0.33 mmol/g determined by UV measurement at 304 nm upon Fmoc deprotection.

After automated Fmoc SPPS, the N-terminal Fmoc was removed, and 5(6)-carboxyfluorescein was manually coupled to the resin using HATU and NMM in DMF for 4 h, followed by washing with DMF and CH₂Cl₂. The dried resin was subjected to cleavage using 95:2.5:2.5 TFA/TIPS/H₂O, filtered and the filtrate evaporated under vacuum. The peptide segment was precipitated with Et₂O; centrifugation and drying afforded the crude peptide.

The crude peptide (100 mg) was dissolved 1:1 H₂O/CH₃CN + 0.1% TFA (4 mL) and after complete dissolution, oxone (541 mg, 0.88 mmol, 50 equiv) was added. The reaction mixture was gently agitated for 5 min and the reaction was quenched with 250 µL of DMS. The excess DMS was evaporated using a nitrogen flow and the resulting mixture was redissolved, centrifuged and purified by preparative HPLC using a Shiseido Capcell pak C18 column (20 x 250 mm) with a gradient of 10% CH₃CN for 5 min and 10% to 65% CH₃CN over 30 min at rt with a flow rate of 10 mL/min to obtain 13 mg of fluorescent-S1-KA 4.38 in an overall yield of 14%.

Analytical RP-HPLC traces with a gradient of 5% CH₃CN for 3 min and 5% to 98% CH₃CN over 20 min at 60 °C with a flow rate of 1 mL/min. a) RP-HPLC trace of the crude Fluorescent-S1-KA 4.38. b) RP-HPLC trace of the purified fluorescent-S1-KA 4.38. c) MALDI-TOF MS of purified Fluorescent-S1-KA 4.38, calcd for C₂₅₉H₃₇₃N₆₆O₇₈ [M+H]⁺: 5659.1278, measured 5659.3712.
6.4.2 Synthesis of IFITM3 via the KAHA-NCL strategy

6.4.2.1 KAHA ligations with a thioester moiety

Synthesis of Opr-S2S3-COSR (4.5).

FmocOpr-S2-PheKA 4.2 (17.5 mg, 7.50 µmol, 2.00 equiv) and Opr-S3-COSR 4.3 (15 mg, 3.75 µmol, 1.00 equiv) were mixed together and dissolved in 250 µL of NMP/H₂O 8:2 and 0.1 M oxalic acid. No TCEP was added. The resulting mixture was mixed by vortex until complete solubilization and placed in a 60 °C shaker for 8 h.

After completion, the reaction mixture was dissolved in 1.5 mL of NMP and directly purified by RP-HPLC using a Shiseido Capcell pak C18 column (20 x 250 mm) with a gradient of 20% CH₃CN for 5.0 min and 20% to 85% CH₃CN over 30 min at rt with a flow rate of 10 ml/min to obtain 14.3 mg of FmocOpr-S2S3-COSR Fmoc-4.6, in an overall yield of 63%.

FmocOpr-S2S3-COSR Fmoc-4.6 (14.3 mg, 2.27 µmol) was dissolved in DMSO (4.50 mL), and DBU (45 µl) was added to the solution. After gently shaking the reaction for 30 seconds, the reaction was quenched by addition of cold 1:1 H₂O/TFA (1.00 mL). The resulting mixture was directly purified by HPLC to afford 7.0 mg of the ligated Opr-S2S3-COSR 4.5 in 50% yield.
Analytical RP-HPLC traces with a gradient of 20% CH₃CN for 3 min and 20% to 85% CH₃CN over 30 min at rt with flow rate 1 ml/min. a) Monitoring of the ligation between FmocOpr-S2-KA 4.2 and Opr-S3COSR 4.3. b) RP-HPLC trace of the purified FmocOpr-S2S3-COSR Fmoc-4.5. c) MALDI-TOF of purified FmocOpr-S2S3-COSR Fmoc-4.5, calcd C₂₈₈H₄₂₇N₇₀O₅₃S₃ [M+H]^+: 6294.0856, measured 6294.2395. d) RP-HPLC trace of the purified Opr-S2S3-COSR 4.5. e) MALDI-TOF of purified Opr-S2S3-COSR 4.5, calcd C₂₇₃H₄₁₇N₇₀O₅₁S₃ [M+H]^+: 6071.8469, measured 6071.7691

Synthesis of H₂N-S1S2S3-COSR (4.6).

H₂N-S1-ArgKA 4.1 (12.2 mg, 2.30 µmol, 2 equiv) and of Opr-S2S3-COSR 4.5 (7.0 mg, 1.15 µmol, 1 equiv) were mixed together and dissolved in 76.5 µL of NMP/H₂O 8:2 and 0.1 M oxalic
acid. No TCEP was added. The resulting mixture was mixed by vortex until complete solubilization and then placed in a 60 °C shaker for 15 h.

After completion, the reaction mixture was dissolved with 1.0 mL of NMP and directly purified by RP-HPLC using a Shiseido Capcell pak C18 column (20 x 250 mm) with a gradient of 15% CH₃CN for 5 min and 15% CH₃CN to 85% CH₃CN over 35 min at rt with a flow rate of 10 ml/min to obtain 4.85 mg of H₂N-S₁S₂S₃-COSR 4.6 in an overall yield of 37%.

![Analytical RP-HPLC traces](image)

Analytical RP-HPLC traces with a gradient of 20% CH₃CN for 3 min and 20% to 85% CH₃CN over 30 min at rt with flow rate 1 mL/min. a) Monitoring of the ligation between H₂N-S₁-KA 4.1 and Opr-S₂S₃-COSR 4.5. b) RP-HPLC trace of the purified H₂N-S₁S₂S₃-COSR 4.6. c) MALDI-TOF of purified H₂N-S₁S₂S₃-COSR 4.6, calcd C₅₁₀H₇₇₈N₁₃₆O₁₅₁S₃ [M+H]^+: 11327.6573, measured 11327.8325.

6.4.2.2 Native Chemical Ligation as the final ligation for the synthesis of IFITM3

With a small amount of H₂N-S₁S₂S₃-COSR in hand, we attempted the NCL in order to generate IFITM3-ArgTAG. However, we obtained a complicated mixture of products by HPLC. In order to understand what the issue was as well as to identify conditions for the final ligation by NCL, we first tested the ligation of segments Opr-S₃-COSR 4.3 and Cys-S₄-ArgTAG 4.4. Unfortunately, here again, we were not able to achieve this ligation. The details are provided below.
Many conditions were tried for this reaction, which are summarized in the following tables. Test reactions were performed on 2 mg scale with 1 equiv of 4.4 and 1.2 equiv of 4.3. The concentration was kept at 2 mM; 1 mM was also tried and gave a similar outcome. The thio-additive used in all these cases was MPAA. We also tried thiophenol but these conditions were showing poor solubility. The mixture of 8 M urea and organic co-solvent was kept the same for both TFE and HFIP with a ratio of 9:1 8 M urea/organic solvent.

For all ligation tests, a similar HPLC pattern was observed. Only the intensity of the peak changed, depending on which side product was formed.

Typical HPLC trace of NCL between Opr-S3-COSR 4.3 and Cys-S4-ArgTAG 4.4. The three peaks rising from the decomposition of 4.3 could be analyze by MALDI-TOF and identified as the MPAA thioester adduct 4.7, the hydrolyzed product 4.8 and the cyclic amide 4.9.
**pH 6.8 2 mM of 4.4** (Similar results at pH 7.0)

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Temp</th>
<th>Solubility</th>
<th>Products after 5 h</th>
<th>Products after 15 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>rt</td>
<td>No</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>40 °C</td>
<td>No</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>rt</td>
<td>No</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>40 °C</td>
<td>No</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>rt</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>40 °C</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>rt</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>40 °C</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
</tbody>
</table>

**pH 7.2, 2 mM of 4.4**

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Temp</th>
<th>Solubility</th>
<th>Products after 5 h</th>
<th>Products after 15 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>rt</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7 and 4.8</td>
</tr>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7, 4.8</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>rt</td>
<td>No</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>rt</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>40 °C</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>rt</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>40 °C</td>
<td>Yes</td>
<td>Starting Material</td>
<td>Starting Material</td>
</tr>
</tbody>
</table>
**pH 7.5, 2 mM of 4.4**

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Temp</th>
<th>Solubility</th>
<th>Products after 5 h</th>
<th>Products after 15 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>rt</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7 and 4.8</td>
</tr>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>rt</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7 and 4.8</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.3 and 4.7</td>
<td>4.3, 4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>rt</td>
<td>Yes</td>
<td>4.3 and 4.7</td>
<td>4.3 and 4.7</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>40 °C</td>
<td>Yes</td>
<td>4.3 and 4.7</td>
<td>4.3 and 4.7</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>rt</td>
<td>Yes</td>
<td>4.3 and 4.7</td>
<td>4.3 and 4.7</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>40 °C</td>
<td>Yes</td>
<td>4.3 and 4.7</td>
<td>4.3 and 4.7</td>
</tr>
</tbody>
</table>

**3.4 pH 7.6, 2 mM of 4.4**

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Temp</th>
<th>Solubility</th>
<th>Products after 5 h</th>
<th>Products after 15 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>rt</td>
<td>No</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7 and 4.8</td>
</tr>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>rt</td>
<td>No</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7 and 4.8</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>rt</td>
<td>Yes</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7 and 4.8</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>40 °C</td>
<td>Yes</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>rt</td>
<td>Yes</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7 and 4.8</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>40 °C</td>
<td>Yes</td>
<td>4.3, 4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
</tbody>
</table>
pH 7.7, 2 mM of 4.4

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Temp</th>
<th>Solubility</th>
<th>Products after 5 h</th>
<th>Products after 15 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>rt</td>
<td>No</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>GnHCl 6 M, Na₂HPO₄ 0.2 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>rt</td>
<td>No</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M</td>
<td>40 °C</td>
<td>No</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>rt</td>
<td>Yes</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/HFIP</td>
<td>40 °C</td>
<td>Yes</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>rt</td>
<td>Yes</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
<tr>
<td>Urea 8 M/TFE</td>
<td>40 °C</td>
<td>Yes</td>
<td>4.7 and 4.8</td>
<td>4.7, 4.8 and 4.9</td>
</tr>
</tbody>
</table>
6.4.3 Synthesis of IFITM3-ArgTAG by KAHA ligation

6.4.3.1 KAHA ligation of the peptide segments

Synthesis of Opr-S2S3-ArgTAG (4.16).

NpecOpr-S2-KA 4.14 (120 mg, 23.2 μmol, 1.50 equiv) and Opr-S3-ArgTAG 4.15 (84 mg, 15.5 μmol, 1.00 equiv) were mixed together and dissolved in 775 μL of NMP/H$_2$O 95:5 and 0.1 M oxalic acid and 0.1 M TCEP. The resulting solution was mixed by vortex until complete solubilization and placed in a 60 °C shaker for 20 h.

After completion, the reaction mixture was dissolved with 15.0 mL of CH$_3$CN/H$_2$O (1:1) and 200 μL of DODT was added. Using a UV lamp, the resulting mixture was stirred under 365 nm UV irradiation for 45 min. After complete deprotection, the mixture was directly purified by HPLC using a Shiseido Capcell pak UG 80 C18 column (50 x 250 mm) with a gradient of 35% CH$_3$CN for 7.5 min and 35% CH$_3$CN to 98% CH$_3$CN over 30 min at 60 °C with a flow rate of 40 ml/min to obtain 60.3 mg of Opr-S2S3-ArgTAG 4.16 in an overall yield of 37%.
Analytical RP-HPLC traces with a gradient of 35% CH₃CN for 3 min and 35% to 95% CH₃CN over 30 min at 60°C with flow rate 1 mL/min. a) Monitoring of the ligation between Npec-S2-KA 4.14 and Opr-S3-ArgTAG 4.15 followed by UV deprotection in a one pot fashion. b) RP-HPLC trace of the purified Opr-S2S3-ArgTAG 4.16. c) ESI-FTICR of purified Opr-S2S3-ArgTAG 4.16, calcd C₄₈H₇₁N₁₃₂O₁₂₁S₃ [M+H]+: 10471.3068, measured 10470.88.

Synthesis of H₂N-IFITM3-ArgTAG (4.17).
H$_2$N-S1-KA 4.1 (37.5 mg, 7.01 µmol, 1.50 equiv) and Opr-S2S3-ArgTAG 4.16 (50.0 mg, 4.71 µmol, 1.00 equiv) were mixed together and dissolved in 314 µL of NMP/H$_2$O 95:5 and 0.1 M oxalic acid and 0.1 M TCEP. The resulting mixture was mixed by vortex until complete solubilization and placed in a 60 °C shaker for 20 h. After completion, the reaction mixture was dissolved with 10.0 mL of NaHCO$_3$ pH 8.5 and vortexed until it became clear. The reaction was stirred for 8 h at rt and directly purified by RP-HPLC Vydac C4 column (20 x 250 mm) with a gradient of 35% CH$_3$CN for 5 min and 35% CH$_3$CN to 98% CH$_3$CN over 30 min at 60 °C to obtain 16.3 mg of IFITM3-ArgTAG 17 in an overall yield of 22%.

**Analytical RP-HPLC traces with a gradient of 5% CH$_3$CN for 5 min, 5% CH$_3$CN to 65% CH$_3$CN over 15 min and 65% to 98% CH$_3$CN over 4 min at 60 °C with flow rate 1.5 mL/min. a) Monitoring of the ligation between H$_2$N-S1-KA 4.1 and Opr-S2S3-ArgTAG 4.16 followed by O-N shift in a one pot fashion. b) RP-HPLC trace with a gradient of 35% CH$_3$CN for 3 min, 35% CH$_3$CN to 95% CH$_3$CN over 30 min at 60 °C with flow rate 1.0 mL/min of the purified IFITM3-ArgTAG 4.17. c) ESI-FTICR of purified IFITM3-ArgTAG 4.17, calcd C$_{72}$H$_{113}$N$_{19}$O$_{19}$S$_3$ [M+H]$^+$: 15727.1172, measured 15726.40.**

**6.4.3.2 CD spectra measurement of IFITM3-ArgTAG 4.17**

Extra pure L-alpha-phosphatidylcholine (10.0 mg, 13 mmols) was dissolved in 1.00 mL CH$_3$Cl. 100 µL of the resulting solution were placed in an Eppendorf tube and a N$_2$ flow was applied for 15 min until complete evaporation of the CH$_3$Cl. IFITM3-ArgTAG 4.17 (0.10 mg, 6.3 nmols) was dissolved in 10.0 µL of TFE and added to the previously prepared Eppendorf. A N$_2$ stream
was applied again to remove excess of TFE, and the Eppendorf was placed under vacuum overnight. Then, 350 µL of phosphate-buffered saline (PBS) solution was added to the Eppendorf and the resulting suspension was placed in a sonicator bath for 1.5 h until the solution became clear. The solution was diluted to 1 mL with H₂O and the CD spectrum was recorded from 195–260 nm.

CD spectra of IFITM3-ArgTAG 4.17 reconstituted in lipid vesicles
6.4.4 Synthesis of IFITM3 by KAHA ligation

6.4.4.1 KAHA ligation with peptide segments

Synthesis of Opr-S2S3-OH (4.18).

PhotoOpr-S2-KA 4.14 (120 mg, 23.2 µmol, 1.50 equiv) and Opr-S3-ArgTAG 4.15 (84.0 mg, 15.5 µmol, 1.00 equiv) were mixed together and dissolved in 775 µL of NMP/H2O 95:5 + 0.1 M oxalic acid + 0.1 M TCEP. The resulting solution was mixed by vortexing until complete solubilization and placed in a 60 °C shaker for 20 h. After completion, the reaction mixture was dissolved in 15.0 mL of CH3CN/H2O (1:1) and 200 µL of DODT was added. The resulting mixture was stirred under 365 nm UV irradiation for 45 min.

After complete deprotection, 3.75 mL of 1.0 M NaOH was added and the reaction mixture stirred for 5 min at rt. Then, 4.00 mL of cold TFA/H2O (1:1) solution was used to quench the reaction and the resulting mixture was directly purified by RP-HPLC on a Vydac C4 column (20 x 250 mm) with a gradient of 35% CH3CN for 5 min and 35% CH3CN to 98% CH3CN over 30 min at 60 °C to obtain 34.2 mg of Opr-S2S3-OH 4.18 in an overall yield of 24% yield.

a)
Analytical RP-HPLC traces with a gradient of 35% CH$_3$CN for 5 min, 35% CH$_3$CN to 95% CH$_3$CN over 30 min at 60 °C with flow rate 1.0 mL/min. a) Monitoring of the ligation between Npect-S2-KA 4.14 and Opr-S3-ArgTAG 4.15 followed by UV deprotection and ArgTAG removal in a one pot fashion. b) RP-HPLC trace of the purified Opr-S2S3-OH 4.18. c) ESI-FTICR of purified Opr-S2S3-OH 4.18, calcd C$_{432}$H$_{677}$N$_{102}$O$_{112}$S$_3$ [M+H]$^+$: 9187.8084, measured 9186.99.

Synthesis of IFITM3 (4.19).

[Diagram of the synthesis process]
H₂N-S1-KA 4.1 (29.6 mg, 5.58 µmol, 1.50 equiv) and Opr-S2S3-OH 4.18 (34.2 mg, 3.72 µmol, 1.00 equiv) were mixed together and dissolved in 310 µL of NMP/H₂O 95:5 and 0.1 M oxalic acid and 0.1 M TCEP. The resulting mixture was mixed by vortexing until complete solubilization and placed in a 60 °C shaker for 20 h.

After completion, the reaction mixture was dissolved in 7.50 mL of NaHCO₃ pH 8.5 and vortexed until it became clear. The solution was stirred for 8 h at rt and directly purified by RP-HPLC Vydac C4 column (20 x 250 mm) with a gradient of 40% CH₃CN for 5 min and 40% CH₃CN to 98% CH₃CN over 30 min at 60 °C in to obtain 8.1 mg of IFITM3 4.19 in an overall yield of 15% yield.

Analytical RP-HPLC traces with a gradient of 5% CH₃CN for 5 min, 5% CH₃CN to 65% CH₃CN over 15 min and 65% to 98% CH₃CN over 4 min at 60 °C with flow rate 1.5 mL/min. a) Monitoring of the ligation between H₂N-S1-KA 4.1 and Opr-S2S3-OH 4.18 followed by O-N shift in a one pot fashion. b) RP-HPLC trace with a gradient of 40% CH₃CN for 3 min, 40% CH₃CN to 98% CH₃CN over 30 min at 60 °C with flow rate 1.5 ml/min of the purified IFITM3 4.19.
6.4.4.2 SDS-PAGE analysis of IFITM3-ArgTAG 4.17 and IFITM3 4.19 and MSMS analysis of IFITM3 4.19.

In order to obtain more data confirming the identity of our proteins we conducted an SDS-PAGE analysis of both the protein synthesized. The difference in mass between IFITM3-ArgTAG 4.17 and IFITM3 4.19 can be observed this way. The SDS-PAGE gel was conducted using a 16.5% precast gel and stained with Coomassie blue.

As mentioned in Chapter 4, we were not able to obtain an MS spectrum showing the mass of IFITM3 4.19. In order to generate more evidence for the identity of the protein we synthesized, the band showing IFITM3 4.19 was cut out of the gel and submitted to the Functional Genomics Center of the University of Zürich for MS/MS analysis. Although trypsin digestion was not successful, because of the lack of appropriate cleaving site, MS/MS data could be obtained after treatment of IFITM3 with acid.

95.5% of the protein sequence could be confirmed by MS analysis, showing that the protein was, indeed, IFITM3 4.19.
6.4.4.3 CD spectra measurement of IFITM3 4.19

Extra pure L-alpha-phosphatidylcholine (10.0 mg, 13 mmols) was dissolved in 1.00 mL of CH₃Cl. 100 µL of the resulting solution were placed in an Eppendorf tube and a N₂ flow was applied for 15 min until complete evaporation of the CH₃Cl. IFITM3 4.19 (0.10 mg, 6.9 nmols) was dissolved in 10.0 µL of TFE and added to the previously prepared Eppendorf. N₂ stream was applied again to remove the excess of TFE and the Eppendorf was placed under vacuum overnight. Then, 350 µL of a phosphate-buffered saline (PBS) solution was added to the Eppendorf and the resulting suspension was placed in a sonicator bath for 1.5 h until the solution became clear. The solution was then diluted to 1 mL with H₂O and the CD spectrum was recorded from 195–260 nm.
6.4.5 Synthesis of phosphorylated-IFITM3 (4.22).

phosphorylated-S1-KA 4.21 (10.0 mg, 1.86 μmol, 1.50 equiv) and Opr-S2S3-ArgTAG 4.16 (13.0 mg, 1.24 μmol, 1.00 equiv) were mixed together and dissolved in 82.7 μL of NMP/H₂O 95:5 + 0.1 M oxalic acid + 0.1 M TCEP. The resulting solution was mixed by vortex until complete solubilization and placed in a 60 °C shaker for 20 h. After completion, the reaction mixture was dissolved in 2.50 mL of NaHCO₃ pH 8.5 and vortexed until it became clear. The reaction was stirred for 8 h at rt and directly purified by RP-HPLC on a Vydac C4 column (20 x 250 mm) with a gradient of 35% CH₃CN for 5 min and 35% CH₃CN to 98% CH₃CN over 30 min at 60 °C to afford 4.3 mg of phosphorylated-IFITM3 4.22 in an overall yield of 17% yield.
CHAPTER 6. Experimental

Analytical RP-HPLC traces with a gradient of 5% CH$_3$CN for 5 min, 5% CH$_3$CN to 65% CH$_3$CN over 15 min and 65% to 98% CH$_3$CN over 4 min at 60 ºC with flow rate 1.5 mL/min. a) Monitoring of the ligation between phosphorylated-S1-KA 4.21 and Opr-S2S3-ArgTAG 4.16 followed by O-N shift in a one pot fashion. b) RP-HPLC trace with a gradient of 35% CH$_3$CN for 3 min, 35% CH$_3$CN to 95% CH$_3$CN over 30 min at 60 ºC with flow rate 1.0 mL/min of the purified phosphorylated-IFITM3-ArgTAG 4.22. c) ESI-FTICR of purified phosphorylated-IFITM3-ArgTAG 4.22, calcd C$_{721}$H$_{1134}$N$_{198}$O$_{194}$PS$_3$ [M+H]$^+$: 15807.0971, measured 15806.1273.

6.4.6 Synthesis of fluorescent-IFITM3 (4.39).
Fluorescent-S1-KA 4.38 (10.0 mg, 1.77 µmol, 1.50 equiv) and Opr-S2S3-ArgTAG 4.16 (12.3 mg, 1.19 µmol, 1.00 equiv) were mixed together and dissolved in 79.3 µL of NMP/H₂O 95:5 and 0.1 M oxalic acid and 0.1 M TCEP. The resulting solution was mixed by vortex until complete solubilization and placed in a 60 °C shaker for 20 h.

After completion, the reaction mixture was dissolved in 2.50 mL of NaHCO₃ pH 8.5 and vortexed until it became clear. Then the reaction was stirred for 8 h at rt and directly purified by RP-HPLC on a Vydac C₄ column (20 x 250 mm) with a gradient of 35% CH₃CN for 5 min and 35% CH₃CN to 98% CH₃CN over 30 min at 60 °C with a flow rate of 10 mL/min to obtain 3.64 mg of fluorescent-IFITM3 4.39 in an overall yield of 19%.

Analytical RP-HPLC traces with a gradient of 5% CH₃CN for 5 min, 5% CH₃CN to 65% CH₃CN over 15 min and 65% to 98% CH₃CN over 4 min at 60 °C with flow rate 1.5 mL/min. a) Monitoring of the ligation between fluorescent-S1-KA 4.38 and Opr-S2S3-ArgTAG 4.16 followed by O-N shift in a one pot fashion. b) RP-HPLC trace with a gradient of 35% CH₃CN for 3 min, 35% CH₃CN to 95% CH₃CN over 30 min at 60 °C with flow rate 1.0 mL/min of the purified fluorescent-IFITM3-ArgTAG 4.39. c) ESI-FTICR of purified fluorescent-IFITM3-ArgTAG 4.39, calcd C₇₄₂H₁₁₄₃N₁₉₈O₁₉₇S₃ [M+H]⁺: 16085.4177, measured 16084.5275.
6.4.7 Synthesis of palmitoylated IFITM3  

6.4.7.1 Synthesis of the different cysteine modifying reagents.

Palmitoylhydrazide (4.24).

In a 25-mL round bottomed flask, palmitic acid N-hydroxysuccinimide ester 4.23 (100 mg, 0.28 mmols, 1 equiv) was dissolved in MeOH (5.0 mL) and hydrazine monohydrate (21 mL, 0.43 mmols, 1.5 equiv) was added to the solution. After 2 h, TLC (9:1 hexanes/EtOAc) showed no more starting material. The reaction mixture was cooled down to 0 ºC and filtered. The solid was dried under vacuum for 1 h and recrystallized in EtOH to afford the desired compound 4.24 as a white solid (57 mg, 75% yield).

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 6.66 (s, 1H), 3.88 (s, 2H), 2.14 (m, 2H), 1.61 (m, 2H), 1.28 (m, 24H), 0.88 (t, 3H) ppm.

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 174.0, 34.6, 31.9, 29.7, 29.7, 29.7, 29.6, 29.6, 29.5, 29.4, 29.3, 29.3, 25.5, 22.7, 14.12 ppm.

HR-MS (ESI) calcd. for C$_{16}$H$_{35}$N$_2$O [M+H]$^+$ 271.2744, measured 271.2697.

$\alpha,\alpha$-Di-bromo-adipyl(bis)amide (4.28).

In a 500-mL round bottom flask, adipic acid (25.0 g, 171 mmol, 1 equiv) was suspended in thionylchloride (75.0 mL, 1.03 mol, 6 equiv). The flask was equipped with a condenser and the reaction was heated to reflux (oil bath temp 80 ºC). The reaction was stirred for 90 min at reflux and then cooled to room temperature. CCl$_4$ (100 mL) was added to the reaction followed by $N$-bromosuccinimide (73.1 g, 411 mmol, 2.4 equiv). The reaction mixture was stirred vigorously, 10 drops of HBr (48% aq.) were added and the reaction was heated to reflux. After 2 h at reflux, the reaction was cooled to room temperature and then to 0 ºC. The mixture was stirred at 0 ºC for an additional 2 h. The solid was removed by filtration and washed with Et$_2$O (50 mL). The filtrate was concentrated by rotary evaporation affording a thick, dark red liquid.

In a 500-mL round bottom flask, NH$_4$OH (25% aq.) (200 mL) was cooled to 0 ºC. The crude acid chloride was added dropwise over 20 min to the ammonia solution with rapid stirring. After the addition was complete, the reaction was stirred vigorously at 0 ºC for 1 h. After 1 h, the dark solid was isolated by filtration and dried.
The solid was suspended in H₂O (100 mL) and MeOH (100 mL) and heated to 60 ºC. The mixture was stirred rapidly at 60 ºC for 30 min and the reaction mixture was allowed to cool to rt. The resulting white solid was isolated by filtration and washed with MeOH (200 mL). The product was dried under high vacuum (21.47g, 42%) affording the desired compound 4.28.

\[ \text{1H NMR (400 MHz, (CD$_3$)$_2$SO)} \delta 2.76 (t, J = 7.4, 1H), 2.62 (s, J = 7.2, 1H), 1.70 (m, 2H), 1.28 (m, 6H) \text{ ppm.} \]

\[ \text{13C NMR (101 MHz, (CD$_3$)$_2$SO)} \delta 170.28, 170.23, 48.97, 48.70, 33.04, 32.94 \text{ ppm.} \]

\[ \text{HR-MS (ESI) calcd. for C$_6$H$_{11}$Br$_2$N$_2$O$_3$ [M+H]$^+$ 300.9182, measured 300.9347.} \]

2-Chloro-N-undecylacetamide (4.32).

were dissolved in CH$_2$Cl$_2$ (3.5 mL). The resulting solution was added dropwise to a 25-mL round bottom flask containing chloroacetyl chloride (75.7 µL, 0.95 mmol, 1.00 equiv) at 0 ºC. After 20 min the reaction was allowed to warm up and was stirred at rt for 45 min. After completion, H$_2$O (10 mL) was added to the reaction and the organic layer was separated, washed with saturated aqueous NaCl (250 mL), dried over Na$_2$SO$_4$ and filtered. The filtrate was concentrated by rotary evaporation and column chromatography (gradient of 0–5% EtOAc in hexanes) furnished 2.32 (206 mg, 0.75 mmol, 79% yield) as a white solid.

\[ \text{1H NMR (400 MHz, CDCl$_3$) } \delta 6.58 (s, 1H), 4.07 (s, 2H), 3.32 (m, 2H), 1.58 (m, 2H), 1.28 (m, 20H), 0.9 (t, 3H) \text{ ppm.} \]

\[ \text{13C NMR (101 MHz, CDCl$_3$) } \delta 165.70, 47.72, 39.92, 31.92, 29.67, 29.64, 29.63, 29.56, 29.50, 29.35, 29.32, 29.24, 26.83, 22.69, 14.12 \text{ ppm.} \]

\[ \text{HR-MS (ESI) calcd. for C$_{15}$H$_{30}$CINO [M+H]$^+$ 276.2089, measured 276.2497.} \]

Hexyldecyl-dithiopyridine 4.34

Dithiopyridine (1.7 g, 7.7 mmol, 10 equiv) was dissolved in 12.5 mL of 0.1 M TrisHCl (pH 8)/ n-propanol (1:1) and was added to 50-mL round bottom flask containing a solution of hexyldecyl mercaptan (0.20 g, 0.77 mmol, 1 equiv) in 10 mL of n-propanol. The resulting mixture was stirred at rt for 1.5 h. The reaction mixture was extracted with EtOAc (2 x 30 mL), dried with Na$_2$SO$_4$, filtered and concentrated by rotary evaporation. The crude product was purified by flash column chromatography 16:1 hexanes/ EtOA, the fractions were
combined and concentrated by rotary evaporation yielding the desired product 4.34 as a white powder (0.26 g, 92 % yield).

$^1$H NMR (400 MHz, CDCl3) δ 0.90 (t, 3H), (d, 1H), 1.27 (m, 24H), 1.40 (m, 2H), 1.71 (m, 2H), 8.00 (m, 1H), 7.96 (m, 1H), 7.35 (dd, 1H).

$^{13}$C NMR (101 MHz, CDCl3) δ 14.12, 22.70, 28.50, 28.95, 29.18, 29.37, 29.48, 29.57, 29.63, 29.66, 29.69, 29.69, 29.70, 29.70, 31.93, 39.08, 119.58, 120.45, 136.96, 149.52, 160.76.

HR-MS (ESI) calcd for C$_{21}$H$_{38}$NS$_2$ [M+H]$^+$: 368.2440, measured 368.2452.
6.4.7.2 IFITM3 cysteine modification using palmitoylhydrazide (4.24).

4.24 (2 mg, 7.6 µmol, 120 equiv (40 equiv per cysteine)) was dissolved in a 6 M GnHCl, 0.2 M NaH₂PO₄ (60 µL) at pH 3 buffer and the reaction was stirred at 37 °C for 20 min in order to get as much 4.24 as possible in solution. After 20 min, although the solution was still cloudy, NaNO₂ pre-dissolved in H₂O (17 mg in 500 µL) was added to the reaction (40 µL) and the resulting mixture gently shaken at 37 °C for 15 min.

IFITM3-ArgTAG 4.16 (1 mg, 63 nmol, 1 equiv) was dissolved in the same buffer (50 µL) and the resulting solution was added to the previously prepared solution of 4.25. The pH of the reaction was adjusted to 6.5 and allowed to react overnight (final concentration 0.4 mM).

The reaction was diluted with Milipore-H₂O (500 µL) and dialyzed with a cut off of 3.5 kDa against the same buffer in order to remove excess 4.24. The solution was then spin filtered and “washed” with H₂O three times.

The resulting solution was lyophilized, and the resulting powder analyzed by MALDI-TOF and SDS-PAGE. However, only the mass of IFITM3-ArgTAG 4.17 could be observed and the band on the gel was identical to that obtained for IFITM3-ArgTAG 4.17.

The reaction was repeated in the presence of SDS (2%, 5% and 10%) with the same outcome.

6.4.7.3 IFITM3 cysteine modification using dehydroalanine and thiopalmitic acid (4.26).

Following the procedure developed by Davis and co workers, IFITM3-ArgTAG 4.17 (1 mg, 63 nmol, 1 equiv) was dissolved in H₂O (1 mL). A 250 µL aliquot of this protein solution (~16
nmol) was further diluted with 250 μL sodium phosphate buffer (pH 8.0, 50 mM). The protein solution was kept on ice until needed. A stock solution of α,α′-di-bromoadipyl(bis)amide 4.28 was prepared by dissolving 16.9 mg in 128 μL of DMF. The protein solution was warmed to room temperature and then a 168 μL aliquot of the di-bromide solution (75 μmol) was added. The reaction was vortexed immediately. Unfortunately, the solution remained cloudy. The reaction was shaken at room temperature for 30 minutes and then 1 hour at 37 ºC. After this time, the reaction was cooled to room temperature. Unfortunately, no signals were observed by MALDI-TOF or SDS-PAGE, although the band where IFITM3-ArgTAG is expected would not give us information about the presence or absence of dehydroalanine.

We tried to perform an Ellman test to see if cysteines were still present in the protein. Unfortunately, because of the solubility issue, when the solution was placed in a 1 mm wide cuvette, we could not see through it and UV measurement gave aberrant values. We nonetheless decided to continue and added thiopalmitic acid (0.5 mg, 1.9 μmol, 120 equiv) and placed the resulting mixture on the shaker at 37 ºC overnight.

The reaction was diluted with H₂O (300 μL) and dialyzed with a cut off of 3.5 kDa against the same buffer in order to remove the excess of 4.26 and 4.28. The solution was then spin filtered and “washed” with H₂O 3 times.

The resulting solution was lyophilized, the resulting powder analyzed by MALDI-TOF and SDS-PAGE but only the mass of IFITM3-ArgTAG 4.17 could be observed and the band on the gel was identical to that obtained for IFITM3-ArgTAG 4.17.

6.4.7.4 IFITM3 cysteine modification using 2-chloro-N-undecylacetamide (4.32).

IFITM3-ArgTAG (1 mg, 63 nmol, 1 equiv) was dissolved in 6M GnHCl with 5% SDS at pH 7.8 (315 μL) and 4.32 (2.1 mg, 7.6 mmol, 120 equiv (40 equiv per cys)) was added in one portion to the solution.

The resulting mixture was vortexed and placed in a heated shaker at 40 ºC overnight.

The reaction was diluted with H₂O (300 μL) and dialyzed with a cut off of 3.5 kDa against the same buffer in order to remove excess 4.32. The solution was then spin filtered and “washed” with H₂O 3 times.
The resulting solution was lyophilized, the resulting powder analyzed by MALDI-TOF and SDS-PAGE but only the mass of IFITM3-ArgTAG 4.17 was observed. The band on the gel was identical to that obtained for IFITM3-ArgTAG 4.17.

6.4.7.5 IFITM3 Cysteine modification with hexyldecyl-dithiopyridine 4.34

4.34 (0.60 mg, 1.60 mmol, 90.0 equiv (30 equiv per cysteine)) was dissolved in 1:1:1 DMSO/n-propanol/CH₃CN (380 µL) and stirred for 20 min. The resulting solution was added to 200 µL of a IFITM3-ArgTAG 4.17 solution (1 mg, 63 nmol, 1 equiv, in 1 mL of 9:1 DMSO/AcOH). The reaction mixture was purged with N₂ and stirred under N₂ for 24 h at rt.

After 24 h, the reaction was diluted with H₂O (1 mL) and spin filtered and "washed" with H₂O 3 times. The IFITM3-containing solution was dialyzed with a cut off of 3.5 kDa against 95:5 H₂O/ AcOH for 24 h at 4 °C – in order to remove excess 4.35 – and lyophilized.

The lyophilized powder was analyzed by SDS-PAGE and showed a single band. When 4.35 was co-spotted with IFITM3 4.17, two distinct bands were observed. However when TCEP was added to a 1:1 mixture of 4.17 and 4.35, incubated at 40 °C for 1 h and analyzed by SDS-PAGE, only the band corresponding to 4.17 was observed.

Here again performing an Ellman test was complicated. Although only a qualitative information was obtained, when 5,5'-dithiobis(2-nitrobenzoic acid) was added to a solution of 4.35, the solution remained white. However, when 4.35 was incubated with TCEP, dialyzed, spin-filtered and 5,5'-Dithiobis(2-nitrobenzoic acid) was added to the solution, the resulting mixture turned yellow.
This result is not a 100% guarantee that the three cysteines formed disulfide bonds, but it strongly supports this conclusion.

### 6.4.8 Vesicles assay and fluorescence microscopy.

#### 6.4.8.1 Preparation of the vesicles.

To a 50 mL round bottom flask was added 600 µL of a solution of egg phosphatidylcholine in CHCl$_3$ (10 mg/mL) and 400 µL of additional CHCl$_3$. The solvent was evaporated on a rotary evaporator and the flask was placed under vacuum for overnight. The lipid film was hydrated with 1.0 mL of 0.1 M phosphate buffer pH 8 and the flask was incubated at rt for 1 h.

#### 6.4.8.2 Incorporation of IFITM3 in lipid vesicles.

To a 1.50 mL Eppendorf tube was added 50.0 µL of a 100 µM the desired protein in 0.2% SDS, 0.2 M phosphate buffer pH 8 and 50.0 µL of the previously prepared vesicle solution. The solution was incubated at rt for 30 min and applied to a Sepharose CL-4B gel filtration resin that had been equilibrated with three column volumes of 0.2 M phosphate buffer. The column was washed with 3 times the volume of the column (0.5 mL) and the flow-through was lyophilized and analyzed by SDS-PAGE.

The test is simple: when the protein is loaded on the resin, it becomes trapped by the size exclusion matrix and will remain inside the resin, thus not showing any band when analyzed by SDS-PAGE.
However, if the protein is incorporated into the preformed liposomes, because of their much larger size, the liposome-protein complex passes through the resin and a band corresponding to the mass of the protein can be observed. This experiment showed that IFITM3 4.17 and Palm-IFITM3 4.35 behaved as expected and are incorporated easily into lipid membranes. Chymotrypsinogen A 4.36 was used as a negative control, as reported by Tilley et al., since it does not get incorporated into the lipid vesicle.

### 6.4.8.3 Fluorescence microscopy of fluorescent-IFITM3 in the membrane of vesicles.

The same procedure as in 6.2 was repeated using fluorescent-IFITM3 4.39. After filtration, the flow through was concentrated – using spin filters with a 3.5 kDa cut off – and analyzed by fluorescence microscopy. Even though it could arise from the sample preparation, fluorescence microscopy shows labeled IFITM3 as compact and defined “dots” inside the lipid bilayer. This could indicate that IFITM3 localizes in lipid membranes as clusters or aggregates.

![Fig. S26. Picture of the Fluorescent-IFITM3 4.39 in the membrane of vesicles after purification by size exclusion resin. (left) un-zoomed. (right) zoomed.](image-url)
CHAPTER 7
APPENDIX: NMR SPECTRA
OH

S

f1 (ppm)
Fmoc

\[
\text{O} \quad \text{N} \quad \text{O} \\
\text{OH}
\]
# Thibault Harmand Curriculum Vitae

## Personal details

<table>
<thead>
<tr>
<th>Last name:</th>
<th>Harmand</th>
</tr>
</thead>
<tbody>
<tr>
<td>First name:</td>
<td>Thibault</td>
</tr>
<tr>
<td>Born:</td>
<td>15.03.1989, Nantes, France</td>
</tr>
<tr>
<td>Address:</td>
<td>ETH Zürich/HCI E324 Vladimir-Prelog-Weg 4 8093 Zurich, Switzerland</td>
</tr>
<tr>
<td>E-mail:</td>
<td><a href="mailto:tharmand@org.chem.ethz.ch">tharmand@org.chem.ethz.ch</a></td>
</tr>
<tr>
<td>Phone:</td>
<td>+41 78 619 05 16</td>
</tr>
</tbody>
</table>

## Education

- **10.2012 – present**  
  Doctoral student, Swiss Federal Institute of Technology (ETH Zürich) under the supervision of Prof. J. W. Bode.

- **2010 – 2012**  
  Master of Sciences in Organic Chemistry, University of Lyon, France.

- **2007 – 2010**  
  Bachelor of Sciences in Chemistry, University of Nantes, France.

## Research Experience

- **10/2012 – present**  
  Doctoral student, Swiss Federal Institute of Technology (ETH Zürich) under the supervision of Prof. J. W. Bode.
  - Chemical synthesis of protein by KAHA ligation and NCL.
  - Large scale (>30g) synthesis of key building blocks for KAHA ligation.
  - Synthesis of unnatural amino acids.
  - Side chain modifications of unprotected peptides by chemistry.
  - Maintenance of the group RP-HPLCs and the group MALDI-TOF

- **10/2015 – 12/2015**  
  3 months stay at the Institute of Transformative Bio-Molecule (ITbM), Nagoya University, Japan, in the Bode group satellite lab.
  - Setting up laboratory after move into our new building
  - Training the new members in protein synthesis.
  - Chemical synthesis of a protein in collaboration with the biology department.

- **02/2012 – 08/2012**  
  Master Thesis at the University of Oxford, UK, under the supervision of Prof. G. W. J. Fleet.
  - Synthesis of azetidine iminosugar derivatives.
  - Synthesis of new glycosidase inhibitors.
  - Optimization of scaffold synthesis.

- **03/2011 – 08/2011**  
  Internship at the University of Cardiff, UK, under the supervision of Prof. M. C. Bagley.
  - Total synthesis of MAP kinase-activated protein kinase 2 inhibitors.

- **04/2010 – 06/2010**  
  Internship at the Institute of Materials, University of Nantes, France, under the supervision of Dr. M. Bujoli-Doeuff.
  - Synthesis of new Metal/organic complexes with optical properties.
Personal skills and competences

Languages:
- French Native
- English Fluent
- German Basic Knowledge

Teaching Experience

Teaching Assistant at ETH Zürich for Organic Chemistry II
Teaching Assistant at ETH Zürich for Organic Chemistry laboratory class.
Supervisor of Bachelor Student Mr. Christian Rüeger, Master Student Mr. Yves Wittwer and Ph.D. Student Mr. Hikaru Takano.

Publications


Conference and presentation

- Chemical Protein Synthesis Meeting, St Augustin, Florida, USA, June 16 – 19, 2015.
  KAHA ligation for the convergent synthesis of betatrophin.