Intein-Mediated Semi-Synthesis and Characterization of Glycosylphosphatidyl-inositol (GPI)-Anchored Proteins

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Renée Fabienne Roller

from Essen

2018

This work was performed between October 2013 and March 2018 under the supervision of Dr. Daniel Varón Silva in the Biomolecular Systems Department of the Max-Planck-Institute of Colloids and Interfaces, Potsdam, and the Institute of Chemistry and Biochemistry, Free University Berlin.

1st reviewer: Dr. Daniel Varón Silva

2nd reviewer: Prof. Dr. Markus Wahl

Date of oral defense: 09.07.2018

Declaration

Herewith I state that I performed this work myself, if not stated otherwise, and that any help and resources used have been fully acknowledged.

---------- ----------

(Date, Place) (Signature)

Acknowledgements

First, I wish to express my deep gratitude for my supervisor, Dr. Daniel Varón Silva, for all his support and advice during the years and for his trust in me. I also want to thank him for the opportunities he offered me to experience scientific conferences and symposia on three continents.

I am grateful for Prof. Markus Wahl for kindly agreeing to review this thesis. I also want to thank Prof. Peter H. Seeberger for giving me the opportunity to perform my dissertation project in his interdisciplinary department of Biomolecular Systems at the Max Planck Institute of Colloids and Interfaces, where I learned so much during the past years.

Special thanks go to current and former members of my group for support in the lab and for the great working environment: Hyunil, Dana, Ankita, Antonella, Monika, Maurice, Maria and Sandra. Without the work Dana, Antonella, Monika and Daniel did to provide me with synthetic peptides and GPIs, this work would not have been possible. I also want to thank my bachelor student Şafak for her work on the IFN-α project and the great time we had.

I am grateful for knowing my colleagues, especially Silvia, Deborah, Jessica, Mara, Alonso, Hannes, Marco, Mauro, Priya, Fei-Fei, Jamal, Mónica, Chandradhish, Mike, Martina, Bart, Andreas, Felix, Andrew, Andreia, Uwe and Falko for making this time such an inspiring part of my life. I want to thank Jonas, Jonas and Eike for lots of help in the beginning of my PhD, and Reka for initial studies and valuable advice.

I wish to thank Eva for making fighting with QTOF more fun and for keeping the HPLCs running. Organizational support by Dorothee is also highly appreciated.

I would like to thank our cooperation partners for providing us with plasmids (Prof Henning Mootz from University of Münster) and fusion proteins (Prof. Joaquín Castilla from BioGUNE, Bilbao, Spain).

Financial support by DFG with SPP 1623, RIKEN-Max-Planck Joint Research Centre for Systems Chemical Biology and Max-Planck-Society is gratefully acknowledged.

Last but not least, I want to thank my parents, my sister, my friends and especially Christian for their never-ending support and encouragement.

List of Publications

Parts of this work have been or will be published.

Scientific Publications

1. Roller, R. F., Michel, D., Rella, A., Garg, M., Seeberger, P. H., Varón Silva, D. Intein-Based Semi-Synthesis of Homogeneous GPI-Anchored Proteins. (in preparation)

Scientific Conferences and Symposia

1. Intein-Based Methods for the Semi-Synthesis of Homogeneous GPI-Anchored Proteins (Oral presentation). Ringberg Conference on Structural and Physical Aspects of Carbohydrates in Glycobiology and Material Science, Ringberg Castle, Rottach-Egern, Germany, 2017.
2. Semi-Synthesis of Homogeneous Glycosylphosphatidylinositol-Anchored Proteins Applying Protein *Trans*-Splicing (Poster). Meeting of the GDCh-Division Biochemistry, Frankfurt, Germany, 2016.
3. Semi-Synthesis of Homogeneous Glycosylphosphatidylinositol-Anchored Proteins Applying Protein *Trans*-Splicing (Poster). 24th Dutch Peptide Symposium, Lelystad, Netherlands, 2016.
4. Semi-Synthesis of Pure Glycosylphosphatidylinositol-Anchored Proteins Applying Protein Splicing (Oral presentation). Ringberg Conference on Glycoproteins and Glycolipids – Synthesis, Analysis and Function, Ringberg Castle, Rottach-Egern, Germany, 2015.
5. Protein *Trans*-Splicing for the Semi-Synthesis of Glycosylphosphatidylinositol-Anchored Proteins (Poster). The Fourth Symposium – RIKEN-Max Planck Joint Research Center, Kobe, Japan, 2015.
6. Semi-Synthesis of Pure Glycosylphosphatidylinositol-Anchored Proteins Applying Protein Splicing (Poster and oral presentation). 6th Chemical Proteins Synthesis Meeting, St. Augustine, Florida, USA, 2015.
7. Intein-Based Ligation Strategies for the Generation of Homogeneous GPI-Anchored Proteins (Poster). 33rd European Peptide Symposium, Sofia, Bulgaria, 2014.

8. Intein-Based Ligation Strategies for the Generation of Homogeneous GPI-Anchored Proteins (Poster). Bioorthogonal Chemistry Meeting of the Biochemistry Division (GDCh), Berlin, Germany, 2014.
9. Intein-Based Strategies for the Generation of Homogeneous GPI-Anchored proteins (Poster). RIKEN-Max-Planck Joint Research Center for Systems Chemical Biology, Ringberg Castle, Rottach-Egern, Germany, 2014.

Summary

Many eukaryotic proteins are attached to the cell membrane via Glycosylphosphatidyl-inositol (GPI) anchors that are added post-translationally to the C-terminus of proteins. These complex structures contain a highly conserved carbohydrate core, variable number of phosphate residues and lipid chains. The functions of GPI-anchored proteins (GPI-APs) are widespread, including participation in signal transduction, immune response regulation, lipid raft partitioning and prion disease pathogenesis. However, the effect of the GPI-anchor itself in these processes still remains unclear. Due to their high complexity and metabolic expense, a function beyond membrane anchoring has been anticipated.

GPI-APs are not accessible in a homogeneous form and high amounts by isolation from natural sources, making investigation of the GPI effects on the function and structure of the protein difficult. Some progress has been made in the field of total GPI synthesis and small GPI-anchored peptides are now accessible by chemical synthesis, but these strategies are limited regarding peptide length by the scope of solid-phase peptide synthesis (SPPS). So far, no method is available for the routine generation of homogeneous GPI-APs for structural or biological studies.

In this work, different semi-synthetic, intein-based methods have been investigated and established for the generation of homogeneous GPI-anchored proteins, Figure 1. The first method, Expressed Protein Ligation (EPL, a), utilizes the GyraseA intein from *Mycobacterium xenopi* for the generation of protein α-thioesters which are then ligated to cysteine-containing GPI-anchors. The second method, Protein *Trans*-Splicing (PTS, b) uses the naturally split DnaE intein from *Nostoc punctiforme*. For PTS, proteins of interest are expressed in *E. coli* as fusion proteins with the larger N-terminal split intein fragment Npu^N, whereas the shorter (39 amino acids) C-terminal fragment Npu^C is synthesized by SPPS and ligated to a cysteine-containing GPI-anchor. Subsequently both fragments are combined and *trans*-splicing takes place initiated by association of the fragments and folding into a full intein structure, ligating both exteins together with a native peptide bond. This method requires the introduction of some extra amino acid residues which are essential for the splicing reaction.

Both EPL and PTS were successfully established for the anchoring of eGFP as a model protein to biotin and a GPI anchor containing the full pseudo-pentasaccharide core structure, phosphates, and a simplified monolipid chain. Application of these strategies to the semi-

synthesis of the naturally GPI-anchored proteins Thy-1 and prion protein (PrP) showed, however, that thioester generation is inefficient under denaturing conditions, making the EPL strategy less useful for our purpose. PTS on the other hand was efficient under denaturing conditions, although slower. In comparison to the natural conditions, this strategy suffered from insufficient availability of the synthetic GPI-coupled peptide due to a high synthetic effort required for its synthesis.

Therefore, a third semi-synthetic, intein-based method was developed: One-Pot-Ligation (OPL, c). In OPL, the same fusion proteins required for PTS can be used, but they are combined with a mutated C-terminal fragment, *Npu*^{C}(AA), in which two essential amino acid residues are exchanged for alanines, rendering the C-terminus of the split intein function-less. Instead, with this system, a protein intermediate can be captured and protein thioesters can be generated *in situ* by the additions of external thiol reagents, with subsequent ligation of Cys-GPI in a one-pot manner. This strategy was successfully applied for anchoring both soluble and denatured proteins (eGFP, Thy-1, PrP and IL-2) to biotin, dimannose and mono- and bilipidated GPI-anchors. It is more versatile towards the generation of GPI-AP libraries, however exhibited some immature protein thioester hydrolysis which could largely be overcome by the use of more stable thiols.

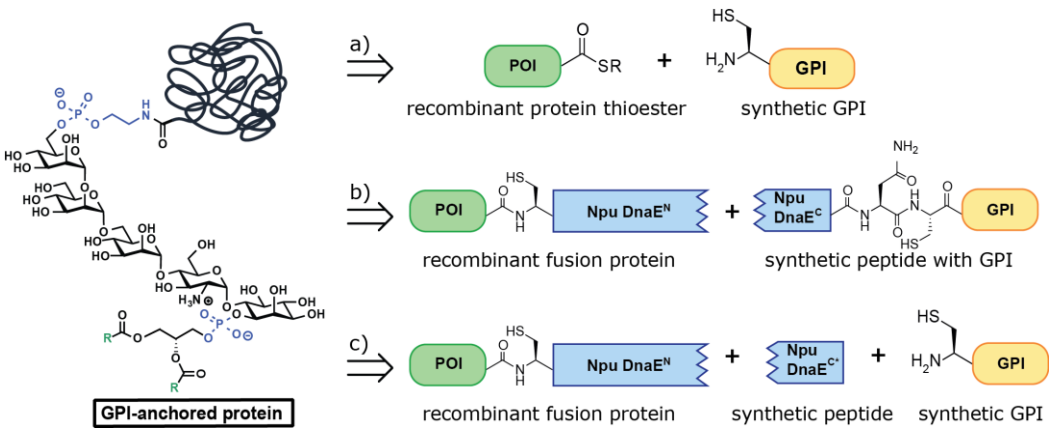


Figure 1. Retrosynthetic Analysis of the Semi-Synthesis of GPI-Anchored Proteins using intein-based strategies. a) Expressed Protein Ligation (EPL), b) Protein *Trans*-Splicing (PTS), c) One-Pot-Ligation (OPL).

Additional challenges arose from characterization of these highly complex protein-carbohydrate-lipid conjugates, carrying multiple charges, in LC-MS. Although progress could be made towards this goal, a complete characterization was not achieved for each protein.

Initial structural characterization of the obtained GPI-APs was performed for eGFP-GPI using circular dichroism (CD). This study showed that using any of the three methods investigated, anchoring to GPIs does not affect the structure of eGFP.

Zusammenfassung

Viele eukaryotische Proteine sind mittels Glycosylphosphatidylinositol (GPI) Ankern an der Zelle verankert, die post-translational an den C-terminus des Proteins angehängt werden. Diese komplexen Strukturen beinhalten einen hochkonservierten Kohlenhydratkern, variable Zahl an Phosphatresten sowie Lipidketten. GPI-verankerte Proteine (GPI-APs) haben sehr diverse Funktionen, unter anderem sind sie beteiligt an der Signaltransduktion, Immunantwortregulation, Lipid-Raft-Partitionierung sowie der Pathogenese von Prionenkrankheiten. Der Effekt des GPI-Ankers selbst auf diese Prozesse ist jedoch immer noch unklar. Aufgrund ihrer hohen Komplexität und metabolisch aufwendigen Herstellung in der Zelle geht man von einer über die reine Membranverankerung hinausgehenden Funktion aus.

GPI-APs können nicht in homogener Form und großen Mengen durch Isolation aus natürlichen Quellen gewonnen werden, was die Erforschung des Effekts des GPI-Ankers auf die Funktion und Struktur von Proteinen erschwert. Durch erzielte Fortschritte auf dem Gebiet der Totalsynthese von GPIs können mittlerweile kleine GPI-verankerte Peptide chemisch hergestellt werden. Diese Strategien werden jedoch durch die Möglichkeiten der Festphasen-Peptidsynthese (SPPS) in Bezug auf die Peptidlänge limitiert. Bisher sind keine Methoden für die routinemäßige Herstellung homogener GPI-APs für strukturelle und biologische Studien verfügbar.

In dieser Arbeit wurden verschiedene semi-synthetische, intein-basierte Methoden für die Herstellung homogener GPI-verankerter Proteine untersucht und etabliert, Abbildung 1. Die erste Methode, Expressed Protein Ligation (EPL, a), nutzt das GyraseA-Intein aus *Mycobacterium xenopi* für die Herstellung von Protein-α-Thioestern, die dann mit Cystein-tragenden GPI-Ankern ligiert werden. Die zweite Methode, Protein-*Trans*-Splicing (PTS, b), nutzt das natürliche gespaltene Intein DnaE aus *Nostoc punctiforme*. Für PTS werden die Zielproteine als Fusionsproteine mit dem größeren N-terminalen Fragment, Npu^N, des gespaltenen Inteins in *E. coli* exprimiert, während das kürzere (39 Aminosäuren lange) C-terminale Fragment Npu^C mittels SPPS synthetisiert wird, gefolgt von Ligation zu einem Cystein-tragenden GPI-Anker. Anschließend werden beide Fragmente kombiniert und die durch Assoziation initiierte *Trans*-Splicing-Reaktion findet statt durch Faltung der Fragmente in eine komplette Inteinstruktur und mit dem Ergebnis der Ligation beider Exteine. Diese

Methode erfordert die Insertion einiger zusätzlicher Aminosäuren in das Endprodukt, die für die Splicing-Reaktion essentiell sind.

Beide Methoden, EPL und PTS, wurden erfolgreich etabliert für die Verankerung von eGFP als Modellprotein an Biotin sowie an einen GPI-Anker, der die vollständige Pseudopentasaccharid-Kernstruktur, Phosphate und eine vereinfachte Monolipidkette enthält. Die Anwendung dieser Strategie auf natürliche GPI-verankerte Proteine Thy-1 und Prion-Protein (PrP), die nur in unlöslicher Form erhalten werden konnten, zeigte allerdings, dass die Thioesterbildung unter denaturierenden Bedingungen ineffizient ist, was die EPL-Strategie für unsere Ziele wenig nützlich macht. PTS war hingegen effizient unter denaturierenden Bedingungen, wenn auch langsamer als unter nativen Bedingungen; jedoch stellte die unzureichende Verfügbarkeit der notwendigen synthetischen GPI-gekoppelten Peptide ein Problem dar, die aus einem hohen synthetischen Aufwand für deren Generierung resultierte.

Daher wurde eine dritte semi-synthetische, intein-basierte Methode entwickelt: One-Pot-Ligation (OPL, c). Für OPL können die gleichen Fusionsproteine genutzt werden wie für PTS, sie werden jedoch mit einem mutierten C-terminalen Fragment, *Npu*^{C}(AA), kombiniert, in welchem zwei essentielle Aminosäuren gegen Alanine ausgetauscht wurden, wodurch der C-terminus des gespaltenen Inteins für ein *Trans*-Splicing funktionslos wird.

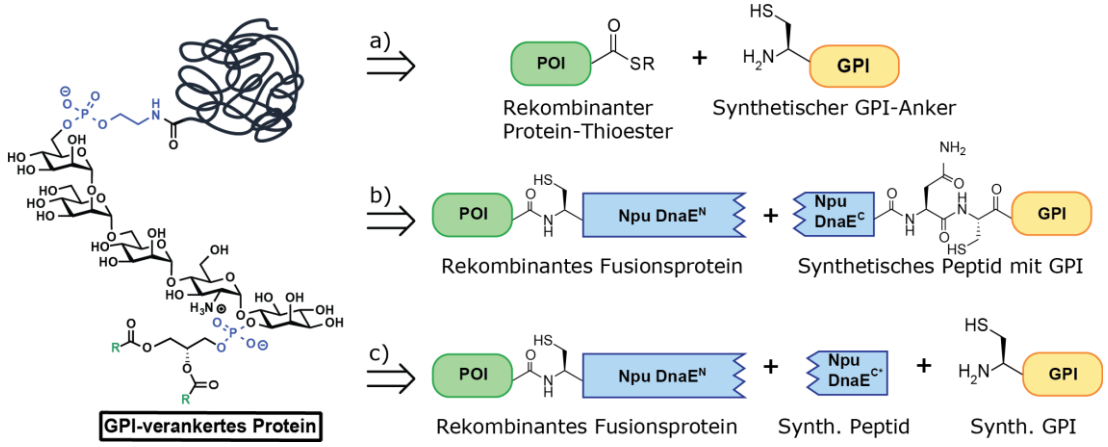


Abbildung 1. Retrosynthetische Analyse der Semi-Synthese GPI-verankerter Proteine mittels intein-basierter Strategien. a) Expressed Protein Ligation (EPL), b) Protein *Trans*-Splicing (PTS), c) One-Pot-Ligation (OPL).

Stattdessen werden generierte Protein-Intermediate in diesem System durch die Zugabe externer Thiol-Reagenzien *in situ* zu Protein-Thioestern umgewandelt, gefolgt von der Ligation zu Cys-GPI in einem Topf. Diese Strategie wurde erfolgreich angewendet, um sowohl lösliche als auch unlösliche Proteine (eGFP, Thy-1, PrP und IL-2) zu Biotin,

Dimannose, sowie mono- und bilipidierten GPI-Ankern zu ligieren. Sie ist vielfältiger anwendbar für die Generierung von Bibliotheken von GPI-APs, zeigte jedoch ein gewisses Ausmaß an vorzeitiger Protein-Thioester-Hydrolyse, was aber durch die Nutzung stabilerer Thiol-Reagenzien zu großen Teilen überwunden werden konnte.

Eine zusätzliche Herausforderung in dieser Arbeit stellte die Charakterisierung dieser hochkomplexen Protein-Kohlenhydrat-Lipid-Konjugate mit mehreren unterschiedlichen Ladungen in LC-MS dar. Obwohl hier große Fortschritte erzielt werden konnten, konnte eine vollständige Charakterisierung nicht für jedes Protein erreicht werden.

Erste Untersuchungen der Struktur von eGFP-Ligationsprodukten unter Nutzung von Zirkulardichroismus (CD) zeigten, dass bei Nutzung keiner der drei verwendeten Strategien die GPI-Verankerung einen Einfluss auf die Struktur von eGFP hat.

Abbreviations

APC	Antigen presenting cell
APS	Ammonium persulfate
BCA	Bicinchoninic Acid
bp	base pair
BSA	Bovine Serum Albumin
CBD	Chitin Binding Domain
CD	Cluster of differentiation
CD	Circular Dichroism
CJD	Creutzfeldt-Jakob-disease
CoA	Coenzyme A
CV	Column Volume
Da	Dalton
DAC	Diammonium hydrogen citrate
DHAP	2,5-Dihydroxyacetophenon
DHB	2,5-Dihydroxybenzoic acid
DIC	Diisopropylcarbodiimide
DNA	Deoxyribonucleic acid
dNTPs	deoxynucleotides (any base)
DTT	Dithiothreitol
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESI-QTOF-MS	Electrospray Ionization – Quadrupole – Time Of Flight Mass Spectrometry
FA	Formic Acid
FPLC	Fast Protein Liquid Chromatography
fw	forward
Glc	Glucose

GlcN	Glucosamine
GlcNAc	*N*-acetylglucosamine
GPI	Glycosylphosphatidylinositol
GPI-AP	GPI-anchored protein
HABA	4'-hydroxyazobenzene-2-carboxylic acid
HRP	Horseradish peroxidase
IFN	Interferon
IgG	Immunoglobulin G
IL	Interleukin
IMPACT	Intein-Mediated Purification with an Affinity Chitin-binding Tag
i.e.	id est
Int	Intein
IPTG	Isopropyl-β-D-thiogalactopyranosid
mAb	Monoclonal antibody
MALDI-TOF-MS	Matrix-assisted Laser Desorption / Ionization – Time Of Flight mass spectrometry
Man	Mannose
MESNA	Sodium 2-Mercaptoethanesulfonate
MMBA	4-(Mercaptomethyl)benzoic acid
MMP	Methyl 3-mercaptopropionate
MPAA	4-Mercaptophenylacetic acid
Nd:YAG laser	Neodymium-doped yttrium aluminum garnet laser
NK cells	Natural killer cells
Npu	*Nostoc punctiforme*
NTA	Nitrilotriacetic acid
OD_{600}	Optical Density at 600 nm
o/n	overnight
PBS(-T)	Phosphate buffered saline (with 0.1% Tween 20)
PEG	Polyethylene glycol

PGAP	Post-GPI-attachment to proteins
PIG	Phosphatidylinositol glycan
PMSF	Phenylmethylsulfonyl fluoride
POI	Protein of interest
PrP	Prion protein
PTS	Protein-*Trans*-Splicing
PVDF	Polyvinylidene difluoride
rev	reverse
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
rpm	rounds per minute
RT	room temperature
SA	Sinapinic acid
SDS-PAGE	Sodiumdodecylsulfate Polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
SeMBA	4-(Selenomercaptomethyl)benzoic acid
TCGF	T-cell growth factor
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	N,N,N',N'-Tetramethylethylendiamin
TBS(-T)	Tris-Buffered Saline (with 0.1% Tween 20)
TFA	Trifluoroacetic acid
TIC	Total Ion Count
TIPS	Triisopropylsilane
t_R	Retention time
TSE	Transmissible spongiform encephalopathy
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Table of Contents

1 Introduction

In living organisms, all information is stored in DNA within the cell. According to the central dogma of molecular biology,[1,2] this information flows from DNA to RNA in a process called transcription that is followed by the flow from RNA to protein via translation,[2] Figure 1.1. Both are directly template-driven processes in which one linear structure is converted into another linear structure, even though two different alphabets are used: on one hand the set of the nucleobases in DNA and RNA (adenine, thymine/uracil, guanine and cytosine) and on the other hand the set of 20 proteinogenic amino acids. Both alphabets are connected to each other via the genetic code.[3] With four bases and 20 amino acids present in the standard repertoire, three bases are at least needed to code for one amino acid ($4^2 = 16$ is not sufficient, $4^3 = 64$ possible translations), forming a codon. The genetic code is non-overlapping, degenerate and almost universal,[2,4,5] meaning that some amino acids are encoded by more than one codon.

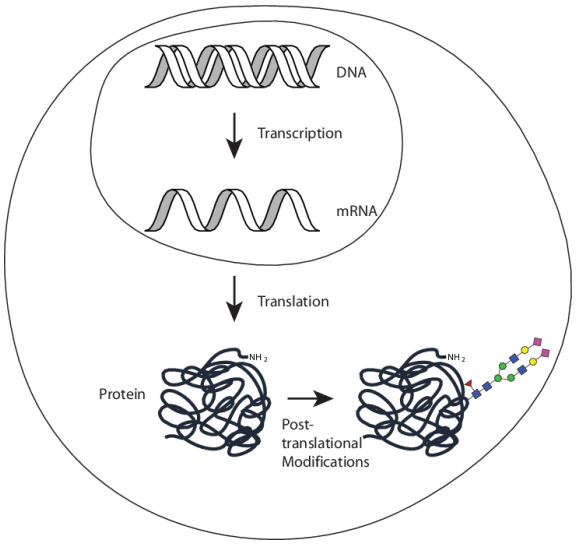


Figure 1.1. The Central Dogma of Molecular Biology: DNA is transcribed into mRNA, which is then translated into proteins. Higher variety is achieved by the addition of posttranslational modification (PTMs) to proteins, such as phosphorylations, glycosylations, glypiation or others.

In prokaryotic organisms the transfer of information from DNA to mRNA and to protein is mostly linear, just as shown in Figure 1.1, although PTMs exist also in prokaryotes.[6] In eukaryotes, however, different variations can occur within these processes. At the DNA level (or epigenetics level), DNA modifications such as methylation can influence gene expression without altering the genome itself by the formation of the chromatin states euchromatin and heterochromatin within the nucleosome.[7]

At the mRNA, or post-transcription level, mRNA splicing can take place.

Pre-mRNA is directly transcribed from the template DNA of protein-coding genes containing introns (**int**ragenic regi**on**) in between exons (**ex**pressed regi**on)**. Introns are removed utilizing Watson-Crick base-pairing between the intron and an snRNA in the spliceosome, generating mature mRNA.[8] A variation of mRNA splicing is alternative splicing, where some exons can be skipped due to weaker exon signals at some sites or variation in exon length, leading to increased complexity in gene expression.[9]

1.1 Post-Translational Modifications

Additionally to the sequence-based diversity, epigenetic regulation and mRNA splicing, post-translational modifications (PTMs) of proteins confer an even higher degree of complexity to the proteome of a cell (Figure 1.1), increasing the number of protein species by two to three orders of magnitude compared to the number expected from the genome.[10-13] These modifications expand the functional scope of the proteome, especially regarding activity, conformation and localization of proteins, as well as interactions with other proteins.[11,13,14] PTMs can also play an important role in addressing various diseases regarding drug development.[15]

A post-translational modification is defined as a covalent modification of an amino acid residue in the protein chain.[13,16] It can be conferred either by enzymatic attachment of functional groups or other fragments, by chemical reactions breaking the peptide backbone of the peptide, i.e. by proteases or autocatalytic cleavage,[10] or by spontaneous chemical reactions via the encounter of a reactive metabolite.[17] These modifications can occur directly after translation next to the ribosome, in special cell organelles, or even outside of the producing cell. In comparison to gene expression, the post-translational modification of proteins is less tightly regulated,[12] making it faster and more dynamic to changes in the environment.[18] PTMs are therefore often used as disease markers.[11] However, due to the vast diversity of PTMs, elucidation of their structure and function is difficult.[12] Nowadays, characterization of PTMs is mostly achieved by western blot or mass spectrometric methods, i.e. MS-based proteomics and LC-MS/MS techniques.[14,18,19] The localization of most PTMs is defined by recognition sites called sequons, of which generally not all are used at all times. This means that PTMs are not stoichiometric, i.e. only small subsets of a protein type carry a certain PTM at a certain time.[14]

The most common PTMs include phosphorylation, acetylation, glycosylation, ubiquitination, hydroxylation, methylation, lipidation and proteolysis of proteins,[6,10-13,16] although many more exist.[14,15,19] Protein phosphorylation takes place at Ser/Thr/Tyr residues, is mediated by protein kinases[13] and is often part of switching protein activity on or off,[16] and signaling by inducing changes in protein conformation.[12] It is a reversible PTM, as protein phosphatases can cleave off the phosphate added by the kinases.[11] Protein acetylation describes the reversible transfer of an acetyl group from acetyl-CoA mainly to lysine residues by lysine acetyltransferases. Acetylation can be reversed by lysine deacetylases, thereby representing a dynamic and reversible mechanism for transcription regulation.[13] It alters the charge distribution in a protein[10] and has been found to be involved in various diseases.[11] In ubiquitination, the 76-residue protein ubiquitin is attached to lysine residues in a three-step process. It is a PTM involved in proteasomal degradation of proteins which are no longer needed,[13] and in apoptosis regulation.[20] Lipidation is another diverse PTM that is also very difficult to analyze due to the lack of suitable antibodies.[13] The two most common lipidations of protein are N-myristoylation at N-terminal glycine residues with simultaneous hydrolysis of the ^{N}Met-Gly peptide bond[10] and S-palmitoylation at cysteine residues, in which acyltransferases attach the lipid from a C_{16} fatty acyl CoA donor to the cysteine residue.[10,11] Palmitoylation is mainly involved in cell trafficking, protein-protein interactions[11] localization and membrane affinity.[13]

1.1.1 **Protein Glycosylation**

One of the most common and most important PTMs is glycosylation, which involves the attachment of carbohydrate structures to proteins.[13] Protein glycosylation takes place in the ER and the golgi apparatus and is performed by glycosyltransferases that attach the carbohydrates to Ser/Thr/Tyr residues in the case of *O*-glycosylation, or to the Asn in the Asn-X-Ser/Thr/Tyr motif (X is any amino acid except for proline) giving the more prevalent and often more complex *N*-glycosylation.[10,13,14] Glycosylation is a non-template-dependent process[13] that leads to significant microheterogeneity of the resulting structures.[21] The complexity of the glycans is especially due to carbohydrate branching, the presence of different possible anomeric linkages and the presence of modifications in the glycans such as phosphates, sulfates,[22] Figure 1.2.

Proteins found in secretory pathways, and on the cell surface are often glycosylated. Glycosylation plays a role in protein folding and quality control.[13,23] Additionally, glycoproteins are also involved in important biological processes such as fertilization, recognition events, immune defense, cell growth, cell-cell adhesion and inflammation.[13,24,25]

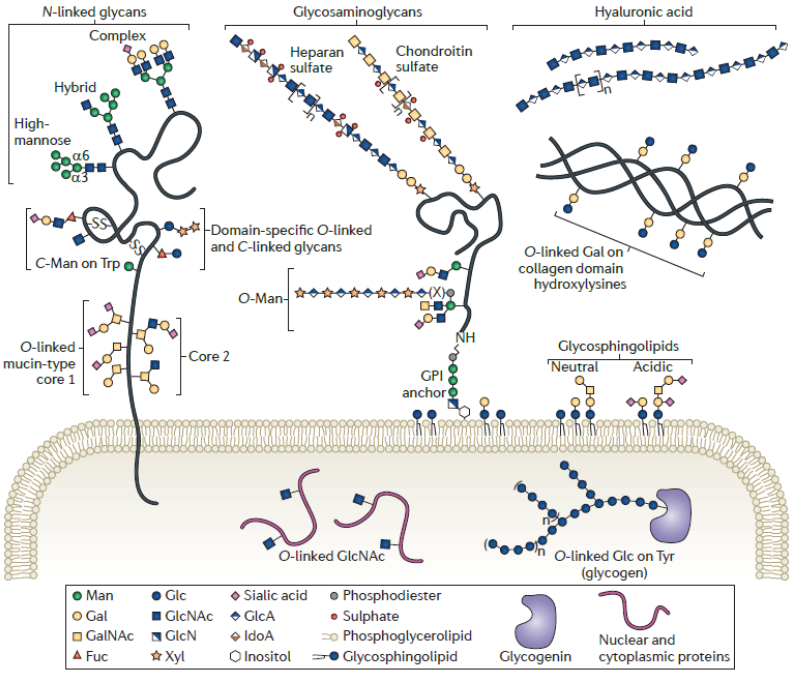


Figure 1.2. Overview of the types of glycoproteins and other glycosylated structures in cells. **(**Figure from Vucic et al., 2011[21]).

1.1.2 GPI-Anchors and Glypiation

A very complex PTM is the attachment of glycosylphosphatidylinositol (GPI) anchors to the C-terminus of proteins, named glypiation.[26-28] In total ca. 0.5 – 1% of all eukaryotic proteins carry a GPI-anchor.[26,27] Since their discovery in the 1980ies,[29] at least 150 proteins have been found to be glypiated in humans.[30] The primary function of the GPI is the anchoring of proteins to the outer leaflet of the cell membrane via their C-terminus.[26,31,32] However, GPI-anchored proteins (GPI-APs) have diverse functions and participate in many important biological processes including cell adhesion, complement regulation, signal transduction and are often receptors.[29,30,33-35] Additional to membrane anchoring, the role of the GPI anchor itself remains still unclear, although their high complexity suggests a large spectrum of functions.[34,36,37]

The GPI anchor contains a highly conserved core structure consisting of a phosphoethanolamine, a trimannoside, a glucosamine and an inositol phospholipid. This structure can also carry additional carbohydrate units, phosphorylations and lipid chains, resulting in heterogeneous structures,[27,32,36,38,39] Figure 1.3.

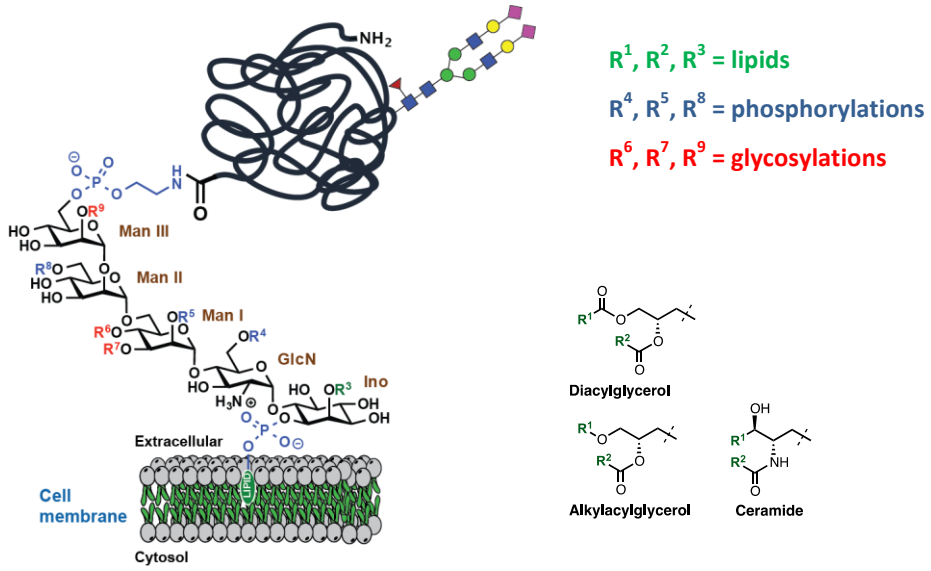


Figure 1.3. Structure of a Glycosylphosphatidylinositol (GPI) anchored glycoprotein. (Figure by Daniel Varón Silva, 2013[40], modified).

GPI-APs are often associated with membrane microdomains rich in cholesterol and sphingolipids, called lipid rafts.[33,41,42] These microdomains are separate, ordered structures within the laterally disordered cell membrane that may facilitate interactions between proteins and lipids with other proteins by clustering them in close proximity to each other.[26,32,33,41] GPI molecules are also found freely on the cell membrane, not attached to a protein, especially on protozoan parasites. In *Plasmodium falciparum*, the parasite causing malaria infection, the free GPI has been described to act as a toxin.[35,40]

1.1.3 **GPI-Biosynthesis**

The biosynthesis of GPIs and GPI-APs is a very complex process involving at least nine enzymatic steps. For this, the gene products of 22 *PIG* (phosphatidyl inositol glycan) genes and four *PGAP* (post-GPI attachment to protein) genes are required, Figure 1.4 a. Biosynthesis of GPI precursors begins on the cytoplasmic side of the endoplasmic reticulum (ER). Following attachment of N-acetyl glucosamine to phosphoinositol and its deacetylation,

the growing GPI-pseudodisaccharide precursor is flipped to the luminal side of the ER by enzymes that still need to be identified.[30] The protein to be glypiated is localized to the ER via an N-terminal signal peptide. Subsequently, the GPI is elongated further and is finally attached to a protein *en bloc* by GPI transamidase in a transamidation reaction, replacing a C-terminal GPI attachment signal peptide, Figure 1.4 b.[28,30,32,43,44] This process takes place within less than one minute from protein localization to the ER.[45] As a final step, the GPI-AP is transported to the golgi apparatus by protein complexes. Structural remodeling of glycan and lipid parts takes place in the golgi apparatus, followed by a final transport to the cell surface.[30,34,46]

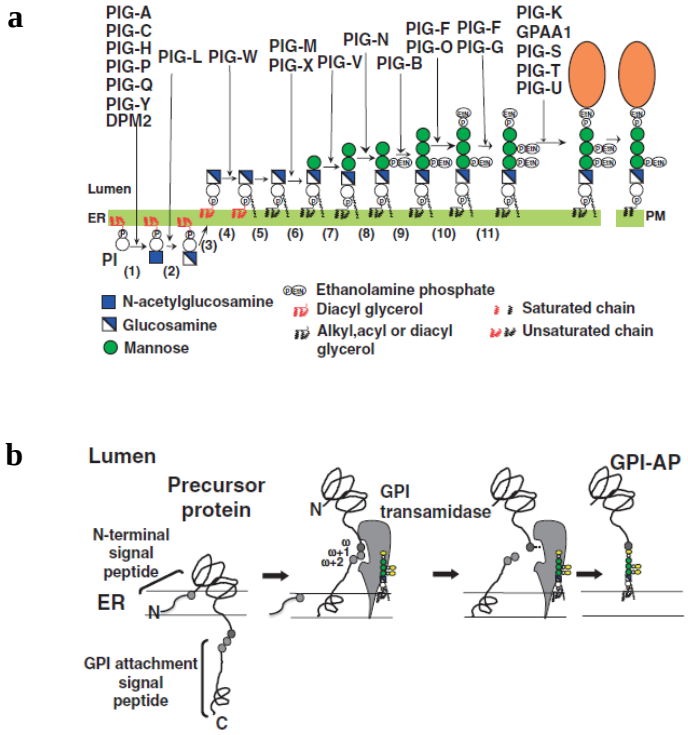


Figure 1.4. Biosynthesis of the GPI anchor and transfer to the C-terminus of a protein. a) Biosynthetic pathway of GPI anchor assembly, flip and protein attachment including all participating enzymes, **b)** Two signal sequences are required for protein glypiation: an N-terminal ER-localization signal and a C-terminal GPI attachment signal that is replaced by the GPI *en bloc* by GPI transferase (GPI-T). (Figures from Kinoshita, 2008[30]).

1.1.4 Chemical Synthesis of GPI Anchors

In order to elucidate the function of the GPI anchor, access to pure, homogeneous material is crucial. Due to their high complexity and difficult isolation, the chemical synthesis of GPI anchors has become the method of choice. However, the synthesis is challenging[34] and only in the two last decades, some routes have been developed for a convergent synthesis of different GPI anchors using a relatively small set of building blocks.[47,48] Generally, these methods

involve a convergent or a linear assembly of a fully protected GPI-glycan, which is modified at a late stage by incorporation of the lipid chains and phosphorylations.

1.1.5 Relevance of GPI-Anchors and GPI-Anchored Proteins in Disease

GPI-anchored proteins are known to be involved in several diseases and lack of GPI anchors is lethal in yeast and in mammals.[26,46,49] Additional to the infection with parasites, the most prominent diseases in context with GPIs are prion disease and Paroxysmal Nocturnal Hemoglobinuria (PNH).[15,46]

Prion diseases are a group of protein borne illnesses that are characterized by a misfolding of the physiological form of the prion protein, PrP^{c}, into the scrapie form PrP^{Sc}. Since this misfolding seems to be transmissible between proteins, it is anticipated that the GPI anchor together with the localization in lipid rafts is crucial for progressing of prion pathogenesis.[46]

PNH is caused by acquired mutations in the *PIGA* gene in a single hematopoietic stem cell and its progeny, rendering them unable to assemble the GPI anchor, as biosynthesis is abrogated in the first step.[46,49] The absence of GPI-biosynthesis leads to a lack of GPI-anchored CD59 protein (a complement inhibitor) on red blood cells and thus to uncontrolled hemolysis mediated by the complement system, inducing the disease.[46,50]

Malaria, Chagas disease and toxoplasmosis are infectious diseases caused by the parasites *Plasmodium falciparum*, *Trypanosoma cruzi* and *Toxoplasma gondii* which are transmitted by different routes and affect large populations, especially in developing countries.[46,51]

GPIs are also abundant in parasite, and their potential use as diagnostics has been highlighted in recent years.[52,53]

1.1.6 Natural GPI-Anchored Proteins

Approximately 0.5-1% of eukaryotic proteins are glypiated.[26,27] Some of the most well-known proteins among them are CD52, CD59, Thy-1 (CD90), the prion protein and the *P. falciparum* Merozoite Surface Protein family. A short overview over their functions and biological implications is given in the following.

CD52 (Campath-1 antigen)

CD52 is a short (12 amino acid) GPI-anchored glycoprotein that is widely expressed on the surface of immune cells (lymphocytes, natural killer (NK) cells, macrophages and others).[54] Due to its large sialylated *N*-glycan, CD52 is negatively charged and it may function as an anti-adhesive molecule on the cell surface, increasing cell mobility. The GPI-anchor of CD52 is cleavable by Phospholipase C, releasing soluble CD52 protein that may act as an immune regulator.[55] It is not completely understood, however, how CD52 is involved in physiology.[54] Since the beginning of synthetic studies regarding GPI-APs, CD52 has been a target of wide interest, largely due to its small size and therefore relatively easy accessibility by synthetic methods.[56,57]

CD59 (MAC-Inhibitory Protein)

CD59 is a GPI-anchored complement regulatory protein which protects the cell from uncontrolled complement-mediated lysis.[58] It is expressed in high levels on cells of each tissue type and is structurally similar to snake neurotoxins. Defects in GPI biosynthesis induce a defective expression of this protein that can lead to Paroxysmal Nocturnal Hemoglobinuria and other acquired conditions.[46] Additionally, CD59 is involved in the infection with and spread of HIV-1 by protecting infected cells and viral particles from destruction by the complement system.[58,59]

Thy-1 (Thymocyte Differentiation Antigen 1)

Thy-1, or CD90, is a conserved, heavily glycosylated GPI-anchored cell surface protein (30% carbohydrate content) with a single V-like immunoglobulin domain, rendering it the smallest member of the immunoglobulin protein family.[60,61] It was the first protein for which the structure of the GPI-anchor was elucidated.[62] It is known as a T-cell marker and is most abundant on murine T-cells. However, the physiological functions of Thy-1 protein are still not completely understood.[62,63] Thy-1 is not only found on T-cells, but also in nervous tissue and is possibly involved in fear inhibition.[64] This protein also plays a role in different types of cancer, especially in chemo-resistance, cancer cell proliferation and metastasis. In some cancers, however, it also has a tumor suppressing function. The mode of action strongly depends on the tumor microenvironment and the signaling mechanisms present in the cancer cells. Due to these implications, Thy-1 is interesting as a cancer biomarker or could possibly be exploited as a therapeutic target.[65]

MSP1 (Merozoite Surface Protein 1)

Malaria is an infectious disease caused by the apicomplexan parasite *Plasmodium falciparum* that was responsible for an estimated 216 million cases and 655,000 deaths in 2010.[51,66] The parasite is transmitted to the human host via the bite of mosquitoes as vectors. Following inoculation in the human blood, the parasite invades the liver (so-called liver-stage) where it reproduces, causing infected hepatocytes to burst. Then, the released parasites infect the host's red blood cells, entering the blood stage of infection, Figure 1.5. In this stage, the MSP1 protein, a GPI-anchored protein present on malaria merozoites, plays an important role during infection.[67] MSP1 is expressed in merozoites as a large (195 kDa) glycoprotein. During infection of red blood cells, the protein undergoes proteolytic processing by subtilisin 1 and 2 proteases, leaving only a small GPI-anchored fragment ($MSP1_{19}$) on the parasite surface.[67,68] $MSP1_{19}$ binds to the red blood cell, initiating contact between parasite and cell.[67] Therefore, MSP1 has been an interesting therapeutic target for malaria treatment[69] and a vaccine candidate.[70]

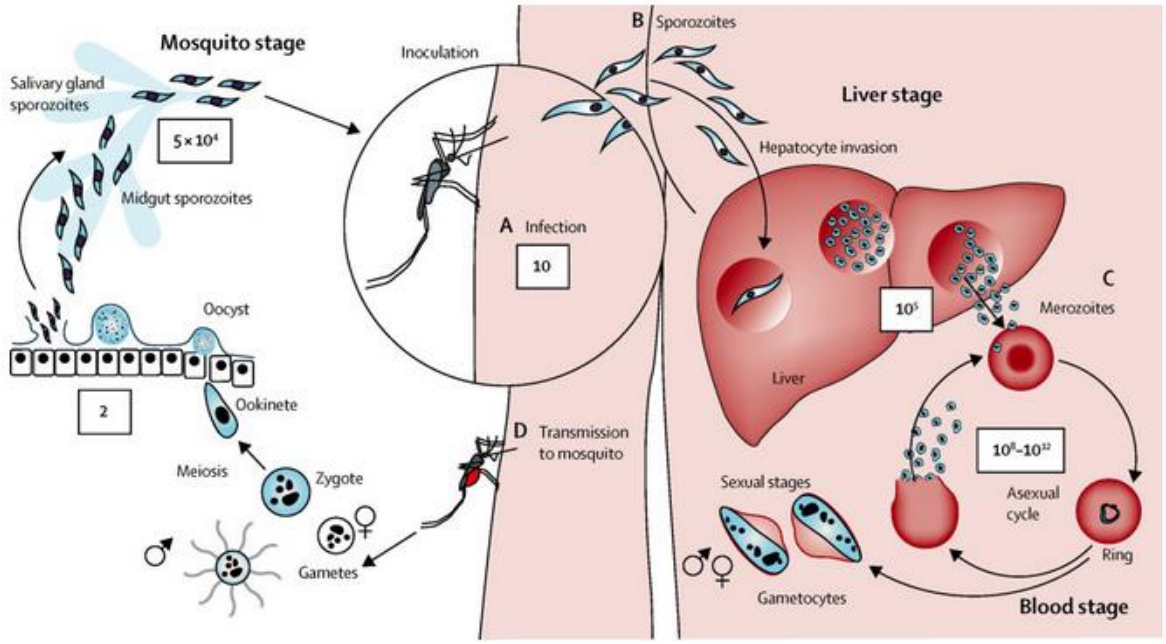


Figure 1.5. Life cycle of *Plasmodium falciparum* during malaria infection. (Figure from White et al., 2013[66]).

Prion Protein

Probably the most notorious example among GPI-anchored proteins is the prion protein, which is responsible for different kinds of transmissible spongiform encephalopathies (TSEs) and the Creutzfeldt-Jakob disease (CJD) affecting the nervous system of humans and many

animals.[71] Differently from initial expectations, the prion protein was found to induce these diseases solely via protein-based mechanisms.[72-74] The term prion was coined by Stanley Prusiner in 1982 and is derived from **pr**oteinaceous and **in**fectious particle (represented by the ending –**on**, from the greek word for particle).[72,73]

This "protein only" hypothesis is widely accepted now. Prion pathogenesis is reported to be the effect of a conformational conversion of the physiological, normal form of the prion protein (PrP^{C}) into a disease-associated form, the so-called scrapie form of PrP (PrP^{Sc}).[75] This misfolding can be transferred from the PrP^{Sc} to a PrP^{C} according to the "refolding" or template assistance model, Figure 1.6 a. Due to an interaction between endogenous PrP^{C} with exogenous PrP^{Sc}, PrP^{C} undergoes a conformational change into PrP^{Sc}. Otherwise, a high energy barrier prevents this transformation. An alternative model of prion pathogenesis is the "seeding" or nucleation-polymerization model (Figure 1.6 b). According to this model, several monomeric PrP^{Sc} molecules can form a highly ordered seed that recruits PrP^{C}, induces its misfolding and causes aggregation to form amyloids. The formation of these aggregates stabilizes the PrP^{Sc} form. Fragmentation of these structures can lead to replication of the so-called prions.[75]

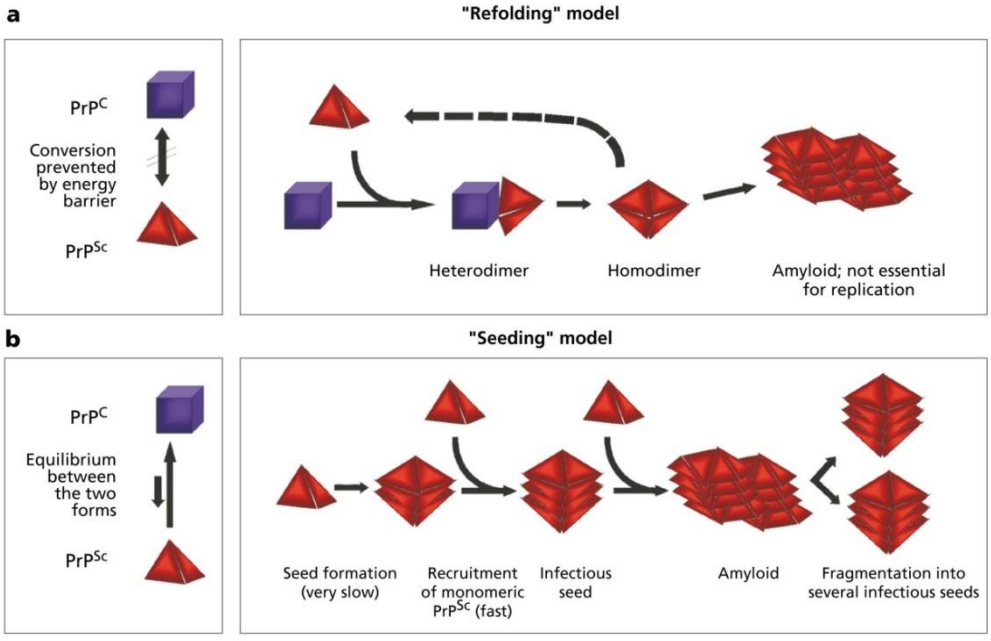
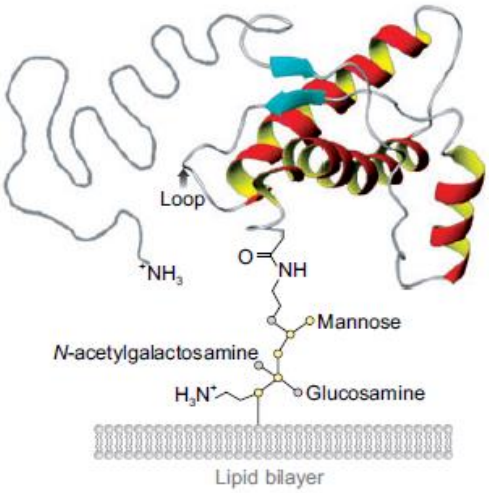


Figure 1.6. Possible mechanisms of prion propagation and pathogenesis. (Figure from Aguzzi & Heppner, 2000[75]).

PrP^{C} and PrP^{Sc} differ structurally and in their biochemical properties. While PrP^{C} consists mainly of α-helical structures with a long, flexible N-terminal tail, is soluble and sensitive to

digestion by proteinase K digestion, PrP^{Sc} has a structure containing mostly β-sheets, is protease resistant and prone to aggregation (Figure 1.7).[73,76] As the misfolding events seem to take place in lipid rafts where many GPI-anchored protein cluster,[33] it is assumed that the GPI-anchor plays an important role in prion propagation and pathogenesis. The membrane composition also seems to play an important role in this process, namely the cholesterol content, which is an important component of lipid rafts. The GPI-anchor enhances lateral protein mobility between cells, a factor that could enable PrP^{Sc} propagation from one cell to another.[45,77]



Biochemical properties

PrP^{C}	PrP^{Sc}
Rich in α-helices	Rich in β-sheets
Soluble	Insoluble
PK sensitive	PK resistant
No aggregation	Aggregation

Figure 1.7. Properties of prion protein (Figure from Aguzzi et al., 2008[76]).

The amyloids formed in the process of prion pathogenesis can aggregate in the brain and lead to cell death. This process creates a sponge-like structure in the brain, causing neurodegenerative symptoms that eventually lead to death.[78] A cure is currently not available for any of the prion diseases.[79,80] As the physiological function of the prion protein is still not completely understood,[73,76] effort towards the understanding of this fatal disease and its molecular mechanisms is required in order to develop effective therapies.[73]

1.1.7 **Therapeutic Proteins**

Linking naturally non-GPI-anchored proteins to GPIs might facilitate their targeting, trafficking and membrane localization using the ability of GPI-APs to spontaneously insert into lipid bilayers.[27,81,82] Nowadays, therapeutic proteins are administered by injection, a

method that is not favorable for most patients.[83] Therefore, new administration ways are necessary. Protein glypiation has been suggested as a potential modification for therapeutic proteins, such as antibodies or cytokines, as it could improve their effectiveness and *in vivo* half-life and at the same time enable lower dosage, resulting in fewer adverse effects or less frequent administration.[84] Two possible examples for small therapeutic proteins are described below.

Interleukin-2 (IL-2), also known as T-cell growth factor (TCGF), is a peptide hormone involved in immune stimulation and regulation. It is secreted by activated T-cells, stimulating proliferation and differentiation of B- and T-cells as well as of cytotoxic and NK cells. It can also act as an immune regulator by promoting homeostasis of regulatory T-cells, maintaining self-tolerance.[85] IL-2 is currently used under the name Aldesleukin for treatment of renal cell carcinoma (it is approved in Germany, Austria and Switzerland as *Proleukin* by Novartis).[86] Here, the immunostimulatory effect of IL-2 is exploited.

In a study published in 2002,[87] Ji and coworkers fused IL-2 to a GPI anchor which was achieved by including a GPI-signal sequence in melanoma cells (B16F0 cell line). It was shown that the protein localized to the tumor cell membrane, resulting in high local concentrations of IL-2 and inhibited melanoma cell growth. This higher activity and localization would enable lower dosage and reduce the severe adverse effects which are common in systemic administration of IL-2.

Interferon-α2a (IFN-α2a) is the most prominent member of the cytokine group of interferons. Interferons are produced in response to viral infection, conferring resistance to virus replication to non-infected cells by activating genes which inhibit viral translation and degrade viral RNA.

Additionally, they activate dendritic cells, macrophages and NK cells, which leads to the killing of infected cells.[88,89] Interferons are used to treat viral infections, such as HIV and hepatitis B and/or C.[83,90] The main problem of interferons as therapeutic proteins is their short half-life time in the blood, making frequent injections necessary to maintain an effect. Following a similar strategy to IL-2, GPI-anchoring of IFN-α2a may be a potential approach in order to increase its half-life in cells.

1.2 Protein Semi-Synthesis

Nowadays there is a high demand for proteins in research and medicine.[91] Most commonly, proteins are obtained by isolation from natural sources or by expression in mammalian cell lines, yeast cells or bacterial cells, i.e. by utilizing protein biosynthesis pathways. This has the advantage of relatively easy access to biomolecules of any size. However, these methods have some limitations like the difficult or impossible installation of site-specific modifications, even if mutagenesis methods are applied.[92] For many modified and unmodified proteins, isolation from biological samples methods often results in heterogeneous products, rendering exact investigation of the function of specific modifications difficult.[12,13] Purely chemical methods, on the other hand, often offer good control over the site of modification but are limited regarding the yield and size of the proteins that can be obtained.[93]

Protein semi-synthesis, defined as the generation of a (modified) protein from a synthetic and a recombinant or native part,[94] is an emerging field. A variety of methods are under investigation including chemical protein synthesis using Expressed Protein Ligation (EPL), cell free expression and genetic code expansion using amber suppression, to name a few.[5,92,95-100] Bioorthogonal chemistry is another option to introduce site-specific modifications or labels to proteins. However these methods often introduce non-natural connections between protein and modification.[95,101] Another possibility to generate proteins with natural modifications at the N- and C-termini is the use of inteins.[102,103]

1.2.1 Inteins and Protein *Trans*-Splicing

Inteins are protein domains capable of cleaving themselves out of flanking protein sequences in a folding dependent, autocatalytic, enzyme-like manner, without requiring any cofactors or energy in the form of ATP, and linking the flanking residues with a natural peptide bond.[102,104] Analogous to mRNA splicing, where introns are spliced out of the RNA sequence to leave only the exons (the mature mRNA), the intein cleavage is called *protein splicing.*[103] Accordingly, protein splicing is a form of post-translational protein processing.

Inteins were first discovered in the early 1990s upon the finding that the final protein product of a yeast ATPase gene (*TFP1*/*VMA1*) was much shorter than expected, however, cDNA and mRNA were found to be exactly as expected. Inteins were first described by Hirata et al. and Kane et al. in 1990.[105,106] Inteins are mainly found in endonucleases and

proteins involved in DNA replication and repair. They exist as large and mini inteins, both of which contain the splicing domain. The large inteins consist of the splicing domain and an endonuclease homing domain which has the role to support lateral transfer of the intein between genomes.[102] This way inteins are replicated within genomes, and since they have no known use for the host organism, they can be considered selfish.[107,108] Inteins are found in all kingdoms of life[103,109] and are present in proteins with diverse biological functions, for example metabolic enzymes, proteases, but mainly in enzymes participating in DNA replication and repair.[110] The enzyme-like mode of action of the protein splicing reaction can be described by a four-step mechanism, Figure 1.8.

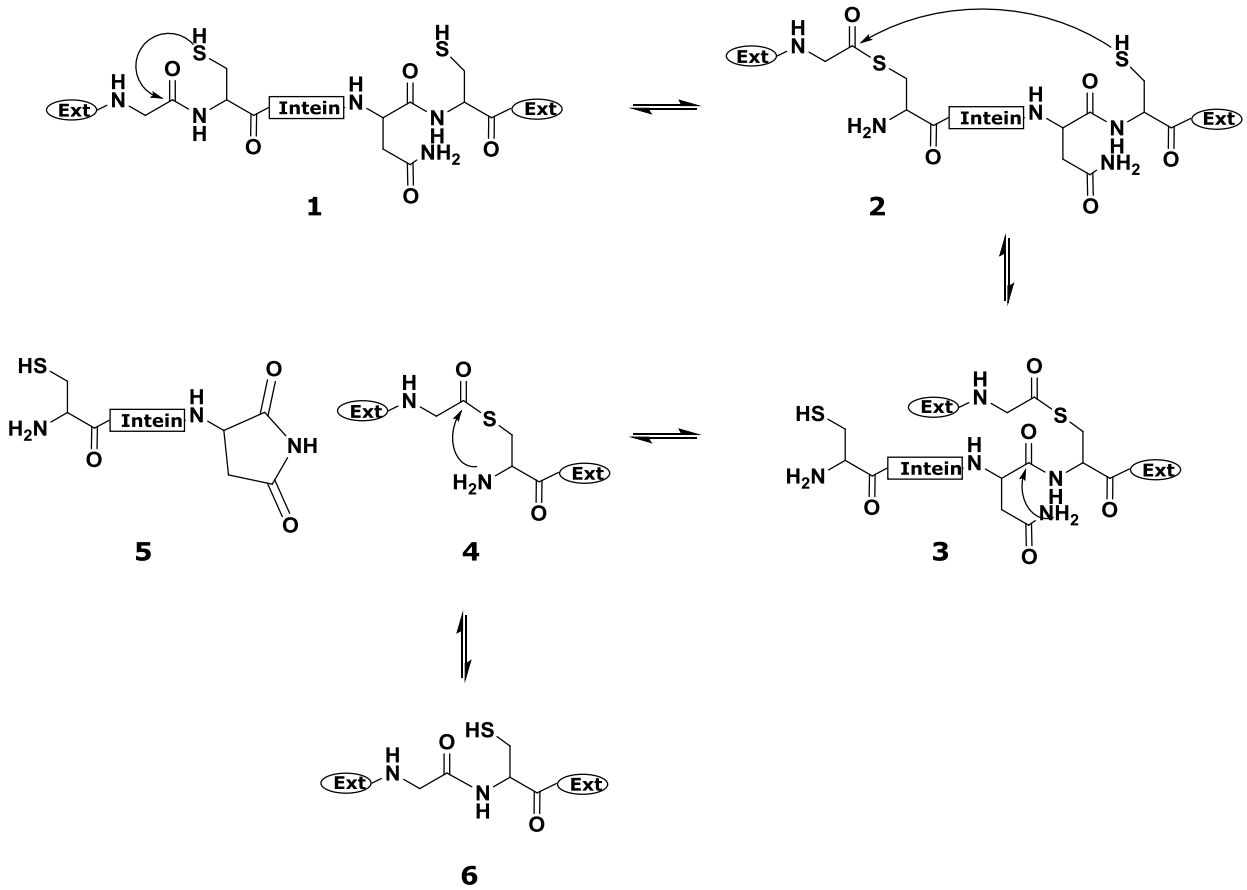


Figure 1.8. Protein Splicing Mechanism. The protein splicing mechanism is a four-step process starting with activation of the N-terminal splice junction by an N-S (or O) acyl shift, giving a thioester intermediate (2). In step 2, a nucleophilic residue from the C-terminal splice junction attacks the thioester, resulting in a branched thioester intermediate (3). Next, intein cleavage is performed by asparagine cyclization, leading to excision of the intein (5). Intermediate (4) finally undergoes a spontaneous rearrangement resulting in the formation of a native peptide bond between the two exteins, giving the splice product (6).[107]

In the protein splicing reaction, initially the side chain group of a conserved Cys or Ser residue in the intein N-terminal splice junction attacks the carbonyl group of the first extein residue inducing an N-to-S/O acyl shift, resulting in formation of a (thio)ester intermediate. In

the second step, a trans(thio)esterification by a nucleophilic attack of the side chain of a conserved Cys or Ser residue in the C-terminal splice junction takes place, giving a thioester intermediate. In the following step, the C-terminal asparagine of the intein undergoes a cyclization reaction, the intein is excised out of the protein sequence upon succinimide formation, and the two exteins are linked together with an ester bond. The final step is a spontaneous rearrangement, in which a peptide bond is formed via an S/O-to-N acyl shift between both exteins.[94,102,104,107] It has been shown that some inteins require only the first downstream extein residue for protein splicing, although this splicing sometimes is inefficient.[103,111]

Intein Structure

Most inteins show low sequence homology and only the terminal regions and some internal motifs required for activation of the peptide bond to be broken in the first step are conserved.[112] The structure of inteins is similar to hedgehog proteins. It contains 12-14 β-sheets forming a U-shape, having the catalytic center in the center of the U on the ends of two neighboring sheets. The splice area forms a separate domain.[110] The complex formed by the two fragments of a split intein has the same fold as contiguous inteins have.[113]

***Trans* Inteins and Protein *Trans*-Splicing**

In 1998, so called *trans* or split inteins were simultaneously discovered in nature and developed artificially.[112] While the classical intein described above splices in *cis*, i.e. within its own protein molecule, the fragments of a *trans* intein are expressed on different positions in the genomes and splice exteins from two distinct proteins to give one final protein. The first naturally split intein discovered was the *Ssp*DnaB intein from the cyanobacterium *Synechocystis species.*[112] Whereas *cis* inteins have been found in all kingdoms of life, split inteins have mainly been identified in different cyanobacteria and archeae.[102] It seems likely that split inteins have evolved from formerly complete inteins that got divided.

Split inteins use the same mechanism as *cis* inteins with an additional initial step involving the association of both intein fragments.[94,114] The intein fragments have very distinct pI values, i.e. the fragments of the *Nostoc punctiforme* intein have pI 4.4 for Npu^N and pI 9.7 for Npu^C, due to which the multi-step split intein folding mechanism is initiated via electrostatic interactions of the two fragments (capture). Subsequently, the so far unstructured intein fragments become stabilized in a “collapse” mechanism, Figure 1.9. The N-intein alone has an extended shape, which seems to collapse upon binding of the C-terminal fragment. The

surface between both intein fragments is mainly hydrophobic, which helps in the stabilization of the complex.[113] Once it is assembled, the intein complex is very stable and shows very low K_D values and very slow dissociation rates.[110]

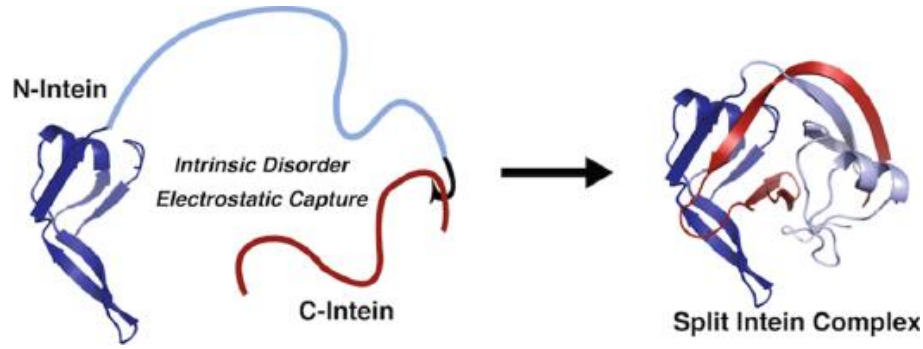


Figure 1.9. Schematic view of the „Capture and Collapse" folding mechanism of naturally split inteins. The N-intein (blue) consists of two lobes, a structured Npu^N_1 and an unstructured Npu^N_2. Only when the N-intein is in close proximity to the C-intein (red), the intrinsically disordered fragment collapses due to an electrostatic capture. This association results in the readily folded complete split intein complex in which both termini are next to each other for the splicing reaction to be performed. (Figure from Shah et al., 2013[115]).

The use of protein *trans*-splicing (PTS) has several advantages over the use of NCL and EPL. Due to the high affinity between Int^N and Int^C, the reaction rates are very high compared to the other selective chemical reactions and no free thiols re required as catalysts.[116] Protein *trans*-splicing is not limited by the size of the fragments, which can be expressed recombinantly. Nevertheless, the smallest known C-terminal intein fragment (Int^C of *Npu* intein) consists of only 36 amino acids, which is accessible by solid phase peptide synthesis. Furthermore, PTS is less concentration dependent than NCL, which relies on random collision of the fragments in the reaction mixture.[117] The cysteine residues flanking the intein fragments on both sides need to be present in reduced form for the splicing reaction.

Split Intein from *Nostoc punctiforme*

The best characterized split inteins are from cyanobacterial sources.[118] The split intein from the filamentous cyanobacterium *Nostoc punctiforme*, *Npu*DnaE intein, is the fastest split intein known. The $t_{1/2}$ of its *trans*-splicing reaction is reported to be ~1 min at 37°C. The reaction conditions needed for PTS (Protein *trans*-splicing) are gentle and very close to physiological conditions and therefore, PTS is possible even in live cells.[116] The *Npu*DnaE intein is also the best characterized of the ultra-fast split inteins.[113]

Applications and Biological Role of Inteins

Inteins have multiple applications in protein chemistry and biotechnology, such as the generation of self-cleaving affinity tags for protein purification, novel ligation systems for protein semi-synthesis, segmental labeling of proteins for NMR analysis,[104] the generation of tag-less proteins and protein thioesters,[117] the segmental construction of proteins or the selective modification of proteins. Non-native exteins are accepted but splicing rates can vary significantly depending on the respective extein.[110,118] Inteins can also be used for protein labeling (fluorophores etc.) with the advantage that the biological function of the proteins is not changed as in most other protein labeling methods.[116] No advantage for the host is known, although inteins are often found in essential proteins.[110] Therefore, inteins are considered selfish protein elements.[107,108]

1.2.2 Expressed Protein Ligation

Native chemical ligation is a widely used technique for the chemoselective reaction of two unprotected peptide fragments, one carrying a C-terminal α-thioester, the other one having an N-terminal cysteine residue.[119,120] After forming a thioester-linked intermediate in a chemoselective transthioesterification, the reaction proceeds with a rapid S-to-N acyl shift, ligating both peptides together with a native amide bond, Figure 1.10.

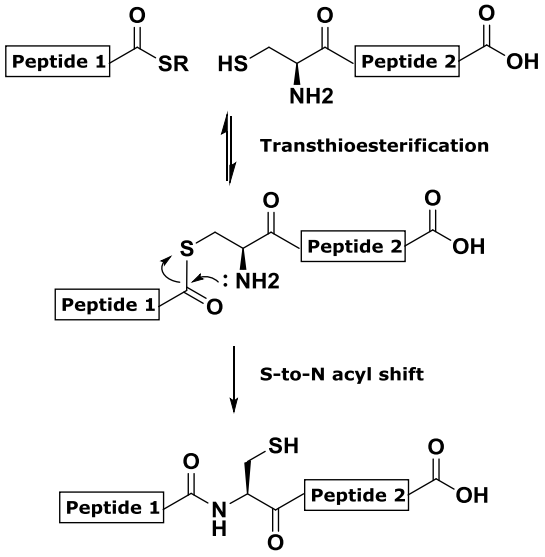


Figure 1.10. Native Chemical Ligation between two unprotected peptides. Peptide 1 has a C-terminal thioester and peptide 2 has a free N-terminal cysteine. The peptides undergo a transthioesterification reaction, forming a thioester-linked intermediate. This step is followed by an S-to-N acyl shift, ligating both fragments together via a native peptide bond.

The NCL strategy is limited by the access to only small protein fragments. Due to the technical limitations of solid phase peptide synthesis (SPPS), only peptides with a maximum size of 50-60 amino acids are generally accessible. An expanded variation of the native chemical ligation reaction is the expressed protein ligation (EPL), in which one or several of the fragments to be ligated are of recombinant origin.[99,120-122] Using recombinant protein expression, larger proteins and protein fragments can be generated.[120,121] Thus, by using EPL, the main limitation of NCL can be overcome. The chemical routes to generate peptide thioesters generally involve the use of protection groups and are not well applicable to large recombinant proteins. Therefore, for recombinant proteins a different strategy was necessary for the generation of reactive peptide-α-thioesters. The protein-splicing activity of inteins (self-excising protein elements), which is known to involve a thioester intermediate, can be applied to generate recombinant protein-α-thioesters.[120,121,123] This technique relies on the capture of the formed protein thioester intermediate by thiols, involving a trans-thiosterification reaction. This is followed by a native chemical ligation reaction, ligating the Cys-bearing peptide fragment to the recombinant protein thioester, Figure 1.11.

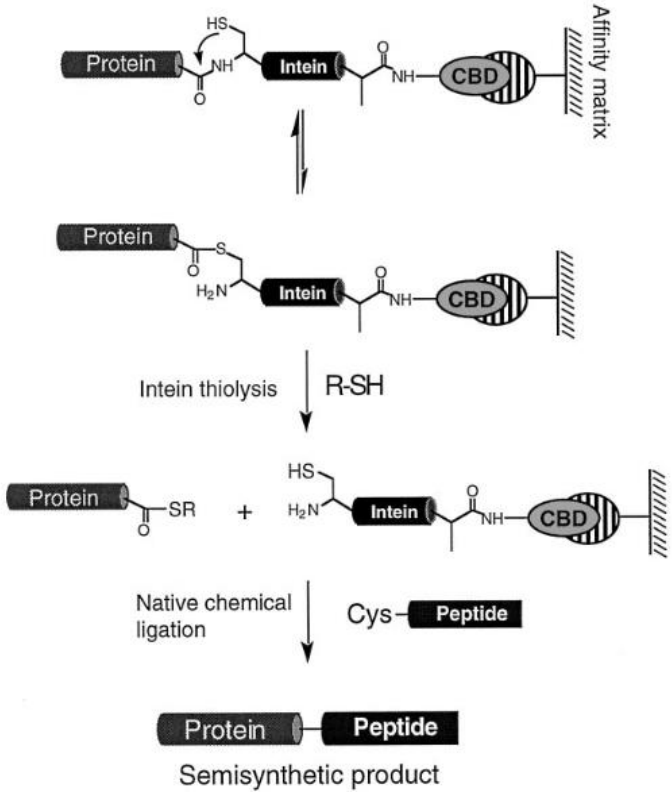


Figure 1.11. Expressed Protein Ligation (EPL). A protein thioester is formed with the using a C-terminally de-functionalized intein that is bound to an affinity matrix. A thiol reagent is added to capture the thioester intermediate formed due to the de-functionalization of the intein. The purified protein thioester can then be ligated to a peptide fragment with an N-terminal cysteine residue (Figure from Muir, 2003[120]).

Thioesters are base-labile and prone to hydrolysis.[123,124] Johnson & Kent (2006)[125] and Reif et al. (2014)[124] performed screenings to find the optimal thiol reagent for thioester formation and protein ligation. Generally aryl thiols, such as thiophenol or 4-mercaptophenylacetic acid (MPAA), are more reactive than alkyl thiols, such as ethanethiol or 2-mercaptoethane sulfonate sodium salt (MESNA). On the other hand, alkyl thioesters are more stable and easier to handle. MESNA has a good balance regarding stability and reactivity is the most commonly used thiol reagent for EPL.[122,125] However, using MESNA, the thioester formation as well as the ligation reaction are slow, requiring several days to go to ~90 % completion.[126] Therefore, NCL and EPL can be improved using a thiol exchange, as described by Dawson et al. in 1997.[127] Thioesters are generated using a less reactive thiol to obtain a stable thioester that can be purified and stored. During the ligation this thioester can be converted to a more reactive species by the addition of a more reactive aryl thiol.

Using the phenyl thioester generated with thiophenol or MPAA, the expressed protein ligation reaction has been shown to obtain reaction up to 90% completion within 2 h.[125] However, thiophenol and its variant MPAA are also significantly prone to hydrolysis.[124] showed that using 4-(mercaptomethyl)benzoic acid (MMBA) the respective thioester is formed in an equally short time and only trace amounts of side products were formed, indicating low levels of hydrolysis. These characteristics suggest MMBA as a useful thiol for the formation of the protein thioester and for protein ligation.

1.2.3 **Semi-Synthesis of GPI-Anchored Proteins – State of the Art**

GPI-anchored proteins are a significant target for protein semi-synthesis. The isolation of these modified proteins generally results in mixtures of glycoforms and low yields and their total chemical synthesis faces several hurdles including availability of large peptides by synthetic methods. So far, there is a lock of suitable semi-synthetic methods for the generation of pure and homogeneous GPI-anchored proteins. A main challenge remains the access to sufficient amounts of homogeneous GPI-anchors. Chemical synthesis of these glycolipids is very challenging as multiple components (carbohydrates, lipids and phosphates) are involved, as described before in 1.1.4.[34,47,48] The high variability of the structure of the carbohydrate portion and the different lipid chains that can be present, including the presence of unsaturated lipids and ceramides makes GPI synthesis even more challenging.[34] Some

progress has been recently made towards both types of GPI modifications in our group,[47,128-130] however there are still many challenges to obtain these molecules.

Due to the amphiphilic character, the handling of GPI anchors is difficult. Therefore, the (semi-)synthesis of GPI-anchored peptides and proteins has been mostly evaluated using GPI fragments, analogues or mimics with apparently similar molecular properties.[34,131,132] In early studies, very simple GPI mimics, protein models and unnatural linkages were employed to connect GPI analogues to peptides and proteins.[133] However, the large difference between these models and the natural structures makes it difficult to elucidate the function of the GPI anchor itself in its native state on the protein's C-terminus.

Total chemical synthesis, mostly by Native Chemical Ligation (NCL), has been an option for the synthesis of small peptides carrying GPI-analogues.[119,134] This strategy requires the generation of peptide thioesters, for example via peptide hydrazides and can be used without limitation such as solubility of the starting molecules.[135,136] One of the first reports in this field was the study of Tanaka et al. (2003)[56] in which NCL was used to generate Cys-GPI-analogues (Cys-dimannose) and peptide thioesters (CD52 peptide). Another study addressing the synthesis of GPI-anchored CD52 peptide was conducted by Xue et al. in 2003.[57] In this study, the two C-terminal amino acids of CD52 were ligated to a GPI mimic containing the carbohydrate core structure and a single C_{16} acyl chain linked to the glycan without phosphate.

Expressed Protein Ligation

As described in 1.2.2, Expressed Protein Ligation (EPL) is commonly used for the semi-synthesis of (glyco-)proteins, and is also applied for studies for the generation of GPI-APs by some groups, an overview over the most successful ones is given in the following.

The Bertozzi group investigated the effect of PEG linkers in comparison to glycans of various lengths on the properties of eGFP protein. In a first report dealing with the attachment of GPI mimics to a larger protein, Paulick et al. (2007)[137] produced synthetic GPI analogues with PEG linkers of different lengths attached to none, one or two mannoses. These mimics were ligated to eGFP. The phosphoethanolamine-inositol-glucosamine moiety was replaced by a simple PEG linker, but two C_{16}-chain lipids were present in the mimic (C16). In the EPL reaction, an excess of GPI analogues was applied. The modified eGFPs were purified by extraction with Triton X-114 as detergent and the effect on protein mobility was investigated. It was found that the GPI mimic can prevent protein-membrane interactions by enhancing protein mobility, probably due to higher rigidity in the presence of higher number of sugars.

Olschewski et al. (2007)[132] reported the semi-synthesis of the Prion Protein (PrP 90-231) fragment with GPI mimics. Lipidated peptides were used as GPI mimics to facilitate membrane targeting. The peptides were equipped with a fluorescent label for detection purposes and a PEG-linker in place of the glycan. Purified proteins were obtained in 18-30% yield, with 60% yield in the EPL reaction. However, due to the lack of natural material to compare them it is difficult to know to which extent these mimics can represent a GPI-anchor in biological studies.

The same group reported the semi-synthesis of a GPI-anchored PrP using the expressed PrP 90-231 fragment that was ligated to a monolipidated Cys-GPI (Becker et al., 2008).[131] Reaction with the crude GPI (1.5 eq.) at pH 7.8 gave the product in 50% yield. The product was purified by RP-HPLC and characterized in LC-ESI-MS, which was reported to be a difficult process due to the amphiphilic nature of the compound that delivered an MS spectrum of low quality. Furthermore, the insertion of the purified PrP-GPI into membranes (SUV) was demonstrated. This partial sequence of PrP, however is missing the N-terminal region (23-89), which has been described to be responsible for the aggregating properties of the Prion Protein.

Protein *Trans*-Splicing

Only few reports describe the use of PTS to generate GPI-APs. Olschewski et al. (2007)[132] applied also the PTS strategy for the semi-synthesis of PrP fragments with lipidated peptides as GPI mimic in 25% yield after purification by HPLC. Furthermore, they also performed PTS with lipidated peptides already inserted in liposomes.

Another report involving PTS for generation of a glypiated protein was reported by the Mootz group (Dhar & Mootz (2011)[138]). In this report, the Npu^C intein fragment was expressed in yeast, carrying a GPI signal sequence to obtain a glypiation *in vivo* prior to its use in PTS with eGFP-Npu^N in live cells. Using this strategy it was possible to show localization of the GPI-AP on the cell membrane. However, the GPI anchor attached in the yeast culture is likely not homogeneous and it is not obvious on which residue of the protein the GPI was attached as analysis in mass spectrometry was not performed.

Sortase A

Sortases are transpeptidases of bacterial origin that naturally transfer proteins to the bacterial cell wall.[139] Some groups employed these enzymes for the chemoenzymatic generation of GPI-APs. Sortase A (SrtA) from *Staphylococcus aureus* recognizes a LPXTG

peptide sequence near the C-terminus of a protein, cleaving the bond between the Thr and Gly, which causes residues in the active site to generate a reactive thioester between Thr of the substrate protein and a Cys residue of SrtA. Subsequently, it transfers the Thr residue of the protein to another substrate protein, or to a synthetic peptidoglycan, especially to fragments having a polyglycine at the N-terminus.[38,81,139]

The Guo group has reported progress using Sortase A for generation of GPI-APs. In Guo et al. (2009)[81], a Cys-Phosphate-Mannose could first not be ligated to a peptide, probably due to the phosphate group that might alter the nucleophilicity of the amino group. Upon addition of two extra Gly residues the N-terminus, product formation was observed. The same GPI-analogue including the two Gly residues was coupled by Wu et al. (2010)[38] to attach a partial sequence of the MUC1 glycopeptide with three glycans and the signal peptide LPKTG. In a following report, Wu et al. (2013)[140] successfully attached eGFP to a GPI derivative containing the GPI core glycan structure, a C_{18}-llipid chain, and phosphate linkages having a GG dipeptide. Due to the Sortase mechanism, however, the final product carries the recognitions sequence LPATGG between the C-terminus of eGFP and the GPI.

Other Methods to Obtain GPI-APs

GPI Glycan-Peptide Conjugates (CD52 peptide) were obtained by Traceless Staudinger Ligation in Zhu & Guo (2017).[141] A Staudinger reaction was carried out between a peptide phosphinoester and a GPI glycan azide, Figure 1.12. This region- and chemoselective reaction was expected to give peptides naturally linked to GPI glycan mimics (containing one to three mannose residues). The reactions were performed in DMF and a DMF-water mixture (1:1) using a fully protected peptide. The CD52 attached to a single mannose and to a trimannose were obtained at 40°C after 1 day reaction in 96% and 78% yield, respectively. The big advantage of this process is its independency of any sequence requirements as the other methods. However, the use of fully protected peptides renders this method useless for the ligation of bigger, recombinant proteins.

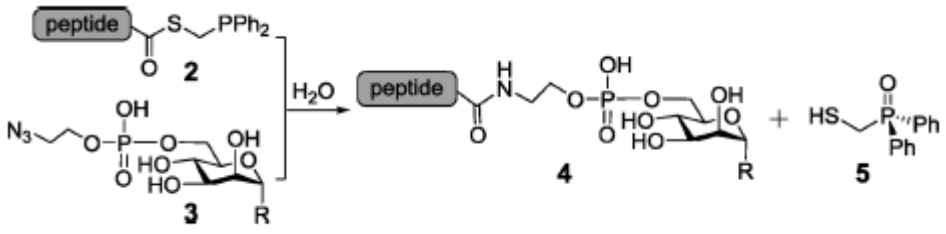


Figure 1.12. Traceless Staudinger Ligation for the generation of GPI-peptide analogues. (Figure from Zhu & Guo, 2017[141]).

A purely biological option to obtain GPI-APs is the direct expression of the GPI-APs in mammalian expression systems such as yeast or other eukaryotic cell lines. In Marbach et al. (2013)[142] PrP was obtained with a native-like glycosylation and glypiation pattern and localized to the yeast plasma membrane via a GPI anchor. However, they observed unexpected insoluble expression of the protein without GPI anchor. eGFP-GPI was also expressed using GPI signal sequences and was found to localize to the cell membrane. High protein expression yields in yeast are given as the main advantage of this method. However, the lack of characterization of the GPI attached to the protein and of a proper determination of the yields of a single product show that this method cannot become a general strategy for the production of GPI-APs. Furthermore, the GPI-glycan attached to the proteins will depend on the glycosylation machinery of the expression host, and could differ from the natural form, like it is the case for yeast and the PrP.

Expression in mammalian systems has rarely been used for the generation of GPI-APs.[143] This, of course, delivers completely natural products, however, the homogeneity cannot be controlled.[12,13] Expression in insect cells has also been reported (Shams-Eldin et al., 2008[144]), which, however suffers from the same problems as mammalian systems.

Most of the methods described above yield unnatural products in one way or the other for many proteins.[34] NCL and EPL involve the addition of at least one extra Cys residue, if this is not the C-terminal residue of the target protein as it is the case for MSP1 and Thy-1 proteins. An additional strategy is the desulfurization strategy, which can be used to convert a cysteine to an alanine residue after ligation.[119] Furthermore, the synthesis of Cys-GPIs is not trivial and the use of SrtA yields products bearing a pentapeptide with unclear implications on the structure and biological functions of the proteins.

Unexplored methods such as the use of GPI transamidase (GPI-T) can be an interesting tool to ligate GPIs to proteins, since this would deliver truly natural GPI-APs.[34] However, it is a very complex enzyme made up of five subunits, which would hamper its application in this challenging task.

1.3 Aim of the Work

The development of strategies for the production of homogeneous GPI-anchored proteins is required to decipher the role of this post-translational modification in eukaryotic proteins.

In order to obtain proteins to decipher the role the GPIs, the aim of this thesis was the development of semi-synthetic strategies to obtain GPI-APs using inteins.

First, native chemical ligation will be applied for the attachment of chemically synthesized GPIs having a cysteine residue to the C-terminus of recombinant proteins. The proteins will be obtained as active protein thioesters by an intein-based strategy using the *Mycobacterium xenopi* Gyrase A intein. Following intein cleavage, the released protein of interest having a C-terminal thioester will be investigated in reactions with GPIs of different structures. A second strategy to be investigated will involve the use of protein *trans*-splicing. Here, the naturally occurring split intein from *Nostoc punctiforme* (NpuDnaE) will be used for the reaction.[145] Recombinant fusion proteins, consisting of the protein of interest and the N-terminal part of a split intein, will be obtained by recombinant expression and will be reacted to the chemically synthesized C-terminal intein part, which will be coupled to a GPI.

Recombinant expression in *E. coli* strains will be used to obtain the fusion proteins of the POIs with the inteins. This commonly used bacterial expression system generally yields high protein amounts relatively easily and it will ensure that the proteins will be present in a homogeneous, non-glycosylated form. For establishing the methods, eGFP will be used as a model protein and fusion partner for the inteins. The expression of this protein is known to be possible in *E. coli* in a soluble form and it is easy to follow in purification and reactions due to its intrinsic fluorescence.

As soon as EPL and PTS methods are established for eGFP, the established strategies will be applied to obtain other proteins coupled to GPI-anchors, creating a GPI-AP library, which can then be used for structural or activity-related studies in order to elucidate further biological functions of the GPIs.

In addition to protein ligation and in order to be able to follow the progress of the reactions, suitable methods for detection will be evaluated and established. Methods such as LC-MS analysis of intact GPI-APs and western blot using anti-GPI antibodies have so far not been used for the characterization of this type of molecule. It is anticipated that ionization of these complex molecules might present a challenge due to the molecular diversity and complexity.

2 Materials and Methods

2.1 Chemicals, Buffers and Consumables

2.1.1 Chemicals

If not stated otherwise, chemicals were purchased from Carl Roth (Karlsruhe, Germany), AppliChem (Darmstadt, Germany), Sigma-Aldrich (St. Louis, USA) or VWR (Radnor, USA). All amino acids were purchased from IRIS Biotech (Marktredwitz, Germany) or Novabiochem (Darmstadt, Germany). Deionized water was obtained by purification with a Milli-Q purification system (Merck Millipore, Billerica, USA).

Chemical Name	Supplier
Acrylamide 4K solution 30 %	AppliChem (Darmstadt, Germany)
BSA	Life Technologies (Darmstadt, Germany)
Deoxycholate, Na-salt	Roth (Karlsruhe, Germany)
D-desthiobiotin	Sigma-Aldrich (St. Louis, USA)
DIC (N,N'-Diisopropylcarbodiimide)	Iris Biotech (Marktredwitz, Germany)
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	Roth (Karlsruhe, Germany)
MESNA (Sodium 2-Mercaptoethanesulfonate)	Sigma-Aldrich (St. Louis, USA)
MMBA (4-(Mercaptomethyl)benzoic acid)	was synthesized as described in Reif et al.[124]
MMP (Methyl 3-mercaptopropionate)	Sigma-Aldrich (St. Louis, USA)
MPAA (4-Mercaptophenylacetic acid)	Sigma-Aldrich (St. Louis, USA)
Octyl β-D-glucopyranoside	Santa Cruz Biotechnology (Dallas, USA)
Oxyma	Iris Biotech (Marktredwitz, Germany)
Thiourea	Sigma-Aldrich (St. Louis, USA)
TIPS (Triisopropylsilane)	Sigma-Aldrich (St. Louis, USA)

2.1.2 Consumables

Product	Supplier
1 kb Plus DNA Ladder	Invitrogen (Carlsbad, USA)
Chitin Resin 50% Slurry	NEB (Ipswith, USA)

Concentrators	
Econo-Pac® Chromatography Columns	Bio-Rad (Hercules, USA)
InstantBlue Stain	Expedeon (Cambridgeshire, UK)
Microtiter plates, clear, flat bottom	Sarstedt (Nümbrecht, Germany)
Mini-PROTEAN® TGX™ Precast Gels Any-kD, 15-well	Bio-Rad (Hercules, USA)
"NucleoBond" Xtra Midi Plasmid DNA Kit	Macherey-Nagel (Düren, Germany)
"NucleoSpin" Plasmid Purification Kit	Macherey-Nagel (Düren, Germany)
Pierce™ BCA Protein Assay Kit	ThermoFisher Scientific (Waltham, USA)
Pierce™ ECL Western Blotting Substrate	ThermoFisher Scientific (Waltham, USA)
PolyPrep® Chromatography Columns	Bio-Rad (Hercules, USA)
Ponceau S solution	AppliChem (Darmstadt, Germany)
Precision Plus Protein™ Dual Color Standard	Bio-Rad Laboratories (Hercules, USA)
PVDF-Star Transfer Membrane 0.45 µm	AppliChem (Darmstadt, Germany)
"QIAquick" Gel Extraction Kit	QIAGEN (Hilden, Germany)
QuickStart Bradford 1x Dye Reagent	Bio-Rad (Hercules, USA)
RotiStore Cryo Vials	Roth (Karlsruhe, Germany)
StrepTactin® Sepharose® 50% suspension	iba (Göttingen, Germany)
Thin Blot Paper	Bio-Rad (Hercules, USA)
ZipTip® Pipette Tips	Merck (Darmstadt, Germany)

2.1.3 **Media & Buffers**

Medium name	**Ingredients**
Luria Broth (LB-Medium)	10 g/l Tryptone
	10 g/L NaCl
	5 g/L yeast extract

All media components were dissolved in 0.9 L/L dH_2O, autoclaved and filled up to 1 L with sterile water.

For LB-agar plates, media components and 15 g/L agar-agar were dissolved in the required amount of water, autoclaved, cooled to < 50°C before the respective antibiotics were added

(generally 100 µg/mL ampicillin or kanamycin) and aliquots of ca. 20 mL were poured into sterile plastic petri dishes. The plates were allowed to cool before storing at 4°C.

Buffer name	**Ingredients**
Buffer W (StrepTactin Wash Buffer)	100 mM Tris/HCl pH 7.5 150 mM NaCl 1 mM EDTA
Buffer E (StrepTactin Elution Buffer)	Buffer W (pH 7.5) + 2.5 mM desthiobiotin
Buffer R (StrepTactin Regeneration Buffer)	Buffer W (pH 7.5) + 1 mM HABA
EPL Buffer (for Expressed Protein Ligation)	20 mM Tris/ HCl, pH 7.5 500 mM NaCl 50 – 200 mM thiol reagent added freshly before use
His-Tag Equilibration Buffer	50 mM Tris/HCl pH 8.0 500 mM NaCl 30 mM Imidazole (can be 10 – 40 mM) optional 6-8 M urea
His-Tag Elution Buffer	50 mM Tris/HCl pH 8.0 500 mM NaCl 500 mM Imidazole optional 6-8 M urea
PBS Buffer	10 mM phosphate, pH 7.4 (Na_2HPO_4 & KH_2PO_4) 137 mM NaCl 0.27 mM KCl

TAE Buffer 1x	40 mM Tris Acetate, pH 8.3 1 mM ETDA
TBS Buffer	50 mM Tris/HCl pH 7.6 150 mM NaCl
Splicing buffer (modified from [145])	50 mM Tris/HCl pH 7.0 300 mM NaCl optional 6 M urea
SDS-PAGE Running Buffer 1x	25 mM Tris, pH 8.8 192 mM Glycine 10% w/v SDS
SDS-PAGE Sample Buffer 4x	10% w/v SDS 50 mM TCEP 20% Glycerol 0.2 M Tris/HCl pH 6.8 0.05% bromophenol blue
Native PAGE Sample Buffer 2x	62.5 mM Tris/HCl pH 6.8 40% glycerol 0.1% bromophenol blue
SEC Running Buffer	50 mM Tris/HCl pH 7.5 700 mM NaCl
Western Blot Transfer Buffer 1x	25 mM Tris, pH 8.8 192 mM Glycine 15 – 20% Methanol freshly added

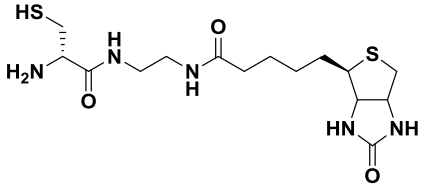
Diazotation Buffer	6 M GdmCl
	200 mM Na_2HPO_4, pH 3.0
	20 mM $NaNO_2$
NCL Buffer (for Native Chemical Ligation)	6 M GdmCl
	100 mM Na_2HPO_4, pH 7.0
	100 eq. thiol (MMBA, MPAA, MMP)
	5 eq. TCEP per Cys residue
Lysis Buffer B1[146]	500 mM NaCl, pH 8.5
	2 mM EDTA
	20 µM PMSF
	0.1 mM TCEP
Washing Buffer IB[146]	1 mg/mL Na-deoxycholate, pH 8.5
	2 mM EDTA
	0.2 mg/mL lysozyme
Denaturation Buffer (modified after[146])	8 M urea, pH 8.0
	2 mM EDTA
	1 mM TCEP

2.2 Synthetic Molecules for Ligation

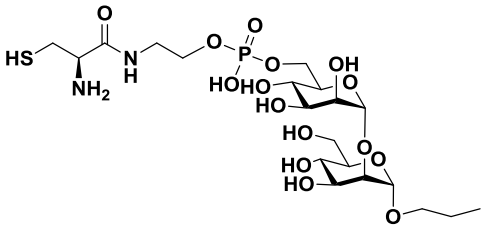
Four synthetic molecules were used for optimization of the ligation reactions and preparation of C-terminally modified proteins: Cys-biotin (Scheme 2.1), Cys-dimannose (Scheme 2.2) and the mono- and bilipidated GPIs (Scheme 2.3 and Scheme 2.4).

These molecules were synthesized by members of the GPI- and Glycoproteins research group of the Biomolecular Systems Department of the MPIKG, Potsdam. The molecules were modified with a cysteine residue for ligation to the proteins.

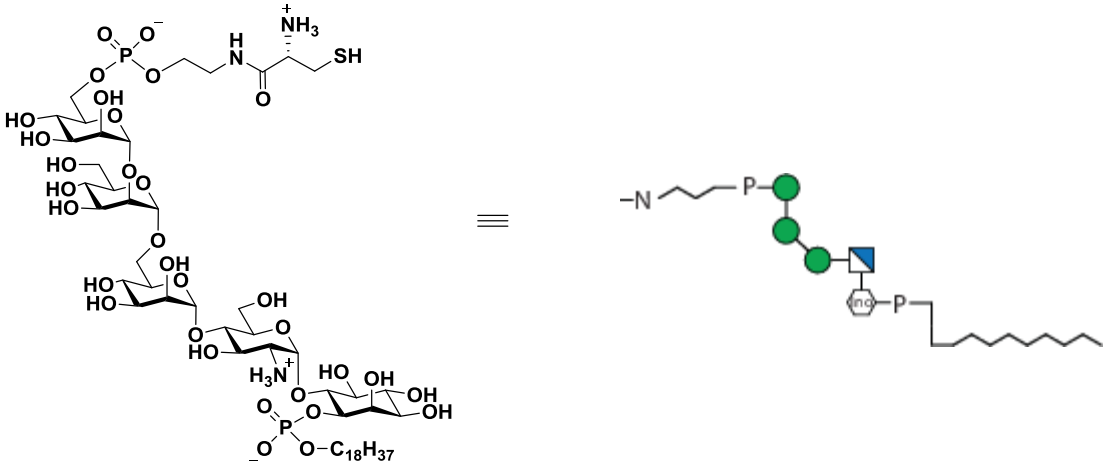
Scheme 2.1. Structure of Cys-biotin. A simple model compound for ligation optimization and easy detection.



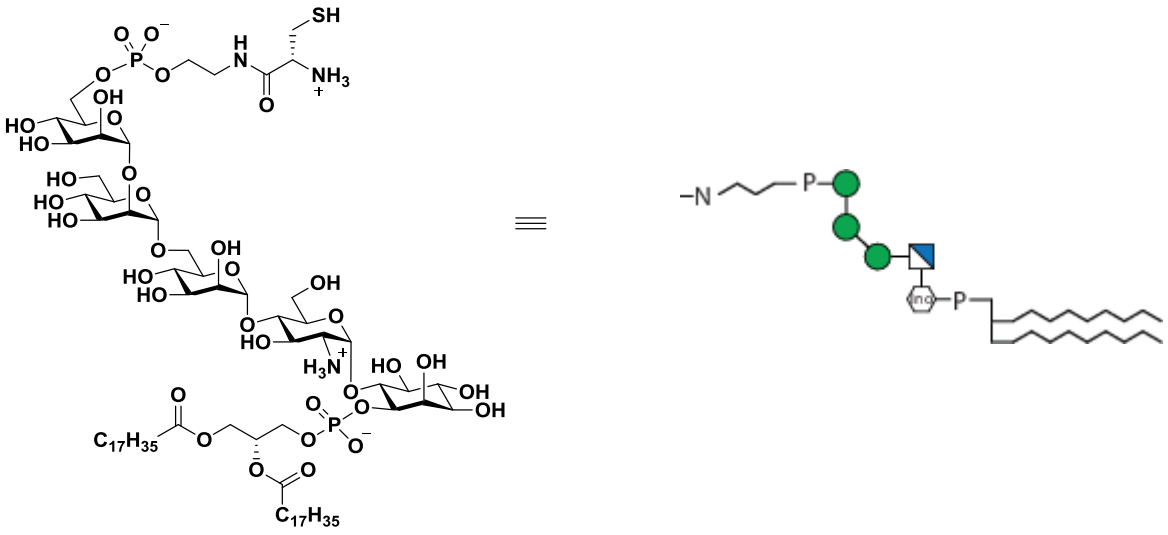
Scheme 2.2. Structure of Cys-dimannose. This molecule contains two mannoses as well as the phosphate link between protein and anchor present in native GPI-anchored proteins. It is therefore a better mimic of GPIs than biotin.



Scheme 2.3: Structure of monolipidated, biphosphorylated GPI. This is the model GPI anchor used in most experiments for method development and optimization. It contains the glycan core structure of all GPI molecules and a simplified lipid moiety having an alkyl chain attached to a phosphate. It contains the natural connection between GPI and protein as well as between GPI and cell membrane anchoring.



Scheme 2.4. Structure of bilipidated, biphosphorylated GPI. This structure is the most simple bilipidated GPI molecule, containing the complete pseudopentasaccharide, phosphate and a bilipid.



2.3 Plasmids, Clones and Primers

Different *E. coli* strains were used for the production of proteins. TOP10 cells were used for cloning since this strain is recombination deficient and has high transformation efficiency. BL21 (DE3) Star cells were mostly used for protein expression as this strain is suitable for use of the T7 promoter and offers enhanced mRNA stability. Some proteins were expressed using Rosetta cells, or Origami cells which provide a better environment for the folding of difficult proteins.

Plasmids were ordered from GenScript (Piscataway, USA) with optimized codons for expression in *E. coli*, if not stated otherwise. Primers were designed using DNAStar Lasergene software (Madison, USA) and ordered from GenScript as well. pTXB1 vector was purchased from NEB (Ipswich, USA). Primers are listed in Table 2.1.

Table 2.1. List of primers used in this work.

No.	Plasmid name	Fusion protein	Origin	Clones
1	pVS071	ST-eGFP-*Npu*^{N}	provided by Prof. Henning Mootz, University of Münster	TOP10 BL21 (DE3) Star
2	pVS41	ST-*Npu*^{C}-eGFP-His$_6$	provided by Prof. Henning Mootz, University of Münster	TOP10 BL21 (DE3) Star
3	pUC57-IL2Npu^N	ST-IL2-*Npu*^{N}	GenScript	TOP10
4	pTXB1-IL2	ST-IL2-*Mxe*GyrA-CBD	created via subcloning	TOP10 BL21 (DE3) Star Origami Origami pLemo
5	pET30a-IL2-Npu^N	ST-IL2-*Npu*^{N}	created via subcloning	TOP10 BL21 (DE3) Star Rosetta
6	pET29a-IFNa2a-Npu^N	IFNα2a-*Npu*^{N}-His$_6$	GenScript	TOP10 BL21 (DE3) Star Rosetta

7	pET29a-Thy1-Npu^N	Thy1-*Npu*^{N}-His$_6$	GenScript	TOP10 BL21 (DE3) Star Rosetta
8	pET29a-eGFP-Mxe	eGFP-*Mxe*GyrA-CBD	GenScript	TOP10 BL21 (DE3) Star Rosetta
9	pET29a-Thy1-Mxe	His$_6$-Thy1-*Mxe*GyrA-CBD	GenScript	TOP10 BL21 (DE3) Star Rosetta

2.4 Antibodies and Enzymes

Antibody	**Supplier**
Strep-MAB-Classic, HRP-Conjugate	iba (Göttingen, Germany)
Mouse-anti-biotin, HRP-Conjugate	Cell Signaling Technology (Danvers, USA)
Mouse-anti-His antibody	Invitrogen (Carlsbad, USA)
Goat anti-Mouse IgG, HRP-Conjugate	Jackson ImmunoResearch Laboratories (West Grove, USA)
Mouse anti-GPI	made in-house by Dr. Reka Kurucz and Dr. Maria Antonietta Carillo

Enzyme	**Supplier**
Antarctic Phosphatase	New England Biolabs (Ipswich, USA)
ConA Lectin	Sigma Aldrich (St. Louis, USA)
Phusion High-Fidelity DNA Polymerase	Thermo-Fisher Scientific(Waltham, USA)
T4 DNA Ligase	New England Biolabs (Ipswich, USA)
Taq Polymerase	New England Biolabs (Ipswich, USA)
Restriction Enzymes: EcoRV, HindIII, NdeI, PvuII, SapI	New England Biolabs (Ipswich, USA)

Protease	**Supplier**
Trypsin, Sequencing Grade	Roche (Rotkreuz, Switzerland)

2.5 Chromatography Columns

Column	Supplier
C4 Jupiter 50 x 2.1 mm, 300 Å	Phenomenex (Torrance, USA)
C4 Jupiter 150 x 10 mm, 300 Å	Phenomenex (Torrance, USA)
C8 ProPack 50 x 3 mm	YMC (Kyoto, Japan)
C18 Hydrosphere 50 x 3 mm	YMC (Kyoto, Japan)
C18 Hydrosphere 150 x 10 mm	YMC (Kyoto, Japan)
Synergi 4 µm Fusion-RP 250 x 4.6mm	Phenomenex (Torrance, USA)
Synergi 4 µm Fusion-RP 250 x 21.2 mm	Phenomenex (Torrance, USA)
HisTrap FF 5 mL	GE Healthcare (Little Chalfont, UK)
HisTrap HP 1 mL	GE Healthcare (Little Chalfont, UK)
Superdex 30 pg HiLoad 16/600	GE Healthcare (Little Chalfont, UK)
Superdex 75 pg HiLoad 16/600	GE Healthcare (Little Chalfont, UK)
Superdex peptide 10/300 GL	GE Healthcare (Little Chalfont, UK)
Yarra SEC-2000 300 x 4.6 mm	Phenomenex (Torrance, USA)

2.6 Software

Software Name	Publisher
Biopharmalynx	Waters (Milford, USA)
ChemDraw Professional 16	PerkinElmer (Waltham, USA)
ChemStation	Agilent (Santa Clara, USA)
DNAStar Lasergene 9	DNAStar (Madison, USA)
FlexControl & FlexAnalysis	Bruker Daltonics (Bremen, Germany)
ImageJ	NIH (Open Source, http://imagej.ni.gov/ij)
MassLynx	Waters (Milford, USA)
MS Office	Microsoft (Redmond, USA)
Origin Pro	Originlab (Northampton, USA)
PurityChrom	Knauer (Berlin, Germany)

2.7 **Equipment**

Instrument	**Manufacturer**
Autoclave HICLAVE HG-80	HMC (Haar, Germany)
Avanti J-E centrifuge with JA-10 and JA-25.50 rotors	Beckman-Coulter (Brea, USA)
Benchtop Orbital Shaker MaxQ™	Thermo Scientific (Waltham, USA)
Bio-Rad Electrophoresis System	Bio-Rad Laboratories (Hercules, USA)
Circular Dichroism Spectrometer Chirascan	Applied Photophysics (Leatherhead, UK)
Electrophoresis Power Supply Power Pack Basic	Bio-Rad Laboratories (Hercules, USA)
ESI-QTOF Xevo G2-XS with Acquity H-class UPLC	Waters (Milford, USA)
FPLC Knauer Azura	Knauer (Berlin, Germany)
HPLC Knauer Azura preparative system	Knauer (Berlin, Germany)
HPLC Waters 600 System with Waters 4187 Dual λ Absorbance Detector	Waters Corporation (Milford, USA)
Herasafe KS Biological Safety Cabinet	Thermo Scientific (Waltham, USA)
HPLC Agilent 1100	Agilent (Santa Clara, USA)
Imager G:Box Chemi-XXX6	Syngene (Cambridge, UK)
Incubator	Memmert (Schwabach, Germany)
Lyophilizer Alpha 2-4 LDplus	Christ (Osterode am Harz, Germany)
MALDI-TOF Autoflex Speed MS	Bruker Daltonics (Bremen, Germany)
Microplate Reader SpectraMax M5	Molecular Devices (San José, USA)
NanoPhotometer NP80	Implen (Munich, Germany)
pH meter pHenomenal 1100L	VWR (Radnor, USA)
Tabletop centrifuge MiniSpin	Eppendorf (Hamburg, Germany)
Tabletop centrifuge Multifuge X3R	Heraeus (Hanau, Germany)
Tabletop centrifuge Micro Star 17R	VWR (Radnor, USA)
Waterbath VWB 6	VWR (Radnor, USA)
ThermoMixer C	Eppendorf (Hamburg, Germany)
Vortex Genie 2	Scientific Industries (Bohemia, USA)

2.8 Methods

2.8.1 Molecular Biology Methods

DNA restriction digests

Restriction enzymes were purchased from New England Biolabs (Ipswich, USA). Digestions were set up according to the manufacturer's instructions using the provided buffers. For test digestions, usually 1 µg DNA was digested using 1 µL restriction enzyme in 1x restriction buffer (total volume 20 µL) at 37°C for 1 h. Digestions were analyzed in agarose gel electrophoresis.

For preparative digestions, usually 5 – 20 µg DNA were digested with 1.5 µL restriction enzyme in 1x restriction buffer (total volume 50 µL) at 37°C for 1 h. Bands were excised under UV Light using a scalpel.

Enzymes were inactivated by heating at 65°C for 15 minutes.

Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments corresponding to their size for analytical as well as preparative purposes. 1% agarose (w/v) was dissolved in TAE buffer, melted in the microwave and poured into a Biozym gel tray. DNA-Dye NonTox (AppliChem) was used as loading buffer and staining solution. Gel electrophoresis was performed at 110 V. GeneRuler 1 kb Plus DNA Ladder (Fermentas) was used as a size marker.

For analytical agarose gel electrophoresis, 200 ng – 1 µg DNA was loaded into the gel wells. The bands were visualized and documented using an UV gel documentation device (INTAS, Göttingen, Germany).

In order to prevent DNA from damage by UV-light, for preparative agarose gel electrophoresis 5 µL of the preparative reaction were mixed with 1 µL DNA-Dye NonTox and loaded into a well as reference lane. In the next well, the remaining 45 µL + 8 µL dye were loaded. Following electrophoresis, the preparative lane was excised from the gel completely. Then, the DNA in the reference band was visualized as mentioned before and the position of the desired fragment's positions was marked. Afterwards, the excised lane was placed again next to the gel and the fragment was cut out at the respective height. It was checked if all DNA had been cut out by shortly switching on the UV-light.

DNA Preparation

Plasmid DNA Isolation from *E. coli* cells

Plasmid DNA was isolated from *E. coli* cells using "NucleoSpin" Plasmid Purification Kit (Macherey-Nagel, Düren, Germany) according to the manufacturers' instructions.

Generation of DNA stocks

Bigger amounts of plasmid DNA were isolated using the "NucleoBond" Xtra Midi Plasmid DNA purification Kit from Macherey-Nagel following the manufacturer's instructions.

DNA Ligation

Vectors were opened using appropriate restriction enzymes. The restriction sites were treated with alkaline phosphatase (New England Biolabs, Ipswich, USA) in order to prevent self-ligation. DNA fragments were ligated into the treated vectors using T4 DNA Ligase (New England Biolabs, Ipswich, USA) in the respective buffer on ice overnight (0°C to room temperature). 50 – 100 ng vector DNA was used and the required amount of insert DNA was calculated using equation (**1**). Vector-to-insert ratios of 1:3 to 1:5 were used.

$$Amount(fragment)\ [ng] = \frac{a * amount\ (vector)[ng] * length\ (fragment)[bp]}{length\ (vector)[bp]} \qquad (1)$$

a – depending on the ratio of vector:insert. a = 3 for ratio 1:3

The ligation reactions were inactivated by heating at 65°C for 20 min. The ligated vectors were purified using agarose gel electrophoresis as described above.

DNA Isolation from Agarose Gel Pieces

Restricted DNA fragments and vectors were isolated from excised agarose gel pieces using the "QIAquick" Gel Extraction Kit from QIAGEN (Hilden, Germany) according to the manufacturer's instructions.

DNA Quantification

DNA concentrations were determined by measuring the absorbance at 260 nm using a NP80 NanoPhotometer (Implen, Munich, Germany) against TAE buffer as reference. DNA purity

was assessed using the ratio of OD_{260}/OD_{230} (expected to be 2.0-2.2) and OD_{260}/OD_{280} (at ~1.8 the sample is considered pure).

Blue-White Screening

In order to assess ligation success Blue-White-Screening was used to detect recombinant DNA in clones. The fragment of interest was ligated into a vector containing the gene for β-galactosidase in the multiple cloning site (pBluescript). Following transformation into competent cells (TOP10 or DH5α), the cells were streaked onto LB-agar plates with X-gal and IPTG and incubated at 37°C overnight. Blue colonies indicate an intact β-galactosidase gene (i.e. no successful ligation): the functional enzyme converts X-Gal into 5,5'-dibromo-4,4'-dichloro-indigo, a bright blue insoluble pigment. White colonies indicate successful ligation: here the β-galactosidase gene has been disrupted by the fragment of interest which was inserted in the multiple cloning site, no functional enzyme can be formed and X-Gal is not converted into the blue pigment.

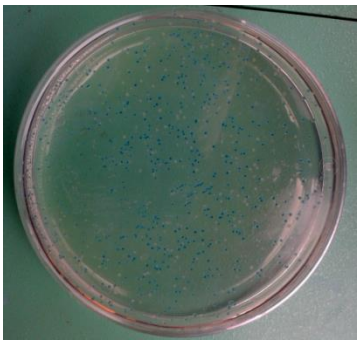


Figure 2.1: Blue-White Screening is used to check whether a DNA fragment of interest was successfully ligated into the vector. This picture shows an LB-agar plate presenting blue as well as white colonies.

PCR – Polymerase Chain Reaction

Preparative and Colony PCR were carried out using a Thermoblock C (Eppendorf, Hamburg, Germany).

Preparative PCR

PCR was applied to generate sufficient amounts of DNA sequences out of existing plasmids for cloning. Therefore, Phusion High-Fidelity DNA Polymerase (Thermo-Fisher Scientific, Waltham, USA) (a highly accurate enzyme) was. Forward and reverse primers were designed according to the desired starting and ending points, as well as required restriction sites. Each PCR reaction (50 µL) contained 1x HF Phusion Buffer, 200 µM dNTPs, 0.5 µM fw primer,

0.5 µM rev primer, 1 – 10 ng template DNA and 1 U Phusion DNA Polymerase. The PCR program used was:

- Start: 5 min, 95°C
- 30 cycles of 30 s, 95°C, 30 s, 45 – 50°C (depending on the primers used), 60 s / 1 kb template, 68°C
- Finish: 5 min, 68°C

The resulting DNA fragment was analyzed in agarose gel electrophoresis.

Colony PCR

For identification of clones containing a desired insert, colony PCR was used. Since here the resulting amplified DNA was not going to be used in further experiments, *Taq* Polymerase (New England Biolabs, Ipswich, USA) was used as it is very fast, but error-prone. For colony PCR, single colonies were picked and resuspended in 100 µL sterile dH_2O. 1 µL of this cell suspension was used as template.

Each PCR reaction (50 µL) contained 1x *Taq* Polymerase Buffer, 200 µM dNTPs, 0.2 µM T7 fw primer, 0.2 µM T7 Term primer and 1.25 U *Taq* Polymerase.

The PCR program used was:

- Start: 3 min, 95°C
- 30 cycles of 30 s, 95°C, 30 s, 50°C, 60 s / 1 kb template, 68°C
- Finish: 5 min, 68°C

The fragment size was checked via agarose gel electrophoresis. The remaining cell suspension of a correct clone was streaked on a LB-agar plate with the respective antibiotic and incubated at 37°C overnight.

DNA Sequencing

DNA samples were sent to MWG-Biotech (Eberberg, Germany) for sequencing in order to check the correctness.

Competent Cells

For the preparation of chemically competent cells, a 2 mL LB preculture of the desired cells was grown (i.e. TOP10, BL21 (DE3) Star, or others) at 37°C overnight, shaking at 220 rpm. The next day, 200 mL of LB medium was inoculated with the preculture and grown to an OD_{600} of 0.4 – 0.6 at 37°C, 150 rpm. As soon as the culture reached the required OD_{600} it was cooled down on ice for 15 minutes in order to stop the growth and cool the bacteria down.

From now on, cells were kept on ice wherever possible. The cells were harvested by centrifugation in a Beckman Avanti Centrifuge at 3000 x *g* for 10 minutes. The cell pellet was resuspended in 10 mL of 100 mM $CaCl_2$ solution and incubated on ice for 1 h. Following centrifugation and incubation, these steps were repeated once. Afterwards, the pellet was resuspended in 2 mL of 100 mM $CaCl_2$ containing 15% (v/v) glycerol and incubated on ice for 30 minutes. Aliquots of 200 µL were transferred quickly into precooled reaction tubes, and frozen immediately in a dry ice / ethanol bath. Competent cells were stored at – 80°C.

To test transformation efficiency, a test transformation was performed using 50 pg of pUC18 vector.

Heat Shock Transformation

An aliquot of 50 – 100 µL of chemically competent cells was thawed slowly on ice. 1 – 2 µL of DNA were added to the cell suspension, and incubated on ice for 30 minutes. Cells were heat shocked in a 42°C water bath for 30 seconds, and rested on ice for 1 – 2 minutes. 1 mL of pre-warmed LB medium was added to the cells and they were incubated at 37°C for 1 h (30 minutes in case of Amp-resistance gene, 1 h for Kan-resistance), shaking vigorously. Cells were centrifuged in an Eppendorf Minispin centrifuge at 3000 rpm, 2 minutes. After removal of the supernatant, the cells were resuspended and plated onto an LB-agar plate containing the required antibiotic. The plates were incubated at 37°C overnight and checked for colonies the next day.

Following the cloning steps, cryo stocks of all clones were prepared by transferring 1 mL of a saturated overnight culture of the respective clone into RotiStore Cryo Vials (Roth, Karlsruhe, Germany), shaking it 5 times and removing the liquid. The cryo stocks were then stored at -80°C.

2.8.2 Bacteria Cultivation and Protein Expression

Cultivation of Bacteria for Protein Expression

Generally, cryo stocks were streaked on a LB-agar plate containing the appropriate antibiotic for the respective clone using a spreader rod and incubated at 37°C overnight. The next day, a liquid preculture (10 – 50 mL LB medium with the appropriate antibiotic) was inoculated with one fresh colony using an inoculation loop and incubated at 37°C overnight, shaking at 220 rpm. If not stated otherwise, the main cultures were inoculated to a starting

OD_{600} of 0.05 in baffled 2 L shake flasks (Roth, Karlsruhe, Germany) filled with max. 0.5 L LB medium with the respective antibiotic. The cultures were grown to the late exponential growth phase (i.e. OD_{600} of ca. 0.6 – 0.8) and the recombinant protein expression was induced with 0.1 – 1 mM IPTG. The induction concentrations as well as expression temperatures were optimized for each clone individually between 12°C and 37°C. Following expression, the cells were harvested by centrifugation at 10.000 x *g*. The cleared medium was discarded.

Measurements of Optical Density

In order to assess bacterial growth, optical density at 600 nm (OD_{600}) was measured against fresh LB medium as reference using half-micro plastic cuvettes with NP80 Nanophotometer (Implen, Munich, Germany). If values above OD_{600} of ~1 were observed or expected, the cell broth was diluted with LB medium, to keep values below 1.

2.8.3 Protein Purification

Cell lysis

After harvesting, the bacterial cells were resuspended in a lysis buffer suitable for the first following purification step. If not stated otherwise, protease inhibitors have been included in the buffers. A first lysis was achieved by incubation with 0.25 – 0.5 mg/mL lysozyme for 30 min – 1 h at 30°C, or at 15 – 18°C overnight. Subsequently, the suspension was sonicated on ice using Digital Sonifier (Branson), with amplitude 20 – 40%, using pulses and pauses of 10 – 30 sec. Special care was taken to avoid heating of the samples. Finally, the suspension was centrifuged at 24,400 x *g* for 1 h at 4°C.

Depending on protein solubility, the resulting supernatant was subjected to further purification methods, or the cell pellet was solubilized using 6 – 8 M urea.

Affinity Chromatography

In this work several affinity tags have been applied for the purification of proteins using affinity chromatography. The methods are listed below, sorted according to the tag used.

Strep-tag

The strep-tag II is an eight amino acid (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) tag that can be used at the N- or C-terminus of proteins for detection or purification using StrepTactin®

resin.[147] For purification, the clarified and filtered (0.2 µm using a syringe filter) cell lysate in buffer W was loaded to StrepTactin® resin according to the manufacturer's instructions, washed with 10 – 15 CV of buffer W, before eluting the protein of interest using 6 x 0.5 CV of buffer E. The fractions including flow-through and wash fractions were analyzed on SDS-PAGE. Fractions containing the protein of interest were pooled and dialyzed in order to remove desthiobiotin, or subjected to further purification steps as required (i.e. size exclusion chromatography).

His-tag

For His-tag affinity purification the protein of interest is fused to six neighboring His residues at its N- or C-terminus.[148,149] The protein of interest can be bound to a Ni^{2+} NTA (nitrilotriacetic acid) resin by metal chelation (see Figure 2.2) whereas host proteins flow through. The tagged protein can be eluted using a gradient of imidazole.[150,151]

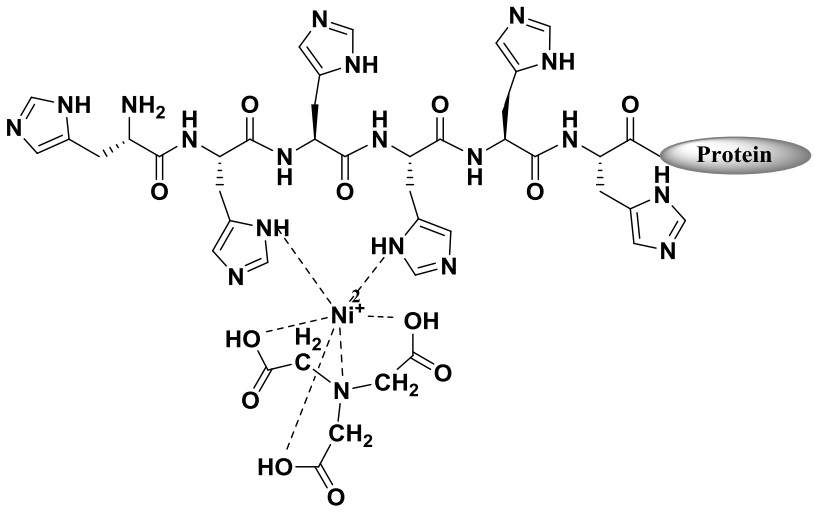


Figure 2.2: Schematic depiction of the interaction of the His_6-tag with Ni^{2+}-NTA resin. Four valencies of the nickel ion are interacting with the immobilized NTA resin while the remaining two valencies are interacting with two His residues of the tag (created in ChemDraw 16, after [150]).

For purification of fusion proteins from cleared cell lysate, HisTrap™ FF crude 5 mL columns (GE Healthcare, Little Chalfont, UK) preloaded with Ni^{2+} ions were used in a Knauer Azura FPLC system. Once column was equilibrated with HisTrap Equilibration buffer (containing 10 – 30 mM imidazole), up to 8.5 mL clarified lysate were loaded using a 10 mL sample loop and a manual injection valve. The flow-through (20 mL) was collected in a 50 mL falcon tube before the column was washed with 10 CV (collected in a 50 mL falcon tube). Subsequently, the bound fusion protein was eluted by a gradient from 10 – 30 mM imidazole

to 500 mM imidazole over 5 CV. The fractions including flow-through and wash fractions were analyzed on SDS-PAGE. Fractions containing the protein of interest were pooled and dialyzed in order to remove imidazole, or subjected to further purification steps as required (i.e. size exclusion chromatography).

Chitin Binding Domain with simultaneous protein thioester formation

The chitin binding domain (CDB) is commonly used in combination with a modified intein, rendering it a self-cleaving affinity tag.[123,152] It is included in the commercially available IMPACT (Intein-Mediated Purification with an Affinity Chitin-binding Tag) system (NEB, Ipswich, USA) for purification of the intein fusion constructs via a chitin resin (also NEB). Different vector systems exist, having the intein-CBD domain either on the N- or the C-terminus. The pTXB1 vector used in this work gives a fusion protein with the protein of interest fused at its C-terminus to the *Mxe*GyrA intein from *Mycobacterium xenopi* and the CBD. If DTT or β-mercaptoethanol is used for intein cleavage, the whole tag will just be removed and the thiol hydrolyzed. The purification via CDB can however also be used with simultaneous intein cleavage and protein thioester formation by using more stable thiol reagents such as MESNA.

In a typical purification, a gravity flow column is prepared with the chitin resin slurry (50% in ethanol); generally a column with CV = 1.5 mL was used for 100 mL culture volume. After washing the resin with 5 CV dH_2O and equilibration with buffer W (pH 7.5, 2 x 5 CV), the clarified lysate was loaded to the column (10 mL). The flow-through was collected. The column was washed with 3 x 10 CV buffer W (collected). Subsequently, the intein cleavage was induced by flushing the column with 3 CV cleavage buffer (buffer W + 50 mM thiol, pH 7.0), closing the column and incubating at room temperature for 16 – 40 h (depending on the thiol used). The pure protein thioester was eluted with 6 x 0.5 CV buffer W, pH 7.0. The fractions including the flow-through and wash fractions were analyzed in SDS-PAGE. Fractions containing the protein thioester (usually elution fractions No. 2 – 5) were pooled and subjected to size-exclusion chromatography on a Superdex peptide 10/300 GL column (GE Healthcare, Little Chalfont, UK) for removal of unreacted thiol.

Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) was generally applied as a polishing step in protein purification, i.e. following an initial capturing step. In SEC, proteins are separated by size rather than by any interaction with the stationary phase. Smaller proteins and other molecules

can enter into the pores of the stationary phase material which usually consists of small porous polymer beads such as agarose or sepharose with different pore sizes. Therefore, smaller molecules are retained longer during the run and elute later than larger proteins.

Depending on the purpose and the molecular size, different SEC columns were used. The separations were carried out using a Knauer Azura FPLC System:

- Superdex 75 pg HiLoad 16/600 for proteins (CV = 120 mL)
- Superdex 30 pg HiLoad 16/600 for peptides (CV = 120 mL)
- Superdex peptide 10/300 GL for peptides or for desalting of proteins (CV = 24 mL)

Generally, all samples were centrifuged at 13.3 k rpm and / or filtered through a 0.2 µm syringe filter prior to loading on SEC columns. Maximum 1.5 mL of sample containing 0.5 – 10 mg protein was loaded onto the Superdex 75 column.

RP-HPLC

Reverse-Phase High Performance Liquid Chromatography (RP-HPLC) was also used for protein purification in some cases. HPLC is based on the principles of distribution (reversed phase) and adsorption (normal phase) of the analyte on the stationary phase. In the case of RP-HPLC the stationary phase generally consists of a porous, often silica-based material which is surface-modified with non-polar alkyl chains (most commonly C18, C8 and C4 chains). The mobile phase is in this case most commonly a combination of water and acetonitrile or methanol. Generally, the analyte is first bound to the stationary phase at a low concentration of acetonitrile (ACN) or methanol before it is eluted by a linear gradient of the respective solvent. When RP-HPLC is used with proteins, the retention of the proteins depends mainly on their hydrophobicity. Low percentages of organic acids are often used in order to decrease the pH, rendering the proteins positively charged and enhancing the separation and peak shape.[153]

For protein purification, a Phenomenex Jupiter C4 column (10 x 150 mm, 5 µm particle size, pore size 300 Å) was used which is specifically designed for the separation of hydrophobic proteins.[154] Separations were performed on a Waters 600 system (Controller and Pump) with a Waters 2487 Dual λ Absorbance Detector and a flow rate of 4 mL/min. Gradients were optimized starting from a standard gradient of 10 to 70% acetonitrile (ACN) in water with 0.1% TFA in 30 minutes for a good separation of the protein of interest from any contaminants. Up to 10 mg of protein was injected using a manual injection valve with a 2 mL sample loop. Fractions were collected manually according to the absorption observed

and analyzed in MALDI-TOF-MS in order to identify fractions containing the protein of interest in high purity. The respective fractions were then combined, frozen and lyophilized.

Protein Concentration Determination

Protein concentrations were generally determined using Pierce BCA Protein Assay Kit (Thermo Scientific, Waltham, USA) according to the manufacturer's instructions (microplate procedure). If use of BCA assay was not possible due to thiol compounds present in the solution, QuickStart Bradford 1x Dye Reagent (Bio-Rad (Hercules, USA) was used by mixing 100 µL dye with 100 µL protein sample in a 96-well plate and measuring the absorption at 595 nm using Microplate Reader SpectraMax M5 (Molecular Devices, San José, USA).

2.8.4 Peptide Synthesis and Native Chemical Ligation

Synthesis of peptide hydrazides using automated SPPS

Peptide synthesis was performed in collaboration with GPI-group members Dr. Dana Michel and Antonella Rella by Fmoc-SPPS (Solid Phase Peptide Synthesis) using a CEM Microwave Peptide Synthesizer (Liberty Blue).

NovaPEG Wang resin (Novabiochem) with a substitution of 0.63 mmol/g was used for peptide synthesis in 0.1 mmol scale (250 mg resin). After introducing a hydrazide linker to the resin, the first amino acid was coupled manually using a fritted single use syringe. The loading was checked using Fmoc quantification as following: approx. 1 mg of dry resin was added to 1.5 mL of 20% piperidine in DMF and shaken for 15 min. The slurry was centrifuged and the supernatant was diluted 1:1 in 20% piperidine in DMF. UV absorption was measured at 290 nm in a quartz cuvette against 20% piperidine in DMF. The substitution grade of the resin was determined using equation **2** (considering an ε=5253 $M^{-1}cm^{-1}$ for the Fmoc-piperidine adduct):

$$Substitution\left(\frac{mmol}{g}\right) = \frac{[Abs.290\ nm]}{mg\ resin * 1.75} \quad (2)$$

According to the substitution grade the amount needed was calculated for the other reagents. The resin was then transferred to the synthesizer and the synthesis was run using 1 M Oxyma with 0.1 M DIPEA in DMA as activator base and 0.5 M DIC in DMF as activator. Fmoc

deprotection was performed with 20% piperidine in DMF after each coupling step. Coupling conditions are listed in Table 2.2.

Table 2.2: Coupling and deprotection cycles for different amino acids.

Amino Acids	Coupling	Deprotection
2nd amino acid	single coupling (90°C for 2 min)	double deprotection
Argo	double coupling (25°C for 25 min, 75°C for 2 min)	double deprotection
His & Cys	double coupling (25°C for 2 min, 50°C for 10 min)	double deprotection
Ile & Glu	double coupling (25°C for 25 min, 75°C for 2 min)	single deprotection
all other amino acids	two minutes single coupling 90°C	single deprotection

The peptide was cleaved from the resin by treatment with a TFA/TIPS/water mixture (90:5:5) 90 min at RT under shaking. The peptide was precipitated and washed with ice-cold ether three times. Peptide samples were analyzed in RP-HPLC (C18 Hydrosphere column 3 x 50 mm, YMC) using a 30-minute gradient of 5 – 70% acetonitrile in water with 0.1% TFA. The sample was analyzed by MALDI-TOF-MS using DHB as matrix and a reflector method in positive ion mode. The peptides were purified using a Knauer Azura HPLC system with a Phenomenex Synergi RP-fusion preparative column (21.1 x 250 mm). Fractions were collected, lyophilized and analyzed in analytical RP-HPLC and MALDI-TOF-MS following lyophilization. Fractions containing the pure Npu^{C}-hydrazide, or Npu^{C*}-hydrazide, respectively, were combined, lyophilized and weighed.

Generation of Peptide Thioesters

Peptide thioesters were generated using a protocol adapted from Zheng at al. (2013).[136] The purified NpuC-hydrazide was dissolved in diazotation buffer at a concentration of 1.5 mM at pH 3 and stirred at -10°C for 20 minutes to form the peptide azide. To induce thioester formation, the solution was warmed to room temperature, 100 eq. thiol (i.e. MMBA, MPAA, MMP) were added and the pH was carefully adjusted to 7.0 using 6 M NaOH. After stirring at RT for 20 minutes, the thioester was immediately purified and analyzed using RP-HPLC as described above.

Native Chemical Ligation of Peptide Thioesters to GPI Molecules and Model Compounds

The purified peptide thioester was dissolved to a concentration of 1.5 mM in NCL buffer containing 100 eq. thiol and 5 eq. TCEP per Cys residue in the sequence. Cys-biotin or Cys-GPI (10 eq. or 0.9 eq., respectively) was added and the mixture was incubated at 37°C, shaking vigorously. The ligation reaction was monitored by RP-HPLC (C18-Hydrosphere, YMC, 3 x 50 mm for biotin ligations or Jupiter C4, Phenomenex, 2.1 x 50 mm for GPI ligations) and LC-MS (C18 or Protein C4 BEH, Waters, 2.1 x 50 mm). Ligations with Cys-biotin were typically completed after 1 – 2 days while ligations to Cys-GPI needed longer incubation (one week). The ligation products were purified using RP-HPLC on a Phenomenex Synergi RP-Fusion column (4.6 x 250 mm) for Npu^C-biotin, and a Phenomenex Jupiter C4 column (10 x 150 mm) for Npu^C-GPI.

2.8.5 Protein Ligation Methods

Expressed Protein Ligation

Fusion proteins of a protein of interest (POI) and the Gyrase A mini intein from *Mycobacterium xenopi* (*Mxe*GyrA)[123] were expressed in *E. coli*. After obtaining the respective protein α-thioesters a solution of 1 mg/mL thioester in Ligation buffer (20 mM Tris/HCl pH 7.0, 150 mM NaCl), TCEP (up to 0.5 mM) and the thiol reagent (MMBA, MPAA, MESNA or MMP) at a final concentration of 50 – 200 mM was prepared. The molecule to be ligated was added to the thioester mixture. Cys-biotin (5 eq.), Cys-dimannose (2 eq.) or Cys-GPI (0.9 – 1.1 eq.), respectively. The pH was readjusted to 7.0 with 1 M NaOH, if necessary. The ligation mixture was incubated shaking at room temperature. Reaction progress was monitored using RP-HPLC, SDS-PAGE and western blotting.

Protein *Trans*-Splicing

For protein *trans*-splicing (PTS), the purified proteins were dialyzed in splicing buffer adapted from Zettler et al.[155] (50 mM Tris/HCl pH 7.0, 300 mM NaCl, optional 6 M urea,). Proteins were combined at a final concentration of 15 – 20 µM with 0.7 – 1.5 eq. of Npu^C-biotin or Npu^C-mGPI. TCEP was added to a final concentration up to 0.5 mM and it was optimized for each POI separately. The reaction mixture was incubated shaking at a

temperature optimized for the respective protein. The reaction progress was monitored using RP-HPLC, SDS-PAGE and western blotting.

One-Pot Ligation

A modified ligation strategy was developed in this work: One-Pot Ligation (OPL) circumvents the necessity to purify a generated peptide or protein thioester ligating it to the target molecule (biotin, dimannose or GPI, respectively) using NCL/EPL *in situ* after formation. Instead, a modified C-terminal intein peptide (Npu^C(AA)) was designed and synthesized in automated SPPS simply as peptide hydrazide (see chapter 2.8.4).

For OPL reactions, the same basic setup as for PTS reactions could be adopted: To a solution of 15 – 20 µM fusion protein in splicing buffer 1.2 eq. of the modified Npu^C(AA) peptide was added together with up to 0.5 mM TCEP (optimized for each protein) and the anchoring molecule (Cys-biotin (5 eq.), Cys-dimannose (2 eq.) or Cys-GPI (0.9 eq.), respectively). Additionally, the thiol reagent (MMP, MMBA or MPAA) was added to a final concentration between 50 and 200 mM. The reaction mixture was then incubated shaking at an optimized temperature between 16°C and 37°C. The reaction progress was monitored in RP-HPLC, SDS-PAGE and western blotting.

2.8.6 Purification of Reaction Products

Purification of the reaction products was carried out using the same chromatography methods that were used for fusion protein purification, such as his-tag or strep-tag affinity chromatography and RP-HPLC.

Furthermore, a phase-separation method using Triton X-114 was tested for the separation of reaction products containing lipidated GPIs from unreacted reaction partners. [156] The method was performed as specified in the protocol.[157] One part of sample solution was mixed with three parts of ice-cold PBS buffer and 1.5 parts pretreated Triton X-114, vortexed vigorously and then the phases were separated using a Pasteur pipette.

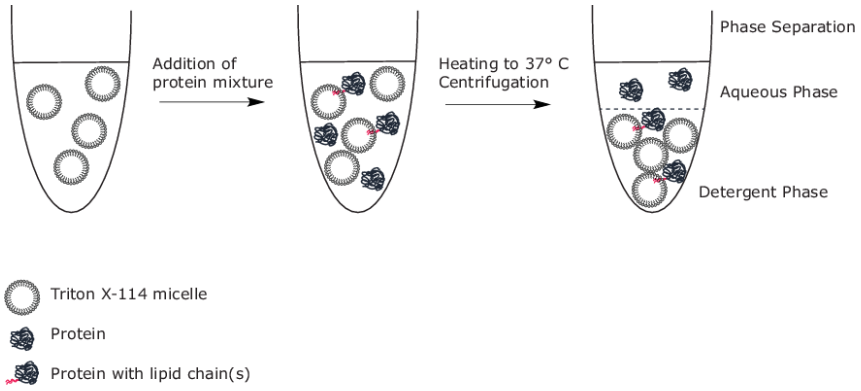


Figure 2.3. Triton X-114 separation of lipidated protein. Proteins carrying a lipidation (or a glypiation) are expected to incorporate into Triton X-114 micelles and by this can be separated from proteins without lipidation by phase separation.

2.8.7 **Product Characterization**

In order to follow the purification of fusion proteins, synthesis of peptides and progress of ligation reactions, following analytical methods were used.

SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a commonly used technique for the separation of proteins or other biomolecules by their molecular weight. Gels are generated from polyacrylamide and placed in a buffer filled tank between a cathode and an anode. SDS is a strong detergent masking the charge of the proteins so by applying a voltage they migrate in the polyacrylamide gel according to their size.[158] To achieve this, the protein samples are treated with a solution of 1% SDS and a reductant (often DTT or β-mercaptoethanol) and heated for 5 minutes at 100°C in order to completely denature them. In 1970, an improved discontinuous system was developed by U. Laemmli in which the proteins are first focused in a stacking gel with low polyacrylamide before migrating at different speeds according to their molecular weight in a separation gel with a polyacrylamide concentration of 8 – 20%, depending on the protein sizes to be separated.[159]

In this work, gels with 15% polyacrylamide were mainly used, if not stated otherwise, which are best suitable for small proteins in the range of 15 – 50 kDa. Gels were prepared in a gel casting chamber (Bio-Rad, Hercules, USA) according to the following protocol and placed in a Mini-PROTEAN Tetra cell (Bio-Rad) with SDS-PAGE running buffer:

- Stacking gels for SDS-PAGE (4%): 125 mM Tris/HCl pH 6.8, 4% acrylamide / bisacrylamide (29:1 mixture), 1% APS, 1% TEMED
- Separation gels for SDS-PAGE (15%): 250 mM Tris/HCl pH 8.8, 15 % acrylamide / bisacrylamide (29:1 mixture), 1% APS, 1% TEMED

In some cases precast gels (Mini-PROTEAN® TGX™ Precast Gels, Any kD, Bio-Rad) were also used for a slightly different separation pattern. The samples were mixed 4:1 with 4x reducing SDS sample buffer, boiled at 99°C for 5 minutes and loaded into the gel. A voltage of 200 V was then applied and the electrophoresis was run until the dye front reached the bottom of the gel (ca. 30 – 45 minutes).

Following electrophoresis, the gels were stained with Coomassie Brilliant Blue using InstantBlue Stain (Expedeon, Cambridgeshire, UK). ImageJ Software was used for relative quantification of calculation of splicing yields in coomassie-stained gels.

Proteins available in very small amounts were detected using silver staining according to the following protocol:

- Fixation of the proteins in the gel by incubation at room temperature for two hours in a solution of 40% ethanol and 10% acetic acid in dH_2O
- Rinsing of the gel to remove excess acetic acid by incubation of the gel in 50% ethanol in dH_2O three times for ten minutes
- Sensitization by incubation in a solution of 0.02% (w/v) sodium thiosulfate ($Na_aS_2O_3$) for 1 – 2 minutes
- Washing in water three times for 20 seconds
- Staining by incubation in a cold solution of 0.2% (w/v) silver nitrate ($AgNO_3$) and 0.075% (v/v 37 % CH_2O) formalin in dH_2O in the dark
- Following another washing step (2 x 20 sec), the gel was developed by incubation in a solution of 6% (w/v) sodium carbonate (Na_2CO_3), 0.05% (v/v 37 % CH_2O) formalin and 2% (v/v) sensitizing solution in dH_2O for up to five minutes, until bands were visible
- The gel was washed in water again (2 x 20 sec) and the staining was stopped by the addition of 5% acetic acid in dH_2O
- The gel can be stored in 1% acetic acid in dH_2O

Sensitizer, staining solution and developer need to be prepared freshly before use.

Native PAGE

Native PAGE is a modification of SDS-PAGE in which proteins can be analyzed in polyacrylamide gels in non-denaturing conditions. In this case the proteins will still have their natural net charges as those are not masked by SDS and their natural conformation. This leads to a separation of the proteins by charge, shape and size rather than only by molecular weight. In the pH of the running buffer (ca. 8.8), most proteins are negatively charged and will migrate towards the anode. This can give insights about the folding of a protein.

The gels were prepared in the same way as described for SDS-PAGE, placed in the gel chamber with Native PAGE running buffer, which does not contain any SDS. The samples were prepared by mixing 1:2 with Native PAGE Sample buffer 2x (no reducing agents) and loaded directly into the wells of the gel without heat treatment. A voltage of 200 V was applied and the proteins were allowed to migrate for 45 minutes.

The gels were stained using coomassie staining as described for SDS-PAGE.

Western Blot

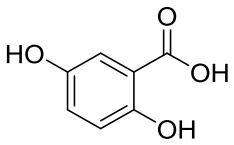
In western blotting, proteins separated using SDS-PAGE are transferred to a nitrocellulose or PVDF (polyvinylidene difluoride) membrane by applying an electric field orthogonally to the gel. The proteins are negatively charged since they are covered in SDS, and migrate towards the anode and the membrane. Following the electrophoretic transfer, the membrane is blocked using BSA to prevent unspecific binding of the antibodies used for detection of the proteins of interest.

The gel with separated proteins by SDS-PAGE was placed in the blotting chamber filled with Western Blot Transfer Buffer together with the prewetted PVDF membrane in between four sheets of blotting paper. The proteins were blotted onto the membrane by applying a current of 350 mA for 1 h. Heating of the buffer was prevented by placing an ice pack in the blotting tank. The proteins on the membrane were stained by placing the membrane in a Ponceau S staining solution for five minutes; excess staining was removed by washing with dH_2O. The membrane was blocked by incubation with 5 % BSA in PBST or TBST (depending on the antibody to be used) for 1 – 2 h.

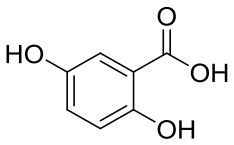
Following blocking, the membrane was incubated with the respective antibody or lectin using the manufacturer's conditions, or using own optimized conditions. Either the primary antibody was already equipped with an HRP-conjugate, or a secondary antibody-HRP-conjugate was used for chemiluminescence detection using ECL substrate together with a chemiluminescence imager (Imager G:Box Chemi-XXX6, Syngene, Cambridge, UK).

MALDI-TOF-MS

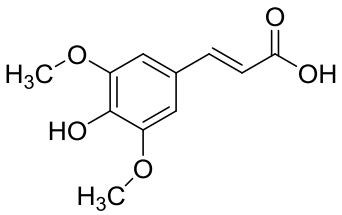
Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry[160] is an MS technology widely applied for the analysis of large biomolecules such as proteins or lipids. It is a soft ionization method that enables big bio-molecules to be transferred into the gas phase. A matrix substance with strong UV absorbance is co-crystallized with the analyte in a ration of ca. 5000:1 (mol/mol). Commonly used matrix compounds are shown in Figure 2.4. The theory of ionization in MALDI is not completely understood but the commonly accepted model is the following as described by Karas & Krüger (2003)[161]: the co-crystals are desorbed from the sample carrier by high-energy laser pulses of few nanoseconds; the high excess of matrix molecules is taking the analyte molecules with them and by a dense plume consisting of neutral matrix molecules, reactive species such as matrix radicals, electrons and hydrogen atoms, and a small but sufficient amount of sample ions. This plume expands into the vacuum of the mass spectrometer source and accelerated according to its mass-to-charge ratio (m/z). Therefore, the MALDI technique is most commonly used in combination with a TOF mass detector due to the pulsed generation of ions. In MALDI, generally the singly charged ion $[M+H]^+$ is the most frequently observed species,[161] although doubly or triply charged ions ($[M+2H]^{2+}$, $[M+3H]^{3+}$) are also sometimes generated.



2,5-dihydroxybenzoic acid (DHB)



2,5-dihydroxyacetophenone (DHAP)



Sinapinic acid (SA)

Figure 2.4: Commonly used matrix compounds in MALDI of biomolecules: 2,5-dihydroxybenzoic acid (DB), dihydroxyacetonphosphate (DHAP) and sinapinic acid (SA).

For analysis in MALDI-TOF-MS, samples were desalted using C18 or C4 ZipTip® pipette tips. Prior to use, the tips were conditioned with 0.1% TFA in acetonitrile and equilibrated with 0.1% TFA in dH_2O. The sample was washed with 0.1% TFA in dH_2O and eluted in 30 – 70% acetonitrile, 0.1% TFA in dH_2O, depending on the sample properties. The salt-free samples were mixed in a 1:1 ratio with a DHB matrix solution, spotted onto a MALDI target plate (AnchorChip™ 400/384, Bruker Daltonics, Bremen, Germany) and dried under ambient conditions at room temperature. If a good ionization could not achieved using these

conditions, a different matrix type (DHAP, SA) as well as different matrix-sample ratios were tested.

Preparations of matrix solutions and sample spotting as described in the Bruker MALDI sample preparation protocols:

- DHB: 20 mg/mL 2,5-dihydroxybenzoic acid in TA30 (30% acetonitrile, 0.1% TFA in dH_2O). Use of dried droplet method, as described above
- DHAP: 7.6 mg 2,5-dihydroxyacetophenone is dissolved in 375 µL ethanol. Then 125 µL of a solution of 18 mg/mL DAC (diammonium hydrogen citrate) is added. The sample solution, matrix solution and a solution of 2% TFA in dH_2O were premixed in a ratio of 1:1:1 and pipetted up and down until crystallization started, then 1 µL of the mixture was spotted onto the target plate and dried
- SA: Matrix solution I: sinapinic acid saturated in ethanol, matrix solution II: sinapinic acid saturated in TA30. For SA, the double layer method was applied: firstly, a drop of matrix solution I is spotted onto the target plate, then 1 µL of a 1:1 mixture of sample solution and matrix solution II are spotted on top and dried

Protein and peptide standards (Peptide Standard IV, Protein Standard I and II, Bruker Daltonics, Bremen, Germany) were also prepared and used for calibration of methods at regular intervals. Following drying, the samples were subjected to MALDI analysis in a Bruker Autoflex™ Speed instrument using Flex Control Software (Bruker Daltonics, Bremen, Germany).

Peptides were generally analyzed in reflector positive ion mode, if not stated otherwise. Mass spectra were acquired in the mass range between 500 m/z to 6,000 m/z. Proteins were analyzed in linear positive ion mode and mass spectra were recorded in the mass range between 10,000 m/z and 50,000 m/z, or as required. Synthetic GPI molecules were analyzed in linear mode, positive as well as negative ion mode; spectra were recorded in the range between 500 m/z and 4,000 m/z. Spectra were analyzed using Flex analysis software (Bruker Daltonics, Bremen, Germany).

RP-HPLC

Analytical RP-HPLC was performed on an Agilent 1100 System with a G1315B DAD detector. Water with 0.1% TFA and acetonitrile were used as mobile phase. All samples were centrifuged and / or filtered prior to injection.

Peptide samples were analyzed using a YMC Hydrosphere C18 column (50 x 3 mm, 3 µm particle size, 12 nm pore size). Protein samples were analyzed using a Phenomenex Jupiter C4 column (50 x 2.1 mm, 5 µm particle size, 300 nm pore size). Generally, 5 – 10 µg peptide or protein dissolved in 1 – 10 µL 20% acetonitrile, 0.1% TFA in dH_2O were injected into the column. As a starting point, a gradient of 5 – 70% acetonitrile in 30 minutes was used for peptide samples. If required, other gradients were also used, for example 10 – 30% acetonitrile in 30 minutes. Chromatograms were analyzed using Agilent ChemStation software.

LC-ESI-QTOF-MS

LC-MS is a valuable tool in the analysis of biomolecules that combines a separation by RP-HPLC with mass spectrometry. In LC-MS, an HPLC instrument with a UV/VIS or DAD detector is coupled to a mass spectrometer and the compounds separated by HPLC are analyzed in mass spectrometry according to their retention time. ESI (electrospray ionization) is most commonly used for LC-MS due to the low energy required for ionization and analysis of large molecules such as proteins without a high degree of fragmentation.[162] However, it generally leads to the generation of multiply charged ions. In case of peptides $[M+2H]^{2+}$ and $[M+3H]^{3+}$ are commonly observed while in case of large proteins ions such as $[M+30H]^{30+}$ are commonly observed. Adduction of counter ions (e.g. sodium, potassium, ammonium ions in pos. mode, formate or acetate in neg. mode) is also observed if salts are present in the sample.[162] In LC-ESI-MS, the solvent coming from the LC part is sprayed through a capillary that is placed orthogonally or off-axis from the mass spectrometer entrance with the help of an inert gas (often nitrogen). At the capillary, a voltage of 3 – 5 kV is present, nebulizing the liquid to form small charged droplets which are then rapidly evaporated by applying heat and nitrogen flow.[162] This reduces the droplet size which following repeated disintegration (Coulomb fissions) yields very highly charged droplets capable of eventually generating gas phase ions. In case of proteins, it is reported to be likely that small droplets containing just one protein molecule are formed and by evaporation of the solvent, the charges of the droplet are transferred to the protein, thereby generating a multiply charged ion. This theory is called Charged Residue Model (CRM). Some of these ions enter the actual mass spectrometer via the interface (either a tiny hole or a capillary at the MS entrance).[163]

In this work, a Waters Xevo G2-XS with an Acquity H-class UPLC was used. In this instrument, the ESI source is coupled to a quadrupole-time-of flight (QTOF) mass analyzer. A quadrupole consists of four parallel metal rods. The two opposite rod pairs are connected

electrically. The quadrupole is used to generate an oscillating electric field which can be used to filter ions of specific m/z ratios. In the TOF mass analyzer, the ions are further separated.

Tryptic Digestions of Proteins

Additional to the before mentioned top-down approach an analyzing the intact proteins and protein conjugates, the classical bottom-up approach was also applied. Proteins in gel pieces were digested by proteolytic enzymes such as trypsin and the peptide fragments were subsequently analyzed in MS.

For tryptic digestions the protocol described in Kolarich et al.[164] was followed in a slightly modified way. The digestions were carried out for only 20 min to 1.5 h at 37°C instead of 16 h as over-digestion was observed using the long incubation time, yielding too short fragments. Furthermore, instead of analyzing the fragments in ESI-MS, they were analyzed in MALDI-TOF-MS.

The expected peptide fragments were calculated by the web tool ExPASy PeptideMass[165] for the respective protease and modifications. The peptide fragments found in MALDI-TOF-MS analysis were then compared to the expected fragments and the coverage of the protein sequence was assessed.

Circular Dichroism (CD)

Circular dichroism utilized circularly polarized light to determine the optical properties of chiral compounds. This is also a commonly used technique to study the secondary structure of proteins.[166] To utilize this effect, plane polarized light made up of two components of equal magnitude, one rotating counter-clockwise (L), the other rotating clockwise (R) is passed through a sample containing the protein in solution (ideally in water). The sample will exhibit absorbance of the L and R light components, and if absorbance of both components is different, a CD signal will be observed, indicating chiral nature of the sample compound. In proteins, many chromophores are present, including the peptide bond (absorbance < 240 nm), aromatic amino acid residues (260-320 nm) and disulphide bonds (around 260 nm). By measuring signals below 240 nm, information about protein secondary structure can be obtained, as α-helices, β-sheets and turn regions each exhibit certain absorbance patterns that are specific for the respective structures. This information can be used to determine the percentage of each structure, or simply to identify the predominant structural features of a protein.

For CD measurements, protein samples were prepared in PBS buffer at concentrations ranging from 0.1 – 1 mg/mL, depending on the amounts available. The protein solution was then filled into a High Precision Cell Quartz cuvette (light path 1 mm) and placed in the Chirascan Circular Dichroism Spectrometer. Circular dichroism was measured between 190 and 260 nm against PBS buffer as a reference and measurements were performed in triplicates. The averages of these measurements were plotted over the wavelength.

3 Results

3.1 Establishing the Analytical Conditions for Proteins

In order to be able to follow the progress of protein purification as well as for monitoring the different reactions and for quality control, it was essential to establish a suitable analytical environment with the respective methods. Standard techniques were used such as SDS-PAGE, Western Blotting, MALDI-TOF-MS, RP-HPLC and tryptic digestions. In the case of HPLC, suitable RP-columns had to be identified for the analysis of the protein-glycolipid conjugates to be analyzed. Furthermore, methods were developed for the analysis of intact proteins using a Waters Xevo G2-XS QTOF which was set up at the MPIKG during the period of this thesis.

3.1.1 HPLC for Glypiated Proteins

Initially, a YMC-Pack PROTEIN-RP column (50 x 3 mm, 200 Å) was used for analytical HPLC. However, the resolving power of this column was not sufficient, and the peak shape was not good enough to detect products eluting with close retention times, Figure 3.1. A new column (Phenomenex Jupiter C4, 50 x 2.1 mm, 300 Å) was used instead, giving peaks of better shape and resolution, Figure 3.2.

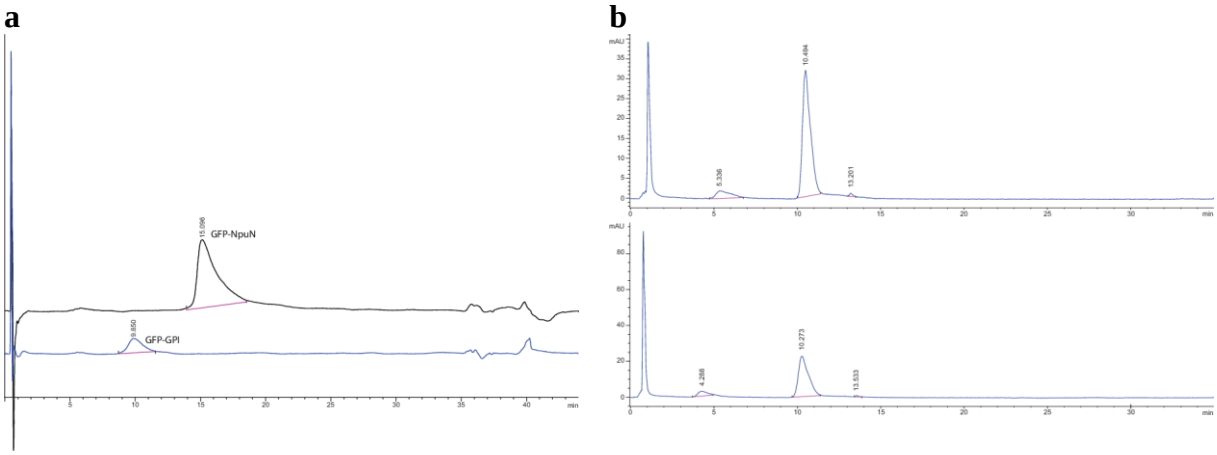


Figure 3.1. HPLC analysis of (a) eGFP-*Npu*^{N} and eGFP-mGPI and **(b)** eGFP-SR / EPL reaction with biotin using YMC-Pack PROTEIN-RP column.

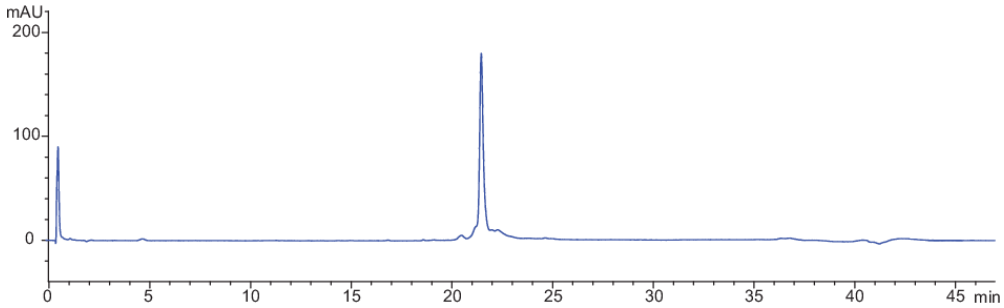


Figure 3.2. HPLC analysis of GFP-*Npu*^{N} using Phenomenex Jupiter C4 column.

3.1.2 Establishing LC-ESI-MS Analysis for Intact Glycolipoprotein Analysis

A mass spectrometer suitable for LC-ESI-MS analysis was set up in the department during the course of this dissertation project. As part of this project the analysis of intact proteins has been established, and especially the analysis of modified proteins was optimized. Initial measurements turned out to yield mass spectra of low resolution not suitable for the determination of the protein mass, see Figure 3.3 a. Optimization of the gradients and the ionization parameters lead to hugely improved mass spectra, see Figure 3.3 b. Fine-tuning of the machine parameters and improved sample preparation delivered more easily deconvoluted spectra such as the one shown in Figure 3.3 c.

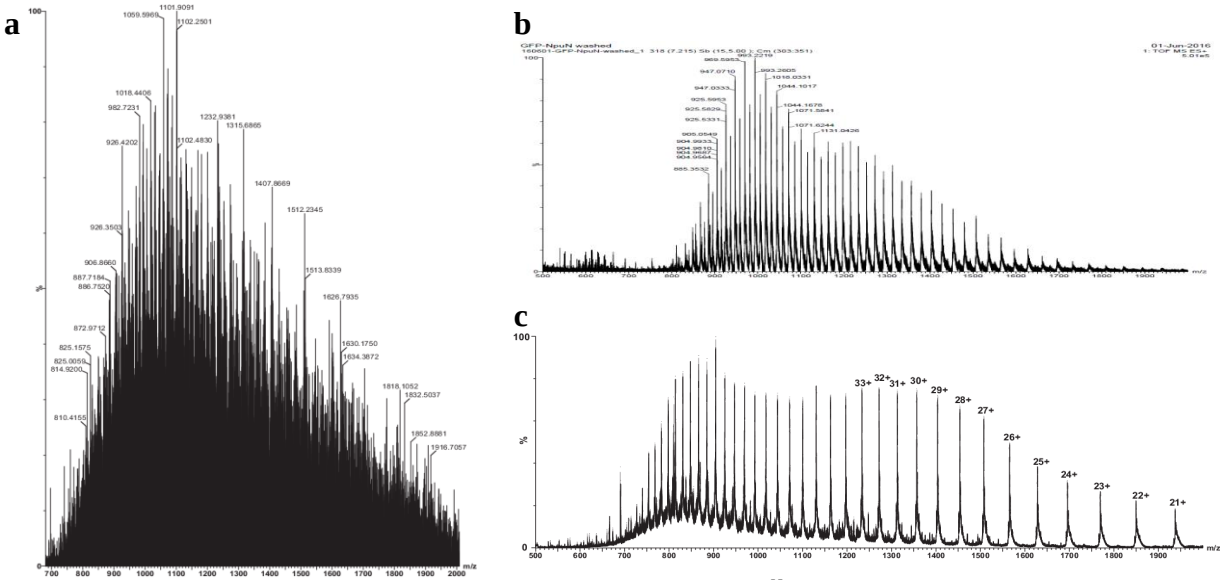


Figure 3.3. ESI-MS spectra of LC-ESI-MS analysis of eGFP-*Npu*^{N} fusion protein. a) Initial analysis using a Waters BEH C4 UPLC column (50 x 2.1 mm, 1,7 µm particle size), gradient 40-60% ACN in water with 0.1% FA in 5 min. **b), c)** Improved LC-ESI-MS mass spectra of eGFP-*Npu*^{N} fusion protein.

3.2 Generation of IL-2 Expression Vectors

3.2.1 Subcloning of IL-2-*Npu*^{N}

The Interleukin-2-*Npu*DnaE$_{N}$ fusion gene in a pUC57Kan vector ordered from GenScript (Piscataway, USA) was subcloned into the pET30a expression vector by digesting both vectors with NdeI and HindIII (simultaneous digestion 4 h at 37°C for the fragment; subsequent digestion, each enzyme 1.5 h at 37°C for linearizing pET30a vector). Linearized pET30a and the 828 bp fragment were purified using agarose gel electrophoresis (1% agarose) and ligated using T4 Ligase and 3:1 and 5:1 vector:insert ratios, incubating at 16°C overnight. Ligated pET30a-IL-2-*Npu*^{N} was transformed into TOP10 cells and streaked onto LB-agar plates containing kanamycin. Colonies were observed for both ratios after incubation at 37°C overnight. Colony PCR was performed with all obtained colonies using T7 fw primer and T7 Term rev primer, Figure 3.4. Ligation was only successful for the 3:1 ratio, but not for the 5:1 ratio. Since the bands were very heterogeneous, DNA of all clones was sent for sequencing, revealing that only clone 13 carried the correct sequence.

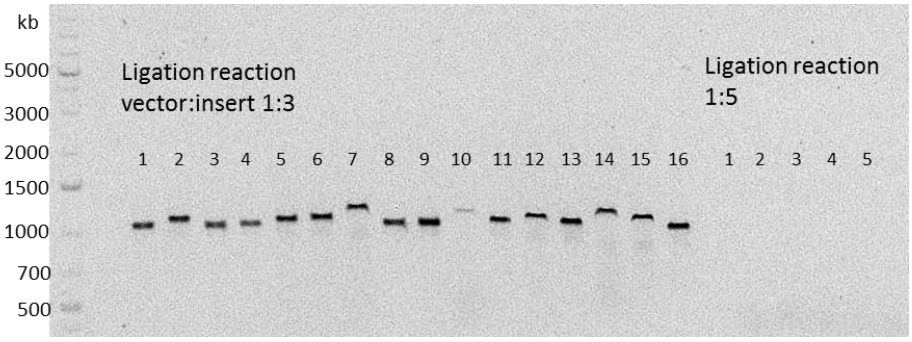


Figure 3.4. Colony PCR of ligations of IL-2-*Npu*^{N} in pET30a. Very heterogeneous bands were observed from the different clones obtained and all clones were sequenced.

3.2.2 Cloning of IL-2 Gene Into pTXB1 Vector for EPL Studies

Generation of an IL-2 Fragment for Ligation into pTXB1 Vector

In order to fuse the IL-2 gene to the *Mxe*GyrA mini intein, the IL-2 gene was excised from the pUC57Kan vector by PCR using the following primers: The T7 Universal Primer was used as forward primer (sequence: 5' TAA TAC GAC TCA CTA TAG GG (20 bp)). As reverse primer the following primer was designed complementary to the gene's end: 5' CCA

GTC AGT GTT GAG ATG ATG C (22 bp). For addition of a 3'-terminal SapI recognition and restriction site, a second PCR was carried out using the purified PCR product as template and again the T7 Universal primer as forward primer. The reverse primer for this PCR was designed to be complementary to the end of the IL-2 gene, with an additional sequence for the SapI site, followed by several bp as spacer, as recommended in Green and Sambrook[167] (protocol 6, chapter 8. Primer sequence: 5' GGT GGT TGC TCT TCC GCA AGT CAG *TGT TGA GAT GAT GC* (38 bp). The resulting PCR product was analyzed in agarose gel electrophoresis which showed the expected fragment size of 566 bp (Figure 3.5). As only one sharp band was observed, the fragment was used without further purification.

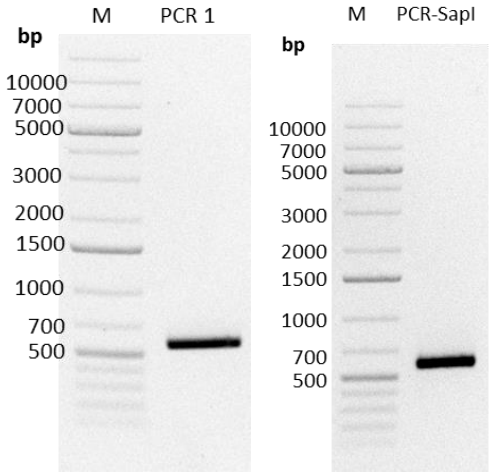


Figure 3.5. Analysis of excised IL-2-fragment (a) and of IL-2 fragment with added SapI restriction site (b) in agarose gel electrophoresis (1%). Sharp bands of the expected sizes were observed (550 bp and 566 bp, respectively).

Blue-White Screening

Direct ligation into the linearized and dephosphorylated pTXB1 vector did not deliver any viable colonies in transformation, thus the fragment was ligated into EcoRV-linearized pBluescript (pBS) vector using a 3:1 ratio (vector:insert) at 16°C overnight as an intermediate step. Undigested pBS vector (#1) as well as unpurified (#2) and purified (#3) samples of the ligations were transformed into TOP10 cells and streaked on LB-agar plates containing ampicillin, IPTG and X-Gal for blue-white-screening. Many blue colonies were observed after overnight incubation at 37°C for the undigested pBS vector, which was the expected result. Unpurified and purified pBS-IL-2 delivered one quarter (90/400) and one third (100/300) white colonies, respectively, indicating successful ligation.

Colony PCR was performed with five white colonies of each screening to verify insertion of the correct fragment using M13 uni primer as the fw primer and pUC19/M13 rev primer (Figure 3.6 a). The expected fragment size of 810 bp was identified in all colonies with unpurified vector and some colonies with purified vector. Clone #2.1 was selected for further experiments and plasmid DNA pBS-IL-2 was prepared from an overnight culture using Miniprep Kit. The correct insertion of IL-2 was confirmed by sequencing.

Ligation of IL-2 Fragment into pTXB1 Vector

pBS-IL-2 from clone #2.1 was digested with SapI and NdeI for 1 h at 37°C. pTXB1 vector was digested with SapI for 1 h at 37°. After heat inactivation of SapI the vector was further digested with NdeI at 37°C for 45 min plus 15 min with addition of fresh NdeI. Subsequently, the linearized vector was dephosphorylated with Antarctic Phosphatase for 15 min at 37°C to prevent self-ligation. Fragment and linearized vector were excised from a 2% agarose gel and purified. Ligation of the fragment into the linearized vector was set up using T4 Ligase and a 5:1 ratio. Following incubation at 16°C overnight, 5 µL of the ligation was transformed into TOP10 cells yielding two colonies after overnight incubation at 37°C. Colony PCR verified the correct fragment size of 719 bp for both colonies (Figure 3.6 b). DNA sequencing which revealed that both clones carried the correct sequence including start codon, IL-2 sequence and SapI restriction region. The plasmid maps are given in the Appendix (7.1.3).

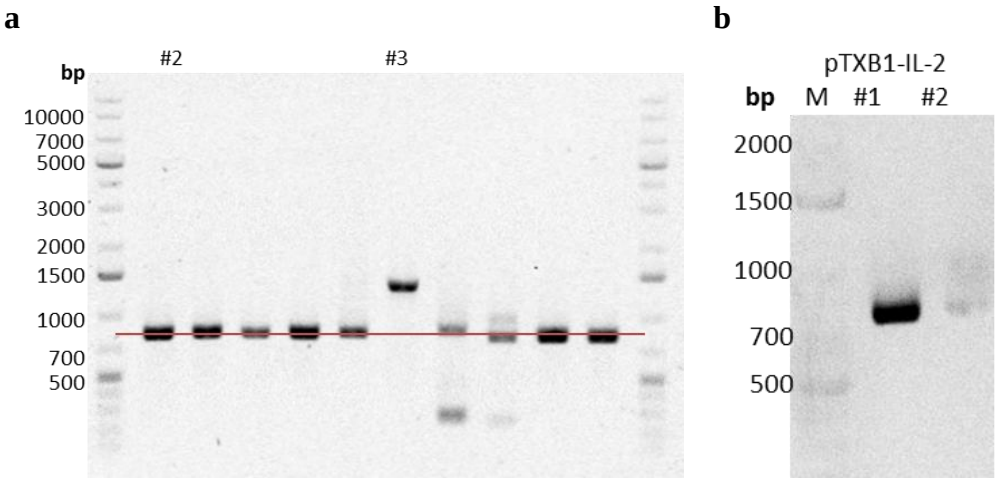


Figure 3.6. a) Colony PCR of IL-2 in pBluescript vector to ensure insertion of the correct fragment (1% agarose gel), #2 – unpurified vector, #3 – purified vector. Clone #2.1 was selected. **b)** Colony PCR of IL-2 in pTXB1 vector (1% agarose gel), #1 – clone 1, #2 – clone 2**.** Both clones have a more or less strong band at the expected size (719 bp) and both were confirmed to carry the correct sequence by DNA sequencing.

3.3 **Quality Control of Synthetic GPI molecules**

In order to ensure the integrity of the synthetic GPI anchor molecules after storage, they were analyzed by MALDI-TOF-MS and LC-ESI-MS. MALDI-TOF-MS was performed using DHB as matrix. Generally, GPI molecules were more efficiently ionized using linear negative mode, Figure 3.7 a and b.

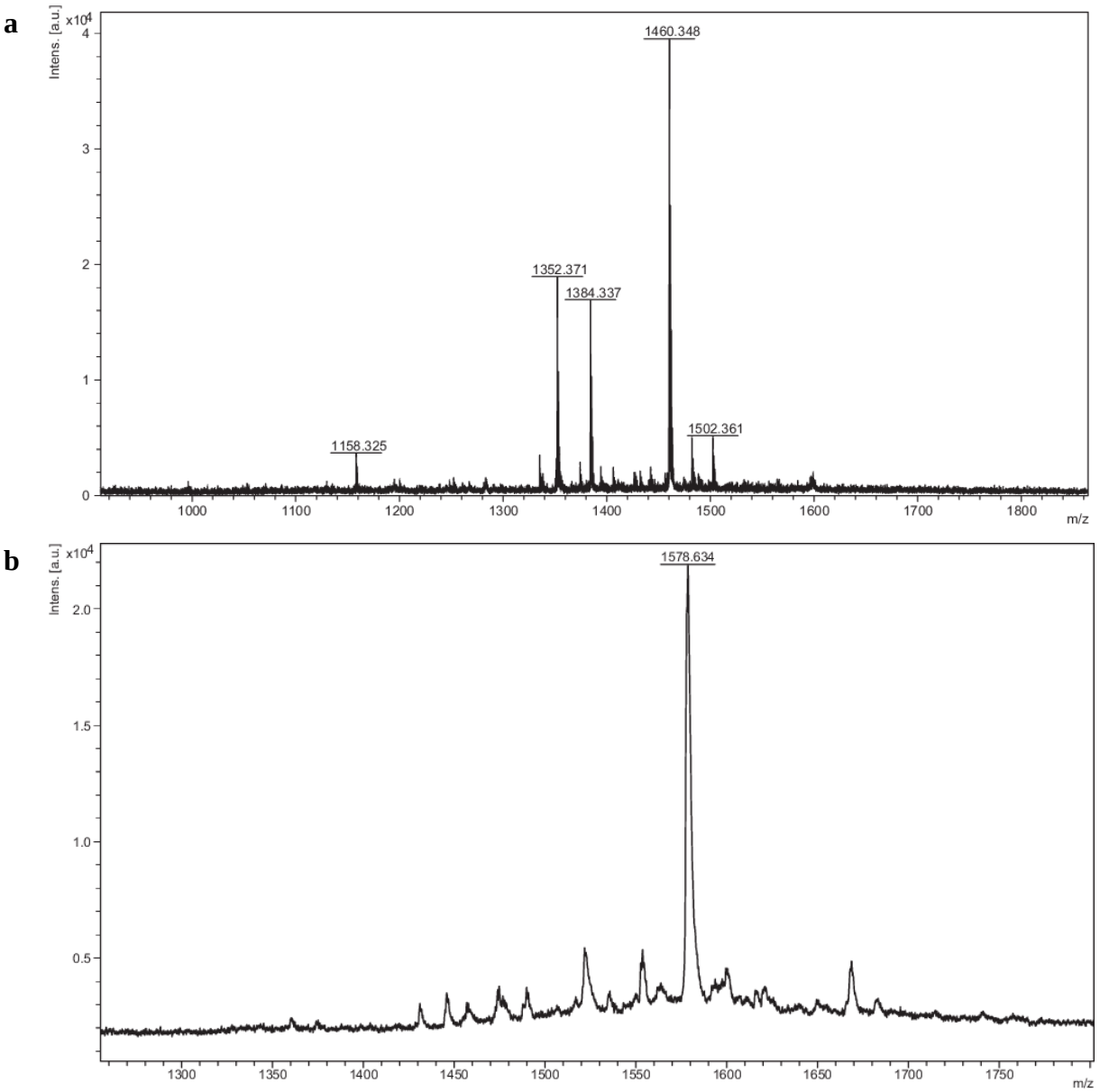


Figure 3.7. MALDI-TOF-MS analysis of GPI molecules. a) MALDI-TOF mass spectrum of mGPI**,** $[M-H]^-_{obs}$=1,460.348 Da, $[M-H]^-_{calc}$=1,458.5 Da, **b)** spectrum of mtGPI, **[**$M-H]^-_{obs}$=1,578.634 Da, **[**$M-H]^-_{calc}$=1,581.5 Da. Both spectra were acquired in linear negative mode.

For LC-ESI-MS analyses, GPI molecules were separated on a YMC C8 Pro-Pack column (50 x 3 mm, 3 µm) using a gradient of 5 – 70% ACN in water (0.1 % FA in 30 min). Similarly to MALDI-TOF-MS analysis, GPI molecules were easier to analyze in negative ion mode (ES^- mode), although ES^+ mode worked as well. LC-ESI-MS analysis of mGPI is shown in Figure 3.8 and for mtGPI it is given in Figure 3.9.

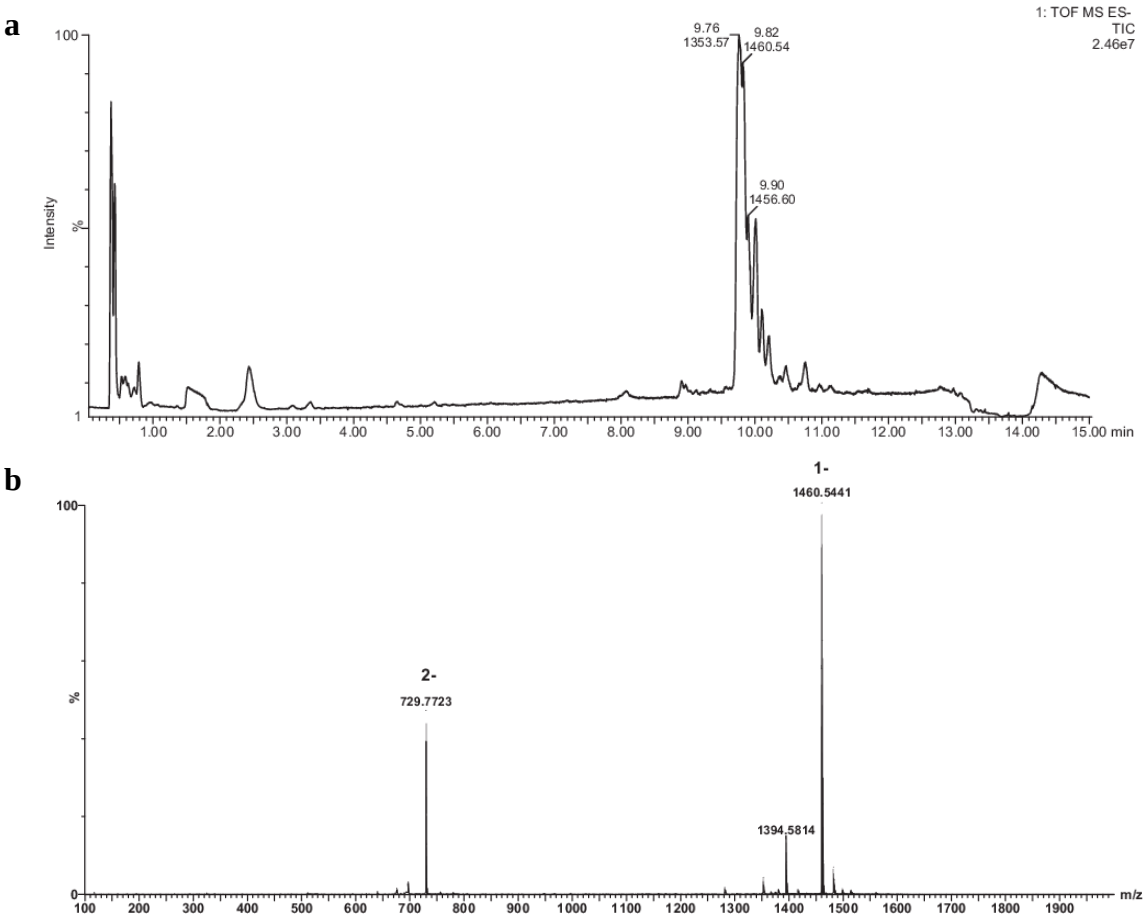


Figure 3.8: LC-ESI-MS analysis of mGPI. a) Total ion chromatogram (TIC) of mGPI analysis (C8), b) mass spectrum acquired at t_R=9.8 min, $[M+H]^+_{obs}$=1,460.5441 Da, $[M+H]^+_{calc}$=1,460.5 .

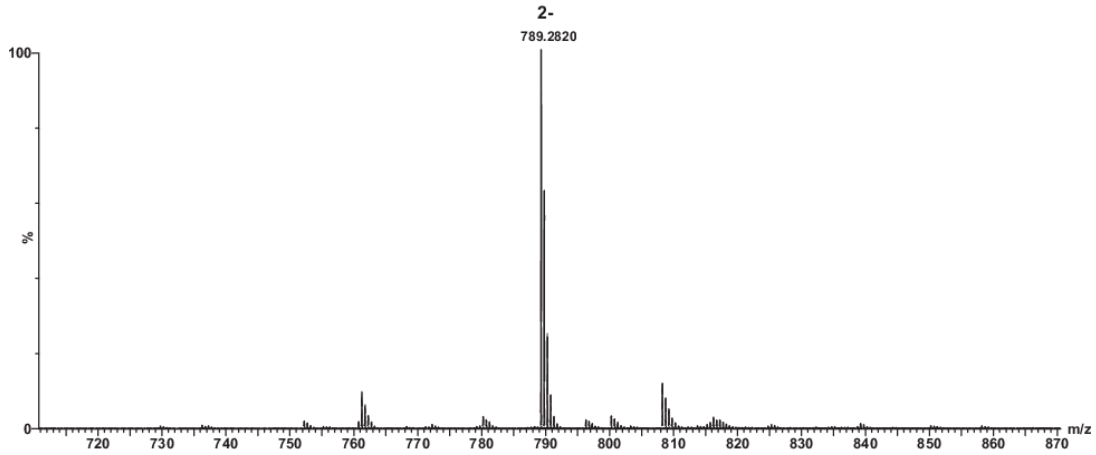


Figure 3.9: LC-ESI-MS spectrum of monolipidated, triphosphorylated GPI (C8, 5-70% ACN), $[M-H]^-_{calc}$=1,581.5 Da, $[M-2H]^{2-}_{obs}$=789.282→1,580.5 Da.

Bilipidated GPI was analyzed in ESI-MS via direct injection without using an HPLC column, with methanol + 0.1% FA as eluent. In ES^+ mode, a double adduct of bGPI was detected (Figure 3.10 a), whereas in ES^- mode, two species were detected that are very close in mass to bGPI ($[M-H]^-_{calc}$=1737.83 Da), Figure 3.10 b.

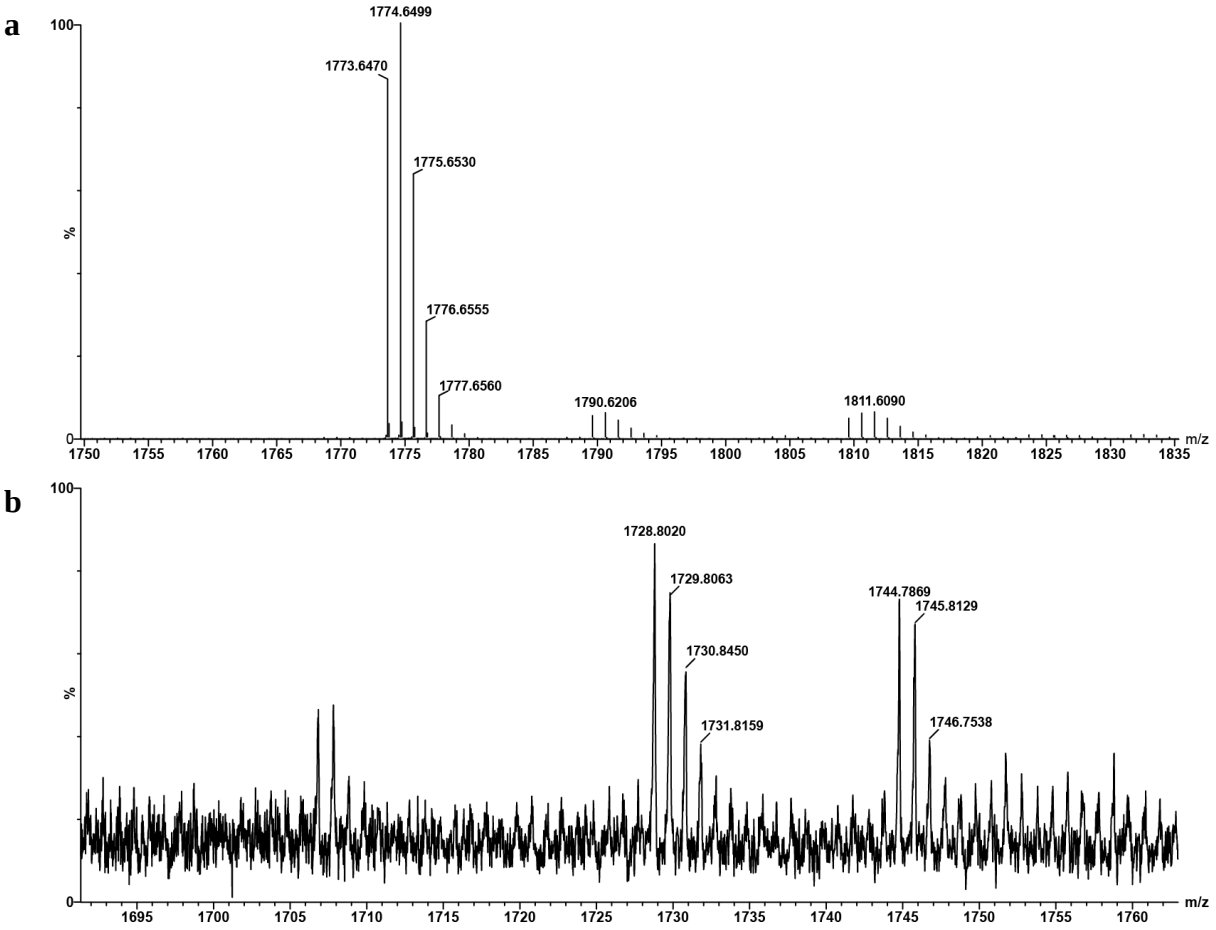


Figure 3.10. ESI-MS spectra of bilipidated, biphosphorylated GPI, $[M+H]^+_{calc}$=1,739.83 Da. a) spectrum acquired in ES^+ mode, $[M+H]^+_{obs}$=1773.6470 m/z, which corresponds to a double ammonium adduct, b) spectrum acquired in ES^- mode, $[M-H]^-_{obs}$=1,728.8020 Da and 1,744.7869 Da, both of which cannot clearly assigned to the product (Δ-9 and +7 Da).

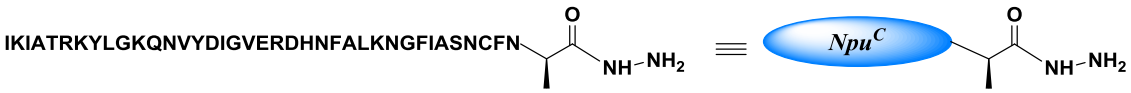
3.4 Synthesis of *Npu*^{C}(WT) and *Npu*^{C}(AA) Peptides and Native Chemical Ligation of *Npu*^{C}(WT)-Thioester

Peptide synthesis was performed by Fmoc-SPPS (Solid Phase Peptide Synthesis) using a CEM Microwave Peptide Synthesizer (Liberty Blue) in collaboration with GPI-group members Dr. Dana Michel and Antonella Rella.

3.4.1 Synthesis of *Npu*^{C}(WT) and Thioester Generation

The C-terminal fragment of the naturally split intein from *N. punctiforme* (*Npu*^{C}) was synthesized as a peptide hydrazide using the conditions described in chapter 2.8.4. The sequence of the peptide is given in Scheme 3.1:

Scheme 3.1 *Npu*^{C} hydrazide.



To obtain the peptide, 250 mg of NovaPEG Wang resin were swollen in DCM for 2 h. After introducing a hydrazide linker to the resin, the first amino acid (Fmoc-alanine) was coupled manually using a fritted syringe. The loading of the resin was determined by quantification of Fmoc removal and was found to be 0.31 mmol/g. The resin was transferred to a reaction vessel and the sequence was elongated on a microwave peptide synthesizer using the optimized conditions (2.8.4). Due to the complexity and the length of the peptide (39 amino acids) the synthesis was split into two parts. First, the first 30 amino acids were coupled using five equivalents of amino acids, DIC and Oxyma. A test cleavage was carried out and the product was checked for major deletion sequences by HPLC. Subsequently, the peptide was elongated with the remaining nine amino acids using six equivalents of amino acids, DIC and Oxyma. The coupling cycles used for every amino acid are listed in Table 2.2. The cycles were optimized according to observed side products occurring during earlier syntheses. Fmoc deprotection was done with 20% piperidine in DMF after each coupling step.

In order to check the completion of the synthesis, a test cleavage was performed using TFA/TIPS/water (90:5:5, v/v/v) solution. The peptide was cleaved from the resin using the same cleavage mixture and purified in several batches using HPLC (C18). A gradient of 10 –

30% acetonitrile in water with 0.1% TFA in 45 minutes was used for purification of *Npu*^{C} hydrazide. Fractions containing the pure desired peptide were combined and lyophilized again. Typically, around 10 mg purified peptide were achieved. The purified peptide was analyzed in RP-HPLC, MALDI-TOF and LC-ESI-QTOF. The respective chromatograms and spectra are given in Figure 3.11.

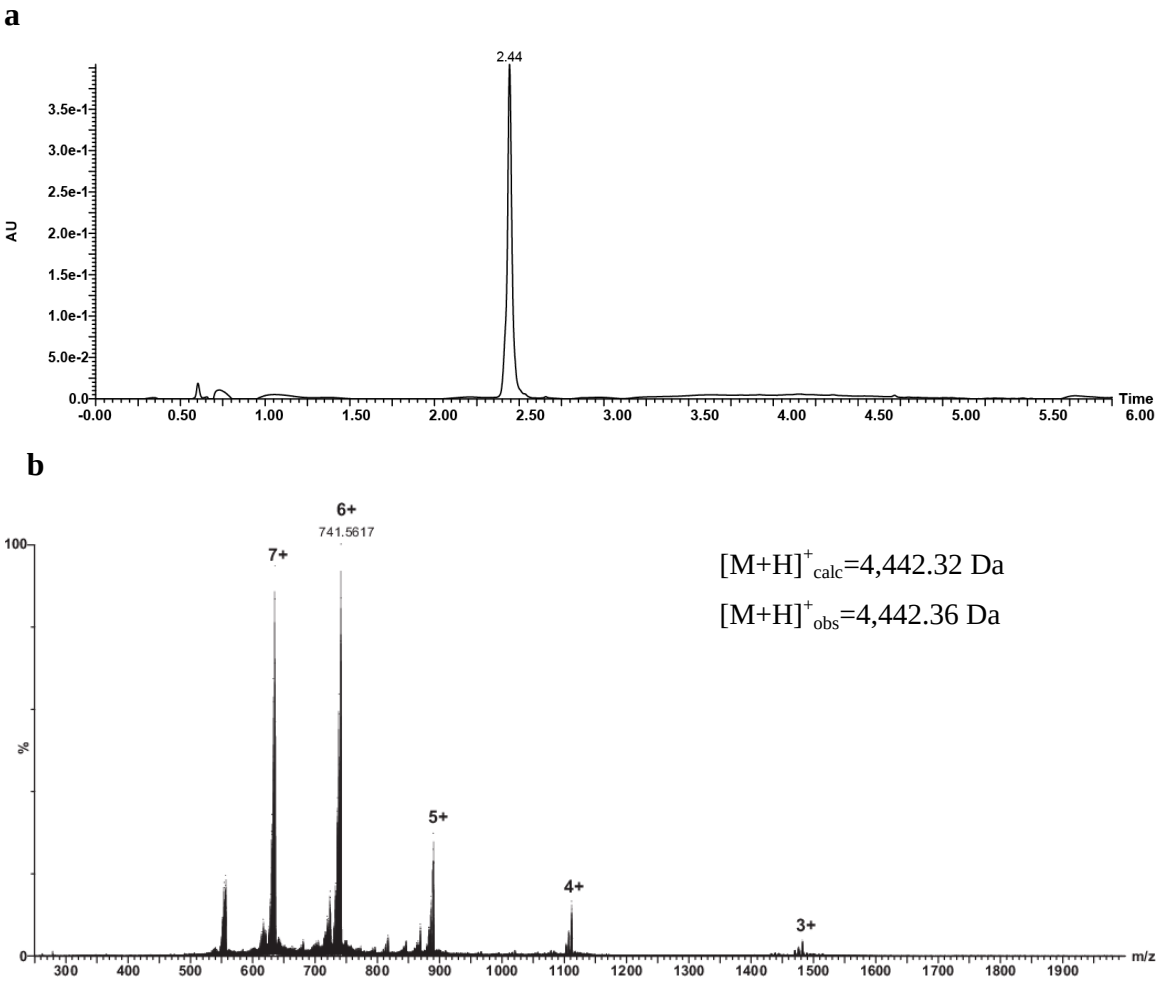
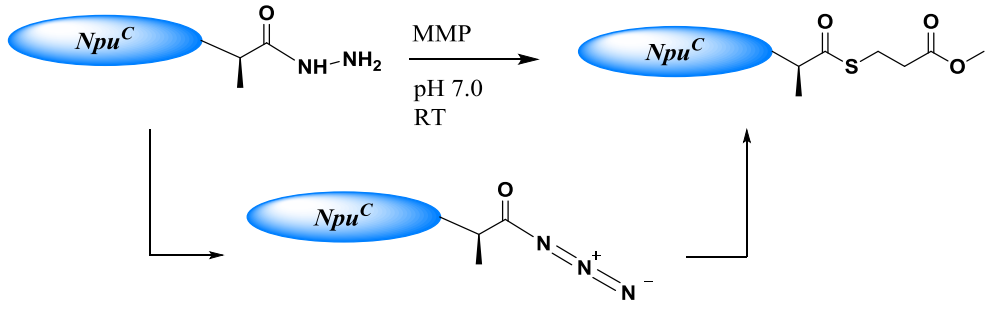


Figure 3.11. Characterization of *Npu*^{C} hydrazide. a) RP-HPLC (C18) chromatogram of purified *Npu*^{C} hydrazide. 5% - 70% ACN in water with 0.1% TFA in 30 min. **b)** Spectrum from LC-ESI-MS analysis of purified *Npu*^{C} hydrazide. $[M+6H]^{6+}_{obs}$: 741.5617 → 4,442.36 Da, $[M+H]^{+}_{calc}$: 4,442.32 Da.

Peptide thioester formation (Scheme 3.2) was carried out using a modification of the protocol in Zheng et al. (2013)[136]. Thioester formation was completed after 20 minutes, monitored by HPLC (C18). The formed *Npu*^{C} thioester was immediately purified in HPLC (C18). Pure *Npu*^{C} MMP thioester could be obtained in 45% yield, and was characterized in HPLC, MALDI-TOF-MS and LC-ESI-MS, Figure 3.12.

Scheme 3.2. Formation of *Npu*^{C}-thioester from the peptide hydrazide via peptide azide.



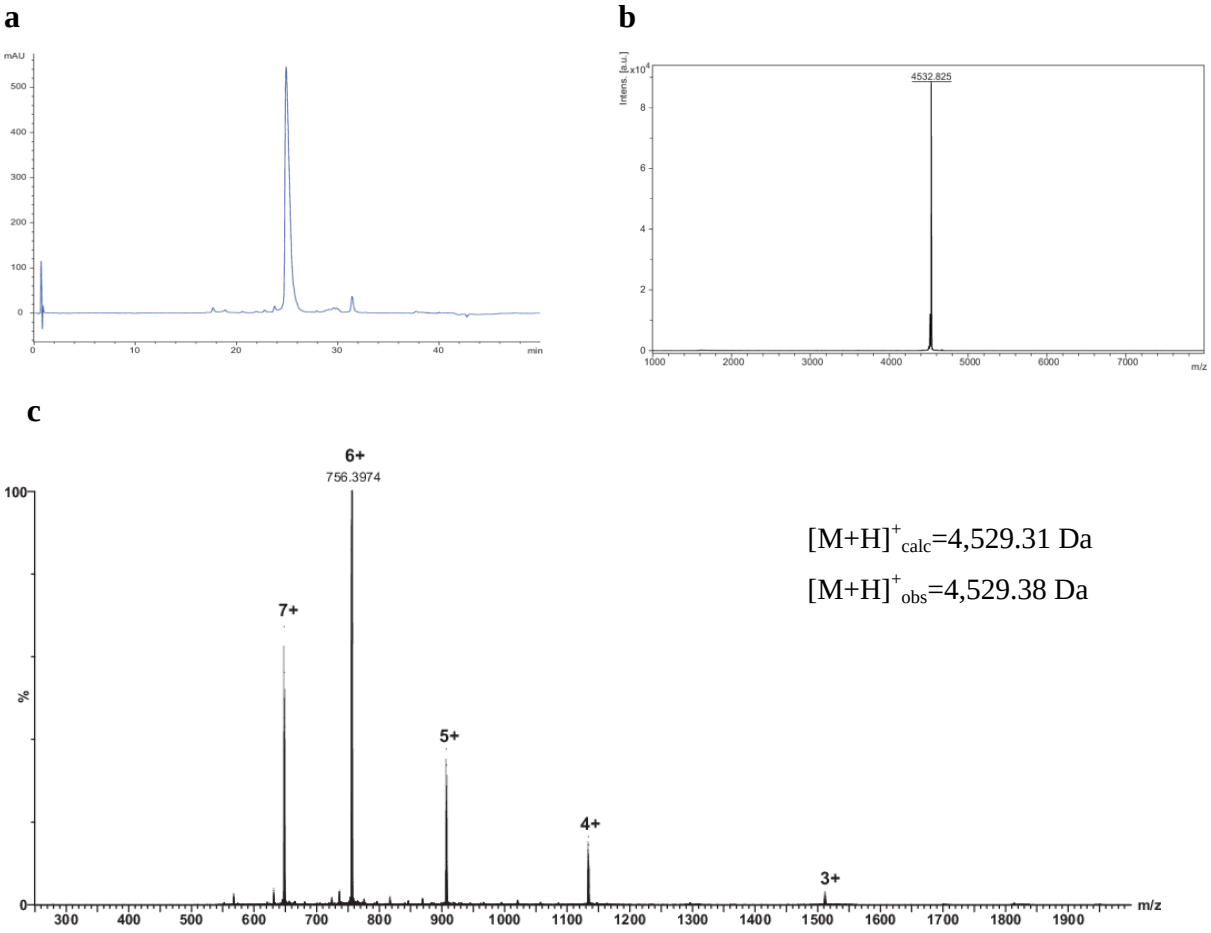


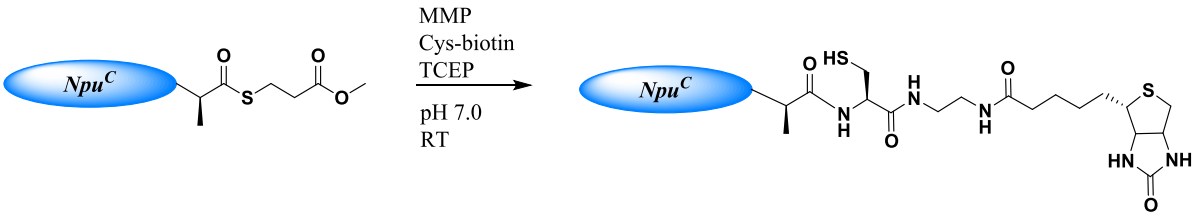
Figure 3.12. Characterization of *Npu*^{C} MMP thioester. a) RP-HPLC chromatogram (C18) of purified *Npu*^{C} MMP thioester. 10% - 30% ACN in water with 0.1% TFA in 30 min. **b)** MALDI-TOF spectrum of, $[M+H]^+_{obs}$: 4,532.82, $[M+H]^+_{calc}$: 4,529.31. **c)** Spectrum from LC-ESI-MS analysis of purified *Npu*^{C} MMP thioester. $[M+6H]^{6+}_{obs}$: 755.8969 → 4,529.38 Da, $[M+H]^+_{calc}$: 4,529.31 Da.

3.4.2 Native Chemical Ligation of *Npu*^{C}(WT)-Thioester

NCL of *Npu*^{C} thioester with Cys-biotin

Native Chemical Ligation (NCL) conditions were evaluated with a model compound before the ligation of more complex GPI molecules. Thus, NCL test reactions to obtain a ligation product with Cys-biotin were completed first. Cys-biotin was ligated to *Npu*^{C} thioester to form *Npu*^{C}-biotin (see Scheme 3.3).

Scheme 3.3. NCL of *Npu*^{C} MMP thioester with Cys-biotin.



For ligation, the *Npu*^{C} thioester was dissolved at a concentration of 1.5 mM in ligation buffer (pH 7.0) as described in 2.8.4, with 5 eq. Cys-biotin. Reaction progress was monitored by LC-ESI-MS (conditions as described in chapter 2.8.7). After three days reaction at 37°C the *Npu*^{C} thioester was consumed. The product *Npu*^{C}-biotin was purified by HPLC (C18). The fraction containing pure *Npu*^{C}-biotin was analyzed in RP-HPLC, MALDI-TOF-MS (Figure 7.3) and LC-ESI-MS, Figure 3.13, and was kept at -20°C for PTS reactions.

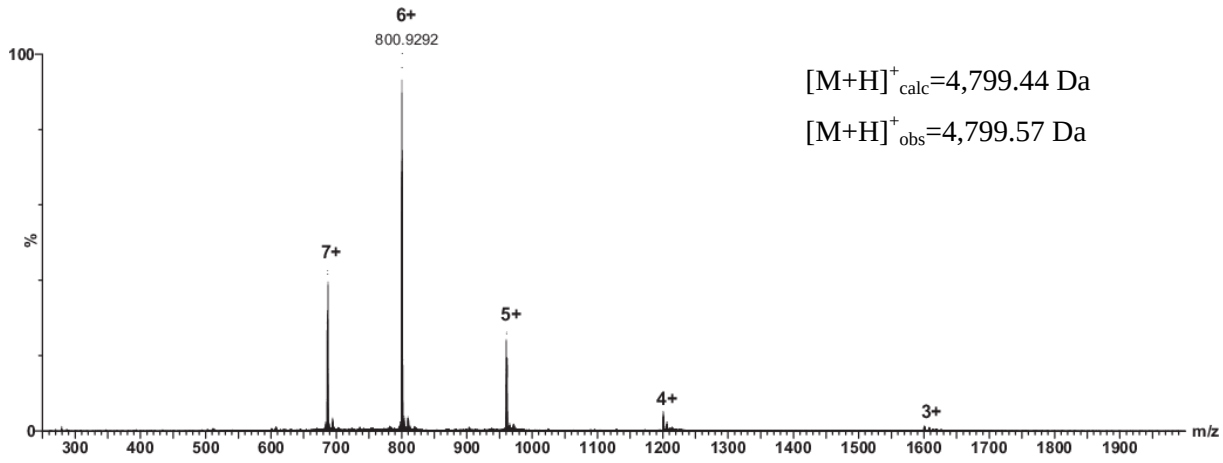
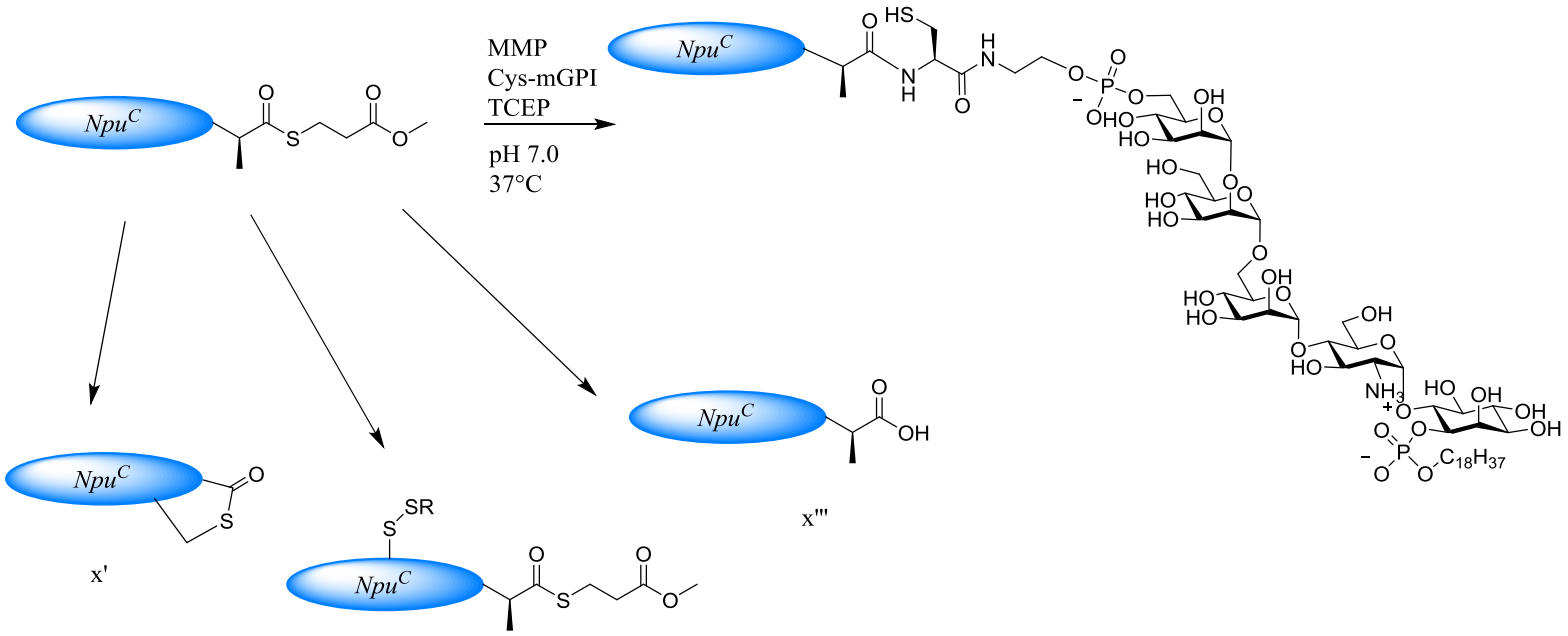


Figure 3.13. Characterization of *Npu*^{C}-biotin. Spectrum from LC-ESI-MS analysis of purified *Npu*^{C}-biotin. $[M+6H]^{6+}_{obs}$= 800.9292 → 4,799.57 Da, $[M+H]^{+}_{calc}$=4,799.44 Da.

NCL between *Npu*^{C} thioester and the monolipidated GPI (mGPI)

By using the optimized conditions with Cys-biotin, the ligation between *Npu*^{C} thioester and Cys-mGPI was investigated, Scheme 3.4.

Scheme 3.4. NCL of *Npu*^{C} MMP thioester with Cys-mGPI including side reactions.



After seven days of ligation at room temperature barely any product could be detected by LC-ESI-MS. Mainly the starting material (*Npu*^{C} thioester) and disulphides of the *Npu*^{C} peptide and the thiols (x''), hydrolyzed thioester (*Npu*^{C}-OH, x''') and the formation of a and a cyclized thiolactame side product of *Npu*^{C} (x') side products were observed, Figure 3.14, Table 3.1.

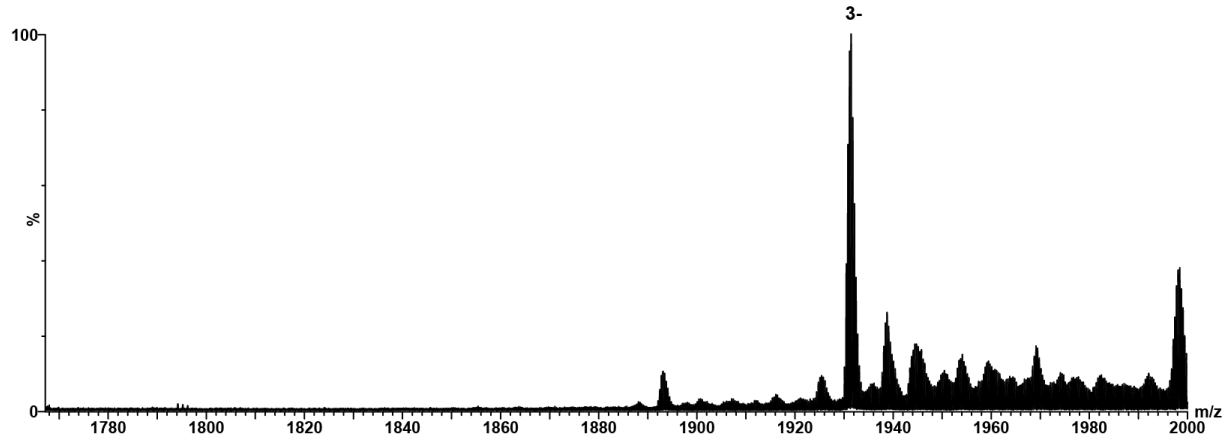


Figure 3.14. Characterization of ligation product *Npu*^{C}-mGPI. Spectrum from LC-ESI-MS analysis of NCL reaction *Npu*^{C}-mGPI (t_R=18.4 min), $[M-3H]^{3-}_{obs}$: 1,930.4391 → 5,794.3173 Da, $[M-H]^-_{calc}$: 5,793.8 Da.

To overcome these side reactions, especially oxidations of the cysteine residues, the ligation was performed under Argon atmosphere. After one day of ligation the product *Npu*^{C}-mGPI was detected in LC-ESI-MS, Figure 3.15, Table 3.1. Additionally, an adduct (+200 Da) was observed, which probably corresponds to a mercury adduct resulting from the deprotection procedure of Cys-mGPI. The reaction mixture was analyzed in LC-ESI-MS on several days and this adduct was decreasing in comparison to the peak corresponding to the exact mass, indicating the exchange of Hg for Na counter ions from the NCL buffer.

Table 3.1. Calculated and observed masses of products and side products of NCL between *Npu*^{C}-MMP thioester and Cys-mGPI. The respective mass spectra are given in Figure 3.15

Product	t_R (min)	$[M+H]^+_{calc}$	$[M+H]^+_{obs}$
***Npu*^{C}-thiolactame (b)**	10.9	4,410.29 Da	735.9197 *6 → 4,410.52
***Npu*^{C}-OH (c)**	11.6	4,427.30 Da	738.9243 *6 → 4,427.54
***Npu*^{C}-MMP (d)**	12.6	4,529.31 Da	756.0955 *6 → 4,530.57
***Npu*^{C}-MMP-MMP (e)**	13.6	4,647.32 Da	775.7636 *6 → 4,648.58
***Npu*^{C}-mGPI (f)**	19.7	5,793.8 Da	1,159.6219 *5 → 5,793.1 1,199.4110 *5 → 5,992.05, Δ 200

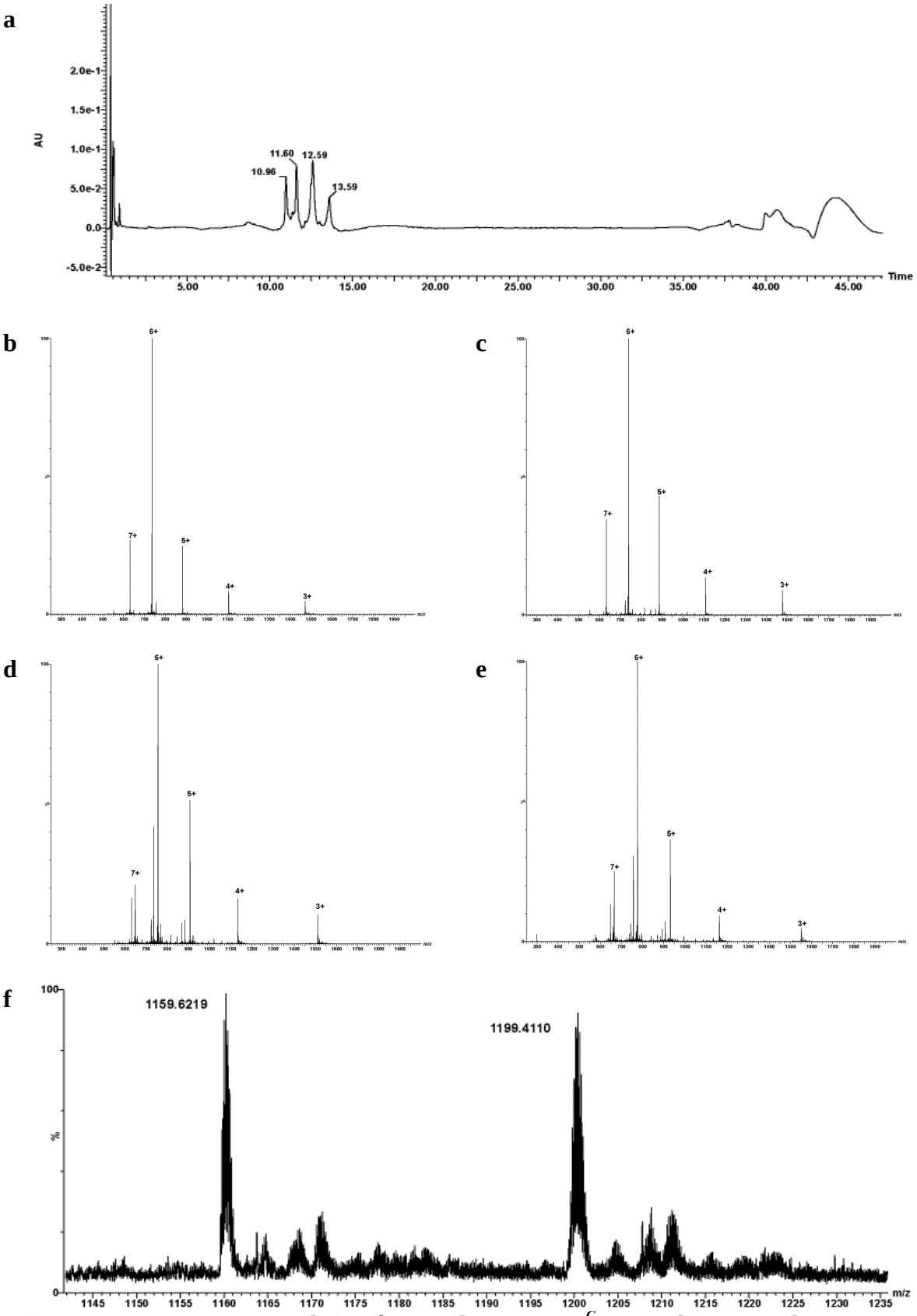


Figure 3.15. LC-ESI-MS analysis of NCL between *Npu*^{C}-MMP thioester and Cys-mGPI. a) chromatogram (C4, 10-70% ACN in 30 min) with four peaks, **b)** mass spectrum of t_R=10.9 min – *Npu*^{C}-thiolactame, **c)** mass spectrum of t_R=11.6 min – *Npu*^{C}-OH, **d)** mass spectrum of t_R=12.6 min – *Npu*^{C}-MMP, **e)** mass spectrum of t_R=13.7 min – *Npu*^{C}-MMP-MMP**, f**) mass spectrum (5x charged ion) of t_R=18.6 min – *Npu*^{C}-mGPI. Calculated and observed masses can be found in Table 3.1.

During seven days reaction, the pH was repeatedly re-adjusted to 7.0 using 1 M NaOH or 1 M HCl, and after one day and four days, one more eq. TCEP was added to the reaction as some oxidation of *Npu*^{C}-MMP was observed (not shown). After seven days the reaction was purified in HPLC (C4, Figure 3.17) since barely any unreacted, non-oxidized peptide thioester was observed anymore and a cleaner mass spectrum was obtained, Figure 3.16 b. The product is expected in the smaller peak at t_R=24.5 min, however could not be detected after purification due to the small amount obtained (140 µg).

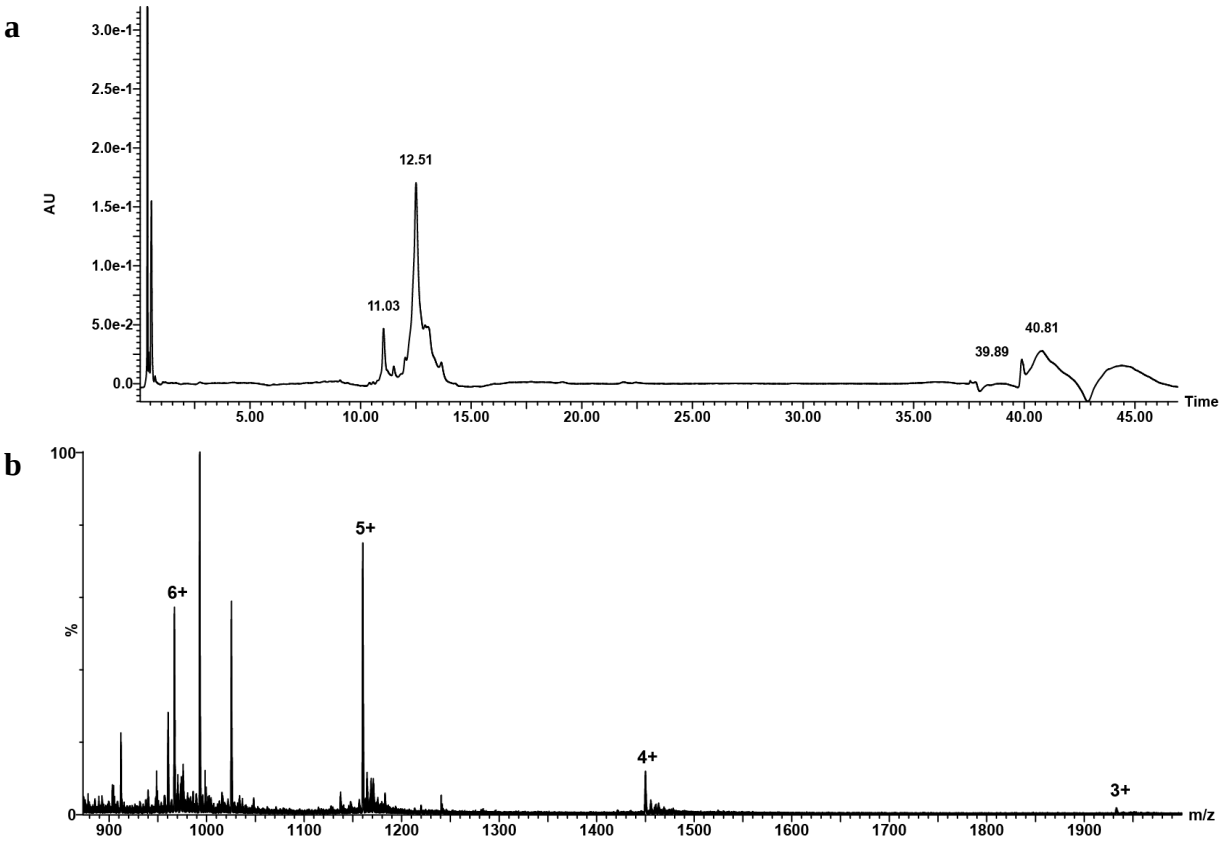


Figure 3.16. LC-ESI-MS analysis of NCL between *Npu*^{C}-MMP after seven days. a) Chromatogram (C4, 10-70% ACN in 30 min), **b)** mass spectrum acquired at t_R=19.1 min – *Npu*^{C}-mGPI: 1,159.7755 m/z, 5x charged → 5,793.8775. Masses at other t_R correlate with *Npu*^{C}-OH (11 min) and an oxidized *Npu*^{C}-thioester (Δ+16, 12.5 min).

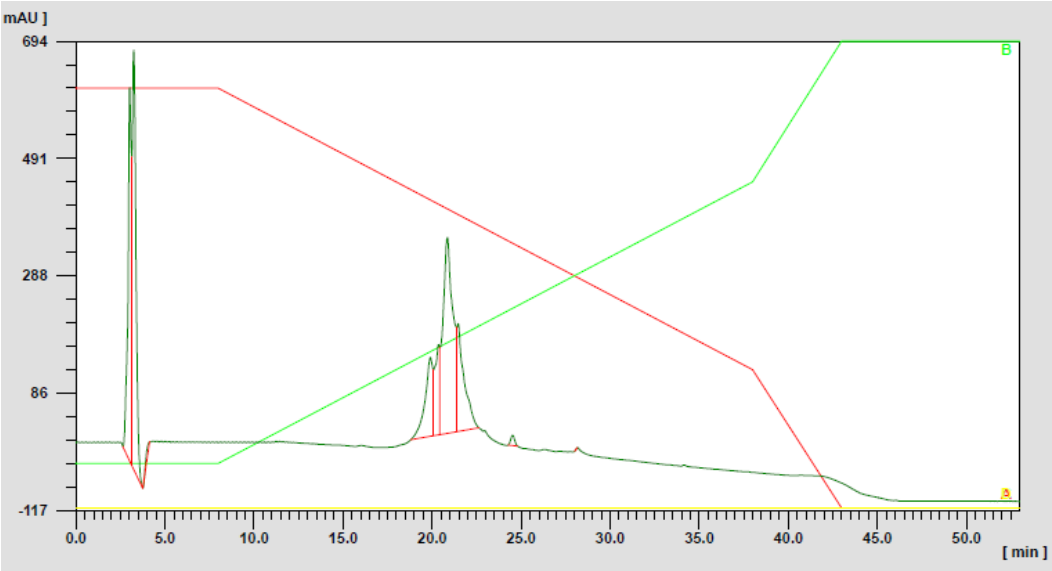


Figure 3.17. Chromatogram including gradient (10 – 70% ACN in 30 min) of the purification of the NCL reaction between *Npu*^{C}-MMP-thioester and Cys-mGPI. The main peak corresponds to thioester derivatives (refer to LC-MS analysis, Figure 3.15) and the peak at t_R=24.5 min probably contains the product *Npu*^{C}-mGPI.

3.4.3 Synthesis of the Mutated *Npu*^{C}(AA) Peptide

The mutated *Npu*^{C}(AA) peptide having Asn^{35} and Cys^{36} exchanged for Ala residues (Scheme 3.5) was synthesized using Rink Amide PEGA Resin (Novabiochem, loading 0.35 mmol/g) under the same conditions as for *Npu*^{C}(WT), 3.4.1. The synthesis of this peptide showed the formation of less side products and better crude yield of the product compared to the synthesis of the WT peptide. The peptide was purified using RP-HPLC (C18) and characterized by MALDI-TOF-MS and LC-ESI-MS (Figure 7.2).

Scheme 3.5. *Npu*^{C}(AA).

3.5 Expressed Protein Ligation

To obtain C-terminally modified proteins by Expressed Protein Ligation (EPL), α-protein thioesters and cysteine tagged molecules were needed: Cys-Biotin and Cys-Dimannose used for method development and Cys-GPIs.

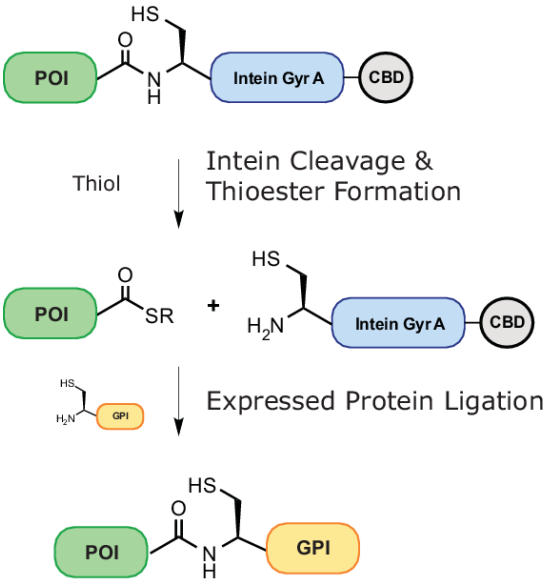


Figure 3.18. Schematic depiction of Expressed Protein Ligation (EPL). The protein of interest (POI) is expressed as a fusion protein with the GyrA intein from *M. xenopi*. Thioester formation is achieved by intein cleavage induced by thiol addition. This is followed by the ligation to a Cys-containing peptide or other fragment (Cys-GPI).

3.5.1 Expression, Purification and Protein-Thioester formation

Expression of eGFP-*Mxe*-CDB and thioester formation

The plasmid pET30a-eGFP-*Mxe*-CBD was obtained from GenScript (Piscataway, USA) with the sequence given in Appendix 7.1.1. To overcome the inefficient intein cleavage with proline as adjacent residue, a mutation in the C-terminus of eGFP was introduced (Pro was exchanged for Ala).[123,168] The plasmid was transformed in *E. coli* BL21 (DE3) and Rosetta cells. Expression was optimized by testing 20°C and 30°C as expression temperatures, 0.2, 0.4 and 1 mM IPTG concentration and induction for 1, 2, 3 and 24 h. *E. coli* BL21 (DE3) Star and Rosetta cells were tested for the expression. As shown in Figure 3.19, the Rosetta cells grew slightly better; they reached the late exponential phase after 3 h whereas the BL21 clone could be induced only after 5 h. Also post-induction the Rosetta clone continued to grow better.

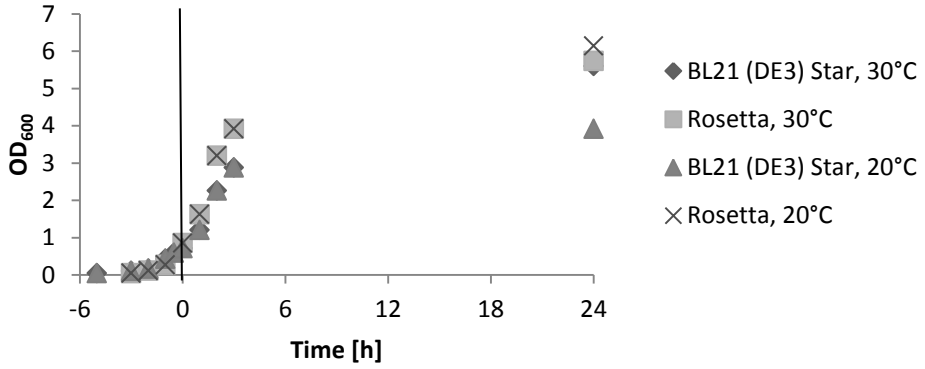


Figure 3.19. Growth curves of the test expression of eGFP-*Mxe*-CDB fusion protein in *E. coli* BL21 (DE3) Star and Rosetta cells. All cells were inoculated to a starting OD_{600} of 0.05. OD was measured frequently until late exponential phase was reached (OD 0.6-0.8). At this point the temperature was changed to 20°C or 30°C, respectively, and this time corresponds to t=0 (induction time, indicated by the vertical line).

To analyze the expression of the eGFP-*Mxe*-CBD fusion protein, samples were taken 1 h, 2 h, 3 h and 24 h post-induction (Figure 3.20). The amount of sample for each lane in the gel was normalized according to the respective OD_{600}. Overall, the overexpression of the fusion protein is good under all tested conditions; however, the best expression of the fusion protein (assessed by band size using Fiji software) was observed using Rosetta cells induced with 0.4 mM IPTG for 24 h at 20°C or 30°C.

In order to assess the expression temperature yielding a higher portion of soluble fusion protein, samples were taken after 24 h from the cultures induced with 0.4 mM IPTG and the soluble fraction was separated from the insoluble fraction by centrifugation. Lysate samples and samples of the pellet resuspended in PBS buffer were analyzed in SDS-PAGE, see Figure 3.21. At both expression temperatures the fusion protein was distributed over the soluble and insoluble fractions, however, better results were obtained at 20°C with ca. 60% of the fusion protein present in the soluble fraction compared to ca. 30% soluble protein observed at 30°C (calculated using Fiji software). Therefore induction with 0.4 mM IPTG and expression at 20°C were chosen as the best conditions for future expressions of the eGFP-*Mxe*-CBD fusion protein.

a – BL21 (DE3) Star **b - Rosetta**

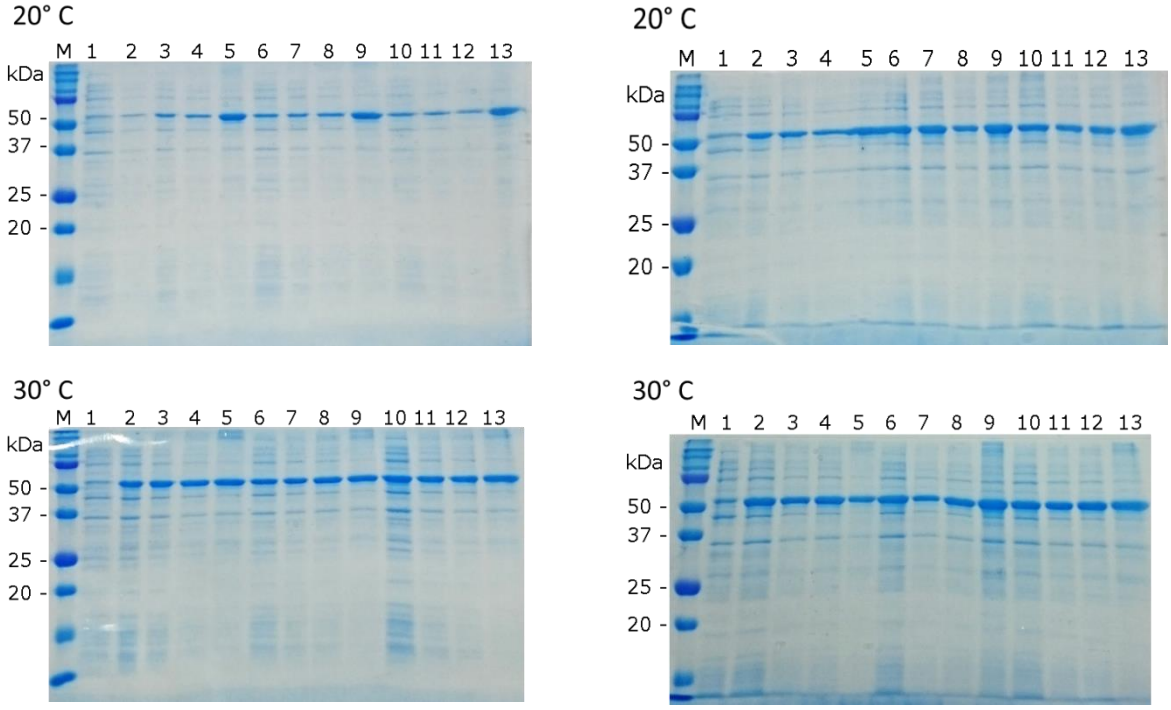


Figure 3.20. SDS-PAGE analysis of the test expressions of eGFP-*Mxe*-CBD. The fusion protein was expressed in *E. coli* BL21 (DE3) Star **(a)** and Rosetta cells **(b)**. Cells were grown to late exponential phase and induced with different IPTG concentrations. Induction temperatures tested were 20°C and 30°C. M – Molecular Weight Marker, 1 – pre induction, 2 – 5 – 0.2 mM IPTG induction 1/2/3/24 h, 6 – 9 – 0.4 mM IPTG induction 1/2/3/24 h, 10 – 13 – 1 mM IPTG induction 1/2/3/24 h.

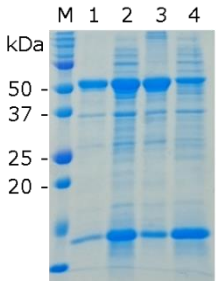


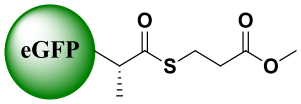
Figure 3.21. Solubility check of the eGFP-*Mxe*-CBD fusion protein using SDS-PAGE. Expression samples induced with 0.4 mM IPTG at 20°C or 30°C, respectively, were lysed and separated. M – Molecular weight marker, 1 – 20°C, insoluble fraction, 2 – 20°C, soluble fraction, 3 – 30°C, insoluble fraction, 4 – 30°C, soluble fraction.

For preparative expression of the eGFP-*Mxe*-CBD fusion protein, the Rosetta clone containing the pET30a-eGFP-*Mxe*-CDB plasmid was plated onto an LB agar plate with 100 µg/mL kanamycin and grown at 37°C overnight. The next day, a culture of 100 mL LB medium with 100 µg/mL kanamycin was inoculated to a starting OD_{600} of 0.05 with the starter culture and grown at 37°C until the late exponential growth phase was reached (OD_{600} 0.6 – 0.8). Protein expression was induced by addition of 0.4 mM IPTG, the temperature was decreased to 20°C and shaking speed was reduced from 220 to 180 rpm. After 24 h induction

the cells were harvested by centrifugation. Cells were lysed using lysozyme and sonication. Lysates were cleared by centrifugation and filtrated through a 0.22 µm syringe filter.

Purification of the protein was performed in one single chromatographic step with simultaneous formation of the protein thioester as described in 2.8.3. Following completion of the cleavage, the pure eGFP-thioester with either MMP (Scheme 3.6), MESNA, MMBA or MPAA was eluted with 6 x 0.5 CV buffer W (pH 7.0). The fractions (flow-through, wash, resin and elution samples) were analyzed in SDS-PAGE. The green elution fractions containing the fusion protein were combined and the pH was adjusted to ca. 6.5 in order to prevent thioester hydrolysis.

Scheme 3.6. eGFP-MMP-thioester



Thiol reagents with different properties regarding reactivity and stability were tested for thioester formation. Thioester formation on-column with the different thiols is shown for eGFP, Figure 3.22. In the presence of the commonly used MESNA thiol (200 mM), intein cleavage reached ca. 80% after 40 h incubation with cleavage buffer at room temperature, Figure 3.22 a. In contrast, using 50 mM MPAA in the cleavage buffer, only ca. 50% intein cleavage was achieved after 24 h incubation at RT (Figure 3.22 b). Excellent results were also obtained by using MMBA and MMP (both 50 mM in cleavage buffer): 80% and 90% cleavage were obtained, respectively, Figure 3.22 c.

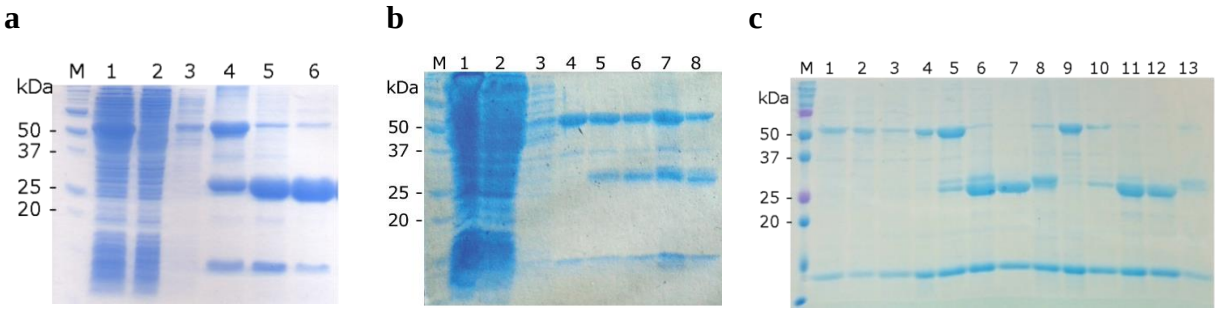


Figure 3.22. Thioester formation on chitin resin using Sodium 2-Mercaptoethanesulfonate (MESNA), 4-Mercaptophenylacetic acid (MPAA), 4-(Mercaptomethyl)benzoic acid (MMBA) and Methyl 3-mercaptopropionate (MMP). a) Thioester formation using **MESNA**. M – Molecular weight marker, 1 – lysate, 2 – FT chitin resin, 3 – wash fraction, 4 – resin t=0 (MESNA), 5 – resin 40 h (MESNA), 6 – elution (MESNA). **b)** Thioester formation using **MPAA**. Molecular weight marker, 1 – lysate, 2 – FT chitin resin, 3 – wash fraction, 4 – resin t=0, 5 – resin 2 h (MPAA), 6 – resin 4 h (MPAA), 7 – resin 6 h (MPAA), 8 – resin 24 h (MPAA), **c)** Thioester formation using **MMBA** and **MMP**. M – Molecular weight marker, 1 – lysate, 2 – FT chitin resin, 3 – wash fraction, 4 – resin t=0 (MMBA), 5 – resin 1 h (MMBA), 6 – resin 16 h (MMBA), 7 – elution pooled (MMBA), 8 – resin after elution (MMBA), 9 – resin t=0 (MMP), 10 – resin 1 h (MMP), 11 – resin 16 h (MMP), 12 – elution pooled (MMP), 13 – resin after elution (MMP).

An analysis of the protein bound to the chitin column showed that the lysozyme used for cell lysis binds to the chitin resin as well, Figure 3.22 c. This is in well correspondence with literature reports describing chitin as a substrate of lysozyme.[169] Therefore, the protein thioesters were re-purified using SEC (Superdex peptide 30/10 GL in EPL buffer) to remove lysozyme and remaining thiol reagent, giving 8 mg pure eGFP-MMP-thioester from 100 mL culture in 81% purity (Figure 3.23).

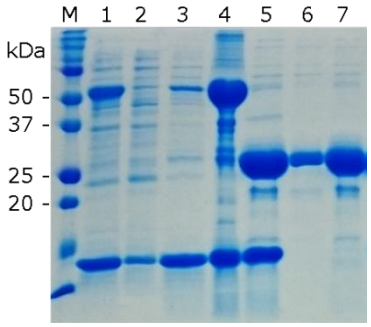


Figure 3.23. Purification and thioester formation of eGFP. The overexpressed fusion protein eGFP-*Mxe* is loaded to a chitin resin, washed and thioester formation is induced by addition of thiol-containing buffer (MMP). Lastly, the thioester is purified using SEC. M – Molecular weight maker, 1 – Lysate, 2 – Flow-through chitin resin, 3 – was fraction chitin resin, 4 – resin sample t=0, 5 – pooled elution fractions, 6, 7 – SEC-purified eGFP-MMP-thioester.

Based on the results of the intein cleavage with MMBA and MMP, these two protein thioesters were selected for ligations with cysteine-containing molecules and analysis in MALDI-TOF-MS (Figure 7.4) and LC-ESI-MS (Figure 3.24 and Figure 7.5).

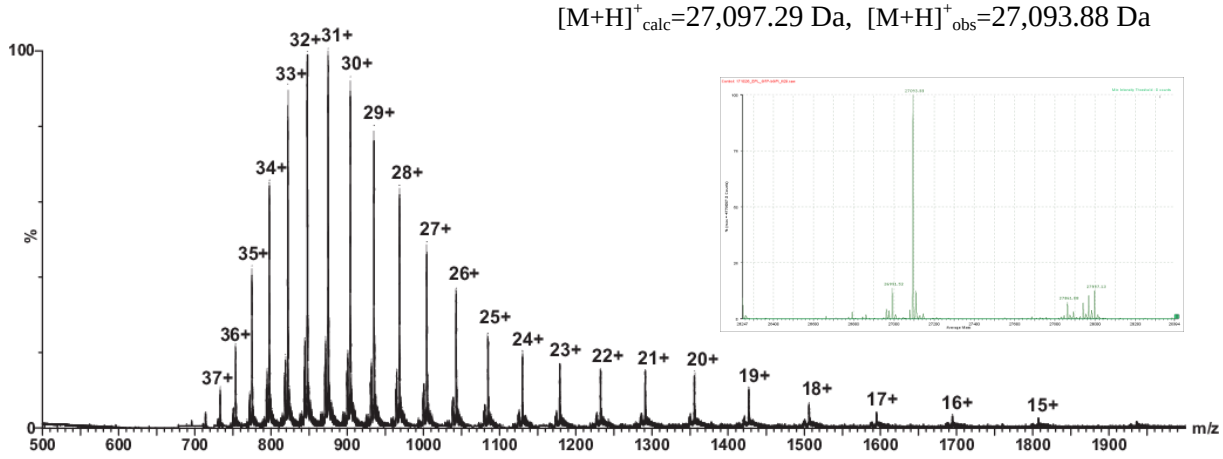


Figure 3.24. Raw and deconvoluted ESI mass spectra for eGFP-MMP-thioester. $[M+H]^+_{calc}$ = 27,097.29, $[M+H]^+_{obs}$ = 27,093.88. The eGFP-thioester eluted from chitin resin was analyzed using a gradient from 10 – 70% acetonitrile in water (0.1% FA in both eluents) in 30 min. The sample was analyzed using ESI-QTOF. At t_R= 18.6 min, the given mass spectrum was recorded. The deconvoluted mass is 27,093.88 Da.

Following the production of the protein thioesters, their stability was assessed by HPLC (C4). The thioester was analyzed immediately following its generation and elution from the chitin resin (i.e. “fresh thioester”). The sample was incubated for one week at 4°C in the elution buffer (pH 6.5-7.0) containing some remaining thiol reagent and subsequently analyzed by HPLC. The chromatogram of the MMBA thioester sample (Figure 3.25 a) showed peaks before t_R=18 min which were assigned to buffer components (mainly MMBA and its disulphides). A peak at t_R=19.5 min corresponds to the lysozyme that was co-purified using the chitin resin (as described before) and the peak at t_R=22.7 min corresponds to the eGFP-thioester. A small shoulder observed besides the eGFP-thioester in the fresh sample increases in size to about 70% of the size of the thioester peak during one week of incubation. LC-MS analysis showed that this shoulder corresponds to the hydrolyzed side product eGFP-OH. In comparison, the chromatogram of the MMP thioester samples were cleaner since MMP and its disulphides do not absorb at the wavelength used for detection (214 nm), Figure 3.25 b.

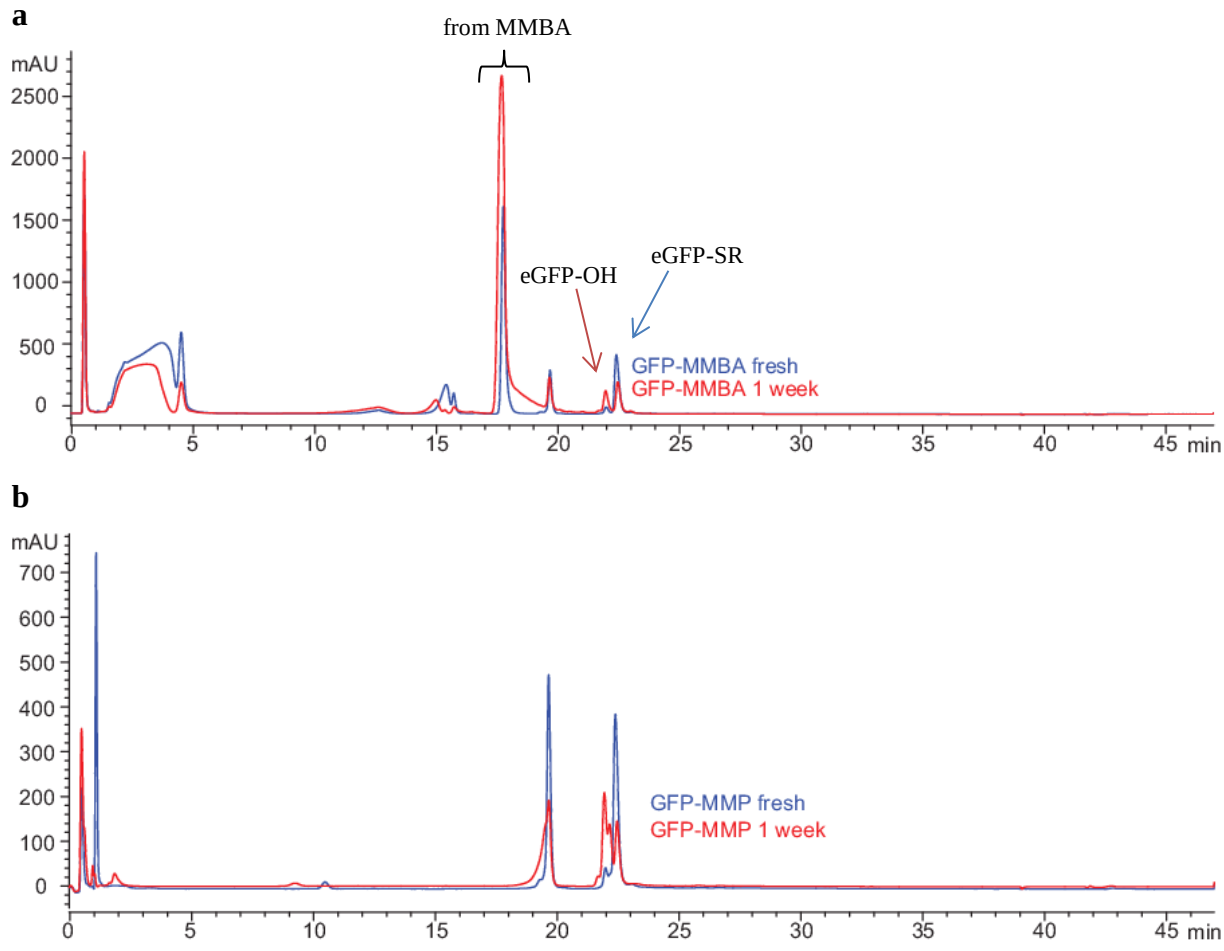


Figure 3.25. Stability test of eGFP-thioesters. Fresh thioesters (with MMBA and MMP) as well as samples stored at 4°C for one week have been analyzed in RP-HPLC (C4, gradient 10 – 70% acetonitrile in water with 0.1% TFA in 30 min). Both thioesters were degraded substantially over one week.

Two peaks corresponding to lysozyme (t_R=19.5 min) and the eGFP-MMP-thioester (t_R=22.7 min) together with a small shoulder were present in the fresh thioester sample. After one week incubation, the thioester peak decreases to about 30% in size compared to the starting sample and the shoulder increases to represent ca. 70% of the eGFP present in the sample. These results showed that both thioesters are not stable enough to be stored for extended periods of time in the ligation buffer. It was therefore decided to always generate the eGFP-thioester freshly before use in EPL reactions.

Expression of Thy-1-*Mxe*-CDB and thioester formation

Plasmid pET29a-Thy1-*Mxe* (Appendix 7.1.2) (GenScript, Piscataway, USA) was transformed into *E. coli* BL21 (DE3) Star and Rosetta cells. Protein expression was tested with both clones at 15°C and 30°C, and induction with IPTG concentrations of 0.2 mM, 0.4 mM and 1 mM.

Expression tests were performed as described for the eGFP-*Mxe*-CBD fusion protein: cultures were grown to late exponential phase (OD_{600} 0.6-0.8) at 37°C before reducing temperature and shaking speed, and induction via IPTG addition. The expression of the fusion protein (expected size = 41.2 kDa) was analyzed in SDS-PAGE. The Rosetta clone showed better growth before and after induction (Figure 7.7). However, almost no fusion protein was expressed (Figure 7.8 b). The best expression was observed with BL21 (DE3) Star cells at 15°C with 0.4 mM IPTG and 30°C with 0.2 mM ITPG. These conditions were also tested to evaluate the portion of soluble protein (Figure 7.9). Unfortunately, under both conditions the fusion protein was expressed only in insoluble form. Therefore, expression at 15°C with 0.4 mM IPTG was selected for larger scale expression.

Preparative expression of Thy-1-*Mxe*-CBD was performed in 100 mL LB medium supplemented with 100 µg/mL kanamycin. The cells were lysed as described before, repeating the sonication and centrifugation steps once. The lysates were discarded and the pellet containing insoluble Thy-1-*Mxe*-CBD was resuspended in HisTrap equilibration buffer containing 8 M urea. Vortexing, ultrasonic bath and shaking overnight at room temperature was necessary to solubilize the protein. The next day, the solubilized inclusion bodies were centrifuged to remove cell debris and the protein was purified via his-tag (HisTrap HP 1mL column). The supernatant was loaded to the column, washed with equilibration buffer for 10 CV and the fusion protein was eluted by a gradient over 5 CV from 30 mM to 500 mM imidazole, Figure 3.26 a. Samples containing the fusion protein were combined and dialyzed

against EPL buffer containing 6 M urea, exchanging the buffer several times over 24 h. 10.5 mg Thy-1-*Mxe* were obtained from 100 mL culture.

Purification with simultaneous thioester formation on-column was not possible for Thy-1-*Mxe*-CBD since the CDB fusion does not bind under denaturing conditions required for solubilization. Therefore, in order to form the thioester, MMP was added to a solution containing purified Thy-1-*Mxe*-CBD in 6 M urea. Unfortunately, no intein cleavage could be observed (not shown). To favor the folding of the intein, thioester formation was tested with MMP and MESNA (Scheme 3.7) in 3 M urea (obtained by slowly diluting the 6 M urea solution with EPL buffer 1:1). After one day incubation at room temperature, some cleavage was observed with both thiols MMP and with MESNA, Figure 3.26 b. However, even with MESNA the cleavage efficiency is very low (12% cleavage by band intensity).

Characterization of the fusion protein and the Thy-1-thioester by HPLC and LC-ESI-MS was challenging, and no mass spectra could be acquired.

Scheme 3.7. Thy-1-MMP-thioester & Thy-1-MESNA-thioester.

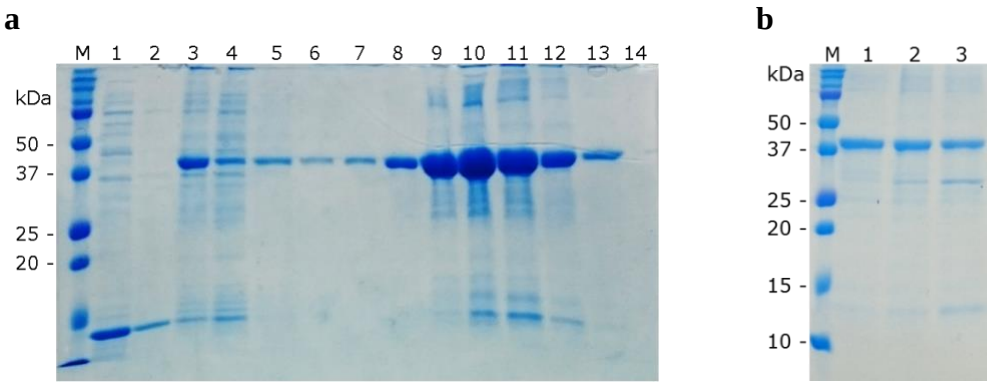


Figure 3.26. a) Expression and His-tag purification of Thy1-Mxe. M – Molecular weight marker, 1 – lysate 1, 2 – lysate 2, 3 – solubilized inclusion bodies, 4 – flow-through HisTrap column, 5 – wash HisTrap column, 6 – 14 – Elution fractions. **b) Thioester formation of Thy-1 in 3 M urea with MMP and MESNA and EPL with Cys-biotin.** Some cleavage is observed with both thiols, however with 100 mM MESNA it is better (ca. 12% by band intensity). M – Molecular weight marker, 1 – Thy-1-*Mxe* in 6 M urea, 2 – thioester in 3 M urea, 50 mM MMP, 1 d, 3 - Thy-1-*Mxe* in 6 M urea, 2 – thioester in 3 M urea, 100 mM MESNA, 1 d.

Expression of IL-2-*Mxe*-CDB and thioester formation tests

The plasmid pTXB1-IL-2 generated by cloning (see 3.2.2) was transformed into BL21 (DE3) Star and Origami cells for expression tests. The Origami cells were tested because it was expected that the intracellular redox conditions they confer could be more suitable for expression of this cysteine-containing protein. Expression was tested at 37, 25, 19 and 12°C.

Ampicillin (100 µg/mL) was used for pTXB1 vector in BL21 (DE3) Star cells, tetracycline (12.5 µg/mL) and kanamycin (15 µg/mL) were used additionally for Origami cells. Fresh transformants were always used for expression. BL21 cells were induced after 2.5 h with 0.4 mM IPTG for 3 h (37 and 25°C) and 24 h (other temperatures). Origami cells grew very slowly and were induced after 4.5 h with 0.4 mM IPTG for 3 or 24 h, respectively. Cells were harvested by centrifugation and lysed by sonication.

The 47 kDa protein was observed only in BL21 cells and mostly in insoluble form. Only at 12°C a portion of the protein was found in the soluble fraction, Figure 3.27 a. In Origami cells, no expression was observed (Figure 3.27 b). Expression at 12°C in BL21 (DE3) Star cells was used for further experiments.

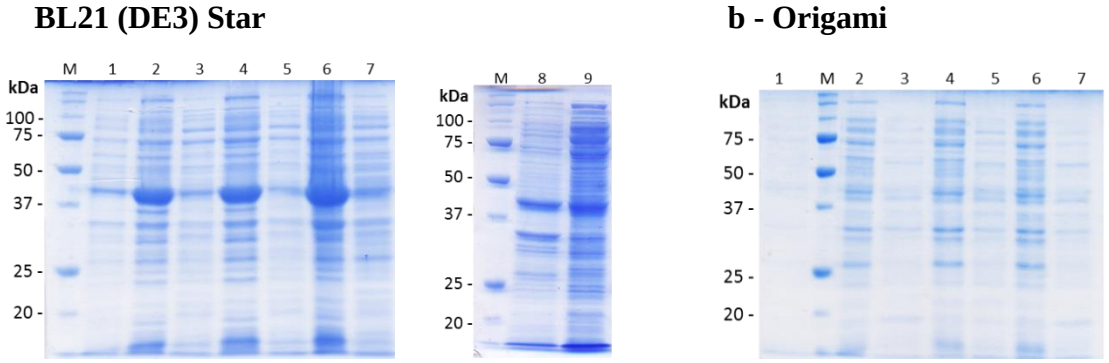


Figure 3.27. Expression test of IL-2-*Mxe*-CBD in BL21 cells. M – Molecular weight marker, 1 – pre induction, 2 – insoluble fraction 37°C, 3 – soluble fraction 37°C, 4 – insoluble fraction 25°C, 5 – soluble fraction 25°C, 6 – insoluble fraction 19°C, 7 – soluble fraction 19°C, 8 – insoluble fraction 12°C, 9 – soluble fraction 12°C.

For the formation of the protein thioester, the lysate of IL-2-*Mxe*-CBD was slowly loaded onto chitin resin and flushed with cleavage buffer containing 100 mM MESNA following washing with 20 CV buffer W and incubation at room temperature for 40 h. Only a small portion of the protein bound to the resin and intein cleavage could not be observed, Figure 3.28 a. Thioester formation was also tested with 100 mM MMBA in the cleavage buffer. However, also under these conditions only a very minor band of the IL-2-thioester could be observed by western blot with an anti-strep antibody-HRP, Figure 3.28 b.

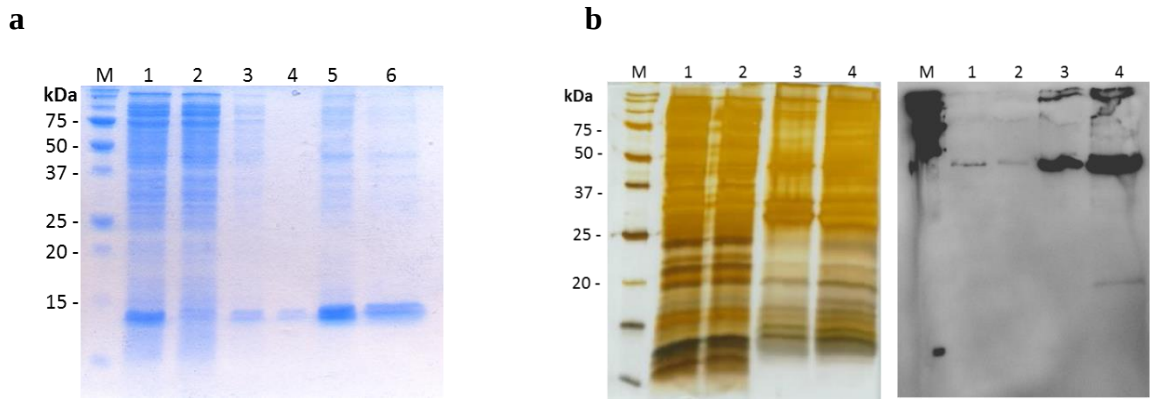


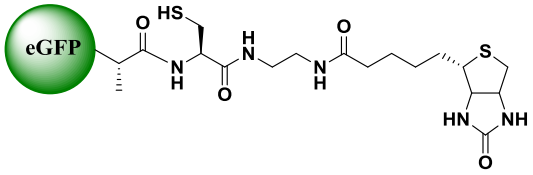
Figure 3.28. Thioester formation with IL-2-*Mxe*-GyrA. a) M – Molecular weight marker, 1 – lysate, 2 – flow-through, 3 – wash 1, 4 – wash 2, 5 – resin t=0, 6 – resin 40 h, **b)** M – Molecular weight marker, 1 – lysate, 2 – flow-through, 3 – resin t=0, 4 – resin 16 h. Gel was stained using silver staining, blot was detected with anti-strep antibody-HRP conjugate.

In order to evaluate thioester formation in solution as described for Thy-1, pure protein was needed. However, any attempt to purify IL-2-*Mxe*-CBD via StrepTactin resin failed due to unsuccessful binding of the protein to the resin (not shown). This strategy was not further pursued for this protein.

3.5.2 Optimization of the Ligation Conditions using eGFP

To test the reactivity of the protein thioesters in EPL reactions, Cys-biotin was used for optimization of the reaction conditions. Initially, purified eGFP-thioester in EPL buffer was mixed with 0.5 mM TCEP, 50 mM thiol (MMP) and 5 eq. Cys-biotin to give the product eGFP-biotin (Scheme 3.8). After one day incubation a sample was analyzed in SDS-PAGE and western blot (Figure 3.29) and LC-ESI-MS (Figure 3.30). A strong signal was observed at the expected size of ~27.3 kDa in the EPL lane (lane 2).

Scheme 3.8. eGFP-biotin.



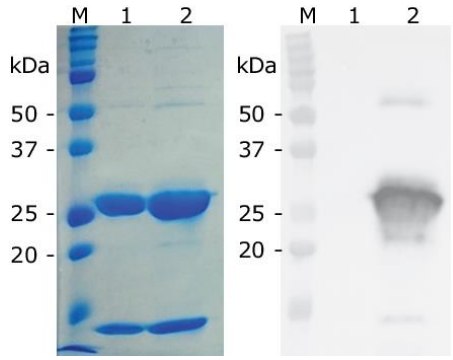


Figure 3.29. EPL of eGFP-MMP-thioester with Cys-biotin. The ligation product eGFP-biotin **xx** was detected in western blot using an anti-biotin antibody-HRP conjugate (1:250 dilution in 5% BSA in PBS-T). M – Molecular weight marker, 1 – eGFP-MMP, 2 – EPL reaction eGFP-biotin 1 day.

LC-ESI-MS analysis of the ligation reaction after one day using a C4 column (Figure 3.30) showed the formation of the product eGFP-biotin as a minor product (t_R=19.2 min, corresponding to the main peak's tailing); however the main peak was still the eGFP-MMP-thioester (t_R=18.9 min). Additionally, a portion of hydrolyzed thioester eGFP-OH was also observed (t_R=18.6 min, shoulder of the main peak).

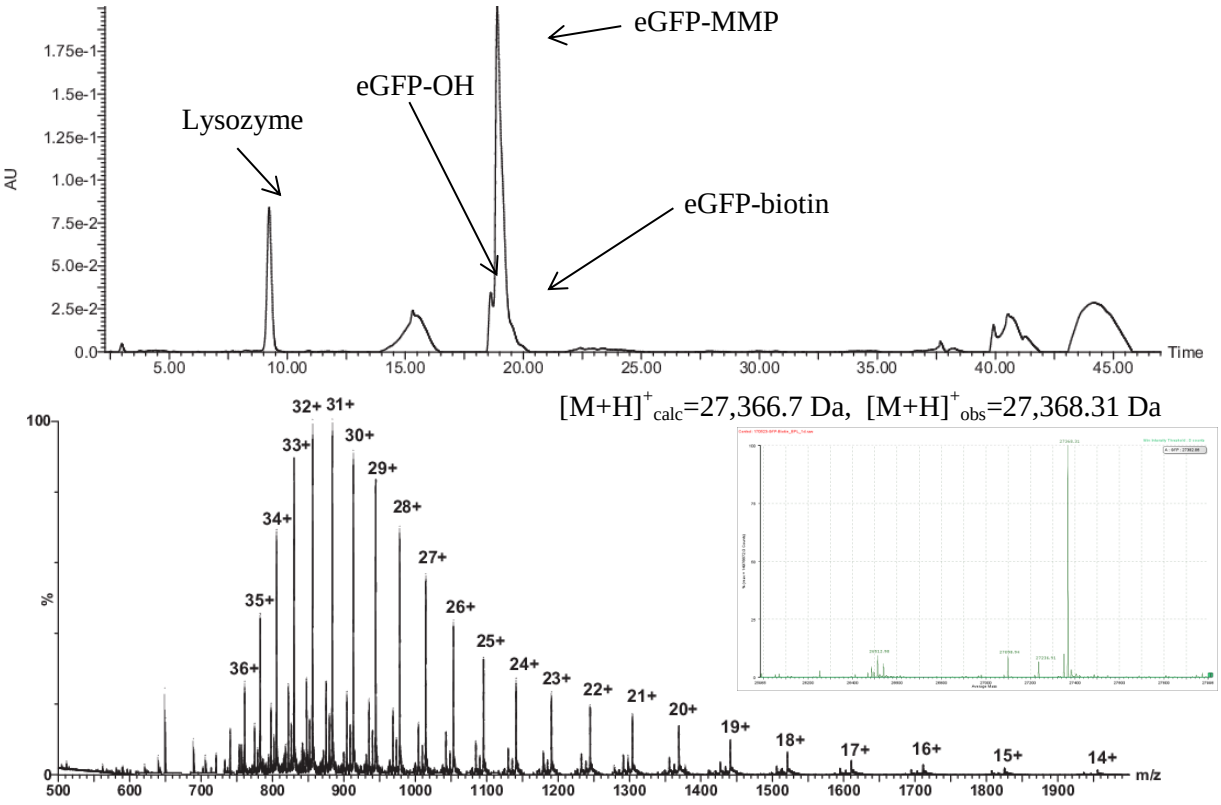


Figure 3.30. Characterization of eGFP-biotin from EPL xx by LC-ESI-MS (after 1 d reaction). a) LC chromatogram (C4, 10 – 70% ACN with 0.1% FA in water with 0.1% FA in 30 min. **b)** Raw and deconvoluted mass spectrum (acquired in ES^+ mode with 1.5 kV capillary voltage), t_R=19.2 min. $[M+H]^+_{calc}$= 27,366.7, obs. $[M+H]^+_{obs}$= 27,368.31. Some hydrolyzed side product GFP-OH is found at t_R=18.6 min. eGFP-MMP-thioester is also observed at t_R=18.9min ($[M+H]^+_{calc}$= 27,097.29, obs. $[M+H]^+_{obs}$= 27,098.94.)

3.5.3 **Kinetic Study of the EPL reaction**

The EPL reaction of eGFP-MMP with Cys-biotin after 24 h incubation delivered only a small portion of the product eGFP-biotin while the starting product thioester was still the main species. Therefore, a kinetic study was performed to gain more insights into the correlation between stability and reactivity of the MMP-thioester. Samples of an EPL reaction were taken at different reaction times and analyzed by HPLC (on a C4 column), Figure 3.31 a. The data from the chromatograms (percentage of peak areas of each compound) is plotted in Figure 3.31 b. Under these conditions, the protein thioester did not hydrolyze and that the ligation reaction proceeds with a linear first order to ca. 47% conversion after seven days. To facilitate trans-thioesterification and faster reaction rates, the more reactive thiol MPAA (1 mM) was added after four days.

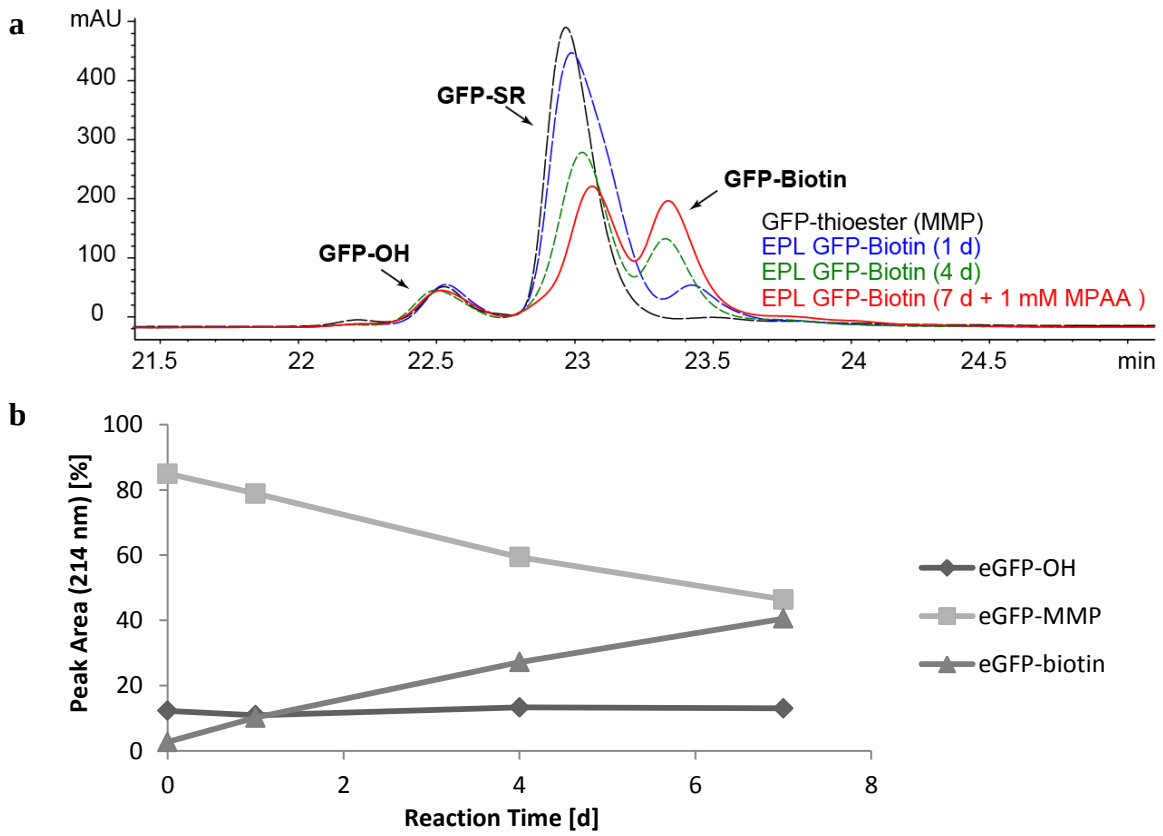


Figure 3.31. Kinetic Study of the EPL reaction of eGFP-MMP with Cys-biotin. Samples of the EPL reaction were taken after 1, 2, 4 and 7 days and analyzed in RP-HPLC (C4), 10 – 70% ACN in water with 0.1% TFA in 30 min. **a)** LC-chromatogram of the different reaction times (zoom to the retention times where the reaction products eluted). At the starting point, some hydrolyzed side product eGFP-OH was already present (t_R=22.5 min), together with the starting material eGFP-MMP (t_R=23 min). Over the time a new peak emerges and increases at t_R=23.5 min: the product eGFP-biotin. **b)** Plot of the peak area in percentage of eGFP-MMP (educt), eGFP-biotin (product) and eGFP-OH (side product). This plot shows that the side product eGFP-OH remains constant over the course of the reaction whereas product formation proceeds with a linear slope until 47% conversion, showing a reaction of first order.

3.5.4 **EPL for Glypiation**

To evaluate the reactivity of a cysteine attached to a phosphoethanolamine as observed in GPIs in a similar environment, the EPL reaction was performed first with Cys-dimannose. The EPL reaction was set up as described before with 50 mM MMP, 0.5 mM TCEP and 2 eq. Cys-dimannose at pH 7.0 and was incubated at room temperature. The reaction mixture was analyzed in LC-ESI-MS, Figure 3.32. The desired product eGFP-dimannose could be detected together with the hydrolyzed side product eGFP-OH. Analysis of the eGFP-thioester showed that significant thioester hydrolysis had already taken place prior to this EPL reaction. This finding lead to the previously described thioester stability studies (Figure 3.25).

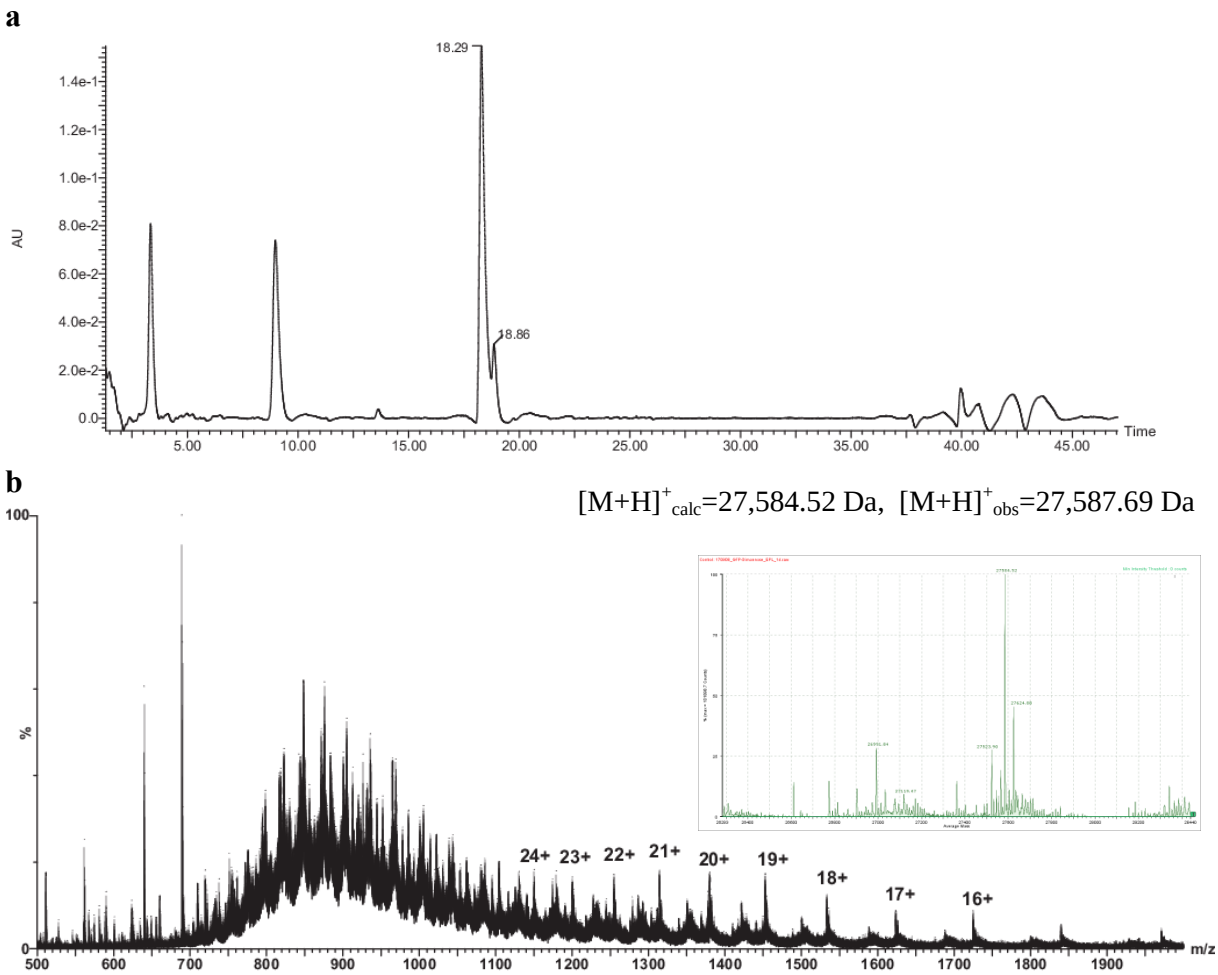


Figure 3.32. Characterization of eGFP-dimannose from EPL. a) LC chromatogram (C4)). **b)** Raw and deconvoluted mass spectra of t_R=18.7 min (acquired in ES^+ mode, 1.5 kV capillary voltage): $[M+H]^+_{obs}$=27,587.69, corresponding to the desired product eGFP-dimannose $[M+H]^+_{calc}$=27,584.52.

The next step was the transfer of the EPL reaction to ligate the Cys-mGPI to eGFP. The reaction was performed using the reaction conditions optimized for Cys-biotin and Cys-dimannose. After one day incubation at room temperature, SDS-PAGE and western blot analysis of the reaction mixture using an in-house generated monoclonal anti-GPI antibody (clone MTG4) showed the formation of the product eGFP-mGPI at the expected size (28.3 kDa, Figure 3.33 a). The sample was also analyzed in MALDI-TOF-MS following desalting using ZipTip (Figure 3.33 b). The size difference found here (1,384 Da) corresponds exactly to mGPI.

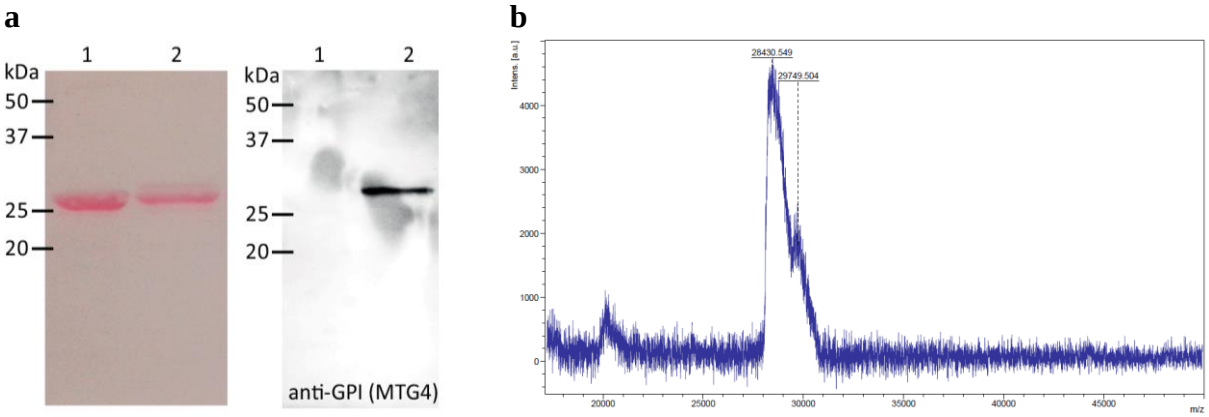


Figure 3.33. EPL reaction of eGFP-MESNA-thioester with Cys-mGPI. a) The product eGFP-mGPI was detected in western blot using an anti-GPI antibody (MTG4) in 1:1,000 dilution in 5% BSA in TBS-T (incubation 1.5 h at room temperature) and a secondary anti-mouse IgG antibody-HRP conjugate in 1:10.000 dilution in 5% BSA in TBS-T for 30 min at room temperature. Lanes: 1 – GFP-MESNA-thioester, 2 – eGFP-mGPI from EPL (1 d). **b) MALDI Spectrum of EPL reaction of eGFP-thioester (MESNA) and mGPI.** It can be seen that only ca. 30 % of the thioester were converted to the product. $[M+H]^+_{calc}$=28,362.1 Da, $[M+H]^+_{obs}$=29,749.504 (old clone, not shown). The size difference of 1,384 Da exactly corresponds to the addition of mGPI to eGFP.

3.5.5 EPL for Glypiation of the Naturally GPI-anchored Thy-1 & Prion Protein

EPL with Thy-1 Protein

After the EPL method was successfully established to ligate the model protein eGFP to biotin, dimannose and a monolipidated GPI-anchor, the next step was the application of these conditions to ligate also more relevant proteins, such as the naturally GPI-anchored protein Thy-1 to natural GPIs.

Although thioester formation with this insolubly expressed mammalian protein was inefficient, Figure 3.26 b, an EPL reaction was tested with Cys-biotin. Cys-biotin (5 eq.) was

added to the freshly formed Thy-1-thioesters of MESNA and MMP. The thiols were present in 100 mM and 50 mM, respectively. MESNA is commonly used in concentrations of 100-200 mM in literature whereas MMP exhibited solubility problems at concentrations above 50 mM. As both thiols are rather unreactive, 1 mM MPAA was added in both cases. After 7 days incubation at room temperature the reactions were analyzed in SDS-PAGE, western blot and LC-ESI-MS. While analysis in mass spectrometry did not reveal the reaction product Thy-1-biotin, it was detected in western blot using anti-biotin antibody-HRP conjugate (Figure 3.34). However, in the blot other unspecific signals were also observed at the sizes of the fusion protein, the cleaved *Mxe* intein as well as in the very top of the gel.

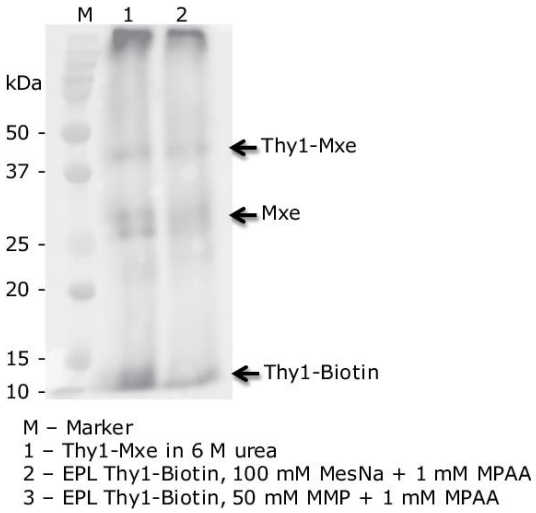


Figure 3.34. Analysis of EPL reaction of Thy-1 thioesters (MESNA or MMP, respectively) with Cys-biotin in western blot using an anti-biotin antibody-HRP conjugate (1:250 dilution in 5% BSA in PBS-T overnight).

As thioester formation was difficult and yields are low, it was decided that the EPL strategy is not well suitable for Thy-1 protein.

EPL with Prion Protein

Another naturally glypiated protein is the infamous Prion Protein (PrP). His-tag purified PrP-*Mxe* fusion protein (~ 51 kDa) in 6 M urea was kindly provided by Prof. Joaquín Castilla from BioGUNE, Bilbao, Spain. Similar to the results with denatured Thy-1 protein, thioester formation of PrP was not possible on-column using chitin resin (not shown). Instead, thioester was formed in solution using 100 mM MMBA and was detected using SDS-PAGE with silver staining and MALDI-TOF-MS, Figure 3.35 a and b. EPL reaction of PrP with Cys-biotin was carried out using the mixture of PrP-MMBA thioester (23 kDa), uncleaved PrP-*Mxe* and cleaved *Mxe* (27 kDa), using 2 mM TCEP and 100 mM MMBA. The ligation product PrP-

biotin was identified in western blot, Figure 3.35 b. To compare the reactivity of the thioester in pure form with the protein mixture, the formed PrP-thioester was purified by HPLC using a C4 column and was directly submitted to EPL with Cys-biotin **xx**. A stronger signal in western blot using anti-biotin antibody-HRP was observed for EPL using MPAA, Figure 3.35 c. The EPL reactions with PrP-thioester and Cys-biotin were furthermore analyzed by HPLC (C4) and MALDI-TOF-MS, Figure 3.36 a and b. By using MPAA as thiol in the reaction, a new peak with earlier t_R was observed in HPLC and MALDI-TOF-MS analysis showed a mass closer to the product PrP-biotin.

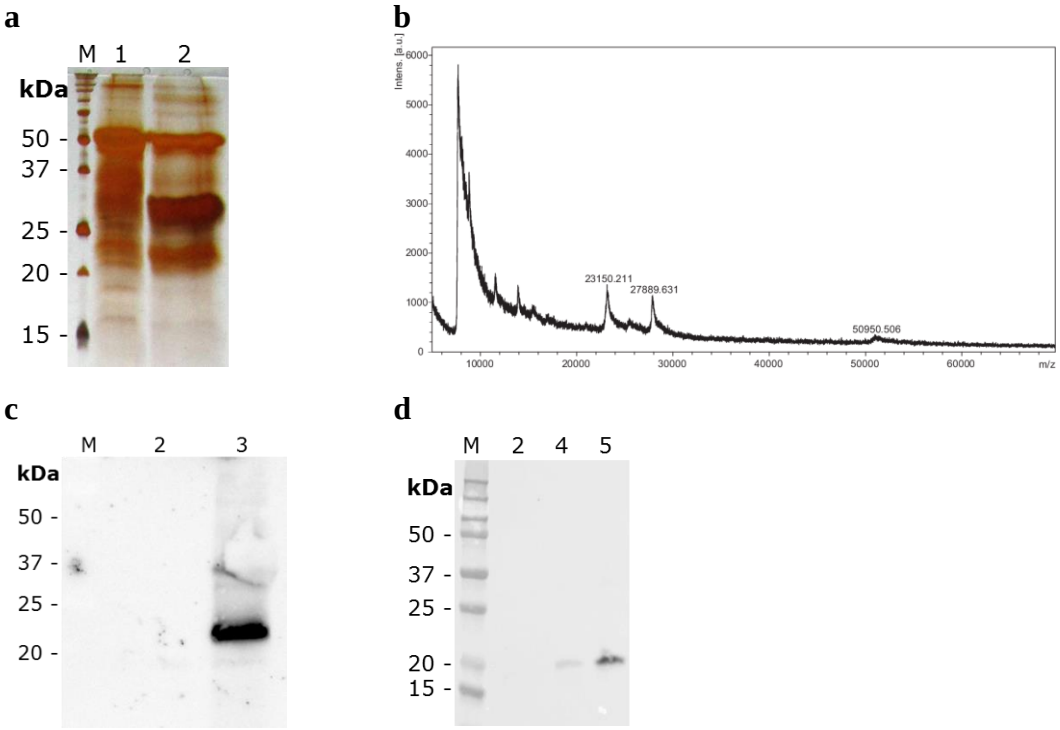


Figure 3.35. Analysis of PrP thioester formation a) analysis in SDS-PAGE with silver staining, b) MALDI-TOF-MS, c, d) EPL to Cys-biotin. PrP-MMBA thioester $[M+H]^+_{obs}$=23,150.211 Da, $[M+H]^+_{calc}$=23,081.81 Da. M – Molecular weight marker, 1 – PrP-*Mxe*, 2 – PrP-MMBA thioester, 3 – EPL PrP-biotin, 4 – EPL with MMBA after C4 purification, 5 – EPL with MPAA after C4 purification. Detection in western blot using anti-biotin antibody-HRP.

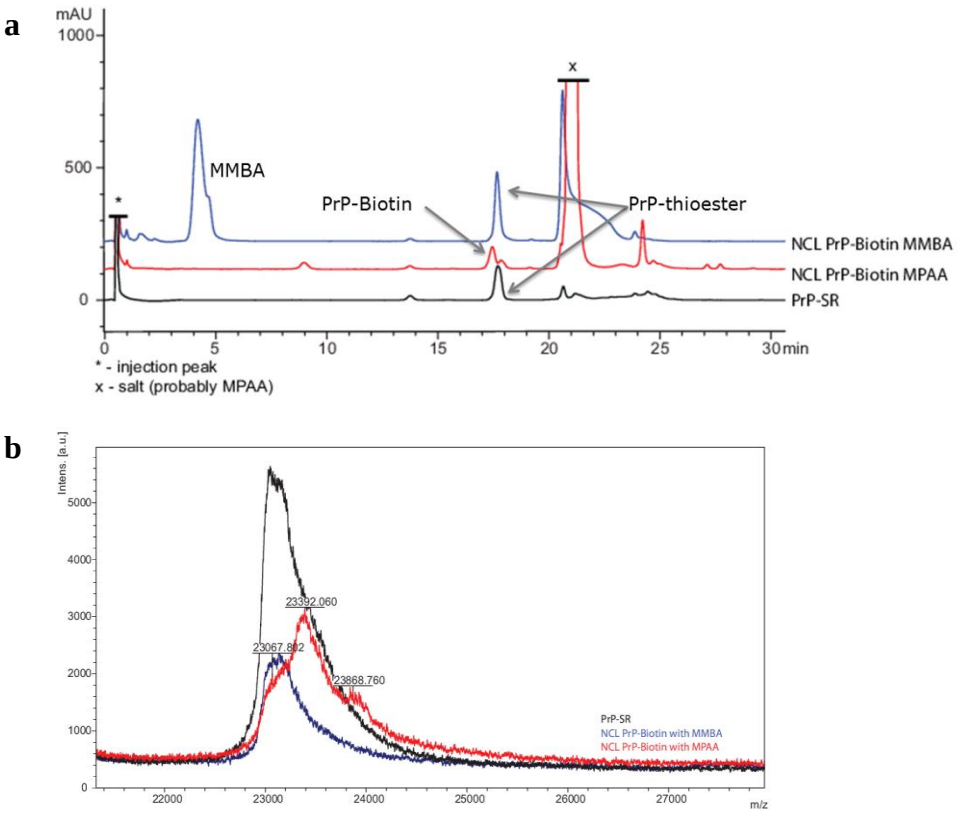


Figure 3.36. EPL of PrP-thioester and biotin with MMBA and MPAA, respectively, analysis in HPLC (C4, gradient 10 – 50 % in 30 min, **a)** and MALDI-TOF-MS (DHAP matrix, pos. ion mode, **b).** $[M+H]^+_{obs}$(with MPAA)=23,392.0 Da, $[M+H]^+_{calc}$(PrP-biotin)=23,334.91 Da.

EPL reactions of PrP-thioester were also performed with Cys-dimannose and Cys-mGPI. The detection of both products was attempted using ConA-HRP in western blot, due to the lack of the anti-GPI antibody at this time (not shown). Unfortunately, only PrP-dimannose was detected. PrP-mGPI was however detected by HPLC (C4) and MALDI-TOF-MS, Figure 3.37 a and b.

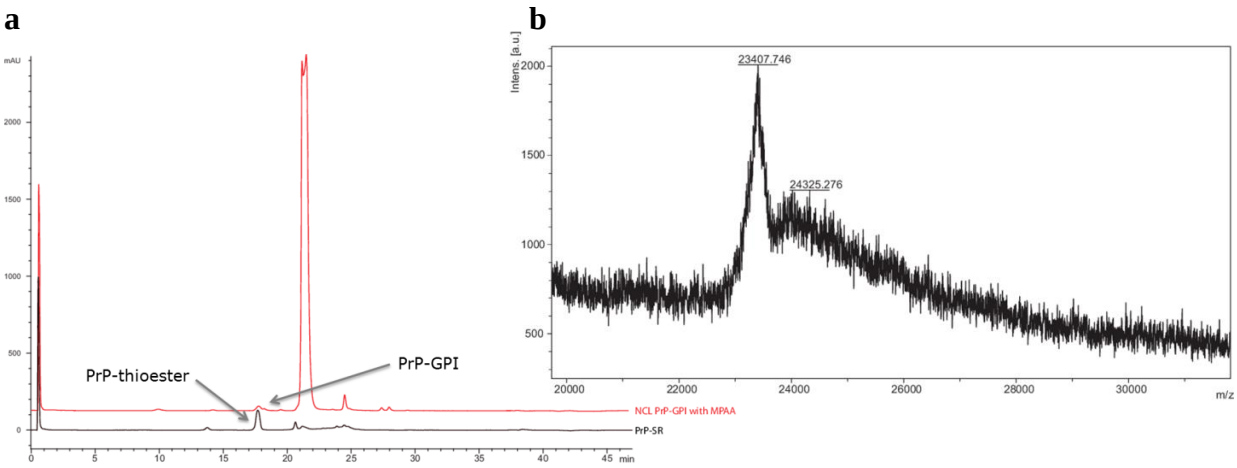


Figure 3.37. EPL of PrP-thioester and mGPI, a) analysis in HPLC (C4, gradient 10 – 50 % in 30 min), **b)** MALDI-TOF-MS spectrum (DHAP matrix, pos. ion mode)**,** $[M+H]^+_{obs}$=24,325.276 Da, $[M+H]^+_{calc}$ (PrP-mGPI)=24,330.30 Da.

3.5.6 **Purification of EPL Products**

EPL reactions with 1 mg eGFP-MMP were carried out with Cys-biotin, Cys-dimannose, Cys-mGPI and Cys-bGPI for three days at RT and purified by SEC using a Superdex 75 column in PBS buffer).

The separation of the product eGFP-biotin from eGFP-MMP thioester and the side product eGFP-OH was not possible on this column (Figure 3.38 a) and they could not be distinguished in SDS-PAGE analysis (not shown). However, analysis of the fractions from different parts of peak #2 from SEC purification in LC-ESI-MS revealed that the shoulder of the peak mainly contains eGFP-OH and eGFP-MMP (not shown) whereas the main peak (later part of the peak) mainly contains eGFP-biotin and a MMP-disulphide of eGFP-biotin, Figure 3.38 c. eGFP-MMP, EPL reaction mixture and purified eGFP-biotin were also analyzed in SDS-PAGE and western blot, Figure 3.38 b.

Similar results were obtained also for the purification of the EPL reaction between eGFP-MMP and Cys-dimannose (not shown). The ligation products of EPL between eGFP and Cys-mGPI and Cys-bGPI were purified in the same way on the Superdex 75 column. However, characterization by LC-ESI-MS was challenging and the anti-GPI antibody was not available at this moment. Fractions expected to contain the products were nevertheless combined and concentrated for CD measurement. Unfortunately, they precipitated during concentration.

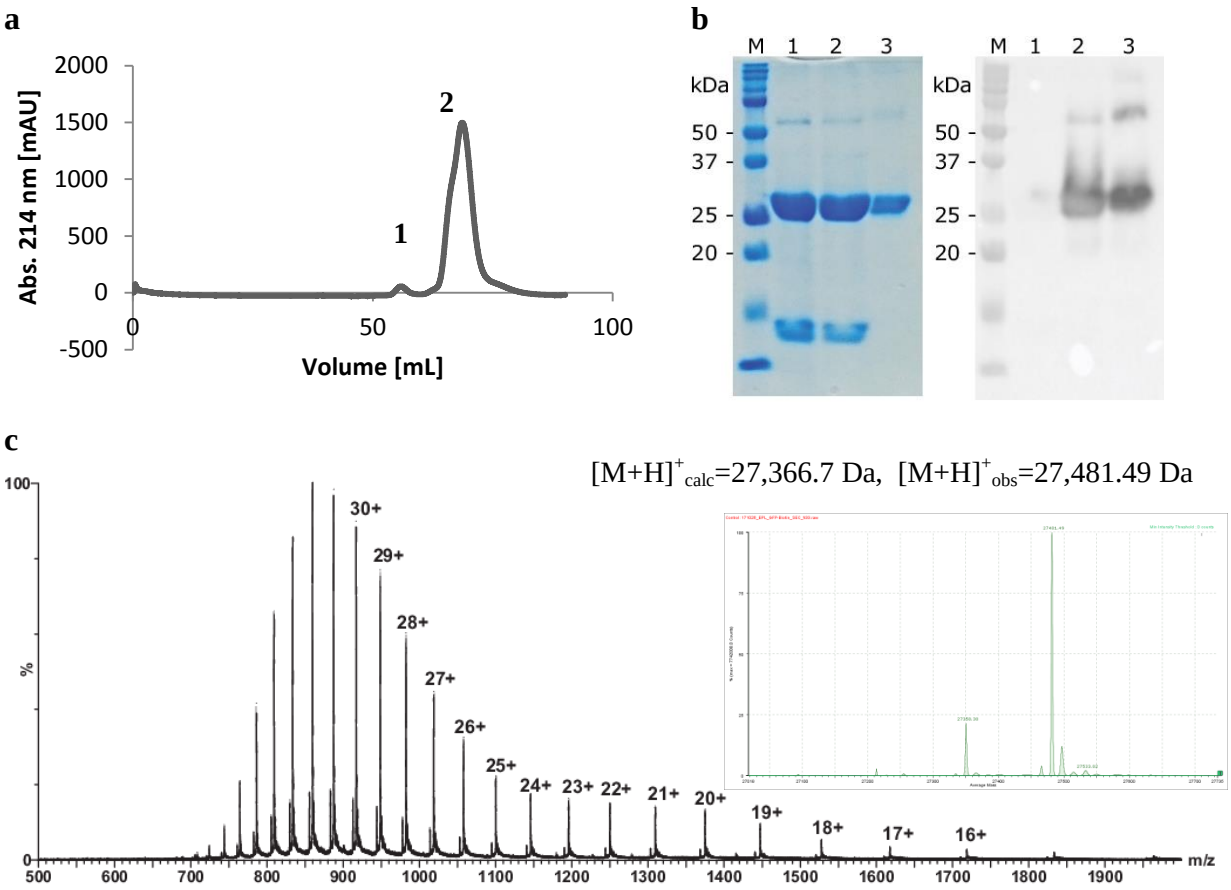


Figure 3.38. SEC purification of eGFP-biotin. a) SEC Chromatogram on a Superdex 75 column of the EPL reaction eGFP-biotin. **b)** SDS-PAGE and western blot analysis of EPL eGFP-biotin as well as purification of the product by SEC**.** M – Molecular weight marker, 1 – GFP-thioester, 2 – EPL reaction GFP-biotin, 3 – eGFP-biotin purified vie SEC. Blot – anti-biotin antibody-HRP (1:250 in 5% BSA-PBST). **c)** Raw and deconvoluted spectra from LC-ESI-MS analysis of SEC fraction #33 (late in peak #2). eGFP-biotin from EPL: $[M+H]^+_{calc}$=27,366.7 Da. The observed mass $[M+H]^+_{obs}$=27,481.49 Da corresponds well to the MMP disulphide of the product. The smaller peak corresponds to eGFP-biotin.

3.6 Protein *Trans*-Splicing

The second strategy evaluated to obtain glypiated proteins was Protein *Trans*-Splicing (PTS) using split inteins. For this strategy two units were needed: fusion proteins of the proteins of interest with the N-terminal fragment of the naturally split intein *Npu*DnaE from *Nostoc punctiforme* and the C-terminal intein fragment, which was chemically synthesized and ligated to the anchor of interest by NCL.

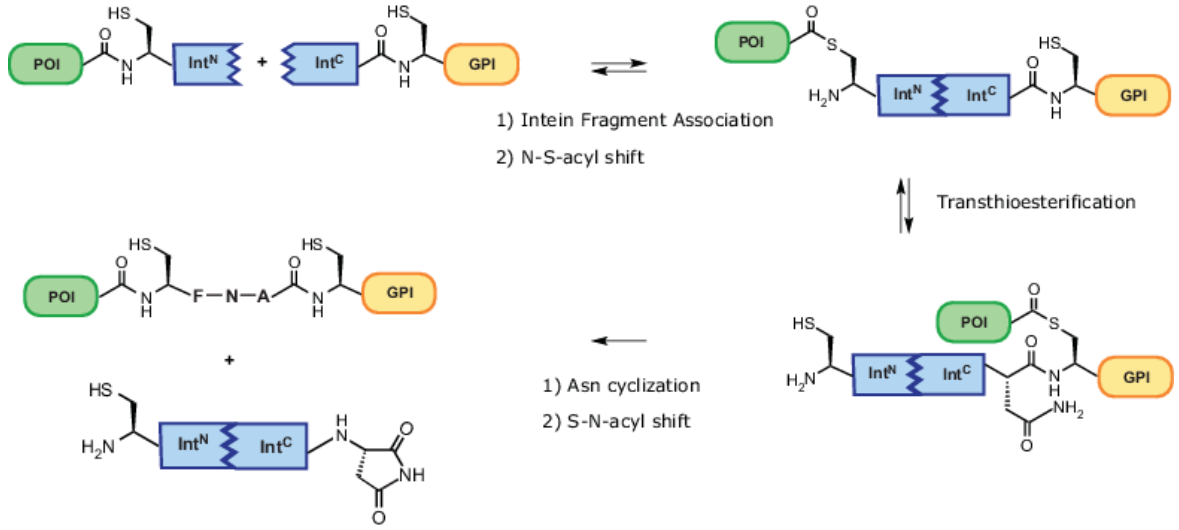


Figure 3.39. Protein *Trans*-Splicing. The protein of interest (POI) is expressed as a fusion protein with the N-terminal fragment of the DnaE intein from *N. punctiforme*. Protein *trans*-splicing is induced upon contact with the C-terminal intein fragment and begins with an N-to-S acyl shift giving a thioester intermediate. This is followed by an intramolecular transthioesterification and asparagine cyclization. The last step is a spontaneous S-to-N acyl shift, resulting in the ligation of both extein fragments via a native peptide bond.

3.6.1 Expression and Purification of Fusion Proteins

Expression and purification of eGFP-*Npu*^{N}

A plasmid containing the sequence of an eGFP-*Npu*^{N} split intein fusion protein (pVS07[145]) was kindly provided by Prof. Henning Mootz from University Münster. The plasmid was transformed into *E. coli* BL21 (DE3) Star cells and expressed in 2 L LB-medium with 100 µg/mL kanamycin with overnight induction (0.4 mM IPTG) at 18°C. Purification was performed via the N-terminal strep-tag II using a Strep-Tactin® resin. As shown in Figure 3.40, this protocol delivered a very good purification of the 40.8 kDa fusion protein. Cell lysis was performed as described for eGFP-*Mxe* using lysozyme and sonication (see chapter 3.5.1). Cleared lysates were loaded very slowly onto StrepTactin® columns equilibrated with buffer

W (pH 7.5, CV = 2 mL). Fractions were analyzed using SDS-PAGE. Fractions containing the fusion protein were combined for further purification using SEC using a Superdex 75 column. The fractions containing the fusion protein concentrated using centricons to deliver 30 mg of pure fusion protein from 2 L culture volume. The protein solution was dialyzed against PTS buffer and stored at max. 2.5 mg/mL at 4°C in order to prevent dimerization.

Figure 3.40 a shows a typical SEC chromatogram of the strep-tag purified fusion protein eGFP-*Npu*^{N} containing two peaks that were analyzed in SDS-PAGE and Native PAGE to get insights into the molecular size and shape of the proteins in the respective fractions (Figure 3.40 b). The comparison of the two peaks from SEC in SDS-PAGE revealed that peak #1 is significantly less clean than peak #2. In Native PAGE most of the protein of peak #1 did not even migrate into the separating gel but remained in the stacking gel, indicating that the proteins are in fact much bigger than expected from their molecular weight, probably due to protein aggregation. Peak #2 showed a main protein band that migrated until ca. the middle of the gel, indicating that this is a properly folded, compact protein. The protein from the first peak was therefore discarded and only the protein eluting in the second peak was kept for PTS reactions. Figure 3.41 shows the complete expression and purification of eGFP-*Npu*^{N} fusion protein.

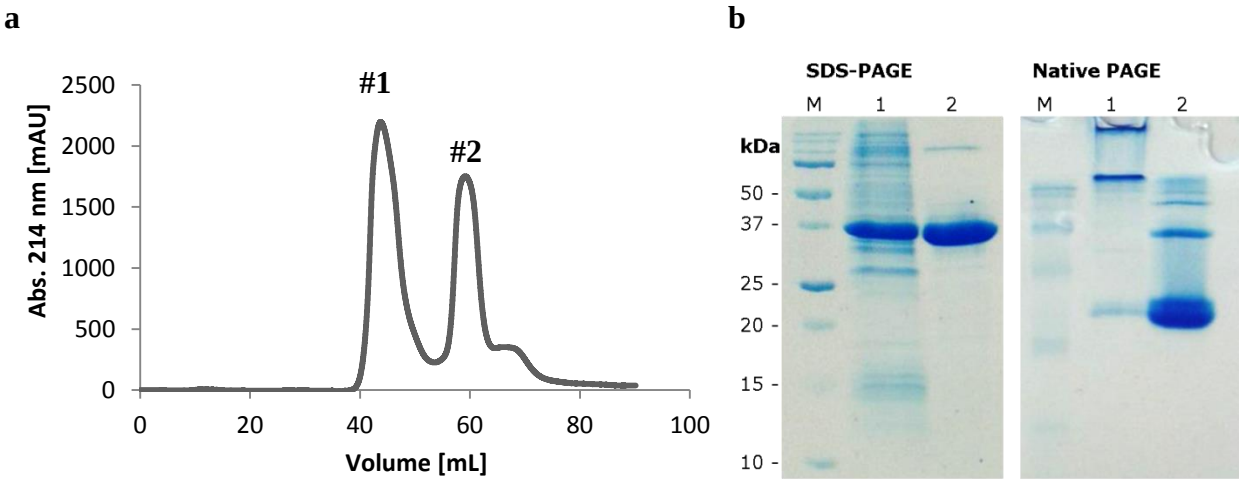


Figure 3.40. SEC purification of GFP-*Npu*^{N}. a) SEC chromatogram from a typical run of strep-tag purified eGFP-*Npu*^{N} fusion protein using Superdex 75 16/600 column on Knauer Azura FPLC. **b)** Analysis of two SEC peaks of eGFP-*Npu*^{N} in SDS-PAGE and Native PAGE. The first peak contains various protein bands, possibly from aggregates while in the second peak, the clean and folded fusion protein is present.

Purified eGFP-*Npu*^{N} was also characterized by HPLC on a C4 column, MALDI-TOF-MS as well as LC-ESI-MS, see Figure 3.41 b and Figure 7.10. The mass observed ($[M+H]^+_{obs}$=40,674.92) corresponds well with the calculated mass of eGFP-*Npu*^{N} lacking the

N-terminal methionine ($[M+H]^+_{calc}$=40,669.16). This is common and well known in literature for recombinant proteins expressed in *E. coli*, especially if alanine is the adjacent residue as is the case here.[170]

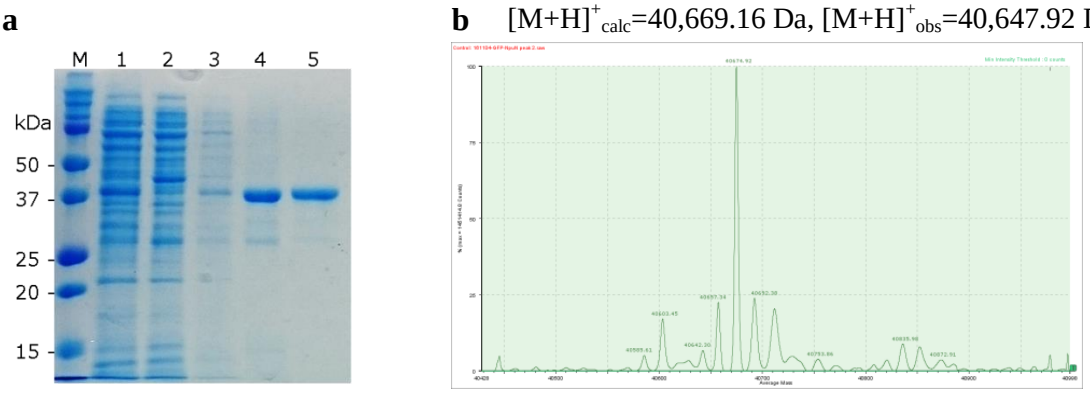


Figure 3.41. Analysis of the expression and complete purification of eGFP-*Npu*^{N}. a) SDS-PAGE: 92 % pure (calculated with Fiji Software). M – Molecular weight marker, 1 – lysate, 2 – flow-through StrepTactin® resin, 3 – wash fraction StrepTactin® resin, 4 – pooled elution from StrepTactin® resin, 5 – SEC-purified GFP-*Npu*^{N} (92% pure, calculated using Fiji software). **b)** Deconvoluted mass spectrum of purified eGFP-*Npu*^{N}.

Expression and purification of Thy-1-*Npu*^{N}

Plasmid pET30a-Thy1-*Npu*^{N} was ordered from GenScript (Piscataway, USA) and transformed into *E. coli* BL21 (DE3) Star and Rosetta cells. Expression was tested with both strains and at different induction conditions (IPTG concentration and temperature, see Figure 7.15). Under the conditions tested the fusion protein was expressed insolubly (see Figure 7.16). However, expression using the Rosetta clone with induction at 30°C and 0.5 mM IPTG for 24 h was found to give the highest expression.

For preparative protein expression, 1 L LB medium were inoculated to OD_{600} = 0.05, grown to late exponential phase (OD_{600} of 0.6 – 0.8) at 37°C, 220 rpm. Protein expression was induced by addition of 0.5 mM IPTG at 30°C, 180 rpm for 24 h. Cells were lysed after resuspension in 25 mL lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM Imidazole) + 0.25 mg/mL lysozyme + protease inhibitors and treated as described before, repeating sonication and centrifugation once. The supernatant was discarded and the inclusion body pellet was solubilized in 20 mL HisTrap equilibration buffer + 8 M urea by vortexing, ultrasonic bath and incubation for 2 h at room temperature (120 rpm). The solubilized inclusion bodies were centrifuged for 1,5 h (24,400 x g, 4°C) and filtered through a 0.2 µm syringe filter prior his-tag purification under denaturing conditions (8 M urea). For this, 8 mL lysate per run were loaded onto a HisTrap FF (5 mL) column equilibrated in HisTrap

equilibration buffer + 8 M urea. The column was washed with 10 CV equilibration buffer + 8 M urea following elution of unbound host proteins, and the fusion protein was eluted using a gradient from 30 mM to 1 M imidazole over 5 CV. The collected fractions were analyzed in SDS-PAGE (see Figure 3.42) and fractions containing the fusion protein were combined and dialyzed into PTS buffer + 6 M urea to give 40 mg fusion protein from 1 L culture volume.

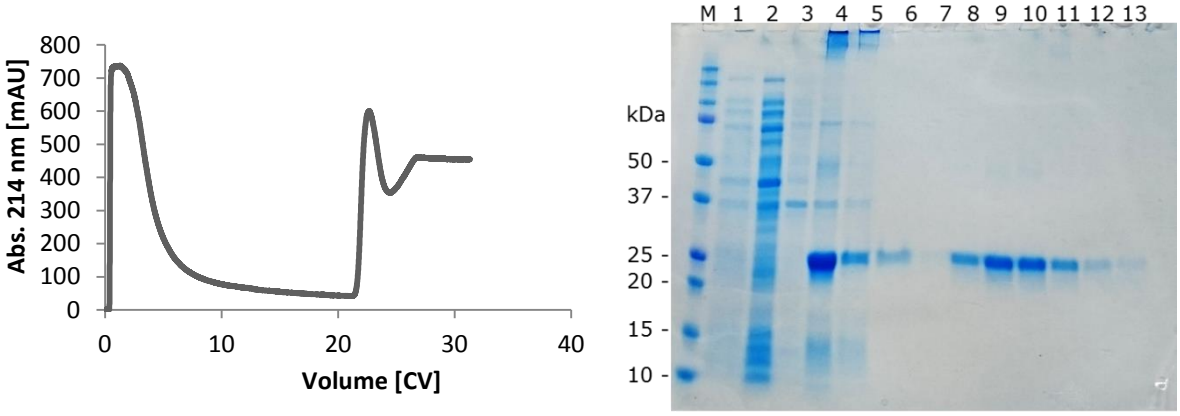


Figure 3.42. Affinity purification of Thy-1-*Npu*^{N} via His-Tag. a) Chromatogram of the purification on a HisTrap column (FF crude, 5 mL), the fusion protein was eluted by a gradient from 30 mM to 1 M imidazole over 5 CV. **b)** Analysis of the HisTrap purification in SDS-PAGE. M- Molecular weight marker, 1 – pre induction, 2 – lysate 1, 3 – lysate 2, 4 – solubilized inclusion bodies (load) 5 – flow-through HisTrap column, 6 – wash fraction, 7 – 13 – elution fractions.

The purified Thy-1-*Npu*^{N} fusion protein was characterized by MALDI-TOF-MS and LC-ESI-MS (Figure 7.17). The mass observed ($[M+H]^+_{obs}$=25,356.21 m/z) corresponds well with the calculated mass ($[M+H]^+_{calc}$=25,226.5 Da), this time with the N-terminal methionine in place, which is also in agreement with literature as here the adjacent residue is a glutamine that inhibits the aminopeptidase from cleaving the methionine.[170]

Thy-1-*Npu*^{N} was also analyzed using tryptic digestion as described in 2.8.7. The coverage of the protein sequence is given in Figure 3.43. Using this protocol a big part of the protein sequence was verified.

MQKVTSLTAC LVDQSLRLDC RHENTSSSPI QYEFSLTRET KKHVLFGTVG VPEHTYRSRT NFTSKYNMKV LYLSAFTSKD EGTYTCALHH SGHSPPISSQ NVTVLRDKLV K-CLSYETEIL TVEYGLLPIG KIVEKRIECT VYSVDNNGNI YTQPVAQWHD RGEQEVFEYC LEDGSLIRAT KDHKFMTVDG QMLPIDEIFE RELDLMRVDN LPNHHHHHH

Figure 3.43. Analysis of Thy-1-*Npu*^{N} fusion protein by tryptic digestion. Underlined sections were identified in MALDI-TOF-MS.

Expression and purification of IL-2-*Npu*^{N}

The plasmid pET30a-IL-2-*Npu*^{N} was prepared by subcloning (3.2.1) and was transformed into BL21 (DE3) Star cells for expression tests. Under all conditions tested the 31.5 kDa protein was expressed only in insoluble form (expression at 37, 25 and 12°C, induction with 1 and 0.4 mM IPTG), Figure 3.44. Since the expression did not deliver soluble protein with any induction temperature, protein production was accomplished at 37°C for 5 h with 0.4 mM IPTG.

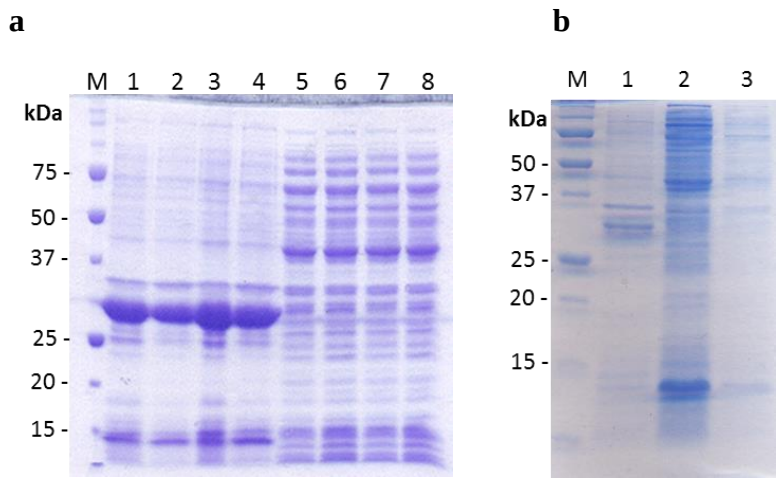


Figure 3.44. Expression test of IL-2-*Npu*^{N} fusion protein. M –Molecular weight marker, 1 – insoluble fraction, 37°C, 1 mM IPTG, 2 – insoluble fraction, 25°C, 1 mM IPTG, 3 – insoluble fraction, 37°C, 0.4 mM IPTG, 4 – insoluble fraction, 25°C, 0.4 mM IPTG, 5 – soluble fraction, 37°C, 1 mM IPTG, 6 – soluble fraction, 25°C, 1 mM IPTG, 7 – soluble fraction, 37°C, 0.4 mM IPTG, 8 – soluble fraction, 25°C, 0.4 mM IPTG.

The inclusion bodies were solubilized using 8 M urea as described for Thy-1 protein. Refolding of the denatured protein was attempted by dialysis in buffer W, dialysis in buffer W + 0.2 M L-arginine as additive (not shown), rapid dilution in buffer W and slow dilution in buffer W. These refolding experiments showed high precipitation during dialysis. In dilution, most of the protein remained soluble, Figure 3.45 a. Therefore, the supernatant from rapid dilution was slowly loaded to StrepTactin® resin for purification via strep-tag. Under these conditions the protein was not separated and coeluted with host proteins, possibly chaperones or similar, Figure 3.45 b.

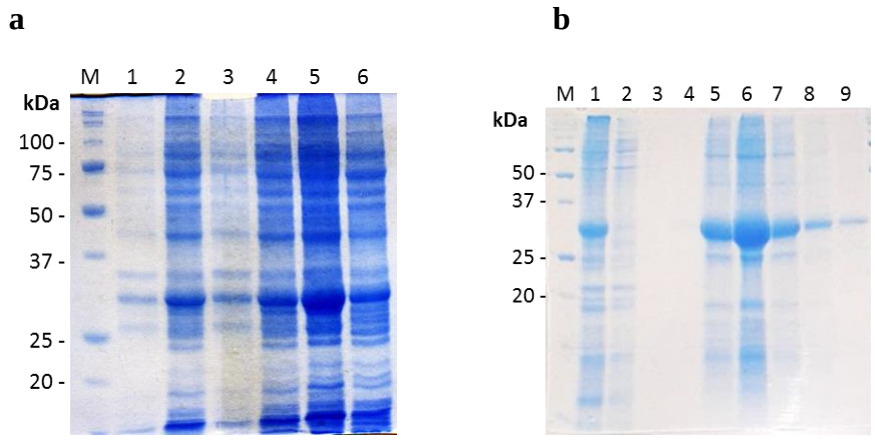


Figure 3.45. Refolding Optimization and Strep-tag purification of IL-2-*Npu*^{N} fusion protein. a) M – Molecular weight marker, 1 – rapid dilution (pellet), 2 – rapid dilution (supernatant), 3 – slow dilution (pellet), 4 – slow dilution (supernatant), 5 – dialysis (pellet), 6 – dialysis (supernatant). **b)** M – Molecular weight marker, 1 – supernatant from rapid dilution, 2 – flow-through StrepTactin® resin, 3 – wash fraction, 4 – 9 – elution fractions.

To overcome the limitation in the purification, a different approach was investigated for IL-2-*Npu*^{N} fusion protein: the inclusion bodies were prepared and reduced using a multi-step washing procedure adapted from Piontek (2007).[146] By this method, protein purity could be improved from 43% to 51% and the insoluble fusion protein was solubilized in 6 M GdmCl, Figure 3.46. Purification by SEC and HPLC (on a C4 column) was also tested for this protein, but these methods did not result in improved purity (not shown). Therefore, the protein was used for PTS experiments as it was after the washing procedure.

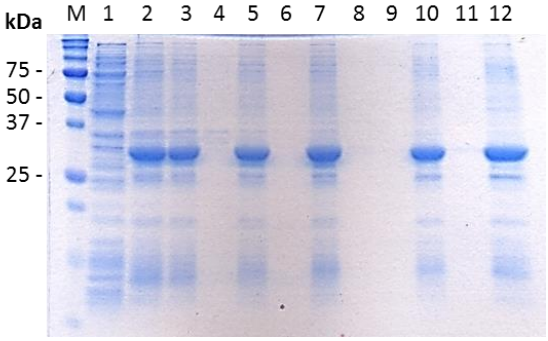


Figure 3.46. Inclusion Body washing procedure. M – Molecular weight marker, 2 – supernatant lysis, 2 – pellet lysozyme wash, 3 – pellet deoxycholate wash, 4 – supernatant deoxycholate wash, 5 – pellet deoxycholate removal, 6 – supernatant deoxycholate removal, 7 – pellet wash with water, 8 – supernatant wash with water, 9 – supernatant reduction with 100 mM DTT, 10 – pellet DTT removal, 11 – supernatant DTT removal, 12 – denatured protein sample in 6 M GdmCl.

3.6.2 PTS of eGFP-*Npu*^{N} with *Npu*^{C} hydrazide and *Npu*^{C}-biotin

In a first attempt, the protein *trans*-splicing reaction was tested using only the eGFP-*Npu*^{N} fusion protein and the *Npu*^{C} peptide hydrazide with 5 mM TCEP under the conditions specified in 2.8.5.[110] Incubation at room temperature and 37°C resulted in 60% and 69% splicing, respectively (see Figure 3.47). Additionally to the temperature effect the PTS reactions between eGFP-*Npu*^{N} and *Npu*^{C} peptide hydrazide was investigated at different concentrations of TCEP (1, 5 and 10 mM TCEP) and different ratios of peptide to protein (1.1, 1.5 and 2.3 eq. of peptide). The highest splicing yields were obtained using 1 mM TCEP, 1.5 eq. of peptide and incubating overnight at 37°C, Figure 7.11 and Figure 7.12.

Next, the PTS reaction between eGFP-*Npu*^{N} fusion protein and *Npu*^{C}-biotin was investigated using the optimized conditions. After 3 days, only 50% conversion was achieved, however the product of the PTS reaction was purified via strep-tag (CV=0.5 mL) to evaluate the success of the ligation. Since eGFP-*Npu*^{N} carries an N-terminal strep-tag, this purification step separates the cleaved intein fragments from the product while unreacted fusion protein co-elutes with the product, Figure 3.47 b. This step should facilitate the characterization of the product as fewer contaminants are present. An analysis of the purified sample by RP-HPLC is given in Figure 7.13. It contains eGFP-biotin as the main product, but another peak is present, most likely corresponding to cleaved *Npu*^{N}.

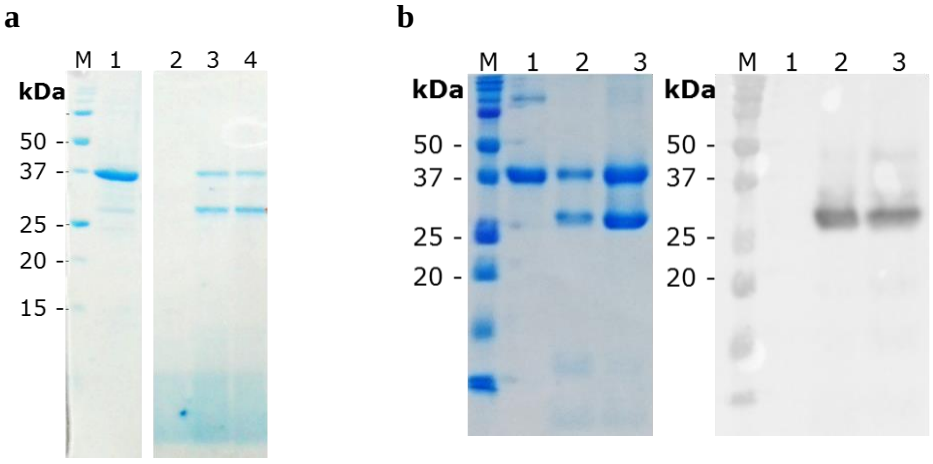


Figure 3.47. a) SDS-PAGE analysis of a first PTS reaction of eGFP-*Npu*^{N} with *Npu*^{C} hydrazide, 5 mM TCEP. M – Molecular weight marker, 1– GFP-*Npu*^{N}, 2 – *Npu*^{C} hydrazide, 3 – PTS 16 h, room temperature, 4 – PTS 16 h, 37°C. **b) PTS reaction of eGFP-*Npu*^{N} with *Npu*^{C}-biotin and purification of eGFP-biotin by SEC.** eGFP-biotin was detected in western blot by anti-biotin antibody (1:250 in 5% BSA in PBS-T, o/n). M – Molecular weight marker, 1 – GFP-*Npu*^{N}, 2 – PTS reaction GFP-biotin 3 d, 3 – GFP-biotin purified via strep-tag.

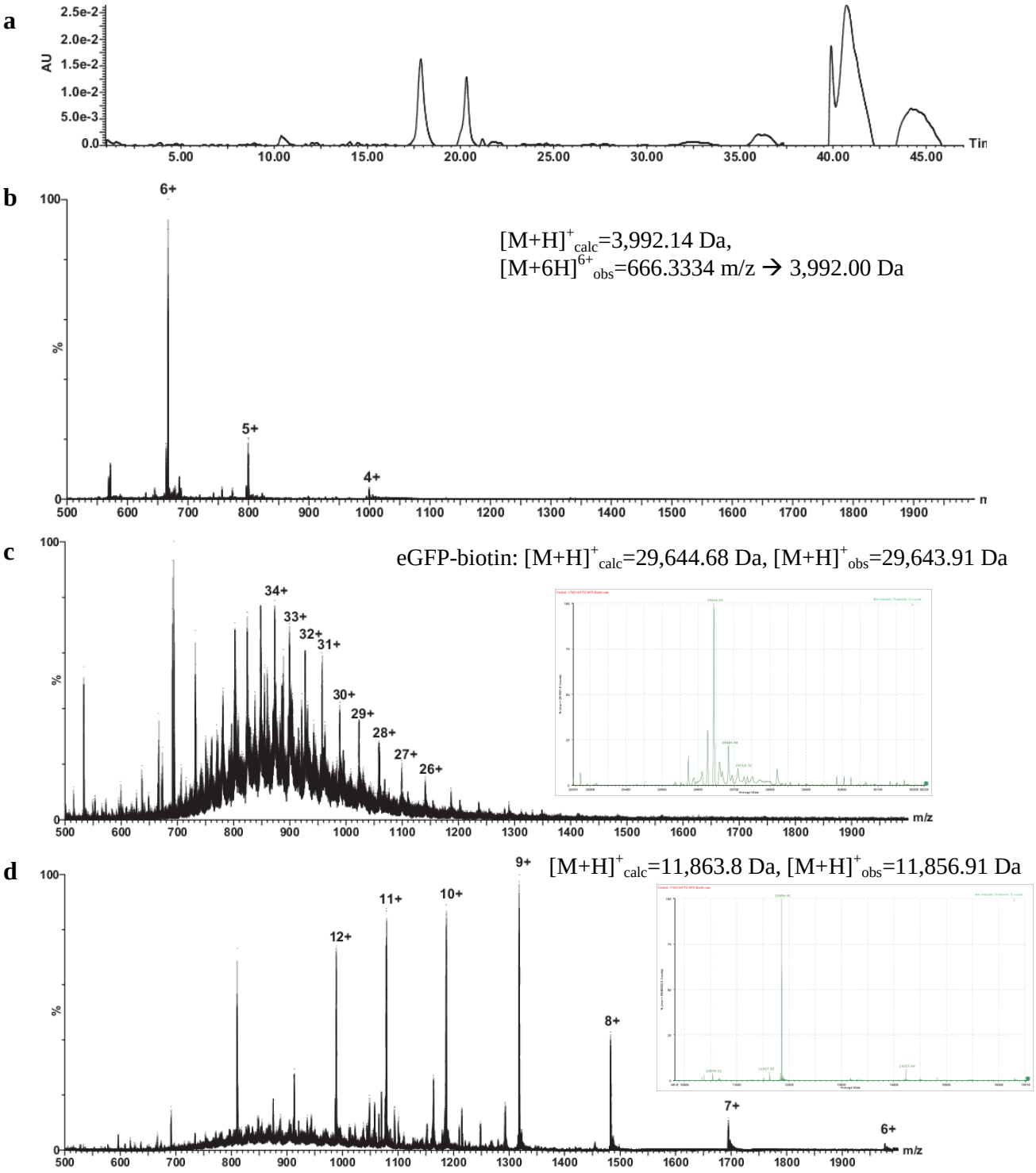


Figure 3.48. Characterization of eGFP-biotin by LC-ESI-MS. Raw and deconvoluted mass spectra are shown: t_R=10.4 min corresponds to the *Npu*^{C} peptide after PTS, i.e. lacking the residues after Cys^{36} ($[M+H]^+_{calc}$=3,992.14 Da, $[M+6H]^{6+}_{obs}$=666.3334 m/z → 3,992.00 Da). At t_R=18 min, the product eGFP-biotin was detected: $[M+H]^+_{calc}$=29,644.68 Da, $[M+H]^+_{obs}$=29,643.91 Da. At t_R=20.4 min, cleaved *Npu*^{N} fragment was detected, $[M+H]^+_{obs}$= 11,856.91Da, $[M+H]^+_{calc}$=11,862.8 Da.

The product eGFP-biotin was detected in LC-ESI-MS with an observed mass of $[M+H]^+_{obs}$=29,643.91 Da. This corresponded with the calculated mass of the desired product $[M+H]^+_{calc}$=29,644.68 Da (see Figure 3.48 for raw and deconvoluted mass spectra) and is the main product according to the chromatogram (t_R=18.1 min). The chromatogram also contains a peak at t_R=10.4 min, corresponding to the *Npu*^{C} peptide after PTS, i.e. lacking the residues after Cys^{36} ($[M+H]^+_{calc}$=3,992.14 Da, $[M+6H]^{6+}_{obs}$=666.3334 m/z → 3,992.00 Da). At t_R=20.4 min, cleaved *Npu*^{N} fragment was detected ($[M+H]^+_{obs}$=11,856.91Da, $[M+H]^+_{calc}$=11,862.8 Da).

3.6.3 **Kinetic Studies of the PTS Reaction**

The *Npu*DnaE split intein is the fastest known split intein.[145] Although these observations always involved reactions with peptides and proteins as the C-extein, it was anticipated that the reactions in this work including a GPI-anchor as extein will also be very fast. To investigate this, a kinetic study with eGFP-*Npu*^{N} and *Npu*^{C}-biotin was performed similarly as described for the EPL reaction, 3.5.3. Reaction samples were starting after a shorter time (5 min, 10 min, 30 min, 1 h, 2 h, 4 h, 6 h, 24 h and after five days) due to the expected faster reaction rates. The samples were analyzed in SDS-PAGE, western blot and HPLC, Figure 3.49. After five minutes, a weak band was already visible in the blot that increases in intensity over the following hours, but not any more after several days (detected using anti-biotin antibody-HRP). However, from the SDS-PAGE the yield still seems to increase over the time, Figure 3.49 a. A high reaction rate at the beginning that stopped after the first day can also be seen in the plot of percentage of conversion over time (Figure 3.49 b). Analysis in HPLC confirmed these findings, Figure 3.49 c.

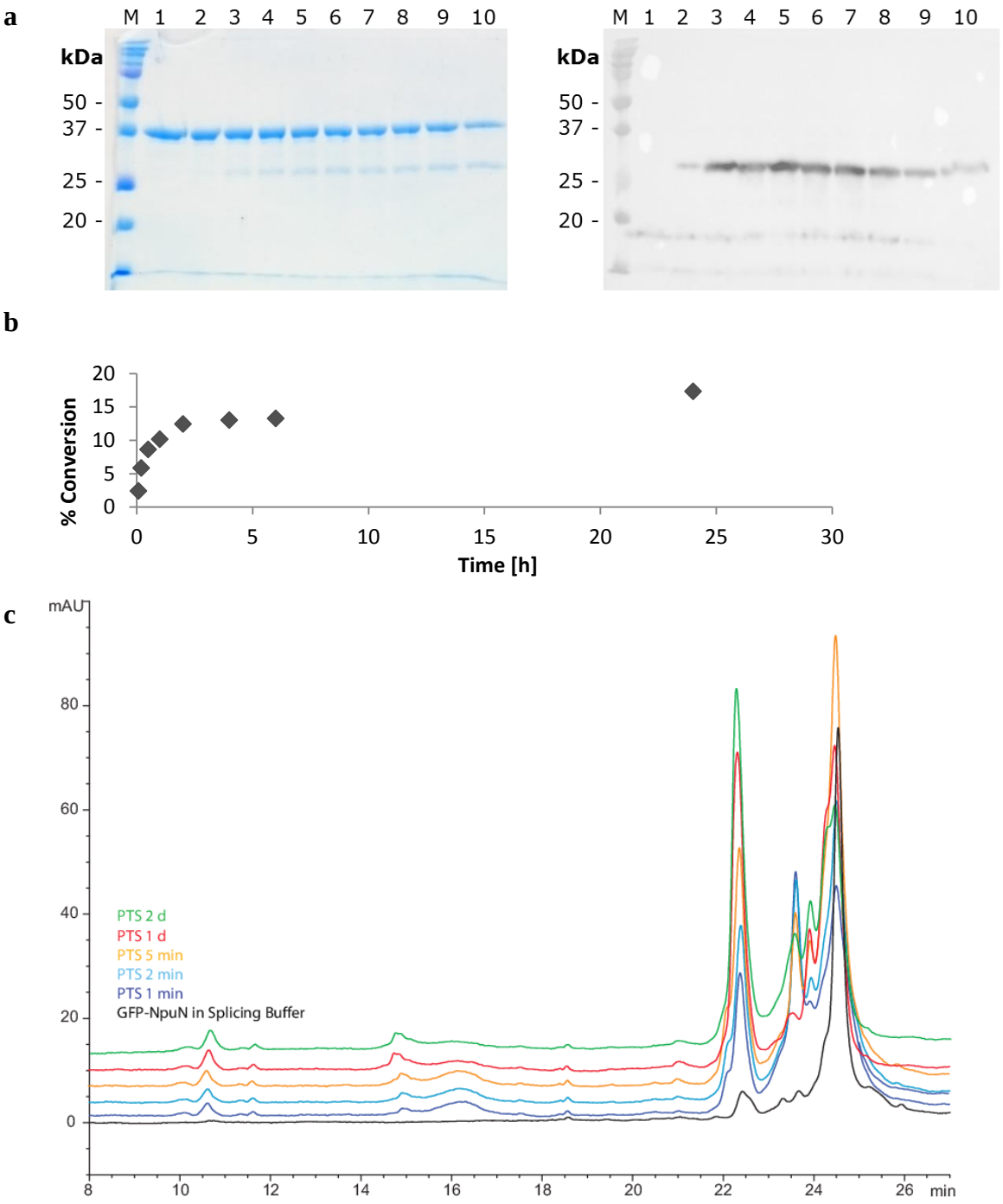


Figure 3.49. Kinetic Study of PTS reaction with eGFP-*Npu*^{N} and *Npu*^{C}-biotin. a) Analysis in SDS-PAGE and western blot. M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – PTS reaction eGFP-biotin 5 min, 3 – 10 min, 4 – 30 min, 5 – 1 h, 6 – 2 h, 7 – 4 h, 8 – 6 h, 9 – 24 h, 10 – 5 d. **b)** Conversion (in percent, calculated using ImageJ software) plotted over the time. **c)** Analysis in HPLC (C4). The fusion protein eGFP-*Npu*^{N} elutes at t_R=25 min. This peak varies in shape over the course of the study as cleaved *Npu*^{N} has a very similar retention time. A new peak emerges and increases at t_R=22.5 min. It corresponds to the product eGFP-biotin, as was confirmed by LC-MS analysis. The flat peak at t_R~16 min corresponds to reacted and unreacted *Npu*^{C}-biotin and the increasing peak just before it (t_R=15 min) corresponds to *Npu*^{C} after PTS (cleaved extein).

3.6.4 PTS of eGFP-*Npu*^{N} with *Npu*^{C}-mGPI

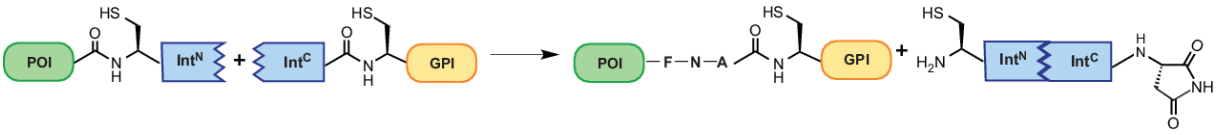


Figure 3.50. Scheme of PTS between a POI and Int^C-GPI. The splicing product will carry some extra amino acids at the splice junction.

Based on the good results obtained in the PTS reaction for the biotinylation of eGFP which delivered the desired product without any hydrolysis of the proteins, the next step was a PTS between eGFP-*Npu*^{N} and *Npu*^{C}-mGPI. The reaction was set up using the conditions used for PTS with *Npu*^{C}-biotin. PTS was tested with 1 mM and 75 µM (5 eq.) TCEP. Better results were obtained using 1 mM TCEP (not shown), when *Npu*^{C}-mGPI was used in 1.5-fold excess. To evaluate the progress of the reaction, samples were also taken at frequent intervals. The product eGFP-mGPI was detected in western blot using an anti-GPI antibody (MTG4) already after 10 minutes reaction at 37°C, Figure 3.51. However, even with prolonged reaction the signal in western blot did not increase in intensity, indicating a similar behavior as described above. The PTS reaction with *Npu*^{C}-mGPI is very fast in the beginning and slows down already after a few hours.

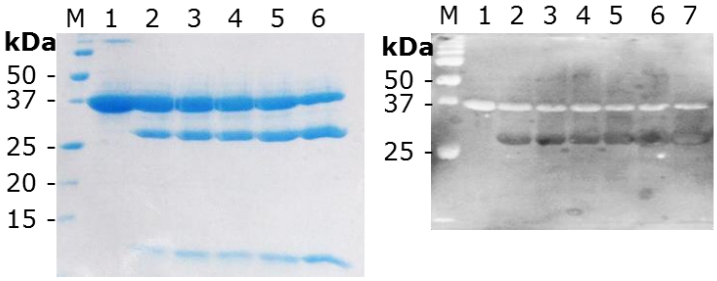


Figure 3.51. PTS reaction between eGFP-*Npu*^{N} and *Npu*^{C}-mGPI. Analysis in SDS-PAGE and western blot. The product eGFP-mGPI was detected using an anti-GPI antibody (MTG4). M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – PTS 10 min, 3 – PTS 30 min, 4 – PTS 1 h, 5 – PTS 2 h, 6 – PTS 8 h, 7 – PTS 24 h.

The product of the PTS reaction to obtain eGFP-mGPI was purified after one day reaction using SEC on a Superdex 75 column. SDS-PAGE analysis and the SEC chromatogram showed a low yield of the reaction, Figure 3.52 a and b. Nevertheless, the fractions collected in SEC purification were analyzed in SDS-PAGE with silver staining and by analytical SEC using Yarra SEC-2000 column, Figure 3.52 b and c. Both methods indicated that the product eGFP-mGPI was present in fraction 4. The purified eGFP-mGPI, the fusion protein eGFP-

Npu^N and the reaction mixture were analyzed in western blot using anti-GPI antibody (MTG4), Figure 3.52 e. The product was detected at the expected size (ca. 30 kDa) in the lanes of the PTS reaction mixture and fraction 4 of the SEC purification.

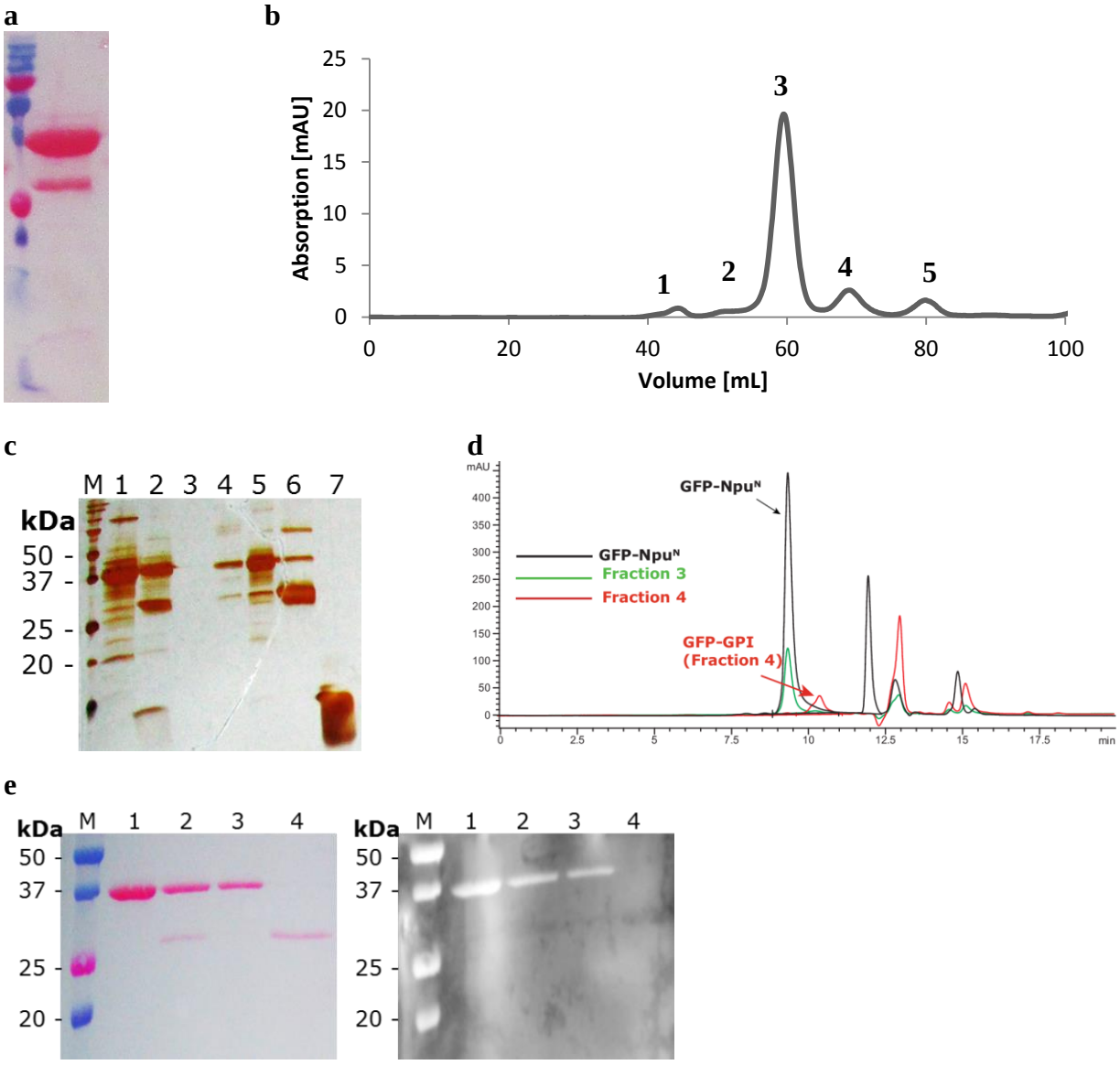


Figure 3.52. Preparative PTS reaction between eGFP-*Npu*^{N} and *Npu*^{C}-mGPI. a) Analysis in SDS-PAGE showed that the conversion was not very good. **b)** Purification of the PTS reaction in SEC (Superdex75). The collected fractions are numbered. **c)** Analysis of the fractions in SDS-PAGE with silver staining (M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – PTS o/n, 3 – SEC fraction 1, 4 – SEC fraction 2, 5 – SEC fraction 3, 6 – SEC fraction 4, 7 – SEC fraction 5) and **d)** analytical SEC (Yarra Bio-SEC 2000). **e)** Western blot analysis of purified eGFP-mGPI using anti-GPI antibody (MTG4). M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – PTS reaction prior to purification, 3 – fraction 3 – containing eGFP-*Npu*^{N}, 4 – fraction 4 – containing eGFP-mGPI. A faint signal is observed at the expected size (~30 kDa) in the lanes of the PTS reaction as well as fraction 4.

Unfortunately, LC-ESI-MS analysis of eGFP-mGPI was not successful, although the cleaved *Npu*^{C} fragment (mass 3,992.14 Da) was observed, indicating that some PTS reaction had occurred.

3.6.5 Application of PTS for the Semi-Synthesis of Naturally Glypiated Proteins

PTS with Thy-1 Protein

PTS was also applied to generate C-terminally modified Thy-1 protein. PTS betweenThy-1-*Npu*^{N} and *Npu*^{C}-biotin was performed at room temperature and 37°C using 0.5 mM TCEP. Thy-1-*Npu*^{N} fusion protein was purified by HPLC (C4) after his-tag purification. Reaction samples were taken after one day and after five days incubation and analyzed in SDS-PAGE and western blot, see Figure 3.53. All reaction partners were found in the gel, i.e. Thy-1-*Npu*^{N} fusion protein (25.2 kDa), *Npu*^{C}-biotin (~5 kDa), cleaved *Npu*^{N}-$_{His6}$ (12.6 kDa) and the product Thy-1-biotin (12.9 kDa). However, Thy-1-biotin and *Npu*^{N} are not well separated due to their small MW difference and they appeared as a double band. In western blot, the biotin moiety was detected using an anti-biotin antibody-HRP. Signals were obtained for Thy-1-biotin and *Npu*^{C}-biotin at the expected positions in the blot; however they are not very strong. Conversion was calculated to be 20 and 23% at room temperature and 37°C after one day, and 28 and 23% after five days. Therefore, incubation at room temperature was selected for further experiments as it furthermore decreases the risk of proteolysis or other protein degradation during the incubation compared to incubation at 37°C.

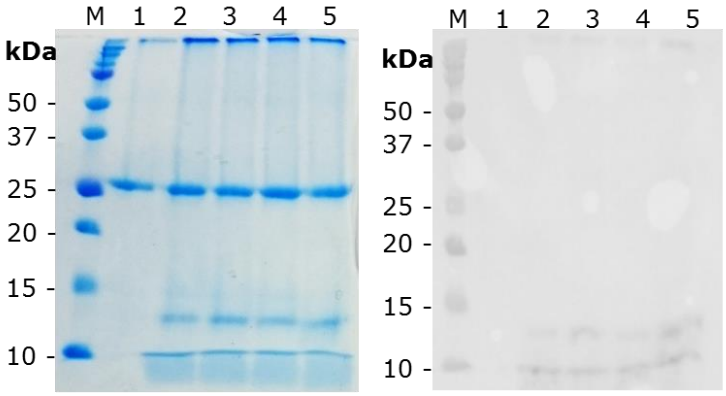


Figure 3.53. Optimization of incubation temperature for PTS with Thy-1-*Npu*^{N} with *Npu*^{C}-biotin. M – Molecular weight marker, 1 – Thy-1-*Npu*^{N}, 2 – PTS Thy-1-biotin RT, 1 d, 3 – PTS RT, 5 d, 4 – PTS 37°C, 1 d, 5 – PTS 37°C, 5 d. The fusion protein is found at the expected position (ca. 25 kDa) as is unreacted *Npu*^{C}-biotin (below 10 kDa). Cleaved *Npu*^{N} and the product are very similar in size (12.6 kDa and 12.9 kDa, respectively), so a separation in SDS-PAGE is not expected. The biotin moiety was detected using an anti-biotin antibody-HRP (1:250 dilution in 5% BSA-PBST).

Due to the low yields of the PTS reaction, characterization of Thy-1-biotin was not possible by LC-ESI-MS. The purification of the Thy-1 fusion protein by HPLC using acetonitrile and

TFA as well as lyophilization might impact the folding of the protein. Although Thy-1-*Npu*^{N} was already present in denatured form, this purification could have an effect in the reaction and hamper intein activity. Therefore, PTS reaction was also tested using Thy-1-*Npu*^{N} fusion protein which was only purified via his-tag and kept in solution all the time prior to PTS. However, under these changes the reaction yield was very low and after one week the product was not visible in SDS-PAGE but was only detected in western blot using anti-biotin antibody-HRP, Figure 3.54. The reaction was initially performed in buffer containing 6 M urea. As this slows down the reaction, the reaction was diluted to 3 M urea after one week of incubation, without any improvement of the yield (not shown).

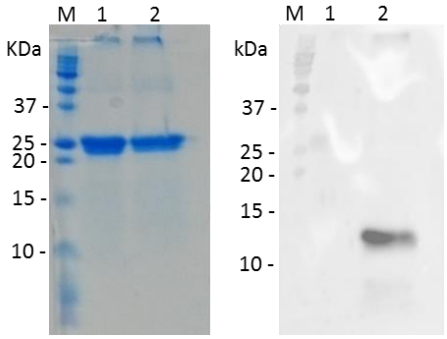


Figure 3.54. PTS reaction between Thy-1-*Npu*^{N} and *Npu*^{C}-biotin using fusion protein only purified via his-tag. Detection in western blot using anti-biotin antibody-HRP. M – Molecular weight marker, 1 – Thy-1-*Npu*^{N}, 2 – PTS Thy-1-biotin 7 days.

PTS was also performed between Thy-1-*Npu*^{N} and *Npu*^{C}-mGPI, however the product could not be detected unambiguously due to difficulties in LC-ESI-MS analysis and lacking availability of the MTG4 antibody. LC-ESI-MS analysis revealed that some PTS reaction was happening as the Post-PTS *Npu*^{C} fragment lacking the extein residues CFNA and mGPI (mass 3,992.14 Da) was detected (not shown).

PTS with Prion Protein

PrP-*Npu*$^N{}_{His6}$ fusion protein (refolded) was kindly provided by Prof. Christian Becker from University of Vienna, Figure 7.21. CD analysis of the refolded protein showed a profile corresponding to an α-helical structure, indicating the PrP^c form of the prion protein.

PTS reaction between PrP-*Npu*^{N} and *Npu*^{C}-biotin was carried out without urea and with 2.5 M urea. After three days incubation at room temperature the reactions were analyzed in SDS-PAGE and western blot (anti-biotin antibody-HRP). The PTS reaction worked in both cases tested, however reaction in 2.5 M urea resulted in a more intense signal in western blot. However, the conversion was overall very low, ca. 10% in both cases.

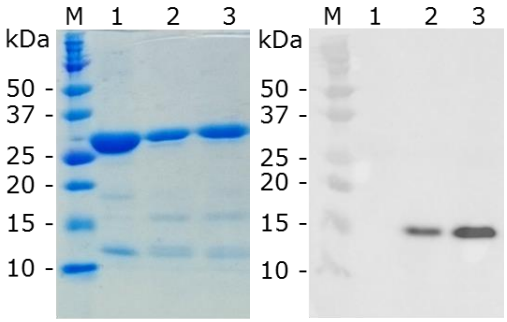
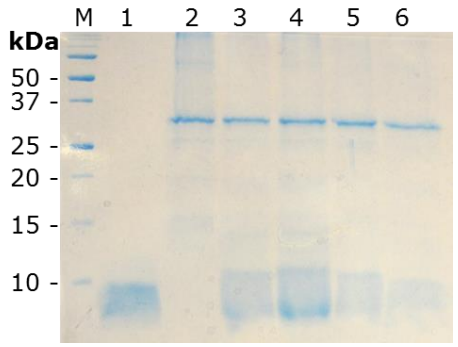


Figure 3.55. PTS between PrP-*Npu*^{N} and *Npu*^{C}-biotin. Detection in western blot using anti-biotin antibody-HRP. M – Molecular weight marker, 1 – PrP-*Npu*^{N}, 2 – PTS reaction w/o urea, 3 d, 3 – PTS reaction in 2.5 M urea, 3 d.

3.6.6 Preliminary Results for PTS with IL-2

To test PTS also for IL-2, a reaction was set up using IL-2-*Npu*^{N} fusion protein (31.5 kDa, denatured as well as refolded) and *Npu*^{C} peptide hydrazide with 1 and 5 mM TCEP. The expected size of cleaved IL-2 is 19.7 kDa. However, at this size no significant new band was observed in SDS-PAGE analysis, Figure 3.56 a. Nevertheless, PTS was also tested between IL-2-*Npu*^{N} and *Npu*^{C}-mGPI, in this case using intermediate urea concentrations (3 and 4 M) as well as refolded protein without urea. Unfortunately, the result was the same as for *Npu*^{C} hydrazide, Figure 3.56 b. Samples were analyzed in analytical SEC without success. Unfortunately, western blot analysis (using anti-GPI antibody, MTG4) resulted in no signal since this antibody generated in-house was losing activity and could not successfully be reproduced during the period of this thesis. Further experiments are still required to evaluate the outcome of this ligation.

a



b

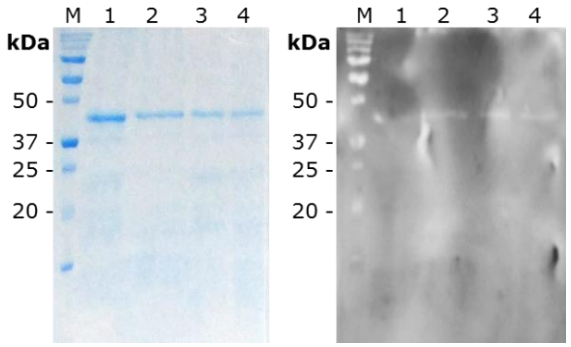


Figure 3.56. a) PTS reaction between IL-2-*Npu*^{N} and *Npu*^{C} hydrazide. M – Molecular weight marker, 1 – *Npu*^{C}, 2 – IL-2-*Npu*^{N}, 3 – PTS without urea, 1 mM TCEP, 4 – PTS without urea, 5 mM TCEP, 5 – PTS with 6 M urea, 1 mM TCEP, 6 – PTS with 6 M urea, 5 mM TCEP. **b) PTS reaction between IL-2-*Npu*^{N} and *Npu*^{C}-mGPI.** M – Molecular weight marker, 1 – PTS without urea, 2 – t=0, 3 – PTS in 3 M urea, 4 – PTS in 4 M urea. Detection in western blot using anti-GPI antibody (MTG4).

3.7 One-Pot Ligation

A new semi-synthetic strategy for generation of GPI-APs was investigated in this work, combining features of both EPL and PTS, which can be conducted in a one-pot manner and was therefore named One-Pot-Ligation. Experiments with EPL and PTS revealed that while EPL was well suitable for soluble proteins, such as eGFP, it mostly failed for insoluble proteins prepared by denaturing conditions by solubilization. PTS on the other hand performed relatively well for both soluble and insoluble proteins, however it suffers from a high technical complexity as the required building blocks are not trivial to obtain (especially glypiated *Npu*^{C}). Therefore, we aimed at the development of a novel strategy that helps avoid the bottleneck of the PTS reaction (the tedious generation of peptides ligated to GPIs) while maintaining its superior robustness.

For this new strategy, a modified C-terminal intein fragment was designed, *Npu*^{C}(AA), in which two essential residues (Asn^{35} and Cys^{36}) are exchanged for Ala residues, de-functionalizing the C-terminus of the fragment for the *trans*-splicing process. Identical to the wildtype *Npu*^{C}, this modified intein fragment can associate with the N-terminal fragment, bringing the POI's C-terminus in close proximity to that of *Npu*^{C}(AA). In order to replace the intramolecular *trans*-thioesterification from Cys^{1} to Cys^{36}, a thiol reagent such as MESNA, MPAA or MMP can be added to form a protein thioester *in situ*. This process is followed by an EPL reaction of the formed protein thioester with the Cys-tagged anchor also present in the reaction mixture.

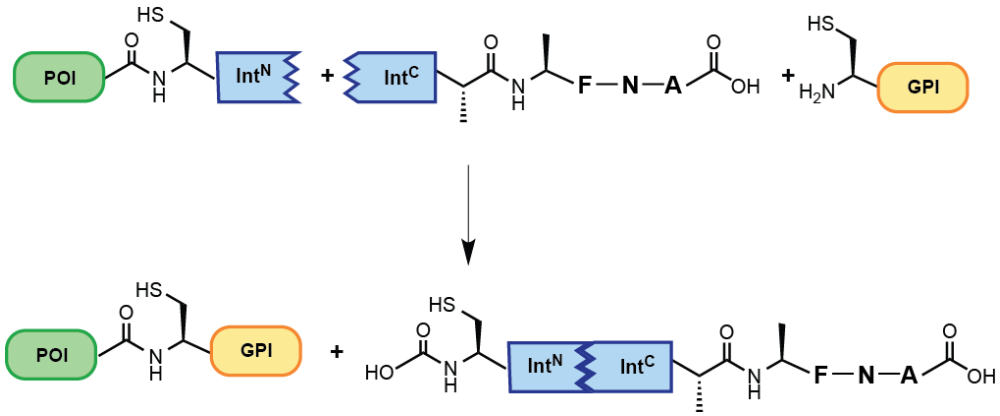


Figure 3.57. One-Pot Ligation. The same fusion proteins as for PTS are combined with a mutated *Npu*^{C} peptide, in which the two essential residues Asn^{35} and Cys^{36} have been replaced by Ala residues. In a one-pot reaction, a protein thioester is formed *in situ*, followed by Native Chemical Ligation of Cys-GPI to the C-terminus of the POI. Only one extra cysteine residue is introduced using this strategy.

3.7.1 Establishing the OPL Method and Optimization of the Thiol Reagent

To prove the applicability of this new strategy for our purposes, OPL was performed with eGFP-*Npu*^{N}, *Npu*^{C}(AA), Cys-biotin and 50 mM MMBA as thiol reagent. Samples for analysis in SDS-PAGE were taken after 2 h, 24 h, 48 h and 72 h reaction. Formation of the product was detected in western blot using anti-biotin antibody-HRP. Under these conditions an efficient intein splicing was observed giving 86% conversion (Figure 3.58).

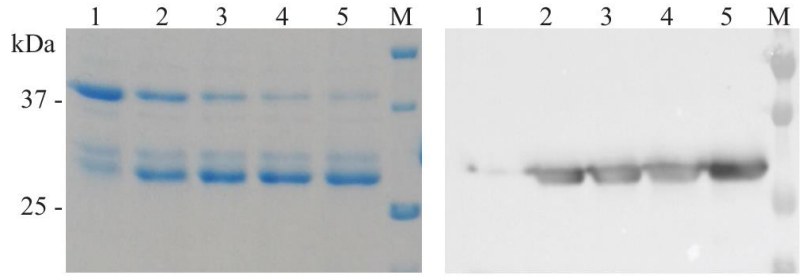


Figure 3.58. SDS-PAGE and western blot analysis of OPL reaction of eGFP-*Npu*^{N}, *Npu*^{C}(AA) and Cys-biotin. M – Molecular weight marker, 1 – GFP-*Npu*^{N}, 2 – OPL 2 h, 3 – 24 h, 4 – 48 h, 72 h.

The OPL reaction mixture was analyzed in LC-ESI-MS. This analysis showed a mixture of products, when using MMBA as thiol reagent, in which only the intermediate product eGFP-MMBA-thioester and the hydrolyzed side product eGFP-OH were detected, but not the product eGFP-biotin, Figure 3.59. This might indicate that although some product was formed and detected in western blot, the amount was too small for detection in LC-MS.

The high efficiency in the formation of the thioester and the poor product formation using MMBA as a thiol suggested that possibly the use of other thiols could help improve the yields. Several thiol reagents with different properties regarding stability and reactivity were tested for OPL reaction, see Scheme 3.9.

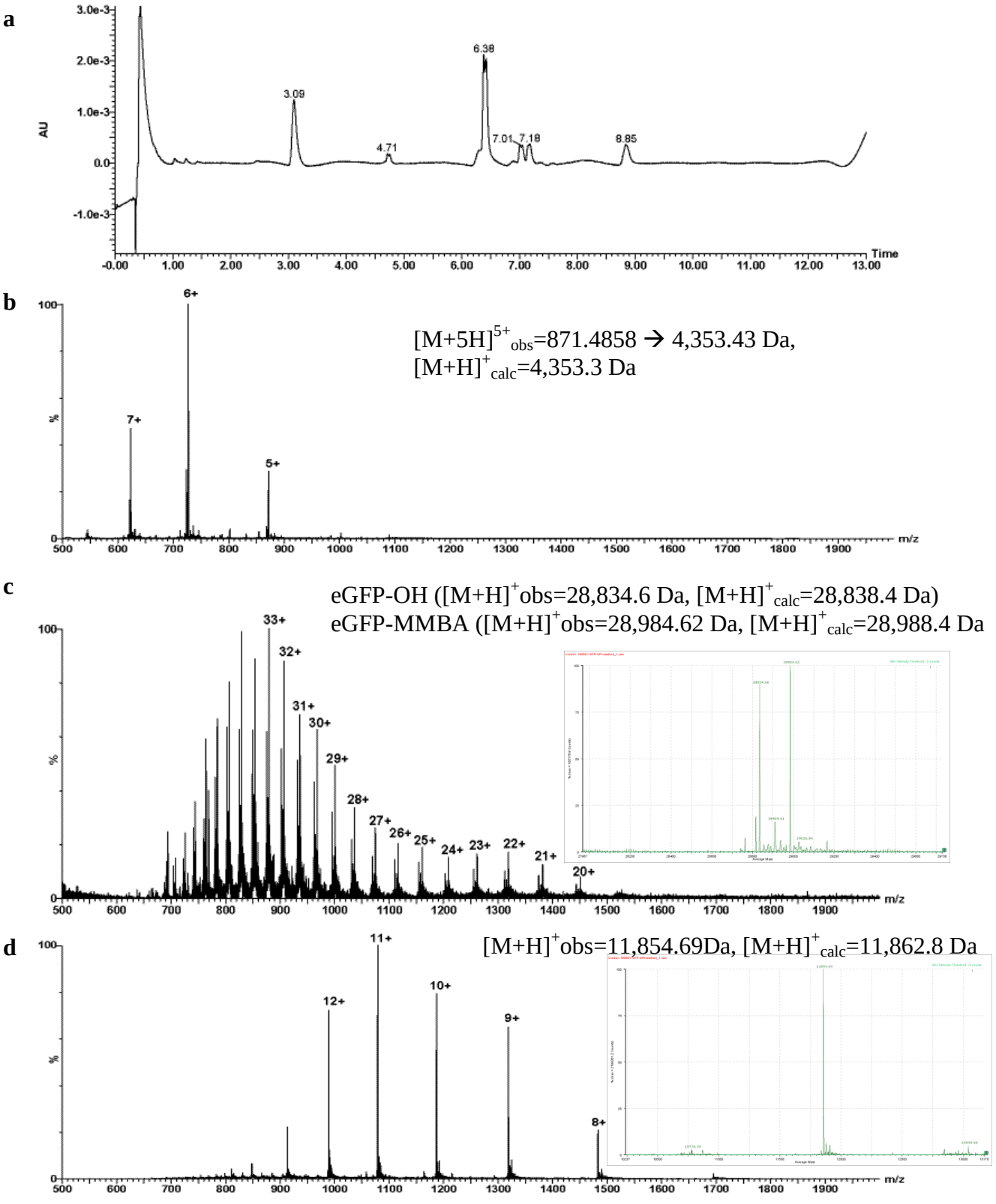
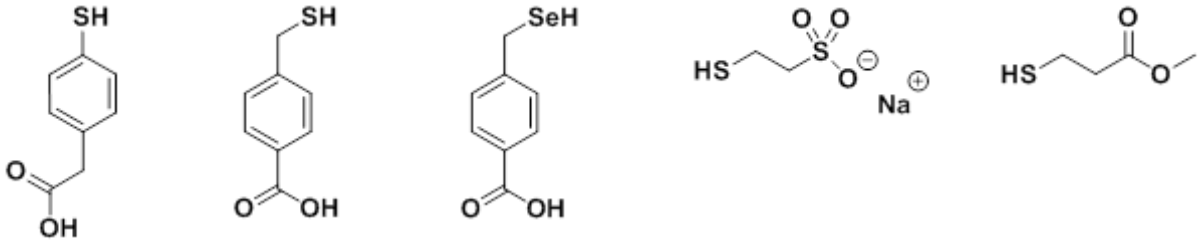


Figure 3.59. Analysis of OPL of eGFP-*Npu*^{N} and mGPI. 50 mM MMBA, 1 d. a) Chromatogram (Waters BEH C4 column, gradient 10 – 70% ACN in 30 min). **b)** Raw spectrum acquired at t_R=4.7 min: *Npu*^{C}(AA) $[M+5H]^{5+}_{obs}$=871.4858 → 4,353.43 Da, $[M+H]^+_{calc}$=4,353.3 Da. **c)** Raw and deconvoluted spectra acquired at t_R=6.5 min: double peak containing GFP-OH ($[M+H]^+$obs=28,834.6 Da, $[M+H]^+_{calc}$=28,838.4 Da) and GFP-MMBA-thioester ($[M+H]^+$obs=28,984.62 Da, $[M+H]^+_{calc}$=28,988.4 Da). t_R=7 min is the fusion protein eGFP-*Npu*^{N} (no clean spectrum was obtained, not shown). **d)** Raw and deconvoluted spectra acquired at t_R=7.2 min: cleaved *Npu*^{N} intein fragment ($[M+H]^+$obs=11,854.69 Da, $[M+H]^+_{calc}$=11,862.8 Da).

Scheme 3.9. Thiol and selenol reagents used in OPL.



MPAA MMBA SeMBA MESNA MMP

Three different thiols (MESNA, MMP and MMBA) and the selenol compound (4-selenomercaptomethyl)benzoic acid (SeMBA) were tested for OPL. SeMBA (10 eq.) was found to be not well suitable for OPL, it generated the selenoester and following reactions in low yield (only 14% after three days, Figure 3.60 a). The comparison of MESNA, MMP and MMBA for OPL showed that MESNA and MMP, but not MMBA are suitable for the ligation reaction and give good yields (ca. 25% after one hour and ca. 40% after one day for both MESNA and MMP, compared to 11% after one hour and 25% obtained after one day using MMBA), Figure 3.60 b.

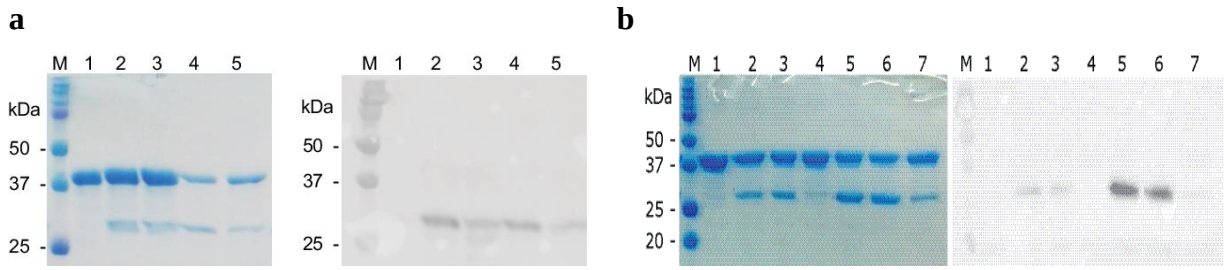


Figure 3.60. Investigation of OPL reaction using three different thiols and one selenol compound (SeMBA). a) Comparison of MMBA and SeMBA: M – Molecular weight marker, 1 – GFP-*Npu*^{N}, 2 – OPL with MMBA, 1 d, 3 – OPL with SeMBA, 1 d, 4 – OPL with MMBA, 3 d, 5 – OPL with SeMBA, 3 d. **b)** Comparison of OPL with MESNA, MMP and MMBA: M – Molecular weight marker, 1 – GFP-*Npu*^{N}, 2 – OPL with MESNA, 1 h , 3 – OPL with MMP, 1 h, 4 – OPL with MMBA, 1 h, 5 – OPL with MESNA, 1 d, 6 – OPL with MMP, 1 d, 7 – OPL with MMBA, 1 d.

Using MMP and MESNA similar results were obtained. However, due to the easier handling of MMP in OPL, this thiol was used for further studies. An OPL reaction was set up with eGFP-*Npu*^{N} and Cys-biotin. Analysis by LC-ESI-MS after one day incubation at room temperature showed the formation of the product. eGFP-biotin was detected at t_R=13.8 min, Figure 3.61 b. However, this product could not be separated from the intermediate product eGFP-thioester and the side product eGFP-OH which was detected as well. Additional to these products, the *Npu*^{C}(AA) peptide (Figure 3.61 a), the fusion protein and the cleaved *Npu*^{N} (not shown) were also identified in the chromatogram analysis.

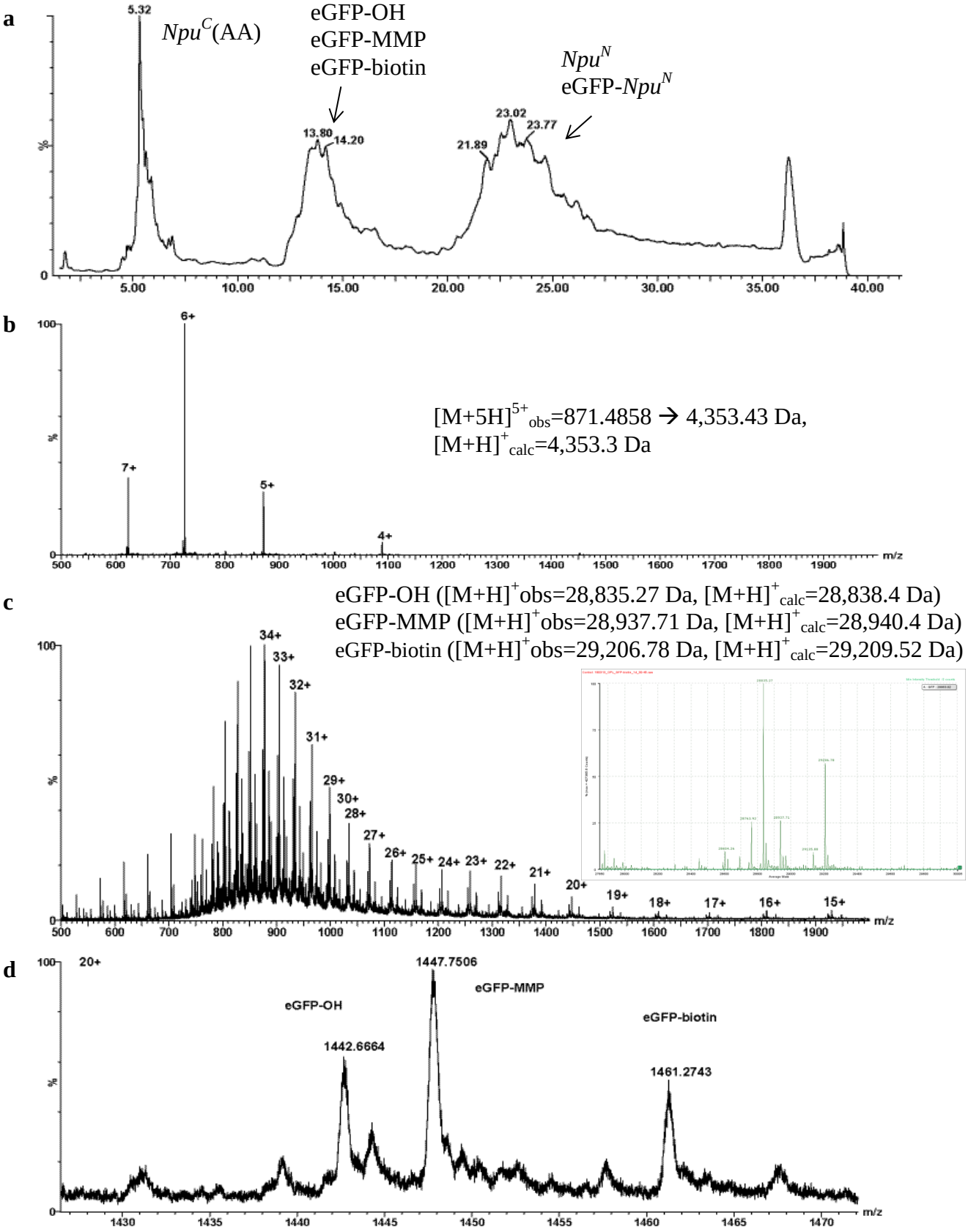


Figure 3.61. LC-ESI-MS analysis of OPL reaction between eGFP-*Npu*^{N} and Cys-biotin. a) TIC profile of the OPL reaction between eGFP-*Npu*^{N} and Cys-biotin, **b)** mass spectrum acquired at t_R=5.3 min *Npu*^{C}(AA) 1,089.3006x5 4,353.20 calc=4,353.3, **c)** Raw and deconvoluted spectrum acquired at t_R=13.8 min. It contains eGFP-OH, eGFP-MMP and eGFP-biotin. **d)** Detailed view of the products from c).

3.7.2 **Kinetic Studies of One-Pot Ligation**

In order to learn about the reaction mechanism as well as the reaction kinetics, One-Pot Ligation reactions were set up and samples were taken at regular intervals (t=0, 1 h, 2 h, 6 h, 24 h, 2 d, 3 d and 6 d). The samples were analyzed in HPLC and SDS-PAGE (Figure 3.62 a and b). The reaction yields at each time point (calculated from the SDS gel using ImageJ software) were plotted over the time, Figure 3.62 c. These results showed that in the beginning the reaction is similarly fast as PTS, however is proceeds longer at intermediate rates, as the yields are still improving after incubation over several days which was not the case for conventional PTS. This in in good agreement with our expectation regarding the mechanism of the process that ranges between EPL and PTS.

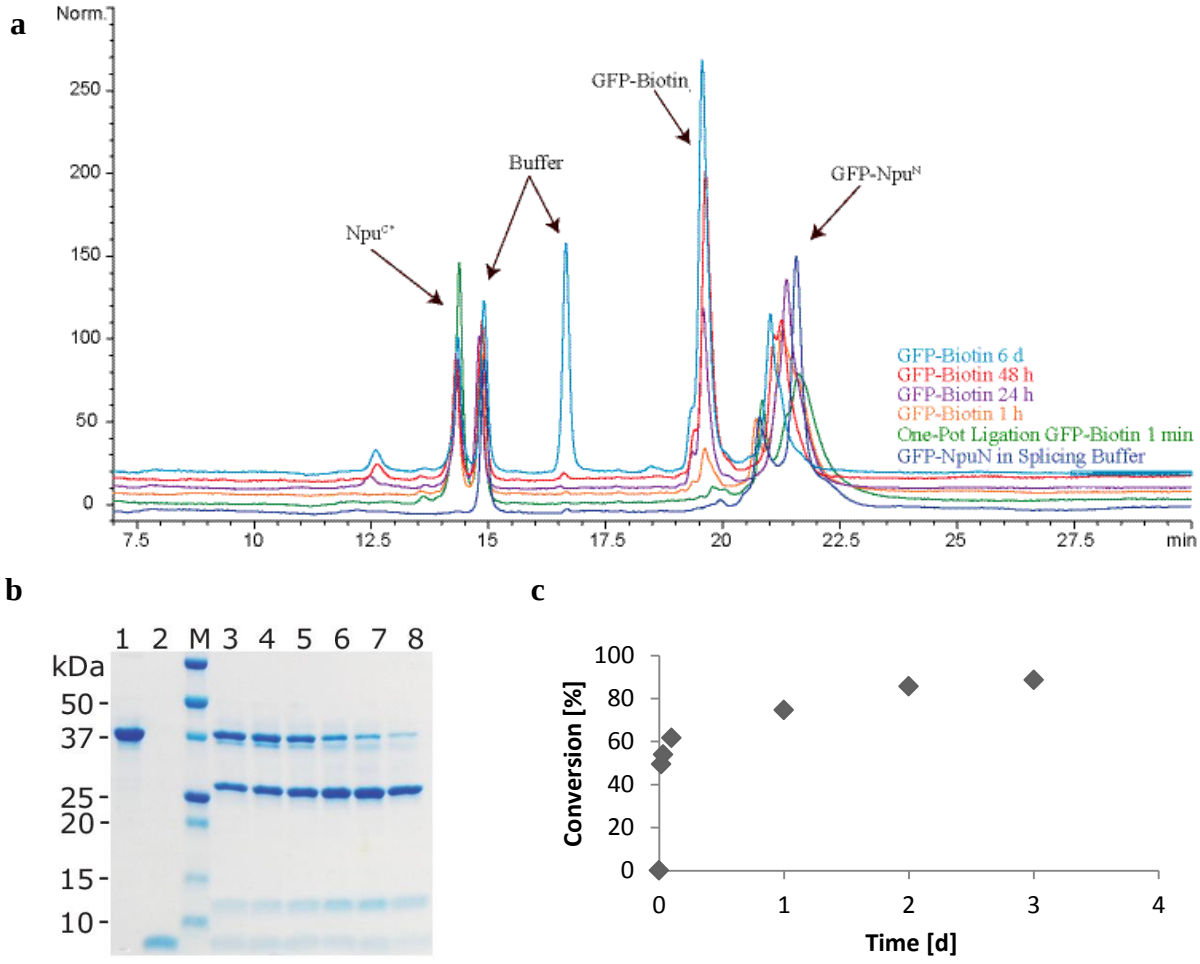


Figure 3.62. Kinetic Study of OPL Ligation between eGFP-*Npu*^{N} and Cys-biotin. a) Analysis in HPLC (C4). The peak of eGFP-*Npu*^{N} (t_R=22 min) decreases with time and changes its shape as the cleaved *Npu*^{N} intein fragment has a similar retention time. The peak of *Npu*^{C}(AA), t_R=14.5 min, decreases as well. A new peak emerges and increases at t_R=19.5 min, corresponding to either eGFP-OH or the product eGFP-biotin. **b)** Analysis in SDS-PAGEM – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – *Npu*^{C}(AA), 3 – OPL 1 h, 4 – OPL 2 h, 5 – OPL 6 h, 6 – OPL 24 h, 7 – OPL 48 h, 8 – OPL 72 h. **c)** Plot of the conversion from SDS-PAGE analysis (calculated using ImageJ software).

3.7.3 **OPL for Protein Glypiation**

Ligation of eGFP to a GPI Mimic by OPL

Before OPL was used for protein glypiation, the reaction was evaluated using eGFP-*Npu*^{N} and Cys-dimannose as a better GPI mimic than biotin; it is a carbohydrate attached to the cysteine via phosphoethanolamine unit, Scheme 2.2. The reaction was set up under the optimized conditions using 0.5 mM TCEP and 50 mM MMP and it was analyzed by LC-ESI-MS. Contrary to observations made with biotin, after ten minutes reaction only the side product eGFP-OH was detected. After one day reaction, the product eGFP-dimannose was detected together with eGFP-MMP thioester, Figure 3.63.

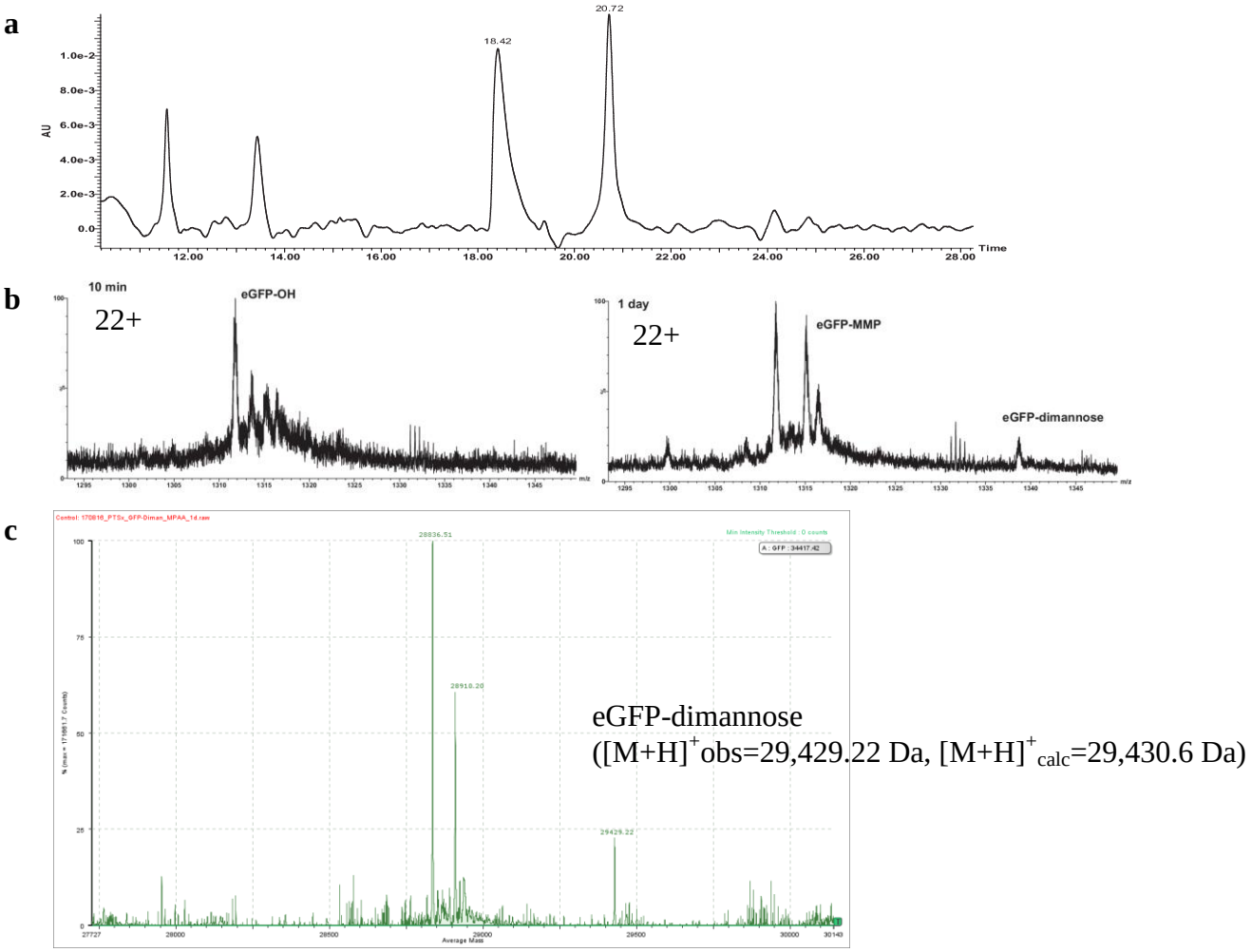


Figure 3.63. OPL reaction between eGFP-*Npu*^{N} and Cys-dimannose, analysis in LC-ESI-MS. a) Chromatogram of the analysis of OPL reaction after 1 day (Waters C4 BEH, gradient 10-70% in 15 min). **b)** raw spectra acquired at t_R=18.3 min from the samples after 10 min and 1 day incubation, zoomed in to the 22+ ion. It shows a small peak corresponding do eGFP-dimannose additionally to the ones of eGFP-OH and eGFP-MMP. **c)** Deconvoluted spectrum from t_R=18.3 min (1 day). The main peak corresponds to eGFP-OH ($[M+H]^+_{obs}$=28,838.51 Da, $[M+H]^+_{calc}$=28,838.4 Da), the second peak represents eGFP-MMP ($[M+H]^+_{obs}$=28,910.2 Da, $[M+H]^+_{calc}$=28,940.4 Da), and the smaller peak corresponds to eGFP-dimannose ($[M+H]^+_{obs}$=29,429.22 Da, $[M+H]^+_{calc}$=29,430.6 Da).

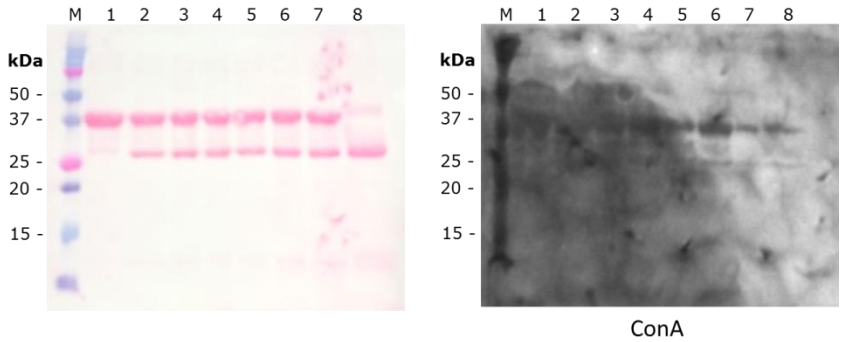


Figure 3.64. Kinetic Study of OPL using Cys-dimannose. Analysis in SDS-PAGE and western blot. 200 mM MMP, 2 mM MPAA. Detection was performed using ConA lectin-HRP. The signals obtained are unspecific, however, and present mainly at the fusion protein size for unknown reasons. M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – OPL 1 min, 3 – 10 min, 4 – 1 h, 5 – 2 h, 6 – 4 h, 7 – 6 h, 8 – 24 h.

eGFP-dimannose obtained by OPL was purified by SEC on a Superdex 75 column in PBS for characterization and Circular dichroism analysis. Three peaks corresponding to unreacted fusion protein eGFP-*Npu*^{N} (peak 1), the product eGFP-dimannose (peak 2) and cleaved *Npu*^{N} fragment (peak 3), Figure 3.65.

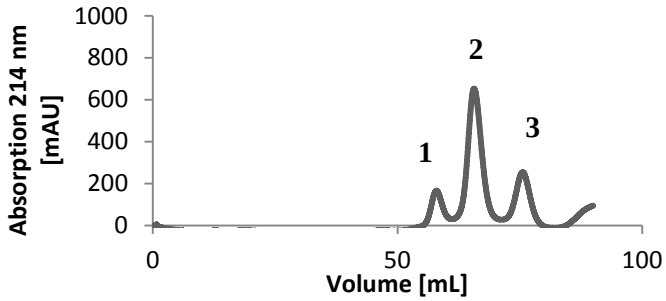


Figure 3.65. Purification of eGFP-dimannose obtained by OPL via SEC (Superdex 75 column, PBS buffer**).** Peak #2 contains the product eGFP-dimannose.

Glypiation of eGFP by OPL

Next, OPL was performed with Cys-mGPI. The reaction (containing 1 mM TCEP and 50 mM MMP) was incubated at 37°C for six days, taking samples after 1 h, 1 day and two days. The samples were analyzed in western blot and by analytical SEC using Yarra SEC-2000 column, Figure 3.66 a and b. Similarly to previous OPL test reactions with Cys-biotin and Cys-dimannose, the reactions required prolonged incubation time for the formation of the product. eGFP-mGPI from OPL was purified by SEC on a Superdex peptide column, Figure 3.66 c and d, and was concentrated for CD measurements.

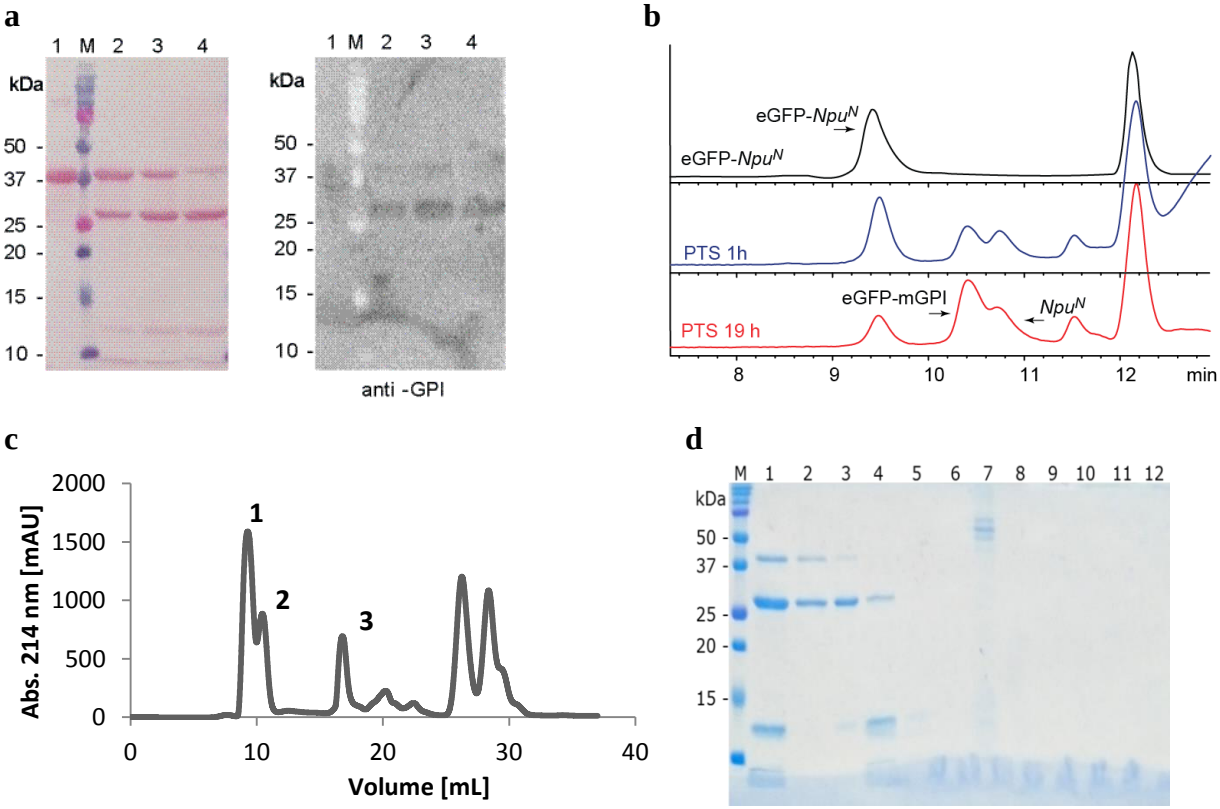
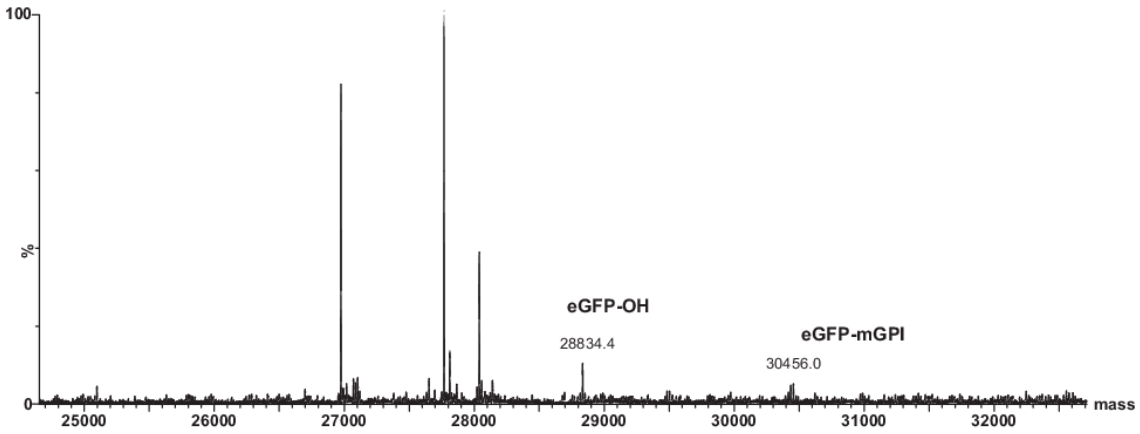


Figure 3.66. OPL reaction between eGFP-*Npu*^{N} and Cys-mGPI. a) Analysis in SDS-PAGE and western blot (detection using anti-GPI antibody MTG4). A signal is already visible after 1 day incubation, but splicing yield increases over the prolonged incubation time. M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – OPL eGFP-mGPI 1 day, 3 – OPL eGFP-mGPI 2 days, 4 – OPL eGFP-mGPI six days. **b)** Analysis in analytical SEC (Yarra SEC-2000) and comparison with eGFP-*Npu*^{N} fusion protein. After 1 h two new peaks emerge, one of which is increasing over longer incubation time (eGFP-mGPI). **c)** SEC chromatogram of purification of eGFP-mGPI from OPL. Peak #1 contains the product, whereas peak #2 contains product as well as *Npu*^{C}(AA). **d)** SDS-PAGE analysis of the fraction from SEC purification. M – Molecular weight marker, 1 – OPL reaction eGFP-mGPI (Load fraction), 2 – 4 – peak #1, 5 – peak #2, 6 – 12 – peak #3.

The product eGFP-mGPI was also detected in LC-MS, measured on a SYNAP HRMS 62-S mass spectrometer at the mass facility of FU Berlin), Figure 3.67. However the peak observed represents only a minor part of the sample. In order to confirm the attachment of the mGPI to the C-terminus, a tryptic digestion of the product separated by SDS-PAGE was performed following the protocol described in 2.8.7. Analysis of the peptide fragments by MALDI-TOF-MS showed only very few parts of the sequence, excluding the C-terminus.

a



b MASWSHPQFE KASGTVSKGE ELFTGVVPIL VELDGDVNGH KFSVSGEGEG DATYGKLTLK FICTTGKLPV PWPTLVTTLT YGVQCFSRYP DHMKQHDFFK SAMPEGYVQE RTIFFKDDGN YKTRAEVKFE GDTLVNRIEL KGIDFKEDGN ILGHKLEYNY NSHNVYIMAD KQKNGIKVNF KIRHNIEDGS VQLADHYQQN TPIGDGPVLL PDNHYLSTQS ALSKDPNEKR DHMVLLEFVT AAGITLGMDE LYKGSRSGSC-mGPI

Figure 3.67. a) Deconvoluted mass spectrum of eGFP-mGPI (measured on a SYNAP 62-S mass spectrometer at the mass facility of FU Berlin): $[M+H]^+_{obs}$=30,456.0 Da, $[M+H]^+_{calc}$=30,336.0 Da with N-terminal Met (in this protein batch with and without Met was found). **b)** Tryptic digestion of eGFP-mGPI from OPL: underlined parts of the sequence have been identified in MALDI-TOF-MS analysis. This does not include the C-terminus of the protein bearing the GPI, this method does not give any information about glypiation.

Finally, OPL was also evaluated for the ligation of a bilipidated GPI-anchor to eGFP. The reaction was carried out using the same conditions as for OPL with mGPI. The product, eGFP-bGPI, was also detected in western blot using anti-GPI antibody (MTG4), Figure 3.68. Weak but distinct signals were observed for each time point investigated (1 d, 2 d and 6 d reaction).

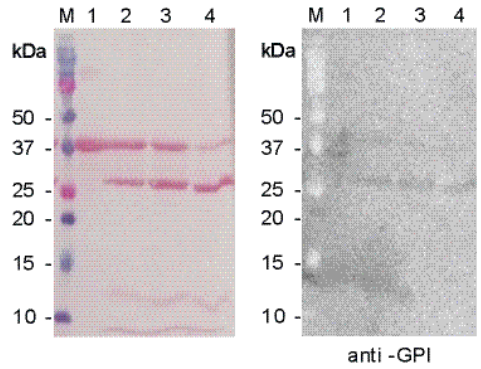


Figure 3.68. Western blot analysis (anti-GPI antibody MTG4) of OPL reaction between eGFP-*Npu*^{N} and Cys-bGPI. M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – OPL eGFP-bGPI 1 day, 3 – OPL eGFP-bGPI 2 days, 4 – OPL eGFP-bGPI six days.

Additionally to chromatographic methods, detergent extraction was evaluated for the purification of glypiated proteins. This method should result in less dilution of the samples and separation of lipidated (glypiated) from unlipidated proteins by phase separation using Triton X-114.[156] This method was tested for eGFP-mGPI and eGFP-bGPI. Unfortunately, due to the lack of an anti-GPI antibody of correct specificity, detection of the extraction was difficult. MTG5 antibody was used for detection, however this antibody gave a signal for glypiated eGFP, but also for eGFP-*Npu*^{N} fusion protein, Figure 7.14. As this result was inconclusive, this separation method was not further investigated.

3.7.4 **OPL for the Semi-Synthesis of Naturally Glypiated Proteins**

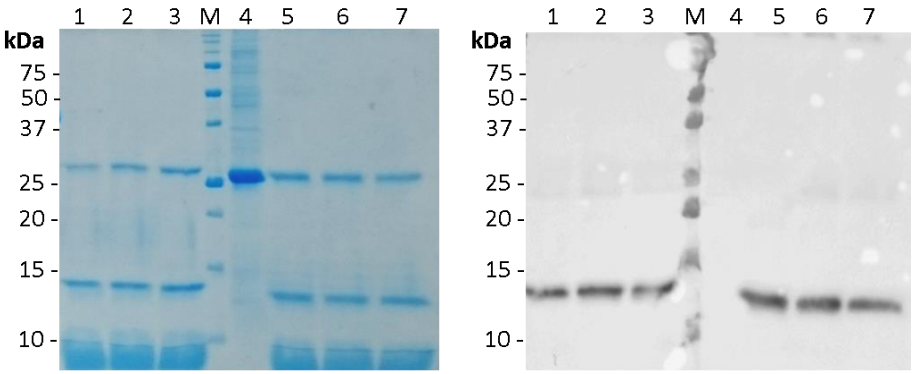
OPL for Thy-1 Protein

After EPL and PTS, OPL was also used to introduce a C-terminal modification to the Thy-1 protein. First, biotinylation of Thy-1 was optimized in buffer containing 6 M urea at room temperature and 37°C and TCEP concentrations of 1 mM, 5 mM and 10 mM. In this case, higher TCEP concentrations than for eGFP were tested as Thy-1 protein was already present in denatured form. The reaction samples were analyzed in SDS-PAGE and western blot after one day incubation, Figure 3.69 a. Reaction at room temperature with 5 mM TCEP gave the highest yield (66% conversion after one day). Longer reaction did not improve the yield, Figure 7.20. The product Thy-1-biotin was analyzed using LC-ESI-MS and MALDI-TOF-MS after tryptic digestion. In tryptic digestion, most of the sequence was covered, however the C-terminal fragment was not found in MALDI-TOF-MS analysis, Figure 3.69 b. LC-ESI-MS analysis did not show the presence of the desired product.

OPL was also performed between Thy-1-*Npu*^{N} and Cys-mGPI. After six days reaction, the mixture was analyzed in LC-ESI-MS, Figure 3.70. The expected mass of Thy-1-mGPI $[M+H]^+_{calc}$=13,939.9 Da was not observed, however, the mass that was found $[M+H]^+_{obs}$=13,590.58, corresponded very well with a delipidated and dephosphorylated species of Thy-1 ($[M+H]^+_{calc}$(Thy-1-GPI$_{delipidated}$)=13,590.63. Therefore, utilizing the fact that unreacted fusion protein and cleaved *Npu*^{N} carry a his-tag, Thy-1-mGPI from OPL was purified using his-tag affinity chromatography (HisTrap HP 1 mL column) under denaturing conditions (6 M urea), Figure 3.71 a. As expected, Thy-1-mGPI (lacking the his-tag) was found in the wash fraction. Thy-1-*Npu*^{N} and *Npu*^{N} bound to the column and were only found in the elution fractions, Figure 3.71 b. *Npu*^{C}(AA) was mainly found in the flow-through

fraction, but also in the same fraction as the product. However, due to its small size it can easily be removed by dialysis.

a



b MQKVTSLTAC LVDQSLRLDC RHENTSSSPI QYEFSLTRET KKHVLFGTVG VPEHTYRSRT NFTSKYNMKV LYLSAFTSKD EGTYTCALHH SGHSPPISSQ NVTVLRDKLV KC-biotin

Figure 3.69. OPL Thy-1-biotin. a) Detection in Western blot with anti-biotin antibody-HRP. M – Molecular weight marker, 1 – OPL RT, 10 mM TCEP, 2 – OPL RT, 5 mM TCEP, 3 – OPL RT, 1 mM TCEP, 4 – Thy-1-*Npu*^{N}, 5 – OPL 37°C, 1 mM TCEP, 6 – OPL 37°C, 5 mM TCEP, 7 – OPL 37°C, 10 mM TCEP. **b) Analysis of Thy-1-biotin from OPL by tryptic digestion.** Underlined sections have been identified in MALDI-TOF-MS.

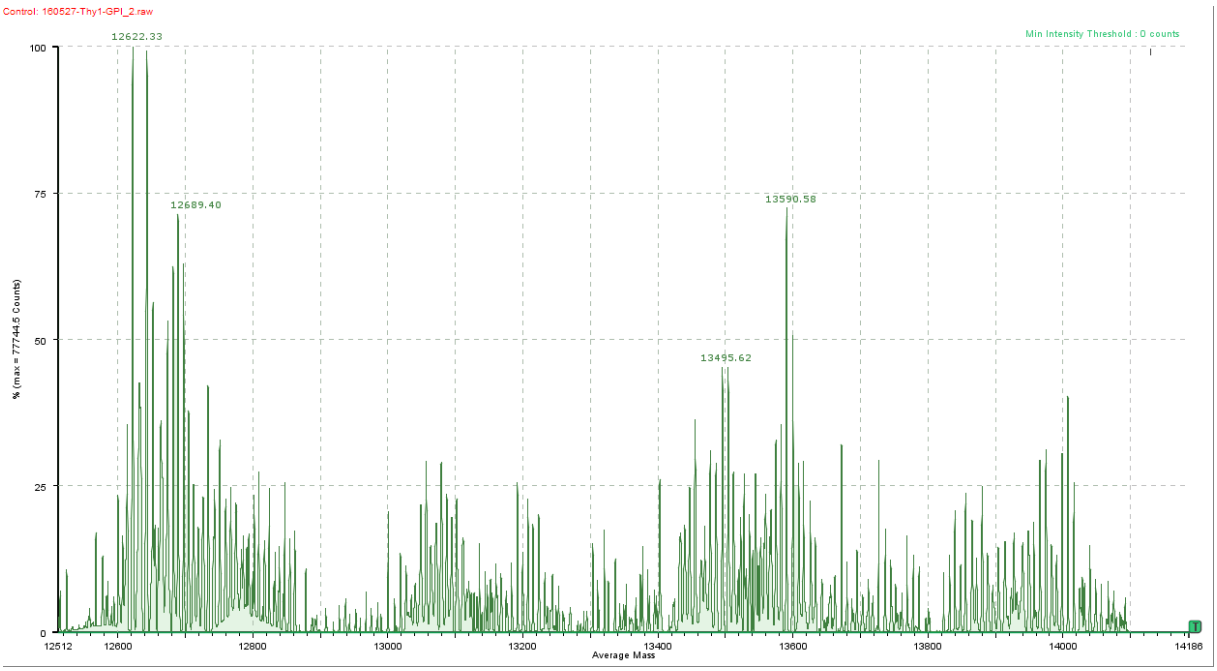


Figure 3.70. LC-ESI-MS analysis of Thy-1-mGPI from OPL. The deconvoluted mass spectrum shows a main peak at $[M+H]^{+}_{obs}$=13,590.58 Da, which does not correspond to the full Thy-1-mGPI ($[M+H]^{+}_{calc}$=13,939.9 Da), but instead to a delipidated, dephosphorylated species ($[M+H]^{+}_{calc}$=13,590.63 Da).

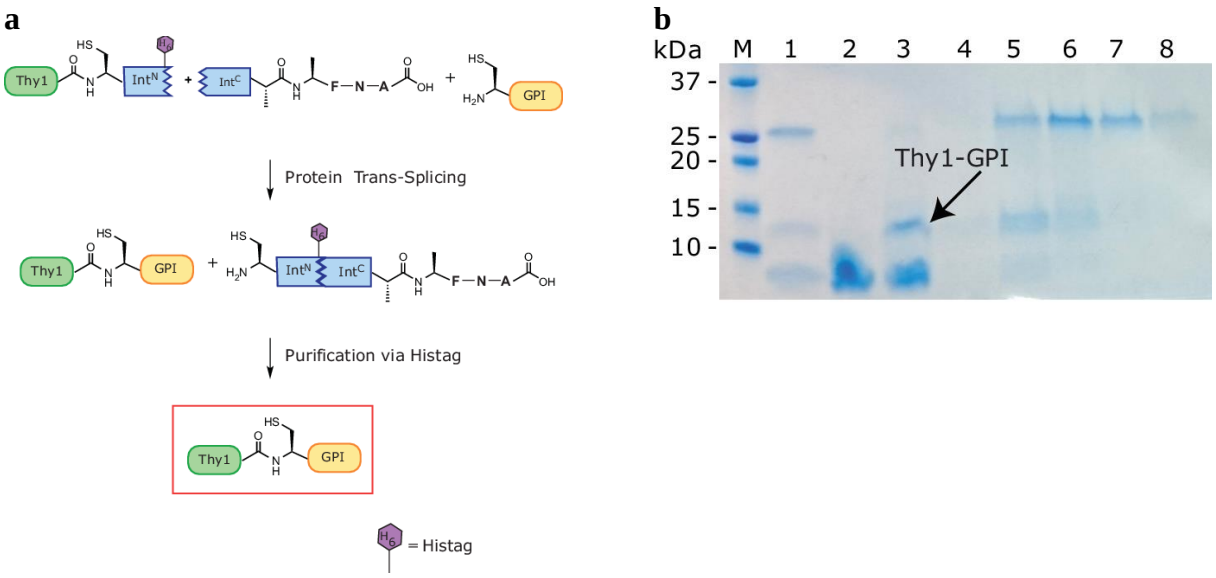


Figure 3.71. Purification of Thy-1-mGPI from OPL using his-tag. a) Strategy. b) SDS-PAGE analysis, M – Molecular weight marker, 1 – OPL reaction Thy-1-mGPI 6 d, 2 – flow-through his-tag purification, 3 – wash fraction, 4 – 8 – elution fractions**.**

OPL for Prion Protein

OPL reaction between PrP-*Npu*^{N} and Cys-biotin was carried out without urea and with 2.5 M urea. After three days incubation at room temperature the reactions were analyzed in SDS-PAGE and western blot using anti-biotin antibody-HRP. In order to compare the efficiency of the OPL, the reaction was performed in parallel with a PTS reaction under the previously mentioned reaction conditions with PrP. In comparison, it can be seen that OPL gave a better intein cleavage than PTS (15% and 13% w/o and with urea, respectively). However, the formation of the product was less efficient giving only a very weak signal in western blot, Figure 3.72.

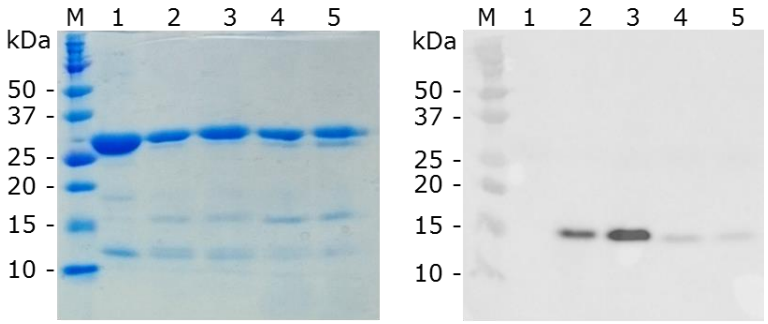


Figure 3.72. OPL between PrP-*Npu*^{N} and Cys-biotin. M – Molecular weight marker, 1 – PrP-*Npu*^{N}, 2 – PTS reaction w/o urea, 3 d, 3 – PTS reaction 2.5 M urea, 3 d, 4 – OPL reaction w/o urea, 3 d, 5 – OPL reaction 2.5 M urea, 3 d.

3.7.5 **Modification of IL-2 by OPL**

Based on the negative results obtained in the PTS for modification of IL-2 protein, OPL was investigated as an alternative strategy. Reaction of IL-2-*Npu*^{N} and Cys-biotin was investigated at room temperature and 37°C. Samples were taken after 2 h, 24 h, 48 h and seven days, and were analyzed in SDS-PAGE and western blot (using anti-biotin antibody-HRP), Figure 3.73 a. Splicing yields were comparable at both temperatures tested (up to 39% and 35%, at RT and 37°C, respectively). However the signal in western blot was stronger in the samples incubated at room temperature, and intensity increased over the whole incubation time (seven days).

Similarly, OPL was performed between IL-2-*Npu*^{N} and Cys-mGPI under the conditions optimized for Cys-biotin. A new band is observed in SDS-PAGE analysis approximately at the expected size (21 kDa), most likely corresponding to the product IL-2-mGPI. However, more extensive characterization was not possible due to lack of anti-GPI antibody and LC-ESI-MS availability. After one day reaction, 20% conversion was obtained, and after seven days conversion increased to only 30%, according to SDS-PAGE analysis, Figure 3.73 b.

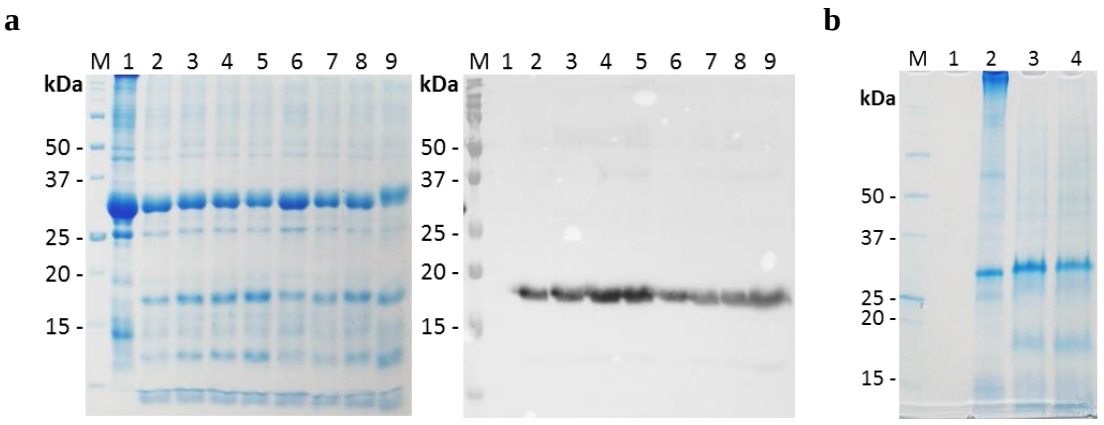


Figure 3.73. OPL IL-2-*Npu*^{N} and a) Cys-biotin and b) mGPI. a) M – Molecular weight marker, 1 – IL-2-*Npu*^{N}, 2 – OPL RT 2 h, 3 – RT 24 h, 4 – RT 48 h, 5 – RT 7 d, 6 – OPL 37°C 2 h, 7 – 37°C 24 h, 8 – 37°C 48 h, 9 – 37°C 7 d. **b)** 1 – *Npu*^{C}(AA), 2 – IL-2-*Npu*^{N}, 3 – OPL RT 24 h, 4 – OPL RT 7 d.

3.8 Investigation of the Effect of C-terminal Modification and Glypiation on Protein Structure Using Circular Dichroism

Circular dichroism (CD) was used to investigate the effect of C-terminal modifications (biotinylation, dimannosylation and glypiation) on the structure of eGFP. In its native fold, eGFP possesses a β-barrel structure.[171] This structure was confirmed in CD measurements for eGFP-MMP thioester and eGFP-OH (made with *Npu*^{C}(AA)), Figure 3.74, which showed a profile corresponding well to previous reports.

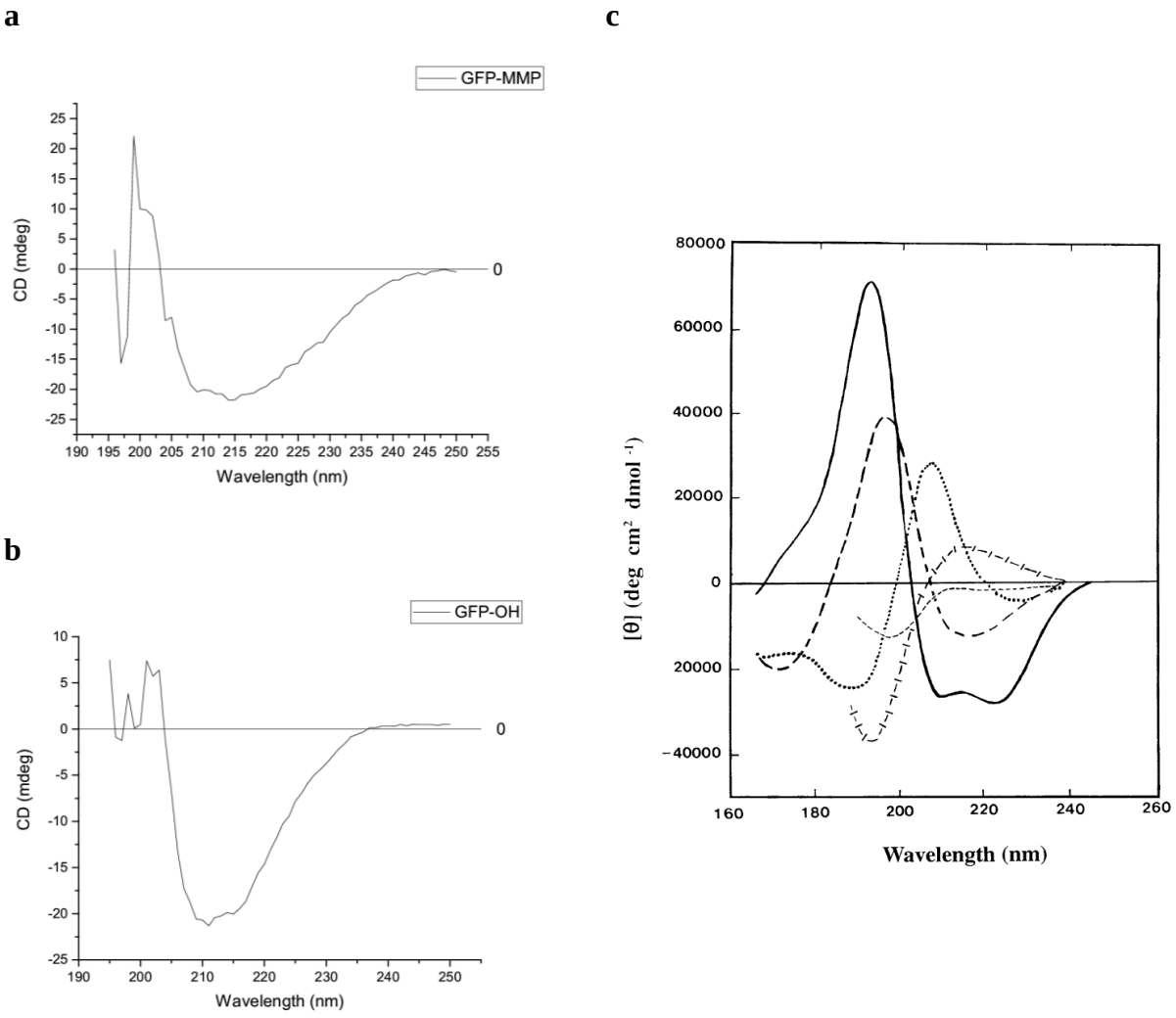


Figure 3.74. CD spectra of eGFP-MMP-thioester (a) and eGFP-OH (b) in comparison to sample far UV CD spectra of different types pf protein secondary structure (c, taken from [166]). The long dotted lone the sample spectrum corresponds to anti-parallel β-sheet structure with a minimum between 210 and 220 nm, which is also found for both eGFP spectra, as expected.[171]

The purified samples from EPL (eGFP-MMP, eGFP-biotin, eGFP-dimannose, eGFP-mGPI and eGFP-bGPI) and from OPL (eGFP-OH, eGFP-biotin, eGFP-mGPI and eGFP-bGPI) were also analyzed in circular dichroism. PTS did not yield sufficient sample amounts of any proteins for analysis in CD. For the samples from both EPL and OPL the β-sheet structure was not affected by C-terminal modification, Figure 3.75, Figure 3.76. For biotinylation and dimannosylation, the spectra are entirely unchanged in comparison to those of eGFP-thioester and eGFP-OH, respectively. In the case of modification with mGPI, the minimum at ca. 216 nm is less pronounced for both EPL and OPL samples in comparison to unmodified eGFP, however this is likely to be due to a lower protein concentration as concentrating the samples was risky regarding possible precipitation. For modification with bGPI however, for the EPL sample not spectrum could be obtained and for the OPL sample of eGFP-bGPI, almost no absorption was observed, Figure 3.76 c.

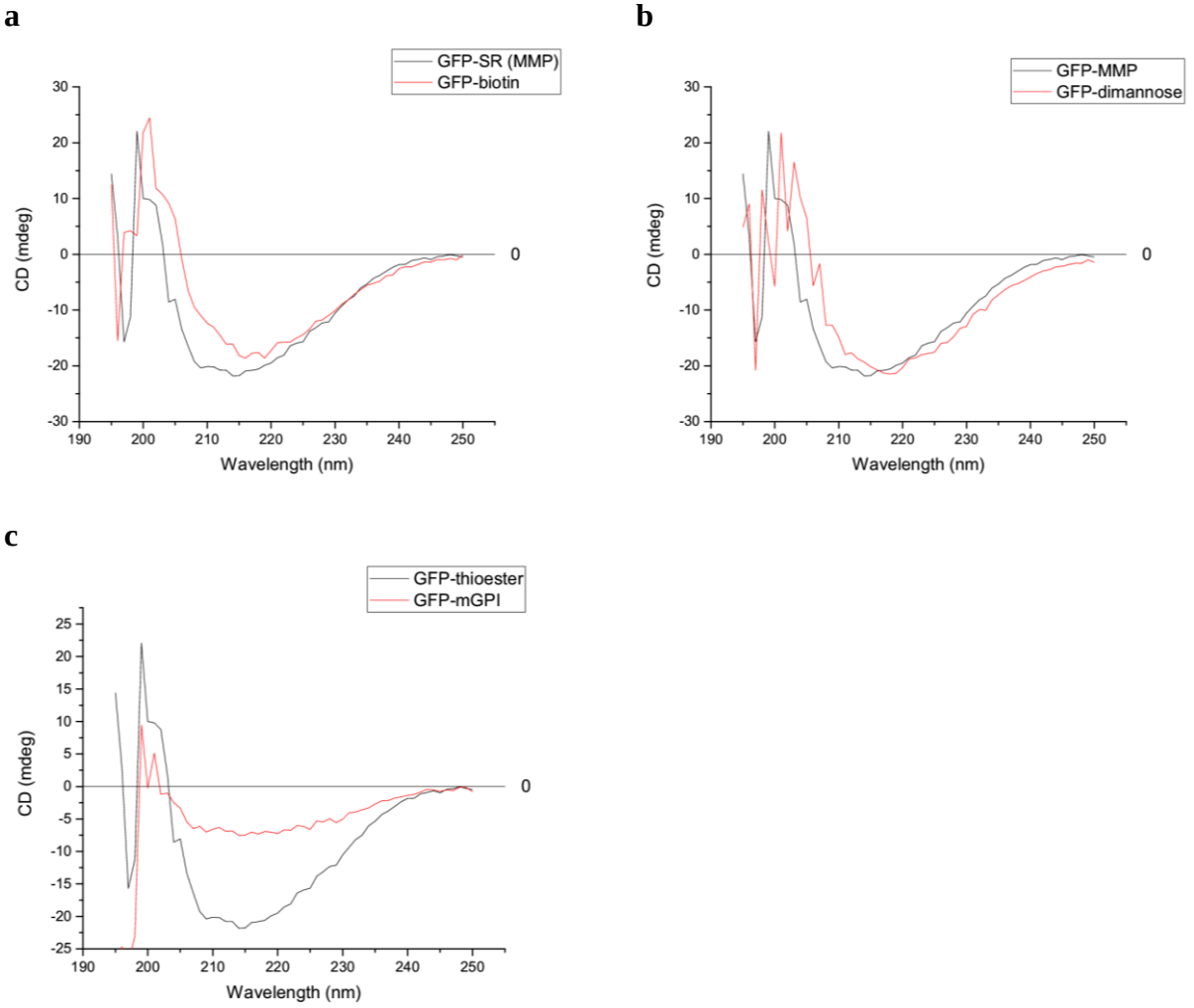


Figure 3.75. CD spectra taken of samples from EPL (eGFP-biotin, eGFP-dimannose and eGFP-mGPI) in comparison with eGFP-MMP thioester. The β-sheet structure of eGFP is not compromised by C-terminal modification with biotin, dimannose or mGPI.

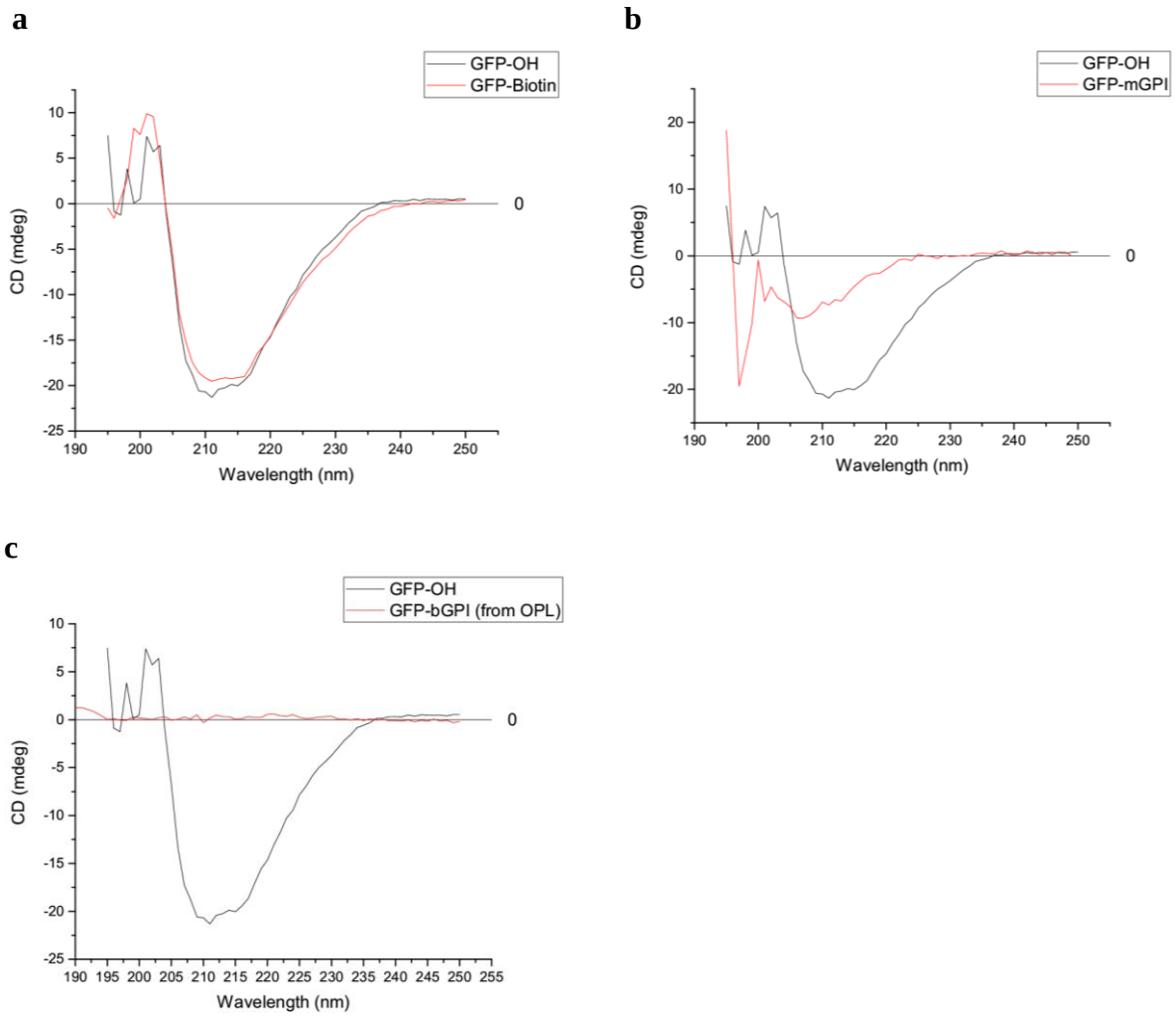


Figure 3.76. CD spectra taken of samples from OPL (eGFP-biotin, eGFP-mGPI and eGFP-bGPI) in comparison to eGFP-OH.

4 Discussion

In this section, the findings of this work will be discussed and evaluated in relation to the obtained improvements, advances and disadvantages, and the potential among the different reported methods for the semi-synthesis of GPI-APs. Section 4.1 deals with the peptide synthesis, thioester formation and NCL reactions. A discussion of the different strategies for the generation of homogeneous GPI-anchored proteins is provided in sections 4.2– 4.4. Additionally, challenges regarding synthesis, generation or characterization of different compounds is addressed in section 4.5. As a general remark it should be noted here that the big challenges resulted from the difficult handling (low solubility) and impaired stability of the highly complex synthetic GPI structures.

4.1 Peptide Synthesis and Native Chemical Ligation

As part of this project, two peptides corresponding to the C-terminal fragment of the DnaE split intein from *Nostoc punctiforme (Npu)* were synthesized to investigate the semi-synthesis of GPI-APs. The Npu^C fragment was synthesized as a peptide thioester for a Native Chemical Ligation with the cysteine-containing biotin and GPI. Although this C-terminal fragment is only a 39 amino acid peptide, its synthesis was not trivial and several challenges were encountered during the process. The sequence contains nearly all proteinogenic amino acids, and for many of them, the coupling and deprotection conditions required separate optimization, Table 2.2. Therefore, the availability of this fragment unexpectedly turned out to represent a significant bottleneck in the establishment and optimization of the PTS method to obtain GPI-APs. After various syntheses and optimization of the process, this peptide was obtained in 10 mg scale under optimized conditions using a resin that released the peptide as a peptide hydrazide as a thioester precursor.

Having the Npu^C peptide hydrazide in hands, the next step was the peptide thioester generation of Npu^C.[136] This conversion was executed following the report of Zheng et al. (2013). Different thiols were used for the formation of this thioester, however only when MMP was used, the reaction proceeded in quantitative yield within 20 minutes and without the formation of side products, such as cyclic thio-lactame between the C-terminus and the side chain of a cysteine residue, or the product of the hydrolysis of the thioester. Although

some material was lost during this conversion, mainly during purification by HPLC, the main constraint was the difficult synthesis of the peptide hydrazide with low yields.

Following the preparation of the thioester, a native chemical ligation reaction between the *Npu*^{C} thioester and the different cysteine-containing molecules was performed. The yields and reaction rates of this process were strongly dependent on the ligated molecule. In the case of Cys-biotin, NCL was almost completed after three days reaction at 37°C, whereas NCL of *Npu*^{C}-MMP with Cys-mGPI was not successful under the same conditions. In order to avoid the oxidation of the cysteine during the NCL with the GPI, the reaction was performed under argon. However, even under these conditions, the reaction was by far not as successful as ligation to Cys-biotin. After one week incubation under argon at 37°C with freshly added TCEP every one to two days, and re-adjusting the pH to 7.0 frequently, the product could be observed in LC-ESI-MS analysis. Unfortunately, together with the ligated product, also the hydrolyzed peptide and some oxidized side products were detected. After different efforts for the optimization of the reaction, the product of the ligation (*Npu*^{C}-mGPI) was obtained in sufficient amounts to investigate the PTS process.

The modified peptide *Npu*^{C}(AA) on the other hand was much easier to synthesize, possibly due to the lack of the two critical amino acids in the sequence, the cysteine and the asparagine residues participating in the PTS, which were replaced by alanines. Additionally, this peptide does not require any modification post-synthesis, avoiding the loss of material after purification. The easy accessibility to the mutated *Npu*^{C} peptide represented a significant advantage for the application of the OPL method over the PTS method in a slightly different way than anticipated: it was assumed that this method would mainly be more convenient because no peptide thioesters had to be generated and ligated to a GPI or GPI analogues.

4.2 EPL is a Suitable Strategy for Soluble Proteins

The first method evaluated for the ligation of proteins and GPIs was Expressed protein ligation. This method has been largely used to ligate proteins with a C-terminal thioester with other molecules. In this case, EPL was found to be an excellent strategy for the semi-synthesis of C-terminally modified proteins, if the proteins of interest are present in soluble form. The process was usually hampered by the reactivity and stability of the active protein thioesters. However, using an appropriate thiol and a freshly prepared protein thioester in the reaction, a faster ligation with the cysteine-bound molecules ensured little to no hydrolysis.

Unfortunately, even with the use of more stable thiols such as MMP, the protein thioester was not very stable (< 1 week), which was not expected in this short time. Fortunately, expression of the eGFP-*Mxe* fusion protein was straightforward and high-yielding and the protein could be used in molar excess to the ligation partners. The EPL strategy represents a very versatile method for the generation of libraries of proteins carrying different GPI-anchors for comparative studies, as no extra ligation step is required for the generation of the GPI-APs. The production of the thioester requires relatively little synthetic effort, so that the main limitation was the difficult access to sufficient amounts of GPI molecules, the synthesis of which is still a challenge, and the low solubility of the GPI molecules.

In addition to the problems associated with protein solubility, the main limitation of EPL was the relatively slow reaction rates as was shown in the kinetic study in section 3.5.3. It is likely that the slow reaction rates were due to low concentration or reactivity of the GPI-anchors, namely by deactivation of the reactive cysteine by the connecting this residue to the GPI, which is involved in the ligation reaction, as it was observed by Guo et al. (2009)[81] for the Sortase A reaction. In contrast to previous reports involving GPI mimics in EPL reactions, the GPIs are structurally far away from these mimics so that this problem would not be observed,[37] especially when peptide-based mimics were applied.[132]

This method could be efficiently used for the semi-synthesis of eGFP-biotin and eGFP-mGPI, in which all the components are soluble. However, an additional limitation of this strategy was the difficult purification of the reaction products, which do not have large size difference. Generally, unreacted protein thioester and the desired product present a fairly small difference in molecular weight in the case of GPI-APs (i.e. only the size of the GPI) and standard purification methods, such as SEC, IEX or affinity purification could not be applied. Only a separation by HPLC was possible, however this involved a denaturation of the folded eGFP, which is not desirable.

The EPL was an inefficient strategy for ligations involving insoluble proteins. Although the ligation reaction should work well under denaturing conditions, as it is done with peptides in NCL, the conditions required for solubilizing the protein of interest are not always compatible with the conditions required to obtain a proper folding of the *Mxe* intein, which was not capable of performing the intein cleavage reactions 6 M GdmCl, and only to a little extent in 6 M urea, which is a weaker denaturing agent. To overcome this limitation, the formation of the thioesters from the insoluble proteins Thy-1, IL-2 and PrP was completed in solution until a certain level, using a dilution process. However, even when the formation of the protein thioesters was achieved, isolation of pure thioesters was difficult as the chitin binding domain

present in the fusion proteins does not bind to the chitin resin under these conditions. Furthermore, the cleavable *Mxe*-CDB domain was relatively close in size to the protein thioesters which made size exclusion chromatography difficult (additional to the fact that it is not well suited for denatured proteins).

4.3 PTS is a Robust Method Limited by Access to Peptide-GPI Conjugates

The second strategy evaluated in this work for the semi-synthesis of GPI-APs was the protein *trans*-splicing (PTS). PTS has scarcely been used for the generation of GPI-APs, and in the few reports described in literature, the products were generally not purified but used *in situ*, mostly to demonstrate the localization of the glypiated proteins, for example on the cell membrane in live cells.[138] Therefore, no full characterization of such products is available to date. Since PTS is very robust and reliable, exhibiting nearly no hydrolysis, it can be assumed that these products have been obtained as reported. Based on these results, one of the aims of this work was to expand the application of this strategy to obtain glypiated proteins in amounts that can be further characterized and used in studies to understand the role of this modification of proteins.

In contrast to EPL, the PTS strategy was successful for both native and denatured proteins. This strategy is characterized by high reaction rates, especially under native conditions, which is in well agreement with literature.[145] Although the kinetic studies carried out here did not reach the results of the literature reports where the observed $t_{1/2}$ is < 1 min,[138,172] it is the fastest of the strategies investigated in this work. This is possibly due to the use of exteins that are structurally very different from the native ones (i.e. Cys-GPI instead of peptides). As it is the case for the *Npu* split intein, a strong extein dependency has also been described as a main drawback of the PTS reaction by Shah et al. (2013).[118] Mutations in the non-essential residues close to essential ones was investigated by Stevens et al. (2017)[172] to generate more promiscuous split inteins with some success. Most of the mutations slowed down the splicing reaction, but some of them, especially the presence of a GXP motif close to the penultimate His125) did result in a wider spectrum of accepted exteins.

One advantage of PTS over EPL pointed out in Shah & Muir (2011)[117] is its independence on protein concentration. The initial step, association of the two intein fragments, does not rely on random collision of the fragments in the reaction solution but can utilize the strong

affinity of both split intein fragments to each other. This strong affinity is probably the reason for the high robustness of the PTS reaction towards chaotropic agents such as urea and can be attributed to: the big difference in pI of the two split intein fragments, which makes the complex very stable and therefore independent of salt concentration.[115]

Due to the nature of the PTS reaction involving the cleavage of a relatively large protein domain, the Npu^N fragment, from the fusion protein and its replacement by the rather small GPI-anchors, that starting fusion protein and the product possess a large difference in molecular weight that offers good opportunity for purification via SEC for native proteins. Furthermore, in this a good option is the introduction of C-terminal tags, because this facilitates easy separation of the products from the cleaved intein fragments. These advantages were demonstrated in the formation of glypiated Thy-1 and Prion proteins, although full conversion was not achieved.

The main drawbacks of the use of the PTS strategy were the high synthetic effort needed to obtain the synthetic fragments and the loss of material in the numerous purification steps after reach reaction. As already discussed, especially the generation of Npu^C-mGPI in sufficient amounts represented a major challenge during this work. Furthermore, the need of multiple steps makes this strategy less versatile to investigate the effect of a GPI structure on a specific protein and for the generation of GPI-AP libraries with different GPI-anchors or other C-terminal modifications, which will require the synthesis of each single structure attached to the Npu^C fragment, involving also purification.

In this context and due to these difficulties, PTS was successfully applied for the generation of eGFP-biotin, Thy-1-biotin and also IL-2-biotin in small amounts, but generation of POI-mGPI was mainly limited by access to Npu^C-mGPI. PTS with other exteins than biotin and the monolipidated GPI-anchor was not achieved so far.

4.4 OPL Represents a Promising Strategy for GPI-AP Semi-Synthesis

Considering the problems and disadvantages observed in the EPL and PTS strategies, as described in section 3.7, a third strategy (One-Pot Ligation) was considered for the ligation of the GPIs to proteins. The OPL was envisioned as a new strategy overcoming the drawbacks of both EPL and PTS by circumventing the tedious process of peptide thioester formation and purification, as well as NCL to GPI-anchors. The strategy is similar to the one reported in

Vila-Perello et al. (2013),[100] where a mutated *Npu*^{C} peptide was employed for protein thioester generation from cell lysate, and ligation of this thioester to a small fluorescent peptide marker was also investigated. This method has not been investigated towards the generation of GPI-APs.

The mutated *Npu*^{C}(AA) peptide used here has a de-functionalized C-terminus lacking the essential residues Asn^{35} and Cys^{36} which were replaced by Ala residues. By using this peptide, the intein fragments undergo the association and the splicing process is arrested following the initial N-to-S acyl shift to form a protein thioester, because the transthioesterification is impeded due to the lack of the thiol group from the cysteine at the C-terminus of the mutated *Npu*^{C} peptide, Figure 4.1. This formed thioester intermediate between the C-terminus of the protein and the cysteine at the N-terminus of the *Npu*^{N} fragment can be captured in the following reaction by adding a reactive thiol which will induce formation of a protein thioester *in situ*. There is no need, however, to purify this protein thioester, which was found to be unstable during EPL investigations, see section 3.5.2. In contrast, the formed thioester can be reacted *in situ*, avoiding its manipulation and preventing its hydrolysis. Protein thioester formation is followed by a transthioesterification with the external thiol reagent and the S-to-N acyl shift with formation of a peptide bond between the extein's C-terminus and the cysteine residue of Cys-GPI. The reaction can proceed over different pathways, i.e. it is also conceivable that Cys-GPI might directly attack the thioester intermediate after the N-to-S acyl shift, although this route seems less likely since Cys-mGPI is not present in excess, other than the thiol reagent and was also less reactive.

One-Pot Ligation was developed to overcome the respective limitations of the EPL and PTS strategies, i.e. slow reaction rates in case of EPL and high synthetic effort in case of PTS. Indeed, it offers some advantages such as the easier synthesis of the mutated *Npu*^{C} peptide than for the wildtype peptide, probably due to the exchange of Cys and Asn residues to alanines. Only one purification step was required following SPPS and deprotection (HPLC) and no further manipulation of the peptide, resulting in significantly easier access to this peptide. The OPL reaction also showed reaction rates that were determined to be in between the ones of the EPL and PTS processes, reaching ca. 80% conversion after one day in comparison to ca. 50% in seven days for EPL.

In order to understand the progress of the reaction, an analysis of the reaction kinetics was carried out using Cys-biotin, see section 3.7.2. This study showed that protein thioester formation *in situ* was very fast, having rates comparable to PTS. This was observed mainly in SDS-PAGE analysis. Analysis of the same samples by HPLC and LC-MS, however, revealed

that this thioester intermediate was present for a long time, indicating that the subsequent *in situ* EPL reaction took places much slower and that indeed the pathway via the thioester intermediate was the predominant one, and not the direct attack of the Cys-GPI, Figure 4.1.

Although OPL did in fact combine advantages of both previously investigated strategies, it also showed a high level hydrolysis of the *in situ* formed protein thioester, which was dependent on the thiol used for the reaction. This also became obvious in the kinetic study mentioned above, which at the same time revealed the reason for this finding, i.e. the fast creation of the protein thioester (which had already been found to be unstable during the PL investigations) and the slow following *in situ* EPL reaction. As in EPL, this might refer to the same problem of reactivity as observed by Guo et al. (2009)[81] who observed activation of the peptide but not the ligation to the GPI mimic directly. The easy setup and purification, requiring only one purification step at the end of the ligation, makes OPL also as variable as EPL regarding its potential for the generation of a library of GPI-APs for further studies.

The product purification after the OPL reaction was a straightforward process due to a sufficient size difference between starting compounds and products. As demonstrated in the synthesis foe GFP-mGPI, the products of the reaction can be isolated using SEC or affinity chromatography. Moreover, under optimized conditions, this strategy can be used to convert all protein thioester into the product, although this was not always the case. Using more stable thiols such as MMP, hydrolysis of the thioesters was strongly reduced and the ligation products could be obtained.

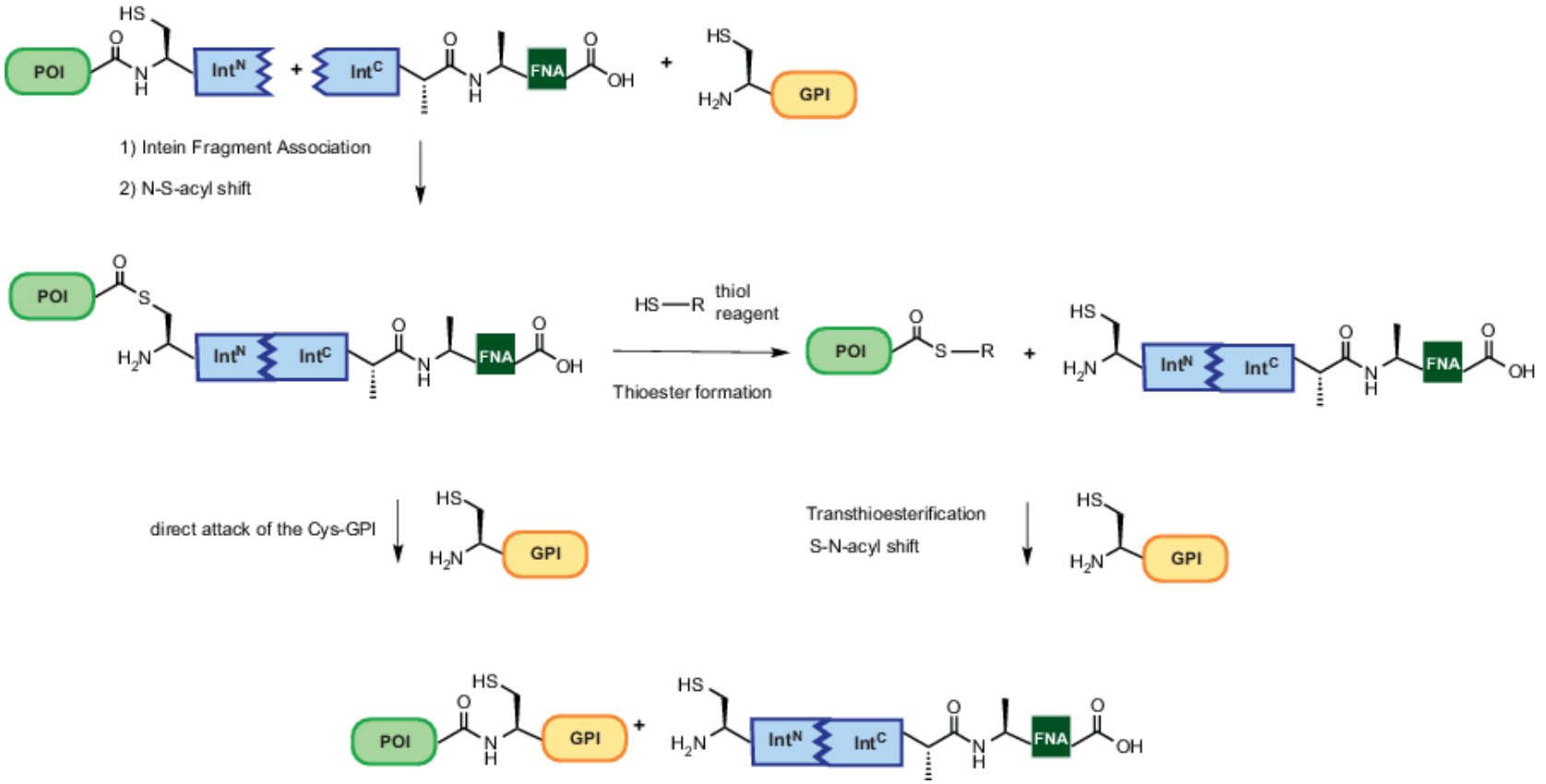


Figure 4.1: Proposed mechanism for One-Pot Ligation (OPL). Following intein fragment association and N-S-acyl shift (steps 1 and 2), two routes are suggested for the reaction progress: mainly, a protein thioester will be formed *in situ* by attack of the thiol reagent, which is present in excess. Subsequently, an EPL reaction takes place in the same reaction mixture, ligating the anchoring molecule (here a GPI molecule) to the C-terminus of the protein. A possible second route is the direct attack of the thio-group of the Cys-GPI, although this seems less likely due to the excess of thiol reagent present.

4.5 Methodological Challenges

In order to fully characterize the proteins, peptides, cysteine-containing molecules and the expected GPI-APs, various methods were applied in this work. This characterization and analysis represented a big challenge, especially for the products, because standard methods to characterize intact GPI-APs are not available. Traditionally, the analysis of GPI-APs is performed separately for the protein and the glycolipid. The analysis generally starts by removing the protein from the cell membrane by cleaving the lipid portion with phospholipases C or D. The released protein part is attached to the GPI glycan, which can then be analyzed in western blot using antibodies against the protein of interest or by proteomic methods.[173,174] Although partitioning using Triton X-114 has been reported for the separation of glypiated from unmodified proteins, the subsequent steps are likely the same as described above.[175]

Digestion by proteases (especially trypsin) with subsequent detection of the peptide fragments in mass spectrometry is a commonly used technique for protein identification and detection of modifications. This technique was also applied in this work for the detection of the generated GPI-APs. However, the C-terminal peptide fragment from the tryptic digestions from the generated proteins was too short (two amino acids) in all cases and could not be identified in MALDI-TOF-MS analysis (neither with nor without a GPI-anchor attached). Significantly shortened digestion duration in comparison to the protocol used was investigated to obtain larger fragments due to more missed cleavages, [164] however, even with this modification the detections of the C-terminus was not possible. This method was therefore also considered not suitable for the identification of our GPI-APs. It was useful, however, to prove the identity of the fusion proteins.

As an alternative to detect the glycan modification on the C-terminus, PAS (periodic acid-Schiff) staining in acrylamide gels was investigated for GPI-AP analysis, anticipating that the carbohydrate part of the GPI would result in staining. PAS staining has mainly been applied for detection of highly glycosylated proteins and is not very sensitive. Therefore, this assay also did not deliver detection of GPIs and it was not possible to stain small amounts of GPI-APs.

Another option was detection of the glycan part of the GPI in western blot using an anti-GPI antibody. Such an antibody was present in the group, resulting from earlier studies on possible anti-malarial therapies on the basis of anti-GPI antibodies.[176] The MTG4 antibody described

in this work was known to exhibit the best recognition of various GPI-derivatives while not being the most specific antibody. Initially, the MTG4 antibody was successfully used for detection of GPI-APs in western blots. However, this antibody showed a low stability and quickly lost its activity. After some successful experiments, it was so far not possible to generate a new functional batch of it from the existing clones.

Electrospray-ionization mass spectrometry in combination with HPLC was anticipated to be the most suitable method for detection and analysis of the products. This technique would allow for monitoring of the reaction progress as well as detection of the final products in an unambiguous way. While characterization of intact proteins in LC-ESI-MS is commonly performed,[177,178] detection of glypiated proteins indeed was a substantial challenge. This is mainly due different properties of the GPI-APs: the amphiphilic nature of the GPIs and presence of different functional groups with different stabilities and with different charges. These problems have been reported and recognized by other groups working towards the semi-synthesis of GPI-APs. Paulick et al. (2007)[137] reported problems with ionization of their eGFP-GPI product in ESI-MS. The hydrolyzed side product GFP-OH ionized primarily and suppressed other ions (such as the desired product). MALDI-TOF was reported to give better results, however. Wu et al. (2013)[140] were not able to acquire mass spectra of their obtained eGFP-GPI-products using SrtA, and claimed this was due to the lipid chains and phosphate groups present, referring to Harvey (2003).[179] Generally, characterization of GPIs is a topic scarcely addressed so far. Harvey (2017)[180] reports a few examples of the analysis of glycolipids in MALDI-TOF, referring mainly to GIPCs. GPIs were barely mentioned also in this extensive review on mass spectrometry of post-translationally modified proteins of all kinds.

Considering the work reported in the field of GPI-AP semi-synthesis (see section 1.2.3), it becomes clear that basically none of the groups achieved both successful generation of the products and complete characterization, displaying the difficulty of this undertaking. Therefore, mostly indirect methods were employed to demonstrate the presence of the GPI on the proteins.[138,140]

This work is no exception in this regard, although significant progress has been made in the ionization of the products and the formation of the Npu^C-mGPI. For this 39 amino acid peptide Npu^C ligated to a GPI-anchor bearing one lipid chain and two phosphate linkages, after long optimization and many efforts, it was possible to obtain decent mass spectra, even though this product was not present in very large amounts. Earlier studies performed by other groups often involved very small peptides (such as dipeptides) and/or very simple GPI-

mimics (such as dimannose or even peptidomimics). Complete characterization of the generated eGFP-GPI constructs still remains elusive, and the final GPI-bearing products, could successfully be detected in ESI-MS.

A big limitation during this work was, however, the availability of the different characterization methods at any given time, as the MTG4 antibody was unfortunately never available at the same time with an appropriate MS system and therefore both methods could not be used for cross-validation.

4.6 Structural Studies

Only a few studies have evaluated the role of glypiation in the structure of proteins. Interestingly, there are various reports in literature claiming both the presence and the absence of a structural effect of GPI-anchoring for different proteins. Becker et al. (2008) found no structural difference between their generated PrP-mGPI and PrP alone, according to CD analysis.[131] Thy-1, on the other hand, is reported to undergo conformational changes upon GPI removal.[60]

Using the ligation products with eGFP obtained by the different strategies, preliminary structural studies with and without GPI-anchors were performed using circular dichroism. As expected, no structural change was observed for eGFP as it is an intrinsically very stable protein having its C-terminus exposed at one end of the typical β-barrel structure and therefore easily accessible for manipulation.[171,181] This was true for both EPL and OPL (samples from PTS could not be analyzed due to a lack of sufficient material), demonstrating the potential of these two strategies, especially the use of OPL for the modification of folded proteins without further manipulation of the ligation products other than purification.

For the naturally GPI-anchored proteins is was not possible to perform CD-measurements as all of them were only obtained in insoluble form, or not in sufficient amounts for its refolding and for further structural studies.

5 Conclusion and Outlook

5.1 Conclusion

There is still a lack of methods for the efficient generation of pure, homogeneous GPI-anchored proteins for structural and functional studies.[34] Recently, enormous progress has been made in the total synthesis of GPIs, however the generation of GPI-anchored peptides and proteins is still a big challenge. As revealed in section 1.2.3, there are only a handful of groups active in this field and some of them are not using native or complete GPI-anchors for attachment to the proteins.

Generally, purely biological methods have either evolved over a long period of time in nature, or utilize such systems *in vitro*, making them very robust. They are, however, not always selective. Chemical Protein Synthesis can offer advantages over these methods regarding the site-specific installation of PTMs and has been the method of choice to investigate post-translational modifications on proteins because it yields more homogeneous products for the analysis of structure-function and activity relationships.[182] Strategies using semi-synthesis are emerging as an excellent possibility to extend the accessibility of all kinds of natural or unnatural modifications in a site-specific manner towards large proteins. These strategies are derived from pure chemical synthesis with the prospect to access proteins of any size, which is only possible by using (recombinant) expression in living systems. There are good examples of successful application of these combined methods, but semi-synthesis also presents some drawbacks, especially regarding the yields that are obtained.

The aim of this work was to generate large naturally glypiated protein by semi-synthetic methods. For this, initially two strategies were investigated using eGFP as a convenient model protein, i.e. Expressed Protein Ligation (EPL) and Protein *Trans*-Splicing (PTS). Both methods were successfully established for eGFP, together with the methods to characterize the reaction products. The implementation of the strategies, however, developed to an even bigger challenge than the semi-synthetic methods themselves, as direct detection (i.e. mass spectrometry and LC-MS) to characterize the highly complex protein-carbohydrate-lipid conjugates with additional phosphate linkages is not trivial. When comparing these results to reported methods (see sections 1.2.3 and 4.5), different improvements were established towards this goal in this work.

When glypiation of other, more relevant proteins was investigated using the EPL and PTS strategies (i.e. Thy-1, PrP and IL-2), it became obvious during the expression of the proteins and the first attempts to obtain the desired products, that for these insolubly expressed proteins, the success from eGFP is not fully transferrable, especially for EPL. The formation of protein thioesters was difficult under denaturing conditions and only low yields were achieved. After optimizing the conditions to obtain these proteins in solution, purification of the protein thioesters was necessary, which turned out to be difficult as well. Additionally, protein thioesters were found to be unstable and prone to hydrolysis, making additional steps (such as purification) undesirable. PTS represented a more reliable method also for insoluble proteins; however this strategy suffered from the high synthetic effort required to obtain the precursor of the ligation reaction, especially the glypiated Npu^C peptide was a major problem.

Therefore, OPL was devised to overcome these limitations, combining the easier setup from EPL with the superior robustness and speed from PTS. Indeed, the synthetic effort required was less than for PTS as the mutated Npu^C peptide was easier to synthesize and no thioester generation and ligation was required. The reaction rates in OPL were found to be intermediate in between those of EPL and PTS. Kinetic studies performed on this strategy revealed that the *in situ* protein thioester formation is indeed a very fast process, resulting from the properties of the ultrafast *Npu* split intein. The following *in situ* EPL reaction was the rate-limiting step. The low rates in this second step resulted in some hydrolysis of the thioesters, even though they did not need to be purified or manipulated in any other way. The use of thiols balancing reactivity and stability helped in overcoming this limitation. Therefore OPL is a promising tool for the semi-synthesis of GPI-anchored proteins for structural and biological studies.

5.2 **Outlook**

Two main limitations were observed during this dissertation project: the challenging detection of GPI-APs and the generation of these molecules in amounts big enough for further studies towards elucidation of the actual function of the GPI-anchor on the proteins. Some initial studies in this direction have been performed by comparing the secondary structure of the glypiated proteins by circular dichroism. However, this was only possible for eGFP and should be pursued for the other proteins, especially for proteins which are naturally glypiated. Therefore, it will be necessary to increase the amount of GPI-APs accessible by the presented

methods and as the next step, to refold the insolubly expressed proteins. Also in this direction some initial studies have been performed in this work, but have not been pursued further due to a lack of time.

It was found that the his-tag represents a more reliable means of purification than the strep-tag at least for the fusion proteins used in the PTS and OPL strategies. Also, the C-terminal attachment of the his-tag (on the C-terminus of the intein fusion) is superior over the N-terminal position, since this facilitates a more elegant and straight-forward possibility to separate reacted and unreacted proteins, as well as the cleaved Npu^N fragment.

As a next step, further studies can be envisioned in which the generated GPI-APs are incubated with cell cultures (i.e. T-cell cultures) to study the effect of the GPI-anchor in their activation (for example via cytokine release). Due to their intrinsic fluorescence, the generated eGFP-GPI products can also be used for insertion into liposomes to perform localization studies and to investigate the behavior of these molecules in their natural environment, such as the formation of lipid rafts.

5. Conclusion and Outlook

6 References

1 Crick, F. Central Dogma of Molecular Biology. *Nature*, **227**, 561-563 (1970).

2 Hinegardner, R. T. & Engelberg, J. Rationale for a Universal Genetic Code. *Science*, **142**, 1083-1085 (1963).

3 Griffith, A. J. F., Miller, J. H. & Suzuki, D. T. in *An Introduction to Genetic Analysis* (W.H.Freeman & Co Ltd, 2000).

4 Koonin, E. V. & Novozhilov, A. S. Origin and Evolution of the Genetic Code: The Universal Enigma. *Iubmb Life*, **61**, 99-111 (2009).

5 Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature*, **550**, 53 (2017).

6 Cain, J. A., Solis, N. & Cordwell, S. J. Beyond Gene Expression: The Impact of Protein Post-Translational Modifications in Bacteria. *Journal of Proteomics*, **97**, 265-286 (2014).

7 Allis, C. D. & Jenuwein, T. The Molecular Hallmarks of Epigenetic Control. *Nature Reviews Genetics*, **17**, 487 (2016).

8 Shi, Y. Mechanistic Insights into Precursor Messenger Rna Splicing by the Spliceosome. *Nature Reviews Molecular Cell Biology*, **18**, 655 (2017).

9 Wang, Y. A. N., Liu, J., Huang, B. O., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J. I. N., Zhang, J., Min, Q.-H., Yang, W.-M. & Wang, X.-Z. Mechanism of Alternative Splicing and Its Regulation. *Biomedical Reports*, **3**, 152-158 (2015).

10 Walsh Christopher, T., Garneau-Tsodikova, S. & Gatto Gregory, J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie International Edition*, **44**, 7342-7372 (2005).

11 Karve, T. M. & Cheema, A. K. Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease. *Journal of Amino Acids* (2011).

12 Davis, B. G. Mimicking Posttranslational Modifications of Proteins. *Science*, **303**, 480 (2004).

13 Chuh, Kelly N., Batt, Anna R. & Pratt, Matthew R. Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chemical Biology*, **23**, 86-107 (2016).

14 Jensen, O. N. Interpreting the Protein Language Using Proteomics. *Nat Rev Mol Cell Biol*, **7**, 391-403 (2006).

15 Basak, S., Lu, C. & Basak, A. Post-Translational Protein Modifications of Rare and Unconventional Types: Implications in Functions and Diseases. *Current medicinal chemistry*, **23**, 714-745 (2016).

16 Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-Translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding. *Wiley interdisciplinary reviews. Systems biology and medicine*, **4**, 565-583 (2012).

17 Harmel, R. & Fiedler, D. Features and Regulation of Non-Enzymatic Post-Translational Modifications. *Nature Chemical Biology*, **14**, 244 (2018).

18 Pagel, O., Loroch, S., Sickmann, A. & Zahedi, R. P. Current Strategies and Findings in Clinically Relevant Post-Translational Modification-Specific Proteomics. *Expert Review of Proteomics*, **12**, 235-253 (2015).

19 Zhao, Y. & Jensen, O. N. Modification-Specific Proteomics: Strategies for Characterization of Post-Translational Modifications Using Enrichment Techniques. *Proteomics*, **9**, 4632-4641 (2009).

20 Vucic, D., Dixit, V. M. & Wertz, I. E. Ubiquitylation in Apoptosis: A Post-Translational Modification at the Edge of Life and Death. *Nature Reviews Molecular Cell Biology*, **12**, 439 (2011).

21 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate Protein Glycosylation: Diversity, Synthesis and Function. *Nature Reviews Molecular Cell Biology*, **13**, 448 (2012).

22 Cummings, R. D. The Repertoire of Glycan Determinants in the Human Glycome. *Molecular bioSystems*, **5**, 1087-1104 (2009).

23 Ruddock, L. W. & Molinari, M. N-Glycan Processing in Er Quality Control. *Journal of Cell Science*, **119**, 4373 (2006).

24 Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chemical Reviews*, **96**, 683-720 (1996).

25 Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the Immune System. *Science*, **291**, 2370 (2001).

26 Orlean, P. & Menon, A. K. Thematic Review Series: Lipid Posttranslational Modifications. Gpi Anchoring of Protein in Yeast and Mammalian Cells, Or: How We Learned to Stop Worrying and Love Glycophospholipids. *Journal of Lipid Research*, **48**, 993-1011 (2007).

27 Heider, S., Dangerfield, J. A. & Metzner, C. Biomedical Applications of Glycosylphosphatidylinositol-Anchored Proteins. *J. Lipid Res.* , **57**, 1778-1788 (2016).

28 Spiro, R. G. Protein Glycosylation: Nature, Distribution, Enzymatic Formation, and Disease Implications of Glycopeptide Bonds. *Glycobiology*, **12**, 43R-56R (2002).

29 Low, M. G. & Saltiel, A. R. Structural and Functional Roles of Glycosyl-Phosphatidylinositol in Membranes. *Science*, **239**, 268-275 (1988).

30 Kinoshita, T. Biosynthesis and Deficiencies of Glycosylphosphatidylinositol. *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, **90**, 130-143 (2014).

31 Maeda, Y. & Kinoshita, T. Structural Remodeling, Trafficking and Functions of Glycosylphosphatidylinositol-Anchored Proteins. *Progress in Lipid Research*, **50**, 411-424 (2011).

32 Ikezawa, H. Glycosylphosphatidylinositol (Gpi)-Anchored Proteins. *Biological and Pharmaceutical Bulletin*, **25**, 409-417 (2002).

33 Munro, S. Lipid Rafts: Elusive or Illusive? *Cell*, **115**, 377-388 (2003).

34 Yu, S., Guo, Z., Johnson, C., Gu, G. & Wu, Q. Recent Progress in Synthetic and Biological Studies of Gpi Anchors and Gpi-Anchored Proteins. *Current opinion in chemical biology*, **17**, 10.1016/j.cbpa.2013.1009.1016 (2013).

35 Tsai, Y.-H., Liu, X. & Seeberger, P. H. Chemical Biology of Glycosylphosphatidylinositol Anchors. *Angewandte Chemie International Edition*, **51**, 11438-11456 (2012).

36 Paulick, M. G. & Bertozzi, C. R. The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. *Biochemistry*, **47**, 6991-7000 (2008).

37 Paulick, M. G., Forstner, M. B., Groves, J. T. & Bertozzi, C. R. A Chemical Approach to Unraveling the Biological Function of the Glycosylphosphatidylinositol Anchor. *PNAS*, **104**, 20332-20337 (2007).

38 Wu, Z., Guo, X. & Guo, Z. Chemoenzymatic Synthesis of Glycosylphosphatidylinositol-Anchored Glycopeptides. *Chemical Communications*, **46**, 5773-5774 (2010).

39 Kinoshita, T., Fujita, M. & Maeda, Y. Biosynthesis, Remodelling and Functions of Mammalian Gpi-Anchored Proteins: Recent Progress. *The Journal of Biochemistry*, **144**, 287-294 (2008).

40 Silva, D. V. Zwischen Protein Und Membran. *Nachrichten aus der Chemie*, **61**, 882-886 (2013).

41 Legler, D. F., Doucey, M.-A., Schneider, P., Chapatte, L., Bender, F. C. & Bron, C. Differential Insertion of Gpi-Anchored Gfps into Lipid Rafts of Live Cells. *The FASEB Journal* (2004).

42 Sezgin, E., Levental, I., Mayor, S. & Eggeling, C. The Mystery of Membrane Organization: Composition, Regulation and Roles of Lipid Rafts. *Nature Reviews Molecular Cell Biology*, **18**, 361 (2017).

43 Kinoshita, T., Ohishi, K. & Takeda, J. Gpi-Anchor Synthesis in Mammalian Cells: Genes, Their Products, and a Deficiency. *Journal of biochemistry*, **122**, 251-257 (1997).

44 Udenfriend, S. & Kodukula, K. How Glycosylphosphatidylinositol-Anchored Membrane Proteins Are Made. *Annual review of biochemistry*, **64**, 563-591 (1995).

45 Puig, B., Altmeppen, H. & Glatzel, M. The Gpi-Anchoring of Prp. *Prion*, **8**, 11-18 (2014).

46 Taylor, D. R. & Hooper, N. M. in *Post-Translational Modifications in Health and Disease* (ed Cecilio J. Vidal) 39-55 (Springer New York, 2011).

47 Tsai, Y.-H., Gotze, S., Vilotijevic, I., Grube, M., Silva, D. V. & Seeberger, P. H. A General and Convergent Synthesis of Diverse Glycosylphosphatidylinositol Glycolipids. *Chemical Science*, **4**, 468-481 (2012).

48 Tsai, Y.-H., Götze, S., Azzouz, N., Hahm, H. S., Seeberger, P. H. & Varón Silva, D. A General Method for Synthesis of Gpi Anchors Illustrated by the Total Synthesis of the Low-Molecular-Weight Antigen from *Toxoplasma Gondii*. *Angew. Chem. Int. Ed.*, **50**, 9961-9964 (2011).

49 Almeida, A., Layton, M. & Karadimitris, A. Inherited Glycosylphosphatidyl Inositol Deficiency: A Treatable Cdg. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, **1792**, 874-880 (2009).

50 Rother, R. P., Rollins, S. A., Mojcik, C. F., Brodsky, R. A. & Bell, L. Discovery and Development of the Complement Inhibitor Eculizumab for the Treatment of Paroxysmal Nocturnal Hemoglobinuria. *Nat Biotech*, **25**, 1256-1264 (2007).

51 Phillips, M. A., Burrows, J. N., Manyando, C., van Huijsduijnen, R. H., Van Voorhis, W. C. & Wells, T. N. C. Malaria. *Nature Reviews Disease Primers*, **3**, 17050 (2017).

52 Gotze, S., Azzouz, N., Tsai, Y. H., Gross, U., Reinhardt, A., Anish, C., Seeberger, P. H. & Varon Silva, D. Diagnosis of Toxoplasmosis Using a Synthetic Glycosylphosphatidylinositol Glycan. *Angewandte Chemie (International ed. in English)*, **53**, 13701-13705 (2014).

53 Gotze, S., Reinhardt, A., Geissner, A., Azzouz, N., Tsai, Y. H., Kurucz, R., Varon Silva, D. & Seeberger, P. H. Investigation of the Protective Properties of Glycosylphosphatidylinositol-Based Vaccine Candidates in a Toxoplasma Gondii Mouse Challenge Model. *Glycobiology*, **25**, 984-991 (2015).

54 Zhao, Y., Su, H., Shen, X., Du, J., Zhang, X. & Zhao, Y. The Immunological Function of Cd52 and Its Targeting in Organ Transplantation. *Inflammation research : official journal of the European Histamine Research Society ... [et al.]*, **66**, 571-578 (2017).

55 Bandala-Sanchez, E., Zhang, Y., Reinwald, S., Dromey, J. A., Lee, B. H., Qian, J., Bohmer, R. M. & Harrison, L. C. T Cell Regulation Mediated by Interaction of Soluble Cd52 with the Inhibitory Receptor Siglec-10. *Nature immunology*, **14**, 741-748 (2013).

56 Tanaka, Y., Nakahara, Y., Hojo, H. & Nakahara, Y. Studies Directed toward the Synthesis of Protein-Bound Gpi Anchor. *Tetrahedron*, **59**, 4059-4067 (2003).

57 Xue, J., Shao, N. & Guo, Z. First Total Synthesis of a Gpi-Anchored Peptide. *The Journal of Organic Chemistry*, **68**, 4020-4029 (2003).

58 Ruiz-Argüelles, A. & Llorente, L. The Role of Complement Regulatory Proteins (Cd55 and Cd59) in the Pathogenesis of Autoimmune Hemocytopenias. *Autoimmunity Reviews*, **6**, 155-161 (2007).

59 Yu, Q., Yu, R. & Qin, X. The Good and the Evil of Complement Activation in Hiv-1 Infection. *Cellular & Moelcular Immunology*, **7**, 331-340 (2010).

60 Barboni, E., Rivero, B. P., George, A. J., Martin, S. R., Renoup, D. V., Hounsell, E. F., Barber, P. C. & Morris, R. J. The Glycophosphatidylinositol Anchor Affects the Conformation of Thy-1 Protein. *Journal of Cell Science*, **108**, 487 (1995).

61 Williams, A. & Gagnon, J. Neuronal Cell Thy-1 Glycoprotein: Homology with Immunoglobulin. *Science*, **216**, 696-703 (1982).

62 Low, M. G. & Kincade, P. W. Phosphatidylinositol Is the Membrane-Anchoring Domain of the Thy-1 Glycoprotein. *Nature*, **318**, 62-64 (1985).

63 Tse, A. G. D., Barclay, A. N., Watts, A. & Williams, A. F. A Glycophospholipid Tail at the Carboxyl Terminus of the Thy-1 Glycoprotein of Neurons and Thymocytes. *Science*, **230**, 1003-1008 (1985).

64 Jasnow, A. M., Ehrlich, D. E., Choi, D. C., Dabrowska, J., Bowers, M. E., McCullough, K. M., Rainnie, D. G. & Ressler, K. J. Thy1-Expressing Neurons in the Basolateral Amygdala May Mediate Fear Inhibition. *The Journal of Neuroscience*, **33**, 10396 (2013).

65 Kumar, A., Bhanja, A., Bhattacharyya, J. & Jaganathan, B. G. Multiple Roles of Cd90 in Cancer. *Tumor Biology*, **37**, 11611-11622 (2016).

66 White, N. J., Pukrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A. & Dondorp, A. M. Malaria. *The Lancet*, **383**, 723-735 (2013).

67 Kadekoppala, M. & Holder, A. A. Merozoite Surface Proteins of the Malaria Parasite: The Msp1 Complex and the Msp7 Family. *International journal for parasitology*, **40**, 1155-1161 (2010).

68 Holder, A. A., Blackman, M. J., Burghaus, P. A., Chappel, J. A., Ling, I. T., McCallum-Deighton, N. & Shai, S. A Malaria Merozoite Surface Protein (Msp1)-Structure, Processing and Function. *Memorias do Instituto Oswaldo Cruz*, **87 Suppl 3**, 37-42 (1992).

69 Moss, D. K., Remarque, E. J., Faber, B. W., Cavanagh, D. R., Arnot, D. E., Thomas, A. W. & Holder, A. A. Plasmodium Falciparum 19-Kilodalton Merozoite Surface Protein 1 (Msp1)-Specific Antibodies That Interfere with Parasite Growth in Vitro Can Inhibit Msp1 Processing, Merozoite Invasion, and Intracellular Parasite Development. *Infection and immunity*, **80**, 1280-1287 (2012).

70 Bisseye, C., Yindom, L. M., Simpore, J., Morgan, W. D., Holder, A. A. & Ismaili, J. An Engineered Plasmodium Falciparum C-Terminal 19-Kilodalton Merozoite Surface Protein 1 Vaccine Candidate Induces High Levels of Interferon-Gamma Production Associated with Cellular Immune Responses to Specific Peptide Sequences in Gambian Adults Naturally Exposed to Malaria. *Clinical and experimental immunology*, **166**, 366-373 (2011).

71 Aguzzi, A. & Calella, A. M. Prions: Protein Aggregation and Infectious Diseases. *Physiological Reviews*, **89**, 1105 (2009).

72 Prusiner, S. Novel Proteinaceous Infectious Particles Cause Scrapie. *Science*, **216**, 136-144 (1982).

73 Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. Prion Protein Biology. *Cell*, **93**, 337-348 (1998).

74 Stahl, N., Borchelt, D. R., Hsiao, K. & Prusiner, S. B. Scrapie Prion Protein Contains a Phosphatidylinositol Glycolipid. *Cell*, **51**, 229-240 (1987).

75 Aguzzi, A. & Heppner, F. L. Pathogenesis of Prion Diseases: A Progress Report. *Cell death and differentiation*, **7**, 889-902 (2000).

76 Aguzzi, A., Sigurdson, C. & Heikenwaelder, M. Molecular Mechanisms of Prion Pathogenesis. *Annual Review of Pathology: Mechanisms of Disease*, **3**, 11-40 (2008).

77 Lin, S. J., Yu, K. H., Wu, J. R., Lee, C. F., Jheng, C. P., Chen, H. R. & Lee, C. I. Liberation of Gpi-Anchored Prion from Phospholipids Accelerates Amyloidogenic Conversion. *International journal of molecular sciences*, **14**, 17943-17957 (2013).

78 Laurén, J., Gimbel, D. A., Nygaard, H. B., Gilbert, J. W. & Strittmatter, S. M. Cellular Prion Protein Mediates Impairment of Synaptic Plasticity by Amyloid-B Oligomers. *Nature*, **457**, 1128 (2009).

79 Ow, S. Y. & Dunstan, D. E. A Brief Overview of Amyloids and Alzheimer's Disease. *Protein science : a publication of the Protein Society*, **23**, 1315-1331 (2014).

80 Belay, E. D. Transmissible Spongiform Encephalopathies in Humans. *Annual Review of Microbiology*, **53**, 283-314 (1999).

81 Guo, X., Wang, Q., Swarts, B. M. & Guo, Z. Sortase-Catalyzed Peptide−Glycosylphosphatidylinositol Analogue Ligation. *Journal of the American Chemical Society*, **131**, 9878-9879 (2009).

82 Müller, G. Novel Applications for Glycosylphosphatidylinositol-Anchored Proteins in Pharmaceutical and Industrial Biotechnology. *Molecular Membrane Biology*, **28**, 187-205 (2011).

83 Antosova, Z., Mackova, M., Kral, V. & Macek, T. Therapeutic Application of Peptides and Proteins: Parenteral Forever? *Trends in biotechnology*, **27**, 628-635 (2009).

84 Lagasse, H. A., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E. & Kimchi-Sarfaty, C. Recent Advances in (Therapeutic Protein) Drug Development. *F1000Research*, **6**, 113 (2017).

85 Bachmann, M. F. & Oxenius, A. Interleukin 2: From Immunostimulation to Immunoregulation and Back Again. *EMBO reports*, **8**, 1142-1148 (2007).

86 *Aldesleukin*, <https://livertox.nih.gov/Aldesleukin.htm>.

87 Ji, J., Li, J., Holmes, L. M., Burgin, K. E., Yu, X., Wagner, T. E. & Wei, Y. Glycoinositol Phospholipid-Anchored Interleukin 2 but Not Secreted Interleukin 2 Inhibits Melanoma Tumor Growth in Mice. *Molecular cancer therapeutics*, **1**, 1019-1024 (2002).

88 Isaacs, A. & Lindemann, J. Virus Interference. I. The Interferon. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, **147**, 258 (1957).

89 Platanias, L. C. Mechanisms of Type-I- and Type-Ii-Interferon-Mediated Signalling. *Nature reviews. Immunology*, **5**, 375-386 (2005).

90 Vazquez, N., Schmeißer, H., Dolan, M. A., Bekisz, J., Zoon, K. C. & Wahl, S. M. Structural Variants of Ifnα Preferentially Promote Antiviral Functions. *Blood*, **118**, 2567-2577 (2011).

91 Assenberg, R., Wan, P. T., Geisse, S. & Mayr, L. M. Advances in Recombinant Protein Expression for Use in Pharmaceutical Research. *Current Opinion in Structural Biology*, **23**, 393-402 (2013).

92 Chen, Z. & Cole, P. A. Synthetic Approaches to Protein Phosphorylation. *Current Opinion in Chemical Biology*, **28**, 115-122 (2015).

93 Krall, N., da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-Selective Protein-Modification Chemistry for Basic Biology and Drug Development. *Nat Chem*, **8**, 103-113 (2016).

94 Mootz, H. D. Split Inteins as Versatile Tools for Protein Semisynthesis. *ChemBioChem*, **10**, 2579-2589 (2009).

95 Chen, M., Heimer, P. & Imhof, D. Synthetic Strategies for Polypeptides and Proteins by Chemical Ligation. *Amino acids*, **47**, 1283-1299 (2015).

96 Guan, X., Chaffey, P. K., Zeng, C. & Tan, Z. in *Protein Ligation and Total Synthesis Ii* (ed Lei Liu) 155-192 (Springer International Publishing, 2015).

97 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science*, **292**, 498-500 (2001).

98 Oza, J. P., Aerni, H. R., Pirman, N. L., Barber, K. W., ter Haar, C. M., Rogulina, S., Amrofell, M. B., Isaacs, F. J., Rinehart, J. & Jewett, M. C. Robust Production of Recombinant Phosphoproteins Using Cell-Free Protein Synthesis. *Nature Communications*, **6**, 8168 (2015).

99 Severinov, K. & Muir, T. W. Expressed Protein Ligation, a Novel Method for Studying Protein-Protein Interactions in Transcription. *Journal of Biological Chemistry*, **273**, 16205-16209 (1998).

100 Vila-Perelló, M., Liu, Z., Shah, N. H., Willis, J. A., Idoyaga, J. & Muir, T. W. Streamlined Expressed Protein Ligation Using Split Inteins. *Journal of the American Chemical Society*, **135**, 286-292 (2013).

101 Patterson, D. M. & Prescher, J. A. Orthogonal Bioorthogonal Chemistries. *Current Opinion in Chemical Biology*, **28**, 141-149 (2015).

102 Elleuche, S. & Pöggeler, S. Inteins, Valuable Genetic Elements in Molecular Biology and Biotechnology. *Applied Microbiology and Biotechnology*, **87**, 479-489 (2010).

103 Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. & Belfort, M. Protein Splicing Elements: Inteins and Exteins--a Definition of Terms and Recommended Nomenclature. *Nucleic acids research*, **22**, 1125 (1994).

104 Anraku, Y., Mizutani, R. & Satow, Y. Protein Splicing: Its Discovery and Structural Insight into Novel Chemical Mechanisms. *IUBMB life*, **57**, 563-574 (2005).

105 Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. & Stevens, T. H. Protein Splicing Converts the Yeast Tfp1 Gene Product to the 69-Kd Subunit of the Vacuolar H (+)-Adenosine Triphosphatase. *Science*, **250**, 651-657 (1990).

106 Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. Molecular Structure of a Gene, Vma1, Encoding the Catalytic Subunit of H (+)-Translocating Adenosine Triphosphatase from Vacuolar Membranes of Saccharomyces Cerevisiae. *Journal of Biological Chemistry*, **265**, 6726-6733 (1990).

107 Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L. & Hilario, E. Inteins: Structure, Function, and Evolution. *Annual Reviews in Microbiology*, **56**, 263-287 (2002).

108 Shah, N. H. & Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chemical science (Royal Society of Chemistry : 2010)*, **5**, 446-461 (2014).

109 Perler, F. B. Inbase: The Intein Database. *Nucleic Acids Research*, **30**, 383-384 (2002).

110 Zettler, J. *Wechselwirkungen Zwischen Adenylierungs- Und Peptidyl Carrier Protein-Domänen in Nicht-Ribosomalen Peptidsynthesen Sowie Biochemische Und Strukturelle Untersuchungen Zu Gespaltenen Inteinen* Doctoral thesis, (2010).

111 Kawasaki, M., Makino, S.-i., Matsuzawa, H., Satow, Y., Ohya, Y. & Anraku, Y. Folding-Dependent *in Vitro* Protein Splicing of the *Saccharomyces Cerevisiae Vma1* Protozyme. *Biochem Bioph Res Co*, **222**, 827-832 (1996).

112 Wu, H., Hu, Z. & Liu, X.-Q. Protein *Trans*-Splicing by a Split Intein Encoded in a Split Dnae Gene of *Synechocystis* Sp. Pcc6803. *Proc Natl Adac Sci USA*, **95**, 9226-9231 (1998).

113 Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Naturally Split Inteins Assemble through a "Capture and Collapse" Mechanism. *J Am Chem Soc*, **135**, 18673-18681 (2013).

114 Wu, Q., Gao, Z., Wei, Y., Zheng, Y., Dong, Y. & Liu, Y. Conserved Residues That Modulate Protein *Trans*-Splicing of *Npu* Dnae Split Intein. *Biochem J*, **461**, 247-255 (2014).

115 Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Naturally Split Inteins Assemble through a "Capture and Collapse" Mechanism. *Journal of the American Chemical Society*, **135**, 18673-18681 (2013).

116 Dhar, T. & Mootz, H. D. Modification of Transmembrane and Gpi-Anchored Proteins on Living Cells by Efficient Protein *Trans*-Splicing Using the *Npu* Dnae Intein. *Chem Commun*, **47**, 3063-3065 (2011).

117 Shah, N. H. & Muir, T. W. Split Inteins: Nature's Protein Ligases. *Israel journal of chemistry*, **51**, 854-861 (2011).

118 Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Extein Residues Play an Intimate Role in the Rate-Limiting Step of Protein Trans-Splicing. *Journal of the American Chemical Society*, **135**, 5839-5847 (2013).

119 Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Stephen, B. H. K. Synthesis of Proteins by Native Chemical Ligation. *Science*, **266**, 776-779 (1994).

120 Muir, T. W. Semisynthesis of Proteins by Expressed Protein Ligation. *Annual review of biochemistry*, **72**, 249-289 (2003).

121 Muir, T. W., Sondhi, D. & Cole, P. A. Expressed Protein Ligation: A General Method for Protein Engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6705-6710 (1998).

122 David, R., Richter, M. P. & Beck-Sickinger, A. G. Expressed Protein Ligation. Method and Applications. *European journal of biochemistry*, **271**, 663-677 (2004).

123 Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H. & Xu, M.-Q. Single-Column Purification of Free Recombinant Proteins Using a Self-Cleavable Affinity Tag Derived from a Protein Splicing Element. *Gene*, **192**, 271-281 (1997).

124 Reif, A., Siebenhaar, S., Troster, A., Schmalzlein, M., Lechner, C., Velisetty, P., Gottwald, K., Pohner, C., Boos, I., Schubert, V., Rose-John, S. & Unverzagt, C. Semisynthesis of Biologically Active Glycoforms of the Human Cytokine Interleukin 6. *Angewandte Chemie (International ed. in English)*, **53**, 12125-12131 (2014).

125 Johnson, E. C. & Kent, S. B. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J Am Chem Soc*, **128**, 6640-6646 (2006).

126 Ingale, S., Buskas, T. & Boons, G.-J. Synthesis of Glyco(Lipo)Peptides by Liposome-Mediated Native Chemical Ligation. *Organic Letters*, **8**, 5785-5788 (2006).

127 Dawson, P. E., Churchill, M. J., Ghadiri, M. R. & Kent, S. B. H. Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives. *Journal of the American Chemical Society*, **119**, 4325-4329 (1997).

128 Grube, M., Lee, B. Y., Garg, M., Michel, D., Vilotijević, I., Malik, A., Seeberger Peter, H. & Varón Silva, D. Synthesis of Galactosylated Glycosylphosphatidylinositol

Derivatives from Trypanosoma Brucei. *Chemistry – A European Journal*, **24**, 3271-3282 (2018).

129 Lee, B. Y., Seeberger, P. H. & Varon Silva, D. Synthesis of Glycosylphosphatidylinositol (Gpi)-Anchor Glycolipids Bearing Unsaturated Lipids. *Chemical Communications*, **52**, 1586-1589 (2016).

130 Swarts, B. M. & Guo, Z. Synthesis of a Glycosylphosphatidylinositol Anchor Bearing Unsaturated Lipid Chains. *Journal of the American Chemical Society*, **132**, 6648-6650 (2010).

131 Becker, C. F., Liu, X., Olschewski, D., Castelli, R., Seidel, R. & Seeberger, P. H. Semisynthesis of a Glycosylphosphatidylinositol-Anchored Prion Protein. *Angewandte Chemie (International ed. in English)*, **47**, 8215-8219 (2008).

132 Olschewski, D., Seidel, R., Miesbauer, M., Rambold, A. S., Oesterhelt, D., Winklhofer, K. F., Tatzelt, J., Engelhard, M. & Becker, C. F. Semisynthetic Murine Prion Protein Equipped with a Gpi Anchor Mimic Incorporates into Cellular Membranes. *Chemistry & biology*, **14**, 994-1006 (2007).

133 Shao, N., Xue, J. & Guo, Z. Chemical Synthesis of a Skeleton Structure of Sperm Cd52—a Gpi-Anchored Glycopeptide. *Angewandte Chemie International Edition*, **43**, 1569-1573 (2004).

134 Dawson, P. E. & Kent, S. B. H. Synthesis of Native Proteins by Chemical Ligation. *Annual review of biochemistry*, **69**, 923-960 (2000).

135 Fang, G. M., Li, Y. M., Shen, F., Huang, Y. C., Li, J. B., Lin, Y., Cui, H. K. & Liu, L. Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angewandte Chemie (International ed. in English)*, **50**, 7645-7649 (2011).

136 Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P. & Liu, L. Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protocols*, **8**, 2483-2495 (2013).

137 Paulick, M. G., Wise, A. R., Forstner, M. B., Groves, J. T. & Bertozzi, C. R. Synthetic Analogues of Glycosylphosphatidylinositol-Anchored Proteins and Their Behavior in Supported Lipid Bilayers. *J. Am. Chem. Soc.* , **129**, 11543-11550 (2007).

138 Dhar, T. & Mootz, H. D. Modification of Transmembrane and Gpi-Anchored Proteins on Living Cells by Efficient Protein Trans-Splicing Using the Npu Dnae Intein. *Chemical communications (Cambridge, England)*, **47**, 3063-3065 (2011).

139 Ling, J. J., Policarpo, R. L., Rabideau, A. E., Liao, X. & Pentelute, B. L. Protein Thioester Synthesis Enabled by Sortase. *Journal of the American Chemical Society*, **134**, 10749-10752 (2012).

140 Wu, Z., Guo, X., Gao, J. & Guo, Z. Sortase a-Mediated Chemoenzymatic Synthesis of Complex Glycosylphosphatidylinositol-Anchored Protein. *Chemical Communications*, **49**, 11689-11691 (2013).

141 Zhu, S. & Guo, Z. Chemical Synthesis of Gpi Glycan–Peptide Conjugates by Traceless Staudinger Ligation. *Organic Letters*, **19**, 3063-3066 (2017).

142 Marbach, J., Zentis, P., Ellinger, P., Müller, H. & Birkmann, E. Expression and Characterization of Fully Posttranslationally Modified Cellular Prion Protein in *Pichia Pastoris*. *Biological Chemistry*, **394**, 1475-1483 (2013).

143 Harrison, P. T., Hutchinson, M. J. & Allen, J. M. A Convenient Method for the Construction and Expression of Gpi-Anchored Proteins. *Nucleic Acids Research*, **22**, 3813-3814 (1994).

144 Shams-Eldin, H., Azzouz, N., Niehus, S., Smith, T. K. & Schwarz, R. T. An Efficient Method to Express Gpi-Anchor Proteins in Insect Cells. *Biochem Biophys Res Commun*, **365**, 657-663 (2008).

145 Zettler, J., Schütz, V. & Mootz, H. D. The Naturally Split Npu Dnae Intein Exhibits an Extraordinarily High Rate in the Protein Trans-Splicing Reaction. *FEBS Letters*, **583**, 909-914 (2009).

146 Piontek, C. *Molekularbiologische Gewinnung Von Rnase 40-124 Fragmenten Zur Synthese Con Einheitlichen Glycoproteinen Durch Native Chemische Ligation* Dr. rer. nat. thesis, Universität Bayreuth, (2007).

147 Schmidt, T. G. M. & Skerra, A. The Strep-Tag System for One-Step Purification and High-Affinity Detection or Capturing of Proteins. *Nat. Protoc.*, **2**, 1528-1535 (2007).

148 Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. & Stüber, D. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio/Technology*, **6**, 1321 (1988).

149 Van Dyke, M. W., Sirito, M. & Sawadogo, M. Single-Step Purification of Bacterially Expressed Polypeptides Containing an Oligo-Histidine Domain. *Gene*, **111**, 99-104 (1992).

150 Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J. & Schäfer, F. in *Methods in Enzymology* Vol. 463 (eds Richard R. Burgess & Murray P. Deutscher) 439-473 (Academic Press, 2009).

151 Hengen, P. N. Purification of His-Tag Fusion Proteins from Escherichia Coli. *Trends in Biochemical Sciences*, **20**, 285-286 (1995).

152 Mitchell, S. F. & Lorsch, J. R. Protein Affinity Purification Using Intein/Chitin Binding Protein Tags. *Methods in enzymology*, **559**, 111-125 (2015).

153 Josic, D. & Kovac, S. Reversed-Phase High Performance Liquid Chromatography of Proteins. *Curr Protoc Protein Sci*, **Chapter 8**, Unit 8 7 (2010).

154 *Phenomenex* *Website*, <http://www.phenomenex.com/Products/HPLCDetail/Jupiter/C4>.

155 Zettler, J., Schutz, V. & Mootz, H. D. The Naturally Split Npu Dnae Intein Exhibits an Extraordinarily High Rate in the Protein Trans-Splicing Reaction. *FEBS Lett*, **583**, 909-914 (2009).

156 Bordier, C. Phase-Separation of Integral Membrane-Proteins in Triton X-114 Solution. *Journal of Biological Chemistry*, **256**, 1604-1607 (1981).

157 Paulick, M. G., Wise, A. R., Forstner, M. B., Groves, J. T. & Bertozzi, C. R. Synthetic Analogues of Glycosylphosphatidylinositol-Anchored Proteins and Their Behavior in Supported Lipid Bilayers. *Journal of the American Chemical Society*, **129**, 11543-11550 (2007).

158 Shapiro, A. L., Viñuela, E. & V. Maizel, J. Molecular Weight Estimation of Polypeptide Chains by Electrophoresis in Sds-Polyacrylamide Gels. *Biochemical and Biophysical Research Communications*, **28**, 815-820 (1967).

159 Laemmli, U. K. Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature*, **227**, 680-& (1970).

160 Karas, M., Bachmann, D., Bahr, U. & Hillenkamp, F. Matrix-Assisted Ultraviolet-Laser Desorption of Nonvolatile Compounds. *Int. J. Mass Spectrom. Ion Process.*, **78**, 53-68 (1987).

161 Karas, M. & Krüger, R. Ion Formation in Maldi: The Cluster Ionization Mechanism. *Chemical Reviews*, **103**, 427-440 (2003).

162 Pitt, J. J. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews*, **30**, 19-34 (2009).

163 Kebarle, P. & Verkerk, U. H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spectrometry Reviews*, **28**, 898-917 (2009).

164 Kolarich, D., Jensen, P. H., Altmann, F. & Packer, N. H. Determination of Site-Specific Glycan Heterogeneity on Glycoproteins. *Nat Protoc*, **7**, 1285-1298 (2012).

165 Bioinformatics, S. S. I. o. <https://web.expasy.org/peptide_mass/>.

166 Kelly, S. M., Jess, T. J. & Price, N. C. How to Study Proteins by Circular Dichroism. *Biochimica et biophysica acta*, **1751**, 119-139 (2005).

167 Green, M. R. & Sambrook, J. *Molecular Cloning: A Laboratory Manual, Fourth Edition*. (Cold Spring Harbor Laboratory Press, 2012).

168 Hackeng, T. M., Griffin, J. H. & Dawson, P. E. Protein Synthesis by Native Chemical Ligation: Expanded Scope by Using Straightforward Methodology. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 10068-10073 (1999).

169 Dahlquist, F. W., Jao, L. & Raftery, M. On the Binding of Chitin Oligosaccharides to Lysozyme. *Proc Natl Acad Sci U S A*, **56**, 26-30 (1966).

170 Sherman, F., Stewart, J. W. & Tsunasawa, S. Methionine or Not Methionine at the Beginning of a Protein. *BioEssays : news and reviews in molecular, cellular and developmental biology*, **3**, 27-31 (1985).

171 Yang, F., Moss, L. G. & Phillips, G. N., Jr. The Molecular Structure of Green Fluorescent Protein. *Nature biotechnology*, **14**, 1246-1251 (1996).

172 Stevens, A. J., Sekar, G., Shah, N. H., Mostafavi, A. Z., Cowburn, D. & Muir, T. W. A Promiscuous Split Intein with Expanded Protein Engineering Applications. *Proceedings of the National Academy of Sciences*, **114**, 8538-8543 (2017).

173 Elortza, F., Mohammed, S., Bunkenborg, J., Foster, L. J., Nuhse, T. S., Brodbeck, U., Peck, S. C. & Jensen, O. N. Modification-Specific Proteomics of Plasma Membrane Proteins: Identification and Characterization of Glycosylphosphatidylinositol-Anchored Proteins Released Upon Phospholipase D Treatment. *Journal of proteome research*, **5**, 935-943 (2006).

174 Nishina, K. A. & Supattapone, S. Immunodetection of Glycophosphatidylinositol-Anchored Proteins Following Treatment with Phospholipase C. *Anal Biochem*, **363**, 318-320 (2007).

175 Doering, T. L., Englund, P. T. & Hart, G. W. in *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., 2001).

176 Kurucz, R. *Generation and Characterization of Monoclonal Antibodies Directed against Synthetic P. Falciparum Glycosylphosphatidylinositol Glycans* Dr. rer. nat. thesis, Freie Universität Berlin, (2014).

177 Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durbin, K. R. & Kelleher, N. L. Analysis of Intact Protein Isoforms by Mass Spectrometry. *Journal of Biological Chemistry*, **286**, 25451-25458 (2011).

178 Whitelegge, J. Intact Protein Mass Spectrometry and Top-Down Proteomics. *Expert Rev Proteomics*, **10**, 127-129 (2013).

179 Harvey David, J. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Carbohydrates and Glycoconjugates. *ChemInform*, **34** (2003).

180 Harvey, D. J. Analysis of Carbohydrates and Glycoconjugates by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: An Update for 2011–2012. *Mass Spectrometry Reviews*, **36**, 255-422 (2017).

181 Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science*, **263**, 802 (1994).

182 Bondalapati, S., Jbara, M. & Brik, A. Expanding the Chemical Toolbox for the Synthesis of Large and Uniquely Modified Proteins. *Nature Chemistry*, **8**, 407 (2016).

7 Appendix

7.1 Sequence Data for Recombinant Proteins

7.1.1 eGFP-*Mxe*-CDB

DNA Sequence:

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc
agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgcc
ctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaag
tccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttc
gagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagt
acaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcg
aggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccacta
cctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggga
tcactctcggcatggacgagctgtacaaggcatgcatcacgggagatgcactagttgccctacccgagggcgagtcggtacgcatcg
ccgacatcgtgccgggtgcgcggcccaacagtgacaacgccatcgacctgaaagtccttgaccggcatggcaatcccgtgctcgcc
gaccggctgttccactccggcgagcatccggtgtacacggtgcgtacggtcgaaggtctgcgtgtgacgggcaccgcgaaccaccc
gttgttgtgtttggtcgacgtcgccggggtgccgaccctgctgtggaagctgatcgacgaaatcaagccgggcgattacgcggtgattc
aacgcagcgcattcagcgtcgactgtgcaggttttgcccgcggaaaacccgaatttgcgcccacaacctacacagtcggcgtccctgg
actggtgcgtttcttggaagcacaccaccgagacccggacgcccaagctatcgccgacgagctgaccgacgggcggttctactacgc
gaaagtcgccagtgtcaccgacgccggcgtgcagccggtgtatagccttcgtgtcgacacggcagaccacgcgtttatcacgaacgg
gttcgtcagccacgctactggcctcaccggtctgaactcaggcctcacgacaaatcctggtgtatccgcttggcaggtcaacacagctt
atactgcgggacaattggtcacatataacggcaagacgtataaatgtttgcagccccacacctccttggcaggatgggaaccatccaac
gttcctgccttgtggcagcttcaatga

Amino acid sequence:

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR
HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVT
AAGITLGMDELYKACITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNP
VLADRLFHSGEHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGD
YAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTD
GRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVS
AWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ.

7.1.2 **Thy-1-*Mxe*-CBD**

DNA sequence:

atgcatcatcatcatcatcagaaagtgaccagcctgaccgcgtgcctggtggatcagagcctgcgtctggattgccgtcatgaaa
acaccagcagcagcccgattcagtatgaatttagcctgacccgtgaaaccaaaaaacatgtgctgtttggcaccgtgggcgtgccgga
acatacctatcgtagccgtaccaactttaccagcaaatataacatgaaagtgctgtatctgagcgcgtttaccagcaaagatgaaggcac
ctatacctgcgcgctgcatcatagcggccatagcccgccgattagcagccagaacgtgaccgtgctgcgtgataaactggtgaaatgc
atcacgggagatgcactagttgccctacccgagggcgagtcggtacgcatcgccgacatcgtgccgggtgcgcggcccaacagtga
caacgccatcgacctgaaagtccttgaccggcatggcaatcccgtgctcgccgaccggctgttccactccggcgagcatccggtgtac
acggtgcgtacggtcgaaggtctgcgtgtgacgggcaccgcgaaccacccgttgttgtgtttggtcgacgtcgccggggtgccgacc
ctgctgtggaagctgatcgacgaaatcaagccgggcgattacgcggtgattcaacgcagcgcattcagcgtcgactgtgcaggttttgc
ccgcggaaaacccgaatttgcgcccacaacctacacagtcggcgtccctggactggtgcgtttcttggaagcacaccaccgagaccc
ggacgcccaagctatcgccgacgagctgaccgacgggcggttctactacgcgaaagtcgccagtgtcaccgacgccggcgtgcag
ccggtgtatagccttcgtgtcgacacggcagaccacgcgtttatcacgaacgggttcgtcagccacgctactggcctcaccggtctgaa
ctcaggcctcacgacaaatcctggtgtatccgcttggcaggtcaacacagcttatactgcgggacaattggtcacatataacggcaaga
cgtataaatgtttgcagccccacacctccttggcaggatgggaaccatccaacgttcctgccttgtggcagcttcaatga

Amino acid sequence:

MHHHHHHQKVTSLTACLVDQSLRLDCRHENTSSSPIQYEFSLTRETKKHVLFGTVG
VPEHTYRSRTNFTSKYNMKVLYLSAFTSKDEGTYTCALHHSGHSPPISSQNVTVLRD
KLVKCITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSG
EHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFS
VDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVA
SVTDAGVQPVYSLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAY
TAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ

7.1.3 **IL-2-*Mxe***

DNA sequence of IL-2-*Mxe* fusion protein:

atggccagttggagccacccgcagttcgaaaaagctagcggtacctacaggatgcaactcctgtcttgcattgcactaagtcttgcac
ttgtcacaaacagtgcacctacttcaagttctacaaagaaaacacagctacaactggagcatttacttctggatttacagatgattttgaatg
gaattaataattacaagaatcccaaactcaccaggatgctcacatttaagttttacatgcccaagaaggccacagaactgaaacatcttca
gtgtctagaagaagaactcaaacctctggaggaagtgctaaatttagctcaaagcaaaaactttcacttaagacccagggacttaatcag
caatatcaacgtaatagttctggaactaaagggatctgaaacaacattcatgtgtgaatatgctgatgagacagcaaccattgtagaatttc
tgaacagatggattacctttttgtcaaagcatcatctcaacactgacttgcatcacgggagatgcactagttgccctacccgagggcgagt
cggtacgcatcgccgacatcgtgccgggtgcgcggcccaacagtgacaacgccatcgacctgaaagtccttgaccggcatggcaat
c

Amino acid sequence:

MASWSHPQFEKASGTYRMQLLSCIALSLALVTNSAPTSSSTKKTQLQLEHLLLDLQ
MILNGINNYKNPKLTRMLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFH
LRPRDLISNINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLTCITGDAL
VALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVE
GLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKP
EFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVY
SLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQLVTYNGK
TYKCLQPHTSLAGWEPSNVPALWQLQ

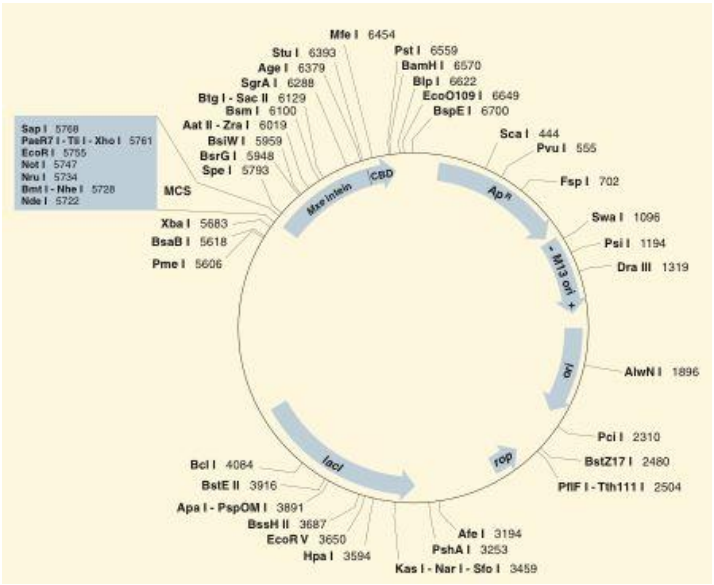


Figure 7.1. Plasmid map of commercially available pTXB1 vector (NEB).

7.1.4 **PrP-*Mxe*-CBD**

Amino acid sequence:

MKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGTWGQPHGGGWGQPHGGSWG
QPHGGSWGQPHGGGWGQGGGTHNQWNKPSKPKTNLKHVAGAAAAGAVVGGLGG
YMLGSAMSRPMIHFGNDWEDRYYRENMYRYPNQVYYRPVDQYSNQNNFVHDCVN
ITIKQHTVTTTTKGENFTETDVKMMERVVEQMCVTQYQKESQAYYDGRSACITGDA
LVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTV
EGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGK
PEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPV
YSLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQLVTYNG
KTYKCLQPHTSLAGWEPSNVPALWQLQ

7.1.5 **eGFP-*Npu*^{N}**

DNA sequence:

agttggagccacccgcagttcgaaaaagctagcggtaccgtgagcaagggcgaggagctgttcaccggggtggtgcccatcctg
gtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacc
ctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagcc
gctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaagga
cgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttc
aaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaa
cggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccc
atcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcg
atcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggatccagatctggatcctgt
ttaagctatgaaacggaaatattgacagtagaatatggattattaccgattggtaaaattgtagaaaagcgcatcgaatgtactgtttatagc
gttgataataatggaaatatttatacacaacctgtagcacaatggcacgatcgcggagaacaagaggtgtttgagtattgtttggaagatg
gttcattgattcgggcaacaaaagaccataagtttatgactgttgatggtcaaatgttgccaattgatgaaatatttgaacgtgaattggattt
gatgcgggttgataatttgccgaattaa

Amino acid sequence:

MASWSHPQFEKASGTVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT
IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMA
DKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDP
NEKRDHMVLLEFVTAAGITLGMDELYKGSRSGSCLSYETEILTVEYGLLPIGKIVEKRI
ECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQM
LPIDEIFERELDLMRVDNLPN

7.1.6 **Thy-1-*Npu*^{N}**

DNA sequence:

atgcagaaagtgaccagcctgaccgcgtgcctggtggatcagagcctgcgtctggattgccgtcatgaaaacaccagcagcagcc
cgattcagtatgaatttagcctgacccgtgaaaccaaaaaacatgtgctgtttggcaccgtgggcgtgccggaacatacctatcgtagcc
gtaccaactttaccagcaaatataacatgaaagtgctgtatctgagcgcgtttaccagcaaagatgaaggcacctatacctgcgcgctgc
atcatagcggccatagcccgccgattagcagccagaacgtgaccgtgctgcgtgataaactggtgaaatgcctgagctacgagaccg
aaattctgaccgtggagtatggtctgctgccgatcggcaagattgttgagaaacgtatcgaatgcaccgtgtacagcgttgacaacaac
ggcaacatttatacccagccggtggcgcaatggcacgatcgtggtgagcaggaagtttttgagtattgcctggaagacggcagcctga
tccgtgcgaccaaggaccacaaattcatgaccgttgatggtcaaatgctgccgatcgacgagatttttgaacgtgaactggacctgatgc
gtgttgacaatctgccgaaccaccaccaccaccaccactaa

Amino Acid sequence:

MQKVTSLTACLVDQSLRLDCRHENTSSSPIQYEFSLTRETKKHVLFGTVGVPEHTYR
SRTNFTSKYNMKVLYLSAFTSKDEGTYTCALHHSGHSPPISSQNVTVLRDKLVKCLS
YETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLE
DGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNHHHHHH

7.1.7 **IL-2-*Npu*^{N}**

DNA sequence:

Atggccagttggagccacccgcagttcgaaaaagctagcggtacctacaggatgcaactcctgtcttgcattgcactaagtcttgca
cttgtcacaaacagtgcacctacttcaagttctacaaagaaaacacagctacaactggagcatttacttctggatttacagatgattttgaat
ggaattaataattacaagaatcccaaactcaccaggatgctcacatttaagttttacatgcccaagaaggccacagaactgaaacatcttc
agtgtctagaagaagaactcaaacctctggaggaagtgctaaatttagctcaaagcaaaaactttcacttaagacccagggacttaatca
gcaatatcaacgtaatagttctggaactaaagggatctgaaacaacattcatgtgtgaatatgctgatgagacagcaaccattgtagaattt
ctgaacagatggattacctttgtcaaagcatcatctcaacactgactggatccagatctggatcctgtttaagctatgaaacggaaatatt
gacagtagaatatggattattaccgattggtaaaattgtagaaaagcgcatcgaatgtactgtttatagcgttgataataatggaaatatttat
acacaacctgtagcacaatggcacgatcgcggagaacaagaggtgtttgagtattgtttggaagatggttcattgattcgggcaacaaa
agaccataagtttatgactgttgatggtcaaatgttgccaattgatgaaatatttgaacgtgaattggatttgatgcgggttgataatttgccg
aattaa

Amino acid sequence:

MASWSHPQFEKASGTYRMQLLSCIALSLALVTNSAPTSSSTKKTQLQLEHLLLDLQ
MILNGINNYKNPKLTRMLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFH
LRPRDLISNINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLTGSRSGSC
LSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYC
LEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

7.1.8 **PrP-*Npu*^{N}**

Amino acid sequence;

MGQGGGTHNQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPMIH
FGNDWEDRYYRENMYRYPNQVYYRPVDQYSNQNNFVHDCVNITIKQHTVTTTTKG
ENFTETDVKMMERVVEQMCVTQYQKESQAYYDGRRSSCLSYETEILTVEYGLLPIGK
IVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTV
DGQMLPIDEIFERELDLMRVDNLPNHHHHHH

7.2 Additional Data on Characterization of Peptides and Proteins

7.2.1 *Npu*^{C} peptides

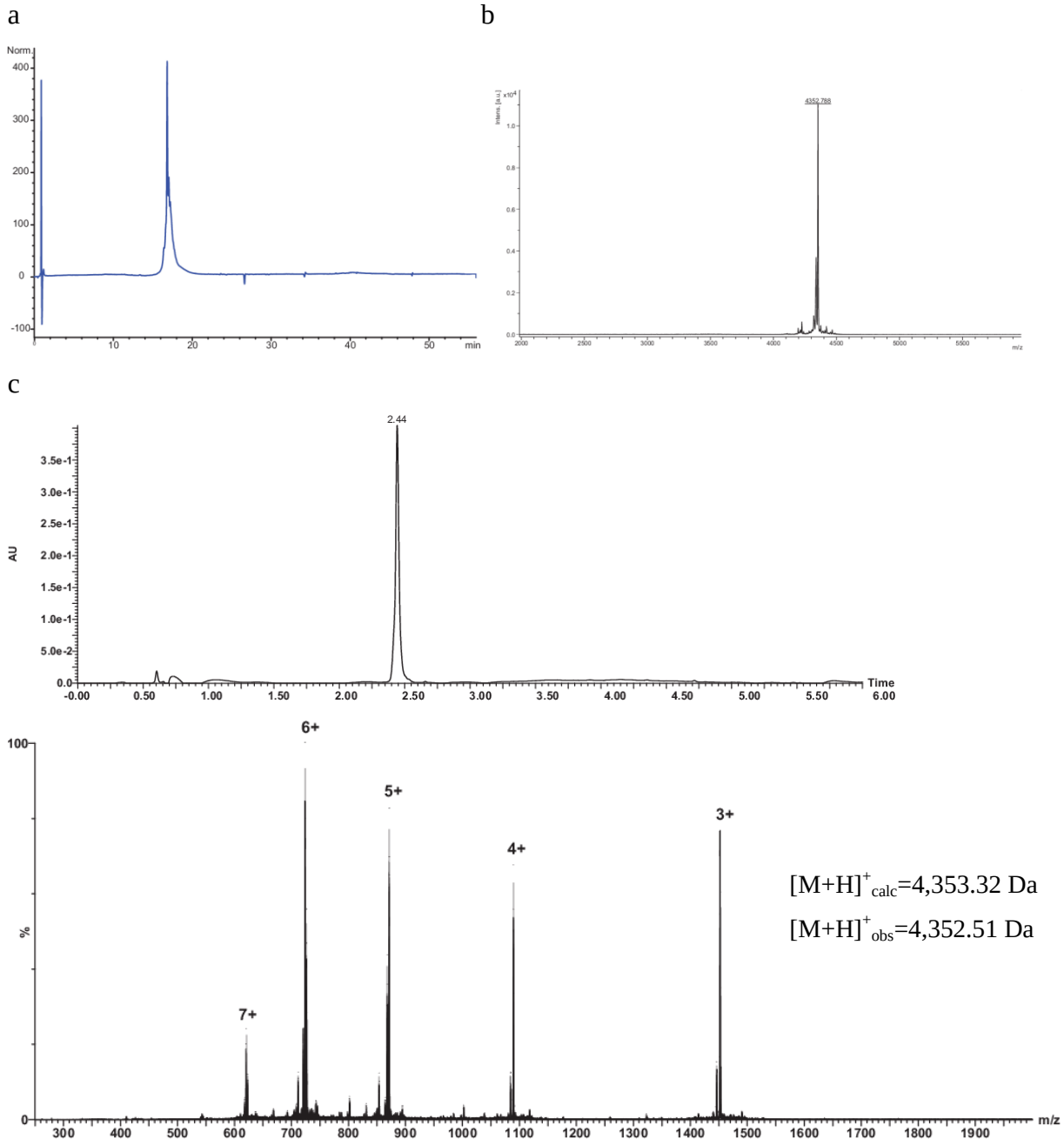


Figure 7.2. Characterization of *Npu*^{C}(AA). a) RP-HPLC chromatogram of purified *Npu*^{C}(AA). LC conditions: Agilent 1100 with YMC Hydrosphere C18 (50 x 3 mm, 3 µm), 5% ACN to 70% ACN in water with

0.1% TFA in 30 min. Flow rate 0.5 mL/min. **b)** MALDI-TOF spectrum of *Npu*^{C}(AA) , DHB was used as matrix, reflector positive mode $[M+H]^{+}_{obs}$: 4,352.788, $[M+H]^{+}_{calc}$: 4,353.2. **c)** Chromatogram (Waters BEH C18 1.7 µm, 50 x 2.1 mm, 5 – 70 % acetonitrile with 0.1%FA in water with 0.1% FA, 0.3 mL/min) and spectrum of the main peak from LC-ESI-MS analysis of purified *Npu*^{C}(AA). $[M+3H]^{3+}_{obs}$: 1,451.5062 → 4,352.5186 Da, $[M+H]^{+}_{calc}$: 4,353.32 Da.

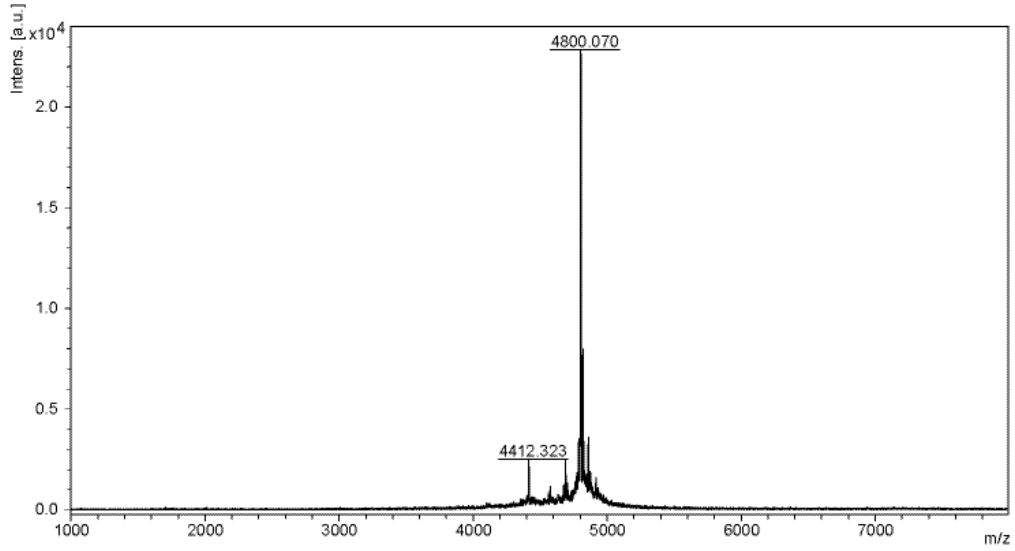


Figure 7.3 MALDI-TOF-MS analysis of *Npu*^{C}-biotin. DHB was used as matrix, mass spectrum was acquired in linear positive mode. $[M+H]^{+}_{calc}$=4,799.44 Da, $[M+H]^{+}_{obs}$=4,800.070 Da.

7.2.2 **eGFP-*Mxe***

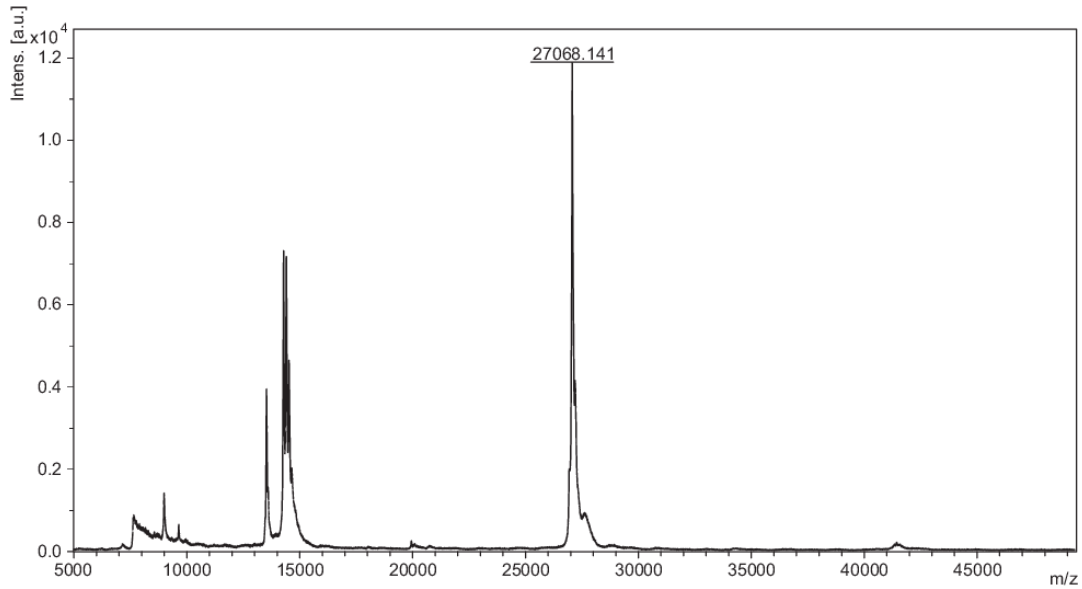


Figure 7.4. MALDI-TOF mass spectrum of eGFP-MMBA-thioester. DHB was used as matrix, mass spectrum was acquired in linear positive mode. $[M+H]^{+}_{calc}$ = 27,145.5 Da, $[M+H]^{+}_{obs}$ = 27,068.141 m/z.

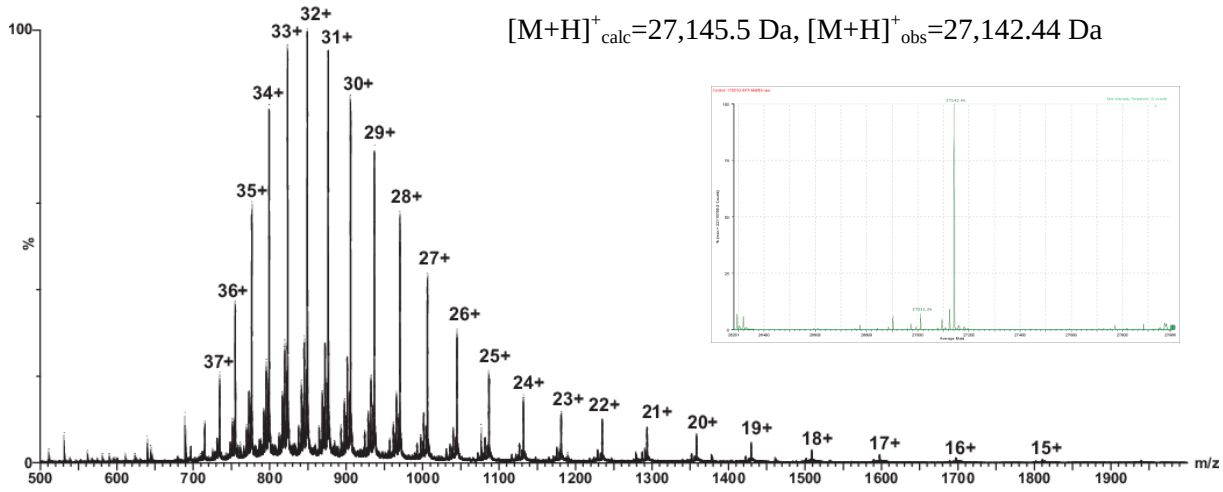


Figure 7.5. Raw and deconvoluted ESI mass spectra for eGFP-MMBA-thioester. $[M+H]^+_{calc}$ = 27,145.5 Da, $[M+H]^+_{obs}$ = 27,142.46 m/z The eGFP-thioester eluted from chitin resin was analyzed using a Phenomenex Jupiter C4 column (50 x 2.1 mm) with a gradient from 10 – 70% acetonitrile in water (0.1% FA in both eluents) in 30 min. The sample was analyzed using ESI-QTOF. At t_R=19.6 min, the given mass spectrum was recorded. This spectrum was deconvoluted using Biopharmalynx software. The resulting deconvoluted mass is 27,142.44 Da.

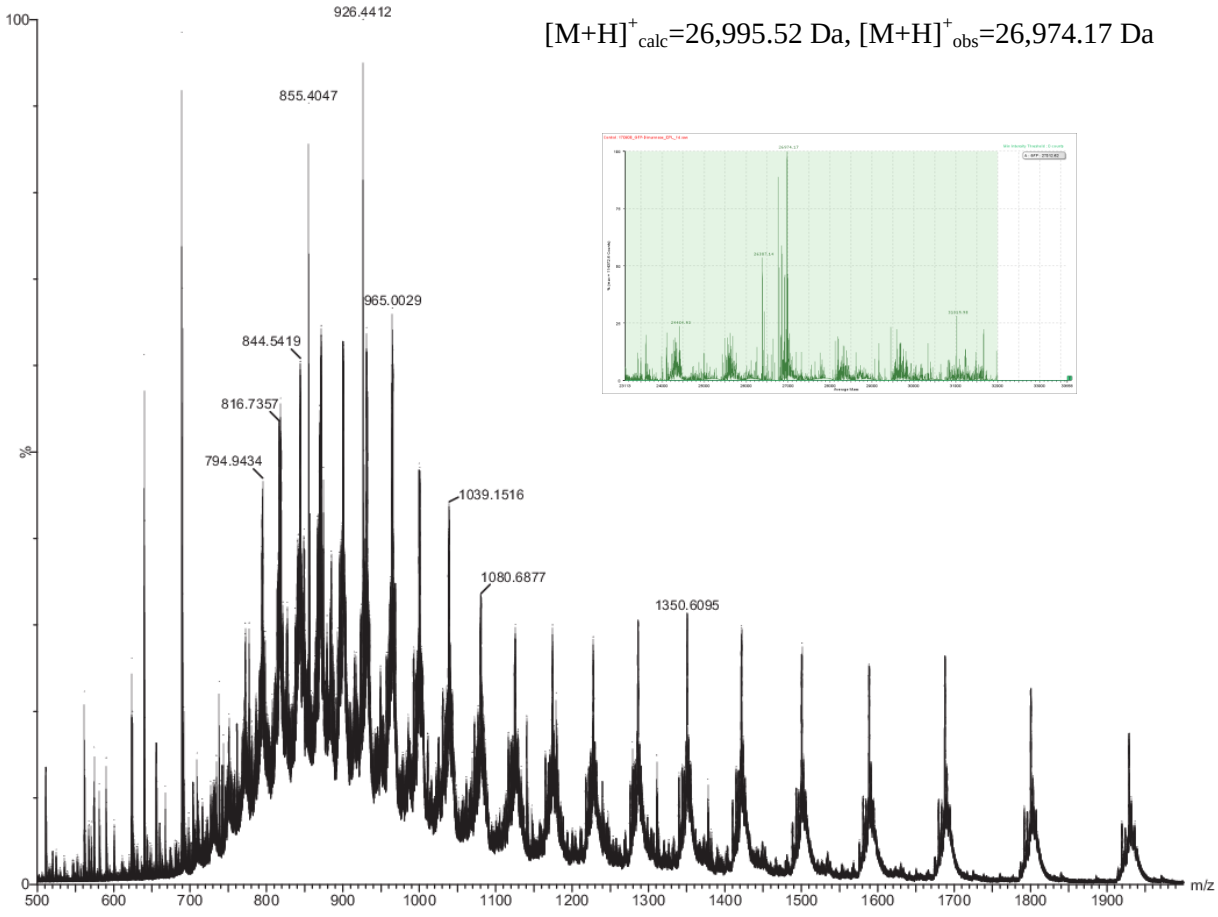


Figure 7.6. Additional data on EPL eGFP-dimannose: hydrolyzed side product Raw and deconvoluted mass spectra of t_R=18.1 min: $[M+H]^+_{obs}$=26,974.17, corresponding to the hydrolyzed side product eGFP-OH, $[M+H]^+_{calc}$ =26,995.52 Da.

7.2.3 **Thy-1-*Mxe***

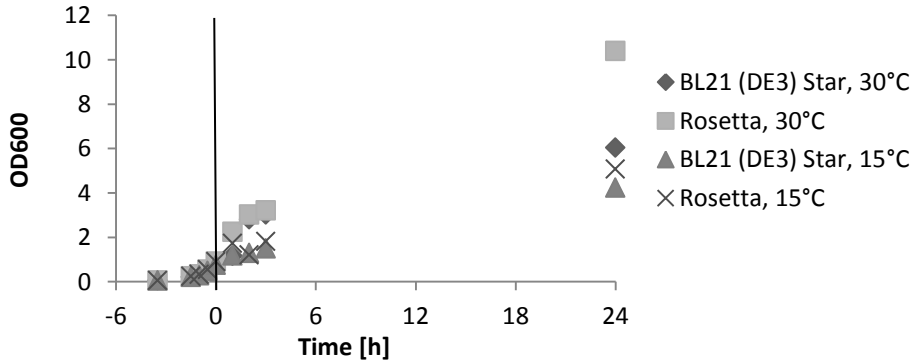


Figure 7.7. Growth curves of the test expression of Thy-1-*Mxe*-CDB fusion protein in *E. coli* BL21 (DE3) Star and Rosetta cells. All cells were inoculated to a starting OD_{600} of 0.05. OD was measured frequently until late exponential phase was reached (OD 0.6-0.8). At this point the temperature was changed to 15°C or 30°C, respectively, and this time corresponds to t=0 (induction time, indicated by the vertical line).

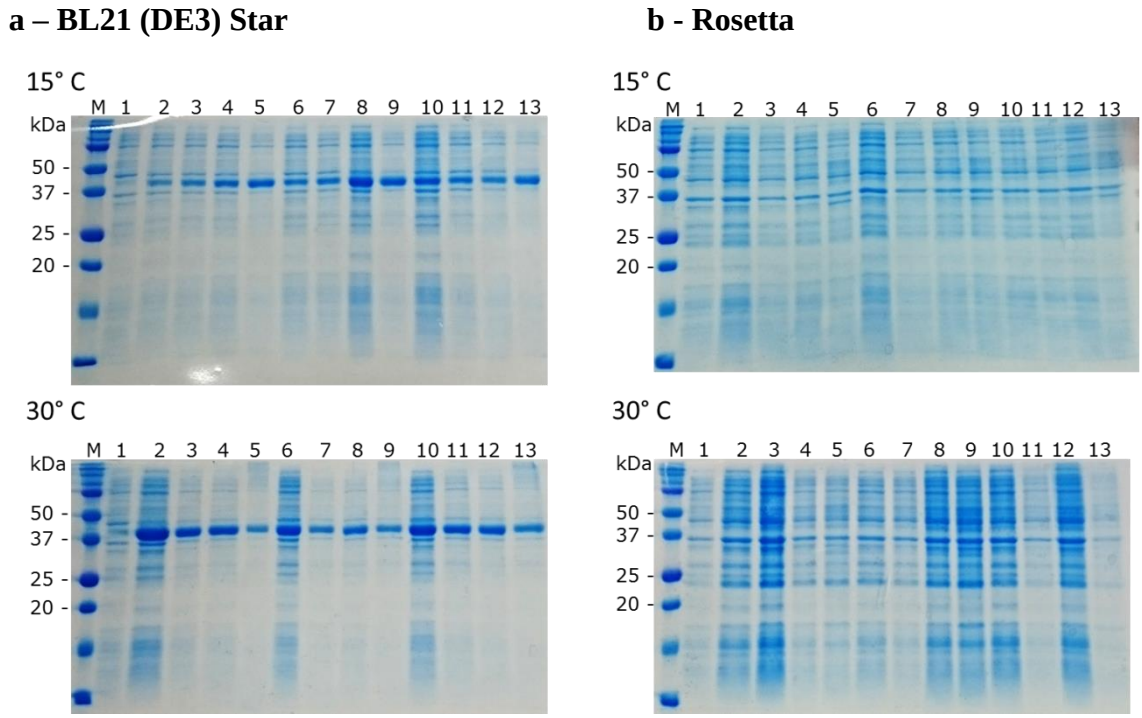


Figure 7.8. Expression test for Thy-1-*Mxe*-CBD fusion protein in *E. coli* BL21 (DE3) Star (a) and Rosetta cells (b). The expression was checked using SDS-PAGE. M – Molecular weight marker, 1 – pre induction, 2 – 5 – 0.2 mM IPTG induction 1/2/3/24 h, 6 – 9 – 0.4 mM IPTG induction 1/2/3/24 h, 10 – 13 – 1 mM IPTG induction 1/2/3/24 h. Expected size = 41.2 kDa.

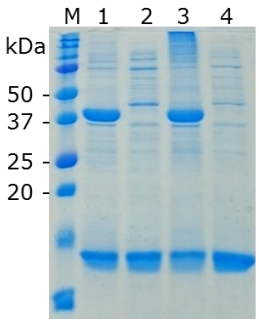


Figure 7.9. Solubility check of the Thy-1-*Mxe*-CDB fusion protein using SDS-PAGE. Samples from expression tests at 15°C, 0.4 mM IPTG and 30°C, 0.2 mM IPTG were lysed and centrifuged. M – Molecular weight marker, 1 – 15°C, 0.4 mM IPTG, insoluble fraction, 2 – 15°C, 0.4 mM IPTG, soluble fraction, 3 – 30°C, 0.2 mM IPTG, insoluble fraction, 4 – 30°C, 0.2 mM IPTG, soluble fraction.

7.2.4 eGFP-*Npu*^{N}

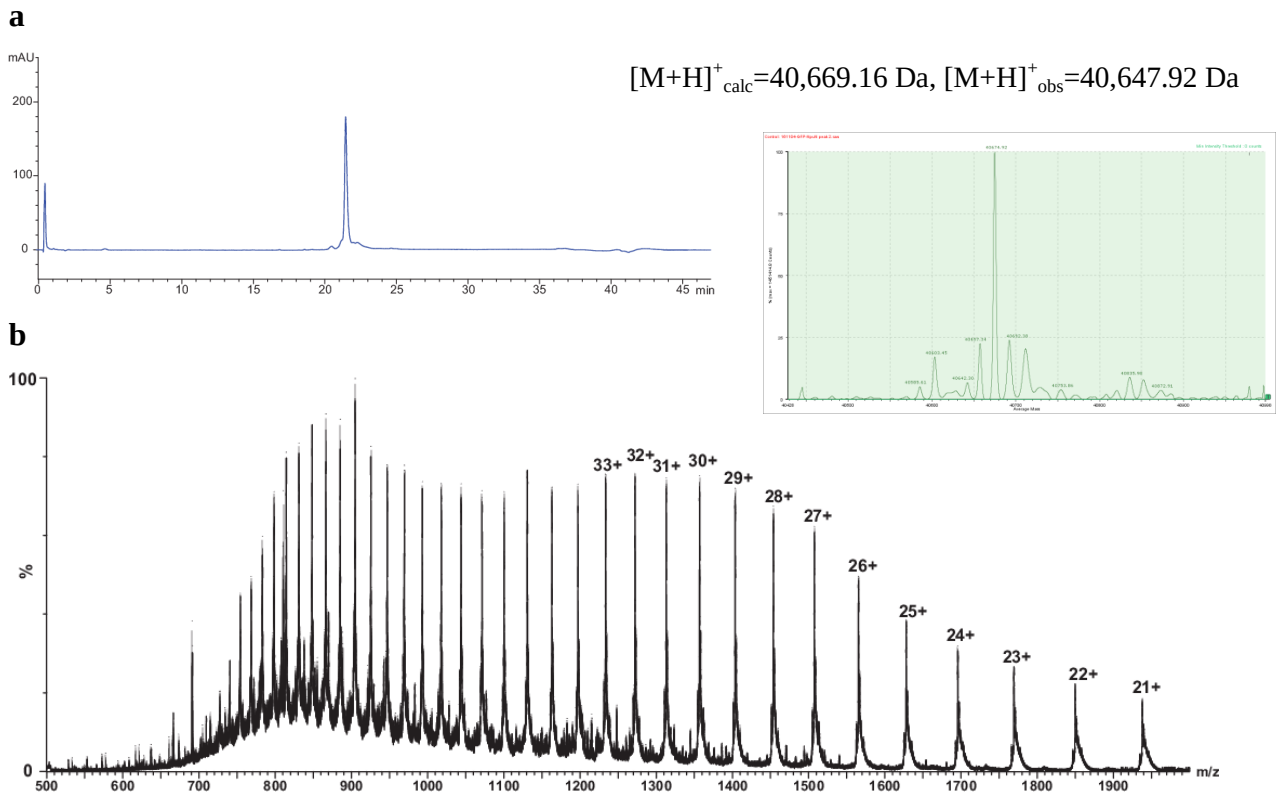


Figure 7.10. Characterization of eGFP-*Npu*^{N} fusion protein. a) RP-HPLC, **b)** LC-ESI-MS, raw and deconvoluted spectra are shown, $[M+H]^+_{calc}$=40,669.16 $[M+H]^+_{obs}$=40,674.92.

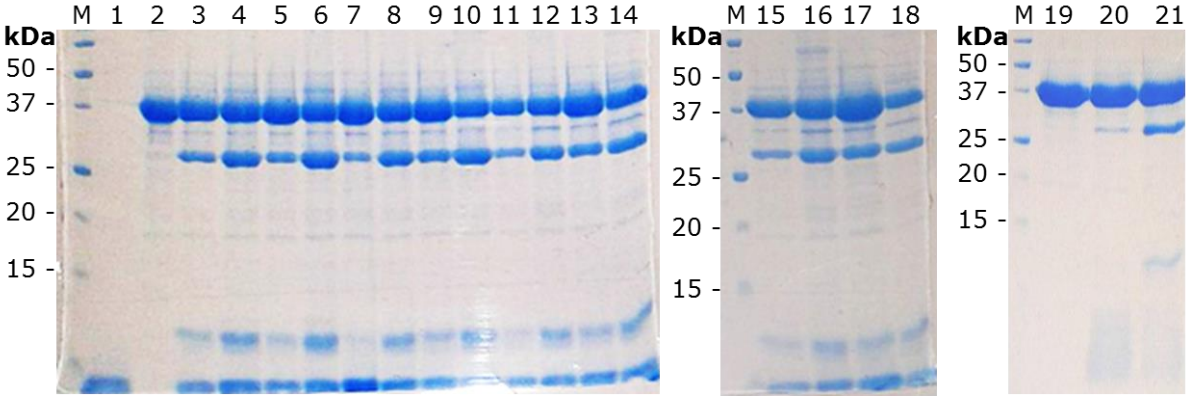


Figure 7.11. Optimization of TCEP concentration and incubation temperature in PTS with eGFP-*Npu*^{N} with *Npu*^{C} hydrazide. M – Molecular weight marker, 1 – *Npu*^{C} hydrazide, 2 – eGFP-*Npu*^{N}, 3 – 1 mM TCEP, RT, 30 min, 4 – 1 mM TCEP, RT, o/n, 5 – 1 mM TCEP, 37°C, 30 min, 6 – 1 mM TCEP, 37°C, o/n, 7 – 2 mM TCEP, RT, 30 min, 8 – 2 mM TCEP, RT, o/n, 9 – 2 mM TCEP, 37°C, 30 min, 10 – 2 mM TCEP, 37°C, o/n, 11 – 5 mM TCEP, RT, 30 min, 12 – 5 mM TCEP, RT, o/n, 13 – 5 mM TCEP, 37°C, 30 min, 14 – 5 mM TCEP. 37°C, o/n, 15 – 10 mM TCEP, RT, 30 min, 16 – 10 mM TCEP, RT, o/n, 17 – 10 mM TCEP, 37°C, 30 min, 18 – 10 mM TCEP, 37°C, o/n, 19 - eGFP-*Npu*^{N}, 20 – 75 µM TCEP (i.e. 5 eq.), 37°C, o/n, 21 – 250 µM TCEP (5 eq. per Cys), 37°C, o/n).

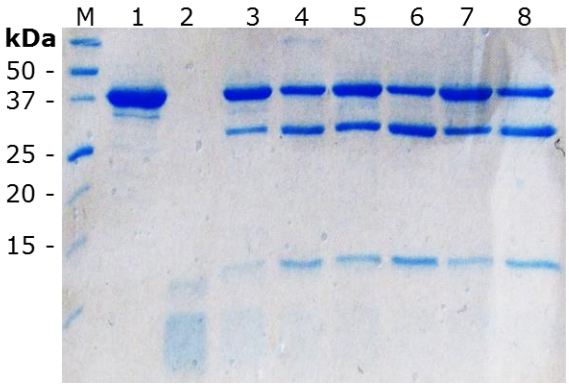


Figure 7.12. Optimization of the protein:peptide ratio in PTS with eGFP-*Npu*^{N} and *Npu*^{C}. M - Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – *Npu*^{C} hydrazide, 3 – PTS with 2.3 eq. of peptide, 1 h, 4 – 2.3 eq. peptide, o/n, 5 – 1.5 eq. peptide, 1 h, 6 – 1.5 eq. peptide, o/n, 7 – 1.1 eq. peptide, 1 h, 8 – 1.1 eq. peptide, o/n.

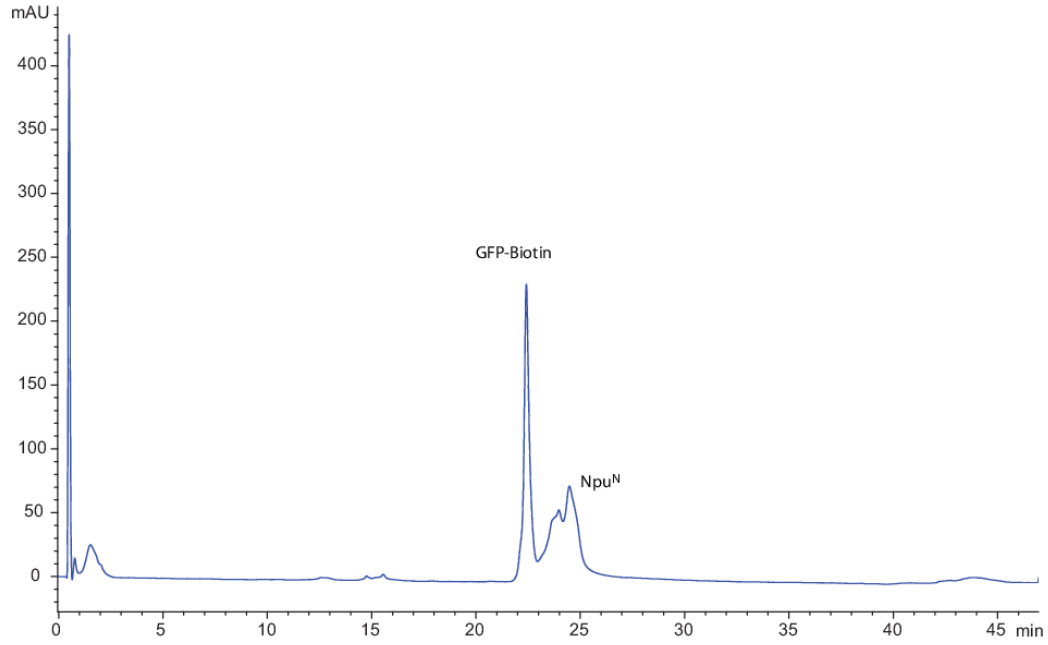


Figure 7.13. Analysis of strep-tag purified eGFP-biotin by RP-HPLC (C4).

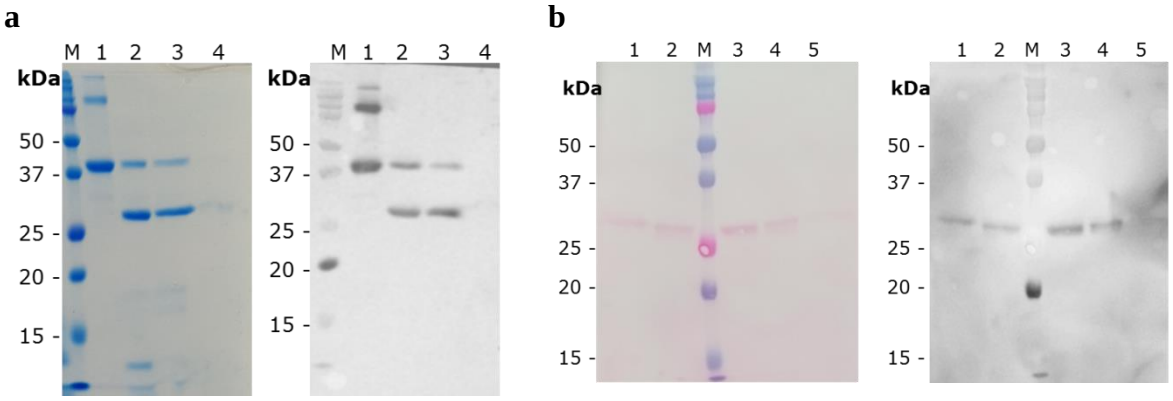


Figure 7.14. Triton separation of (a) eGFP-mGPI and (b) eGFP-bGPI. Detection was performed using MTG5 anti-GPI antibody, however, this clone gave a signal also for the eGFP fusion protein. A) M – Molecular weight marker, 1 – eGFP-*Npu*^{N}, 2 – OPL eGFP-mGPI 4 d, 3 – aqueous phase, 4 – detergent phase. B) M – Molecular weight marker, 1 – separation 1, aqueous phase, 2 – separation 1, detergent phase, 3 – eGFP-bGPI SEC purified, 4 – separation 2, aqueous phase, 5 – separation 2, detergent phase.

7.2.5 **Thy-1-*Npu*^{N}**

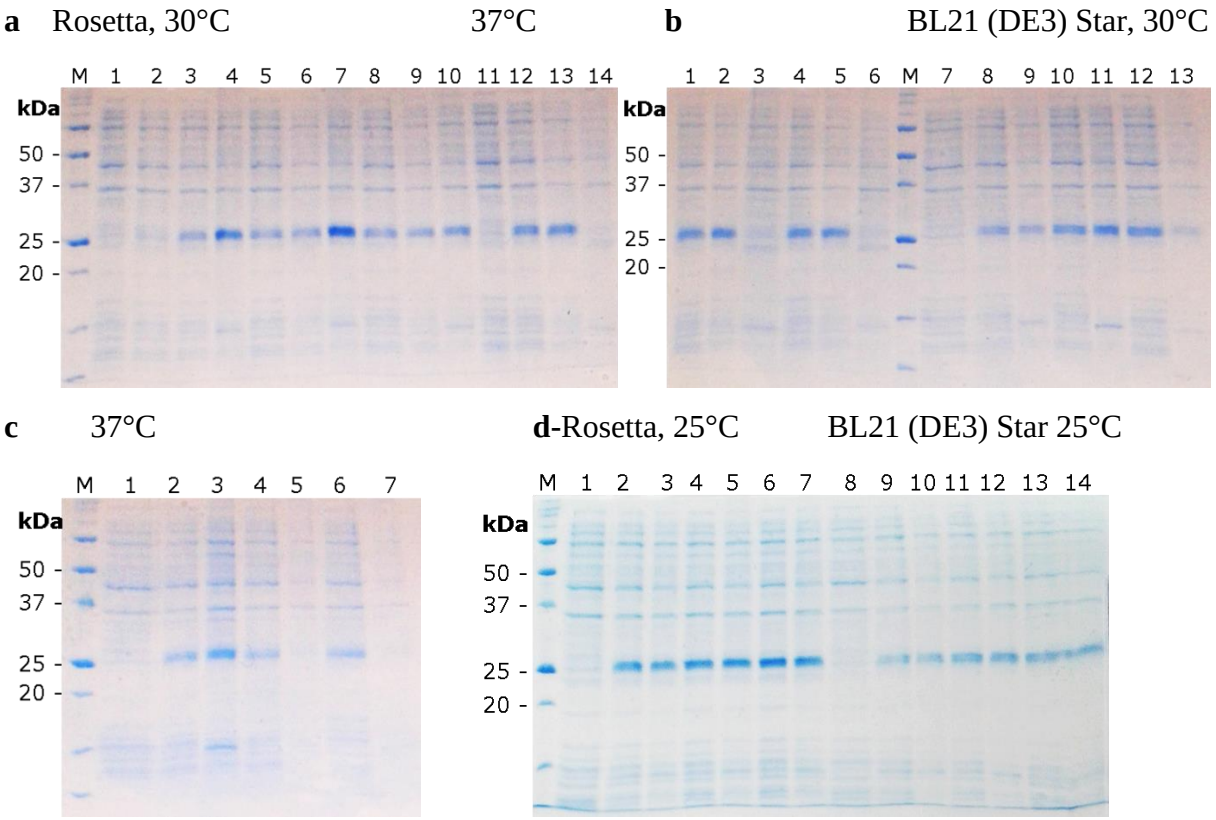


Figure 7.15. Expression tests for Thy-1-*Npu*^{N} analyzed in SDS-PAGE. Sample amounts were normalized to the OD_{600}. **a)** M – Molecular weight marker, 1 – Rosetta 30°C, pre induction, 2, 3, 4 – 0.2 mM IPTG, 1, 2, 24 h, 5, 6, 7 – 0.5 mM IPTG, 1, 2, 24 h, 8, 9, 10 – 1 mM IPTG, 1, 2, 24 h, 11 – Rosetta 37°C, pre induction, 12, 13, 14 – 0.2 mM IPTG, 1, 2, 24 h, **b)** 1, 2, 3 – 0.5 mM IPTG, 1, 2, 24 h, 4, 5, 6 – 1 mM IPTG, 1, 2, 24 h, 7 – BL21 (DE3) Star 30°C, pre induction, 8, 9, 10 – 0.2 mM IPTG, 1, 2, 24 h, 11, 12, 13 – 0.5 mM IPTG, 1, 2, 24 h, **c)** 1 – BL21 (DE3) Star, 37°C, pre induction, 2, 3 – 0.2 mM IPTG, 2, 24 h, 4, 5 – 0.5 mM IPTG, 2, 24 h, 6, 7 – 1 mM IPTG, 2, 24 h, **d)** 1 – Rosetta, pre induction, 2, 3 – Rosetta 0.2 mM IPTG, 3, 24 h, 4, 5 – Rosetta 0.5 mM IPTG, 3, 24 h, 6, 7 – Rosetta 1 mM IPTG, 3, 24 h, 8 – BL21 (DE3) Star, pre induction, 9, 10, 0.2 mM IPTG, 3, 24 h, 11, 12 – 0.5 mM IPTG, 3, 24 h, 13, 14 – 1 mM IPTG, 3, 24 h. Expression in Rosetta clone is better than in BL21 (DE3) Star clone; at 37°C there is almost no expression, the best expression was found with 30°C, 0.5 mM IPTG, 24 h induction; 25°C with 1 mM and 3 – 24 h looks also good.

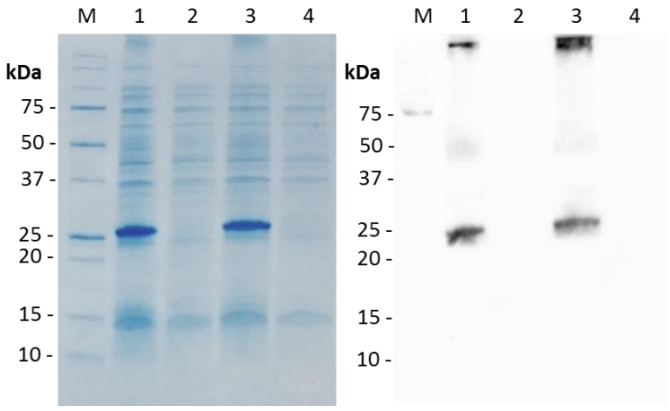
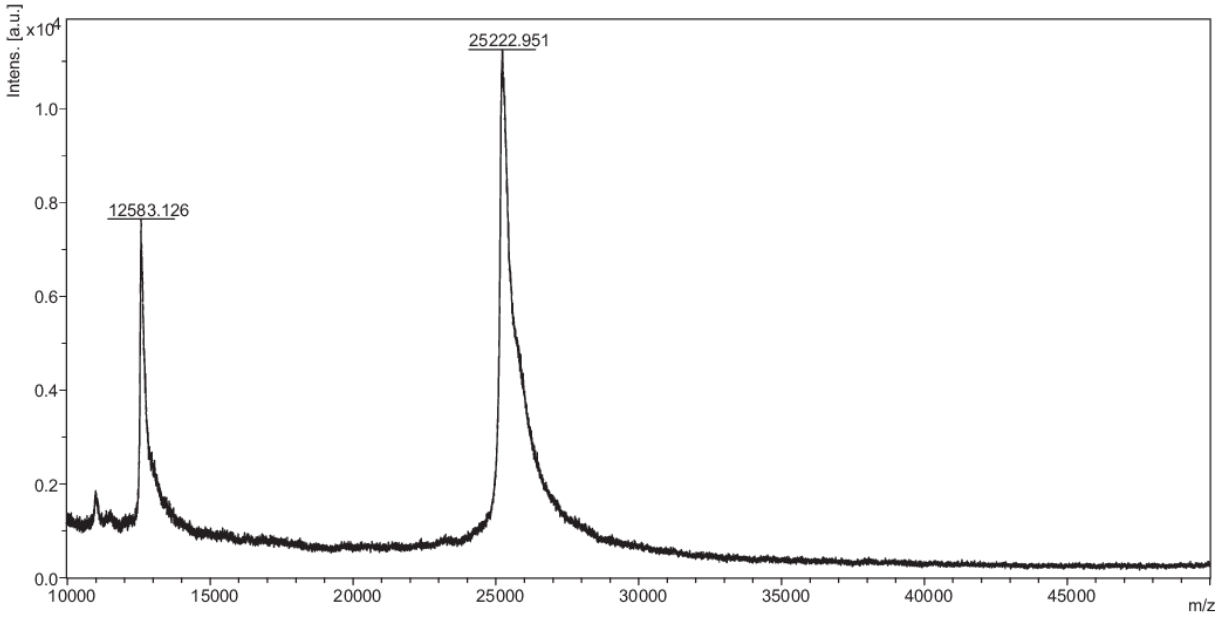


Figure 7.16. Solubility check of Thy-1-*Npu*^{N} fusion protein in Rosetta using SDS-PAGE and western blot. For detection an anti-his-tag antibody (mouse) was used in 1:5,000 dilution in 5% BSA in PBS-T (incubation 1 h at room temperature) and a secondary antibody anti-mouse IgG-HRP conjugate (1:10,000 in 5% BSA in PBS-T, 20 min at room temperature). The fusion protein was found only in the total fraction but not in the soluble fraction, indicating that it was expressed insolubly. M – Molecular weight marker, 1 – 30°C, total lysate, 2 – 30°C, soluble fraction, 3 – 25°C, total, 4 – 25°C, soluble fraction.

a



b

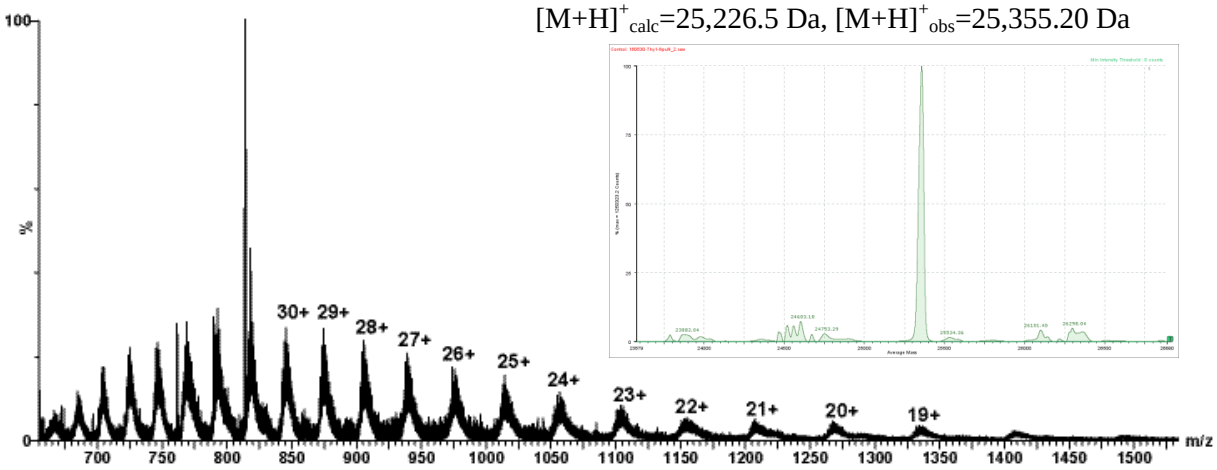


Figure 7.17. Characterization of Thy-1-*Npu*^{N} fusion protein by mass spectrometry. a) MALDI-TOF-MS, DHB as matrix substance, linear positive mode was used, $[M+H]^+_{calc}$=25,226.5 Da, $[M+H]^+_{obs}$=25,222.9 m/z. **b)** Raw and deconvoluted spectra from LC-ESI-MS analysis. Waters BEH C4 UPLC column (50 x 2.1 mm, 1.7 µm) on Waters Acquity H-Class UPLC with Waters Xevo G2-Xs mass spectrometer, gradient 10 – 70% ACN with 0.1% FA in water with 0.1% FA in 30 min, flow rate 0.5 mL/min, ES^+ mode with 1.5 kV capillary voltage. $[M+H]^+_{calc}$=25,226.5 Da, $[M+H]^+_{obs}$=25,356.21 m/z.

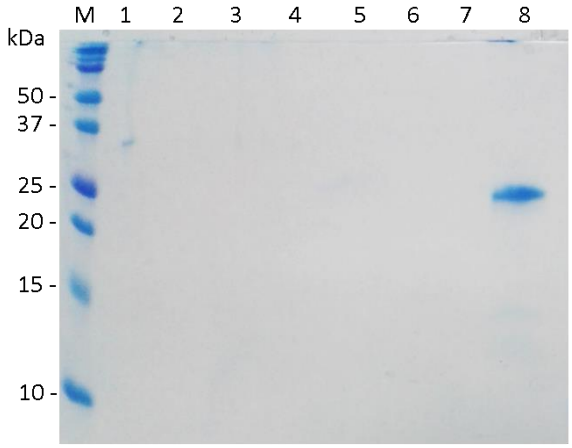


Figure 7.18. Solubility test of Thy-1-*Npu*^{N} in different common buffers for HIC. M – Molecular weight marker, 1 – 1.8 M ammonium sulfate, 2 – 1.5 M ammonium sulfate, 3 – 1 M ammonium sulfate, 4 – 4 M NaCl, 5 – 3 M NaCl, 6 – 1 M sodium sulfate, 7 – sodium sulfate saturated, 8 – PTS buffer + 6 M urea (control). Unfortunately, the fusion protein did not remain soluble in any HIC buffer tested. Therefore, HIC could not be applied for purification.

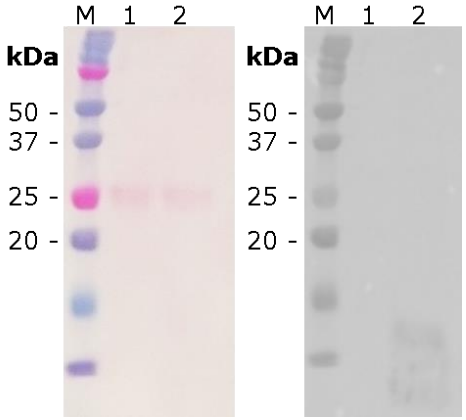


Figure 7.19. PTS Thy-1-biotin. M – Molecular weight marker, 1 – Thy-1-*Npu*^{N}, 2 – PTS reaction Thy-1-biotin (3 days).

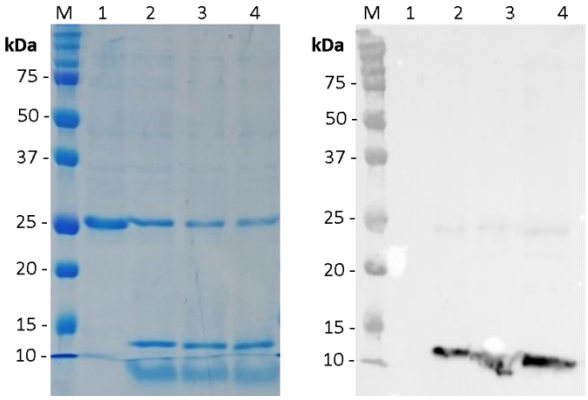


Figure 7.20. OPL reaction over a prolonged time. M – Marker, 1 – Thy1-NpuN, 2 – OPL 1 d, 3 – OPL 3 d, 4 – OPL 6 d. 58% after 3 d, no increase after 6 d.

7.2.6 **PrP-*Npu*^{N}**

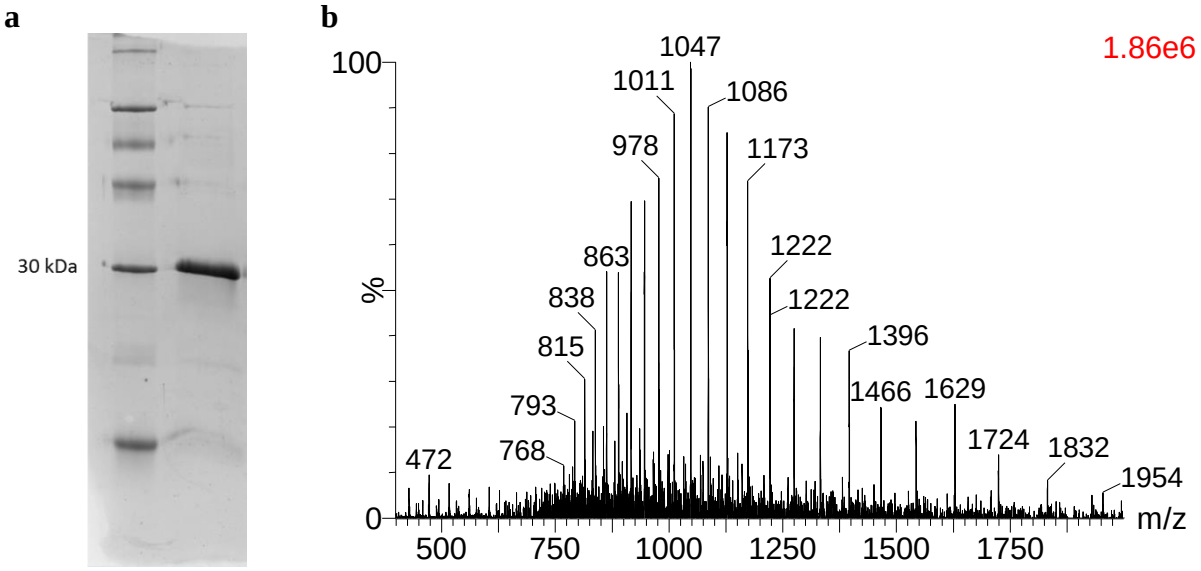


Figure 7.21. PrP-*Npu*^{N} fusion protein provided by Prof. Christian Becker, University of Vienna, analysis in (a) SDS-PAGE and (b) LC-MS. $[M+H]^+_{calc}$=29,128.65 Da, $[M+H]^+_{obs}$=29,297 Da: Arg adduct due to refolding buffer. Images were provided by Stefanie Hackl, University of Vienna.