



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„Investigating the mechanism of immunoglobulin M
assembly using protein semisynthesis“

verfasst von / submitted by

Leopold Dürrauer, BSc. MSc. MSc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt /
degree program code as it appears on the student record
sheet:

A 796 605 419

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student record sheet:

Chemie

Betreut von / Supervisor:

Univ.-Prof. Dr. Christian Friedrich Wilhelm Becker

Disclaimer

This thesis is the result of work performed in the years 2019 to 2023 in the research group of Prof. Dr. Christian Becker at the Institute of Biological Chemistry at the Faculty of Chemistry of the University of Vienna and was further supported by the Vienna Doctoral School in Chemistry of the University of Vienna.

Parts of the work presented were performed in collaboration with Prof. Buchner and Tuan Nguyen at the TUM Department of Chemistry of the Technical University of Munich.

Chemical structures were drawn in ChemDraw (Revvity Signals Software). Schematic figures were created using BioRender (BioRender) and Microsoft PowerPoint (Microsoft).

Data was analyzed using Chromeleon (ThermoFisherScientific), MassLynx (Waters), MagTran, Microsoft Excel (Microsoft), and Origin (OriginLab Corporation).

Spellchecking and grammatical corrections were performed *via* Microsoft Word, incl. Editor (Microsoft) and Grammarly (Grammarly).

Parts of this thesis were presented at the following conferences:

- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*IgM Assembly on the Molecular Level*” International DoSCChem Student Symposium, **2023**
- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*Investigating IgM Assembly on the Molecular Level*” Austrian Peptide Symposium, **2022**
- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*IgM Assembly on the Molecular Level*” International DoSCChem Student Symposium, **2022**
- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*IgM Assembly on the Molecular Level*” 36th European Peptide Symposium, **2022**
- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*IgM Assembly on the Molecular Level*” Annual Retreat of the Vienna Doctoral School in Chemistry, **2022**
- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*Exploring IgM Assembly via Semisynthetic Strategies*” Austrian Peptide Symposium, **2021**
- ❖ L. Duerrauer, O. Gajsek, A. Kravchuk, T. Nguyen, J. Buchner, C.F.W. Becker „*Investigating IgM assembly via C μ 4tp semisynthesis*” ÖGMBT annual meeting, **2021**

1 Acknowledgement

I am grateful to Christian F. W. Becker for entrusting me with this multifaceted project and providing invaluable feedback in person and in writing. I want to express my gratitude to Alexander Kravchuk for his preliminary work, which served as the foundation for this thesis. He provided invaluable and efficient practical and theoretical support throughout the entire project. I also want to thank Johannes Buchner and Tuan Nguyen for their efforts, input, and support.

I also want to highlight the outstanding efforts of the other members of our group in supporting this project, with special emphasis on the following: My gas bottle buddy, Oliver Gajsek, not only performed vital preliminary work on this project during his master's thesis but also demonstrated exceptional patience and competence in helping me overcome numerous challenges faced throughout this project. Alanca Schmid, who first established an SGP isolation and transfer strategy within our research group and still found time to share her insights with me despite being in the last months of a very difficult Ph.D. project. Dominik Vogl, who endured long discussions on project structure and management. Somnath Mukherjee for his excellent guidance and input on scientific and personal matters. Julia Kriegesmann who generously provided formidable feedback and practical input. Hannah Back and Stefan Moser for fruitful discussions on data and providing emotional support.

I also want to thank the heroes providing technical support throughout my years at the Institute of Biological Chemistry: Gerhard Niederacher, Manuel Felkl, Florian Exler, and Laura Holzapfel. In addition, I also want to thank the numerous students who assisted me

Acknowledgement

throughout the last few years and provided valued feedback on how to improve my teaching skills. I want to thank all members of the IBC without exception for making me feel welcome and appreciated at the institute.

No project exists in isolation, and so I want to reiterate my deepest and most humble gratitude to all the wonderful people around me who shared in my successes and defeats and who were there for me when I needed them: My parents, siblings, friends, relatives, and loved ones.

Thank you.

I could not have done this without you.

2 Abstract

Antibodies, also known as immune globulins (Ig), are key players in the human immune system. In particular, the multivalent, polymeric IgM represents an early response against pathogens by the adaptive immune system. Generally, immune globulins consist of two covalently linked light (L) and two heavy (μ) chains, that form μ 2-L2 subunits. In case of IgM, these subunits are further assembled into (μ 2-L2) penta- or (μ 2-L2) hexamers depending on the presence or absence of a J chain, resulting in a complex of over 1.200 kDa. Previous studies have shown that extending the C μ 4 domain of μ chain with an 18-aa C-terminal tailpiece (μ tp) is sufficient and necessary for the formation of hexamers, and that disulfide bond formation between C575 of two C μ 4tp fragments is a prerequisite for assembly.

We aimed to further explore the underlying mechanism of IgM assembly dimerization via creating and investigating μ tp-extended C μ 4 domains. To achieve this goal, we have developed strategies for the semisynthetic production of modified C μ 4tp domains to facilitate the incorporation of non-proteinogenic amino acids into our constructs. We successfully produced three C μ 4tp variants, each carrying a p-benzoyl-L-phenylalanine (BPA) photocrosslinkers in a different position of the tp. These variants (L561Bpa, Y562Bpa and V564Bpa) display folding behavior in line with wild-type C μ 4tp and may allow for searching yet unknown interaction sites during IgM assembly.

Additionally, we present work toward the semisynthetic production of glycosylated C μ 4tp variants, including a minimally glycosylated variant carrying a GlcNAc in position Asn563 and a variant carrying a glycan derived from sialylglycopeptide (SGP) in the same position. Furthermore, we show a new protocol for isolating SGP from egg yolk powder and discuss our

Abstract

results and experiences performing direct transglycosylation reactions using SGP as glycan donor.

3 Zusammenfassung

Antikörper, auch bekannt als Immunglobuline (Ig), sind essenzielle Bestandteile des menschlichen Immunsystems. Eine besondere Bedeutung kommt hierbei dem multivalenten, polymeren IgM zu, welches besonders in der frühen Phase der Immunantwort des adaptiven Immunsystems auf Pathogene eine entscheidende Rolle spielt. Allgemein bestehen Immunglobuline aus zwei kovalent verbundenen leichten (L) und zwei schweren (μ) Ketten, die μ 2-L2-Untereinheiten bilden. Im Fall von IgM werden diese Untereinheiten weiter zu (μ 2-L2) Penta- oder (μ 2-L2) Hexameren zusammengesetzt, in Abhängigkeit der Gegenwart der sogenannten J-Kette, was zur Bildung eines Komplexes von über 1.200 kDa führt. Bestehende Studien konnten zeigen, dass die Verlängerung der C μ 4-Domäne der μ Kette mit einem 18 Aminosäure-langem C-terminalen Tailpiece (μ tp) ausreichend und notwendig für die Bildung von IgM-Hexamern ist, und dass die Bildung einer Disulfidbrücke zwischen C575 von zwei C μ 4tp-Fragmenten eine Voraussetzung für die Assemblage ist.

Unser Ziel war es, den zugrundeliegenden Mechanismus der IgM-Dimerisierung durch die Erstellung und Untersuchung von μ tp-erweiterten C μ 4-Domänen weiter zu erforschen. Um dieses zu bewerkstelligen, haben wir Strategien für die semisynthetische Produktion von modifizierten C μ 4tp-Domänen entwickelt, um die Einfügung von nicht-proteinogenen Aminosäuren in unsere Konstrukte zu ermöglichen. Wir haben erfolgreich mehrere C μ 4tp-Varianten produziert, die einen p-Benzoyl-L-Phenylalanine Photocrosslinker in ihrem tp tragen. Diese Varianten (L561Bpa, Y562Bpa und V564Bpa) zeigen ein Faltungsverhalten, das mit dem wildtypischen C μ 4tp weitgehend übereinstimmt. Diese könnten es ermöglichen, bisher unbekannte Interaktionsstellen während der IgM-Oligomerisation zu suchen.

Zusammenfassung

Zusätzlich präsentieren wir Fortschritte bei der semisynthetischen Darstellung von glykosylierten C μ 4tp-Varianten, einschließlich einer minimal glykosylierten Variante mit einem GlcNAc in Position Asn563 und einer Variante, die ein komplexes Glykan an derselben Position trägt, welche aus Sialylglykopeptid (SGP) gewonnen wurde. Wir zeigen außerdem ein neues Protokoll für die Isolierung von SGP aus Eigelbpulver und diskutieren unsere Ergebnisse und Erfahrungen bei der Durchführung direkter Transglykosylierungsreaktionen unter Verwendung von SGP als Glykanspender.

4 Table of Contents

1	Acknowledgement.....	7
2	Abstract.....	9
3	Zusammenfassung.....	11
4	Table of Contents.....	13
5	Table of Figures.....	19
6	Abbreviations.....	27
7	Introduction.....	31
7.1	Antibodies.....	31
7.2	Antibody structure.....	32
7.3	Antibody isotypes in humans.....	35
7.3.1	IgA.....	36
7.3.2	IgD.....	37
7.3.3	IgE.....	37
7.3.4	IgG.....	37
7.3.5	IgM.....	38
7.4	IgM.....	38
7.4.1	Structure of the IgM monomer.....	39
7.4.2	Membrane-bound IgM and the B cell receptor.....	40
7.4.3	Secretory IgM.....	43
7.4.4	The structure of polymeric IgM.....	46

Table of Contents

7.4.5	C μ 4tp as hexamerization aid.....	50
7.5	Antibody Glycosylation.....	51
7.5.1	Antibody glycosylation.....	51
7.5.2	IgM glycosylation	53
7.6	Methodological considerations regarding the semisynthesis of C μ 4tp.....	54
7.6.1	Overview of protein modification strategies.....	54
7.6.2	Solid phase peptide synthesis – history and limitations.....	55
7.6.3	Native chemical ligation.....	56
7.6.4	Protein semisynthesis.....	57
7.6.5	Photo-affinity labelling.....	59
7.6.6	Glycosylation in protein semisynthesis.....	61
7.6.7	Sialylglycopeptide as a glycan donor.....	63
7.7	Aim	65
8	Materials and Methods.....	67
8.1	Production of N-terminal C μ 4tp segments via protein expression.....	67
8.1.1	Protein expression.....	67
8.1.2	Bacterial cell lysis <i>via</i> high-pressure disruption	68
8.1.3	Bacterial cell lysis <i>via</i> sonication	68
8.1.4	Workup of bacterial cells after lysis.....	68
8.1.5	Affinity column purification	69

Table of Contents

8.1.6	Dialysis of expressed protein solutions.....	69
8.1.7	Intein cleavage using Mesna	69
8.1.8	Thioester purification <i>via</i> preparative HPLC	70
8.1.9	Quality control of synthetic N-terminal C μ 4tp segments.....	70
8.2	Synthesis of C-terminal C μ 4tp segments.....	71
8.2.1	Solid Phase Peptide Synthesis.....	71
8.2.2	Removal of acetyl protecting groups from GlcNAc modified peptides	75
8.2.3	Peptide purification <i>via</i> preparative HPLC	75
8.2.4	Quality control of synthetic C-terminal C μ 4tp segments	76
8.3	Ligation of C μ 4tp segments.....	76
8.3.1	Native chemical ligation.....	76
8.3.2	Reaction monitoring.....	77
8.3.3	Purification of semisynthetic C μ 4tp variants.....	78
8.3.4	Quality control of C μ 4tp variants	78
8.4	Isolation of sialylglycopeptide from egg yolk.....	79
8.4.1	Initial SGP isolation from whole eggs.....	79
8.4.2	Updated SGP isolation protocols.....	83
8.5	Transglycosylation of C μ 4tp with endo-cc N180H	84
8.6	Analytical and structural methods.....	85
8.6.1	SDS-PAGE.....	85

Table of Contents

8.7	Folding of C μ 4tp variants	86
8.8	Reducing and non-reducing PAGE at the TUM	88
8.9	Circular dichroism spectroscopy and Thermal transition	88
8.10	Crosslinking.....	89
9	Results and Discussions.....	91
9.1	Semisynthetic strategy	91
9.2	Semisynthesis of C μ 4tp	94
9.2.1	Production of recombinant C μ 4tp 1-91 thioester.....	94
9.2.2	Synthesis and purification of C μ 4tp 92-132 wt.....	101
9.2.3	Ligation and purification of C μ 4tp wt.....	103
9.2.4	Characterization of C μ 4tp folding behavior.....	106
9.2.5	Discussion of C μ 4tp wt semisynthesis and characterization.....	108
9.3	The semisynthesis of photocrosslinking C μ 4tp variants	111
9.3.1	Synthesis and purification of C μ 4tp 92-132 Bpa variants.....	111
9.3.2	Ligation and purification of IgM C μ 4tp Bpa variants.....	117
9.3.3	Characterization of C μ 4tp Bpa variant folding behaviors	122
9.3.4	Photoaffinity labelling of C μ 4tp Bpa variants.....	126
9.3.5	Discussion of the semisynthesis and characterization of photocrosslinking C μ 4tp variants.....	129
9.4	Towards the semisynthesis of C μ 4tp glycovariants.....	133
9.4.1	Synthesis of C μ 4tp 92-132 N119AsnGlcNAc.....	134

Table of Contents

9.4.2	Ligation of C μ 4tp N119AsnGlcNAc	138
9.4.3	Isolation of and transfer of SGP glycan from egg yolk	140
9.4.4	Direct, enzymatic transfer of the SGP glycan onto C μ 4tp 1-132 N119NGlcNAc	149
9.4.5	Direct, enzymatic transfer of the SGP glycan onto C μ 4tp 92-132 N119GlcNAc .	150
9.4.6	Discussion of the semisynthesis of glycosylated C μ 4tp variants.....	152
10	Conclusion and Outlook.....	155
10.1	Semisynthesis of C μ 4tp and photocrosslinking variants.....	155
10.2	Structural implications regarding the C μ 4tp monomer	156
10.3	Semisynthesis of glycosylated C μ 4tp variants.....	157
10.4	Outlook.....	159
11	References.....	161
12	Appendix.....	173
12.1	Sequences.....	173
12.1.1	DNA Sequence of MxeIntein-H7-CBD.....	173
12.1.2	AA Sequence of MxeIntein-H7-CBD	174
12.1.3	DNA Sequence of C μ 4tp 1-91-MxeIntein-H7-CBD.....	175
12.1.4	AA Sequence of C μ 4tp 1-91-MxeIntein-H7-CBD.....	176
12.1.5	AA Sequence of C μ 4tp 1-132 wt.....	177
12.1.6	AA Sequence of C μ 4tp 1-132 L117Bpa.....	177
12.1.7	AA Sequence of C μ 4tp 1-132 Y118Bpa.....	177
12.1.8	AA Sequence of C μ 4tp 1-132 V120Bpa	178

Table of Contents

12.1.9	AA Sequence of C μ 4tp 1-132 Y132Bpa.....	178
12.1.10	AA Sequence of C μ 4tp 1-132 N119GlcNAc.....	178
12.2	Additional Photocrosslinking Data	179
12.2.1	40 min Photocrosslinking	179

5 Table of Figures

Figure 1: Generalized antibody structure. Variable region (red), constant region (blue). Disulfide bonds are highlighted in yellow.	33
Figure 2: Schematic overview of IgG. Light chains (L) are depicted in green, and heavy chains (H) are depicted in blue. Antigen binding domains are highlighted with lighter shading. Disulfide bonds are highlighted in yellow.	34
Figure 3: Immunoglobulin fold represented by a mouse IgM C μ 4 domain as cartoon model. Disulfide bond depicted in yellow. Capital letters identify β -strands (PDB ID: 4JWV) ⁴⁹⁻⁵¹	35
Figure 4: Schematic overview of human antibody isotypes.	36
Figure 5: Schematic overview of the IgM monomer. Light chains (L) are depicted in purple, and heavy chains (H) are depicted in red. Antigen binding domains are highlighted with lighter shading. Disulfide bonds are highlighted in yellow. Dashed yellow lines indicate disulfide bonds formed with neighboring IgM monomers in the oligomer.....	40
Figure 6: Schematic representation of the BCR structure.	42
Figure 7: Schematic view of hexameric IgM ⁴⁹	46
Figure 8: A) Cryo-em structure of IgM pentamer (grey with an individual μ -chain highlighted in red) with J-chain (green, SC hidden to improve clarity) and B) SC (blue) as cartoon model (PDB ID: 6KXS) ^{51,114,140}	47
Figure 9: Amino acid sequence of C μ 4 tail piece.	48

Table of Figures

Figure 10: A) Cryo-EM structure of IgM pentamer (grey with an individual μ -chain highlighted in red) with J-chain (green) and B) detailed view of the tailpiece assembly, including N-acetylglucosamine (GlcNAc, depicted in blue) attached to Asn563 (PDB ID: 6KXS) ^{51,114,140}	49
Figure 11: Schematic structure of Glc3Man9GlcNAc2	52
Figure 12: A) Cryo-EM structure of IgM pentamer (grey with an individual μ -chain highlighted in red) with J-chain (green) and B) detailed view of the tailpiece assembly, including N-acetylglucosamine (GlcNAc, depicted in blue) attached to Asn563 (PDB ID: 6KXS) ^{51,114,140}	54
Figure 13: Native chemical ligation (NCL), schematic overview	57
Figure 14: Intein cleavage for expressed protein ligation.....	59
Figure 15: Bpa (p-Benzoyl phenylalanine) (A), and Fmoc-Bpa building block for SPPS (B);	60
Figure 16: p-Benzoylphenylalanine photocrosslinking reaction ¹⁹⁸	61
Figure 17: direct transfer of the glycan from sialylglycopeptide to AsnGlcNAc by endo-CC N180H.	64
Figure 18: Amino acid sequence of C μ 4tp. Underscored black amino acids are amino acids to be substituted with Bpa. Underscored red amino acid represents a conserved glycosylation site. Expressed segment highlighted in blue, synthetic segment in green.	91
Figure 19: Semisynthesis scheme for C μ 4tp.....	92
Figure 20: Amino acid sequence of C μ 4tp 1-91 Mesna thioester. Expressed segment highlighted blue. Mesna-thioester highlighted yellow and depicted in cursive. MW _{calc} : 10263.64 Da.....	95
Figure 21: Amino acid sequence of C μ 4tp 1-91-MxeIntein-H7-CBD. C μ 4tp 1-91 highlighted in blue. Intein-H7-CBD highlighted in beige. MW _{calc} : 39197.13 Da.....	96

Table of Figures

Figure 22: Photography of representative SDS-PAGE monitoring of C μ 4tp-Intein fusion construct expression in E coli. The arrow points at the protein of interest. Expected mass ~39 kDa; M: protein marker; S: supernatant; P: pellet; Gel stained with Coomassie blue. Timepoints represent hours after induction via addition of IGPT.....	97
Figure 23: Photography of representative fraction analysis after affinity chromatography using SDS-PAGE. The arrow points at the protein of interest. Expected mass ~39 kDa; Protein marker (M), fractions (f1-6). Gel stained with Coomassie blue.	98
Figure 24: Photography of representative SDS-Page monitoring of Mesna induced intein cleavage of C μ 4tp-Intein fusion protein. MW(fusion construct): 39 kDa, MW(intein): 39197.23 Da, MW(C μ 4tp thioester): 10263.64 Da, MW(hydrolysis): 10139.47 Da, M=protein gel marker (check specifics). Gel stained with Coomassie blue. Timepoints represent hours after addition of Mesna.	99
Figure 25: Representative characterization of C μ 4tp 1-91-thioester A) structure of the C μ 4tp 1-91 Mesna thioester B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; C μ 4tp 1-91-thioester MW _{obs} : 10263 Da, MW _{calc} : 10263.64 Da.....	100
Figure 26: Amino acid sequence of C μ 4tp 92-132 wt. MW 4394.02 Da.	101
Figure 27: Characterization of C μ 4tp wt A) C μ 4tp 92-132 wt structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 4393 Da, MW _{calc} : 4394.02 Da;.....	102
Figure 28: Amino acid sequence of C μ 4tp wt. Expressed segment highlighted in blue, synthetic segment highlighted in green. Expected mass 114515.48 Da.....	103

Table of Figures

Figure 29: Reaction monitoring of the C μ 4tp 1-132 wt ligation using LCMS. A) 214 nm trace during LCMS; the arrow indicates the ligation product as well as the overlapping thioester starting material B) deconvoluted MS of the peak indicated by the arrow in panel B. MW _{obs} : 14514 Da, MW _{calc} : 14515.48 Da;.....	104
Figure 30: Characterization of C μ 4tp wt main pool A) C μ 4tp 1-132 wt structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 14514 Da, MW _{calc} : 14515.48 Da;.....	105
Figure 31: UV-VIS traces of folded C μ 4tp wt during analytical SEC display formation of oligomers and dimers.....	106
Figure 32: CD spectrum of C μ 4tp wt; ssC4 ox represents folded, semisynthetic C μ 4tp, C μ 4 ox represents folded, recombinantly expressed C μ 4tp, and C μ 4 red represents unfolded, recombinantly expressed C μ 4tp	107
Figure 33: Thermal transition spectrum of C μ 4tp wt, C μ 4 ox represents folded, recombinantly expressed C μ 4tp, and C μ 4 red represents unfolded, recombinantly expressed C μ 4tp.....	108
Figure 34: Amino acid sequence of C μ 4tp. Underscored black amino acids are amino acids to be substituted with Bpa. Underscored red amino acid represents a conserved glycosylation site. Expressed segment highlighted in blue, synthetic segment in green.	111
Figure 35: Characterization of C μ 4tp 92-132 L117Bpa via MS and analytical HPLC A) C μ 4tp 92-132 L117Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; C μ 4tp 92-132 L117Bpa MW _{obs} : 4531 Da, MW _{calc} : 4532.15 Da;.....	113

Table of Figures

Figure 36: Characterization of C μ 4tp 92-132 Y118 via MS and analytical HPLC A) C μ 4tp 92-132 Y118Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 4485 Da, MW _{calc} : 4482.13 Da;.....	114
Figure 37: Characterization of C μ 4tp 92-132 V120Bpa via MS and analytical HPLC A) C μ 4tp 92-132 V120Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 4546 Da, MW _{calc} : 4546.17 Da;.....	115
Figure 38: Test cleavage of C μ 4tp 1-132 Y132Bpa via MS A) C μ 4tp 92-132 Y132Bpa structure B) UV-VIS trace of analytical HPLC analysis of C μ 4tp 109-132Bpa after test cleavage C) mass spectrum of C μ 4tp 123-132Bpa after test cleavage D) mass spectrum of C μ 4tp 109-132Bpa after test cleavage; MW _{obs} (C μ 4tp 123-132Bpa): 1021 Da, MW _{calc} (C μ 4tp 123-132Bpa): 1022 Da; MW _{obs} (C μ 4tp 109-132Bpa): n.a., MW _{calc} (C μ 4tp 109-132Bpa): 1022 Da;.....	117
Figure 39 Characterization of C μ 4tp L117Bpa main pool 1 A) C μ 4tp 1-132 L117Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 14650 Da, MW _{calc} : 14653.60 Da;.....	119
Figure 40: Characterization of C μ 4tp Y118Bpa main pool A) C μ 4tp 1-132 Y118Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 14606 Da, MW _{calc} : 14603.58 Da;.....	120
Figure 41: Characterization of C μ 4tp V120Bpa main pool A) C μ 4tp 1-132 V120Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 14667 Da, MW _{calc} : 14667.63 Da;.....	121
Figure 42: UV-VIS traces of folded C μ 4tp variants during analytical SEC	123

Table of Figures

Figure 43: CD spectrum of C μ 4tp wt and variants. A) wt; B) L117Bpa; C) Y118Bpa; D) V120Bpa	124
Figure 44: Thermal transition data of C μ 4tp wt and variants. A) wt; B) L117Bpa; C) Y118Bpa; D) V120Bpa	125
Figure 45: SDS-PAGE analysis of C μ 4tp variants subjected to photocrosslinking.	126
Figure 46: Mass spectrum of variant L117Bpa 20 min after UV irradiation. MW _{obs} : 14650 Da, MW _{calc} : 14654 Da; MW _{obs} : 29301 Da, MW _{calc} : 29298 Da	128
Figure 47: Mass spectrum of variant Y118Bpa 20 min after UV irradiation. MW _{obs} : 14605 Da, MW _{calc} : 14905 Da; MW _{obs} : 29200 Da, MW _{calc} : 29206 Da	128
Figure 48: Mass spectrum of variant V120Bpa 20 min after UV irradiation. MW _{obs} : 14665 Da, MW _{calc} : 14668 Da; MW _{obs} : 29329 Da, MW _{calc} : 29333 Da	129
Figure 49: Amino acid sequence of C μ 4tp 92-132 N119AsnGlcNAc. X = AsnGlcNAc building block. MW 4597.22 Da.	135
Figure 50: Characterization of manually synthesized N119AsnGlcNAc N119AsnGlcNAc, deprotected with sodium methoxide, via MS and analytical HPLC. A) C μ 4tp 92-132 N119AsnGlcNAc structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 4599 Da and 4611 Da, MW _{calc} : 4597.22 Da	136
Figure 51: Characterization of automatically synthesized C μ 4tp 92-132 N119AsnGlcNAc, deprotected at high pH, via MS and analytical HPLC. A) C μ 4tp 92-132 N119AsnGlcNAc structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 4599 Da Da, MW _{calc} : 4597.22 Da	137

Table of Figures

Figure 52: Amino acid sequence of C μ 4tp N119AsnGlcNAc. X = AsnGlcNAc. Expressed segment highlighted in blue, synthetic segment in green. Expected mass 14718.67 Da.....	138
Figure 53: Characterization of C μ 4tp N119AsnGlcNAc main pool 2 via MS and analytical HPLC A) N119AsnGlcNAc structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW _{obs} : 14717 Da, MW _{calc} :14718.67 Da	139
Figure 54: Characterization of SGP isolated with the initial protocol, A) schematic representation of SGP B) UV-VIS trace during analytical RP-HPLC C) MS direct injection D) deconvoluted mass MW _{obs} : 2865 Da, MW _{calc} :2865.78 Da.....	141
Figure 55: Schematic representation of SGP fragmentation during ESI-MS.....	142
Figure 56: Characterization of SGP isolated with the updated protocol, A) schematic representation of SGP B) UV-VIS trace during analytical RP-HPLC C) MS direct injection D) deconvoluted mass MW _{obs} : 2865 Da, MW _{calc} :2865.78 Da.....	148
Figure 57: Schematic representation of SGP fragmentation during ESI-MS.....	149
Figure 58: Characterization of the glycosylated C μ 4tp 92-132 after direct transfer. A) schematic structure B) UV-VIS trace of analytical RP-HPLC C) direct injection mass spectrum D) deconvoluted mass spectrum MW _{obs} : 6597 Da, MW _{calc} : 6600.02 Da.....	151
Figure 59: Deconvoluted MS of C μ 4tp variants after 40 min of irradiation A) L117Bpa B) V120Bpa C) Y118Bp	179

Table of Figures

6 Abbreviations

ACN	Acetonitrile
BCR	B-cell receptor
Boc	<i>Tert</i> -Butyloxycarbonyl
BSA	Bovine serum albumin
CD	Circular dichroism
CDMBI	2-Chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride
C μ 4	IgM constant domain 4
C μ 4tp	IgM constant domain 4 plus the C-terminal tailpiece
DCM	Dichloromethane
DIC	<i>N,N</i> -Diisopropylcarbodiimide
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMF	<i>N,N</i> -Dimethylformamide
DNA	Deoxyribonucleic acid
DoSChem	Vienna Doctoral School in Chemistry
DTT	1,4-dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDT	1,2-ethanedithiol
ER	Endoplasmic reticulum
FDA	US Food and Drug Administration
Fmoc	9-Fluorenylmethoxycarbonyl
Gal	Galactose
Gdn.HCl	Guanidine hydrochloride

Abbreviations

GlcNAc	<i>N</i> -Acetylglucosamine
GSH	Glutathione in its reduced form
GSSG	Glutathione in its oxidized form
H ₂ O	All H ₂ O referred to throughout this thesis was filtered through a Milli-Q system.
HATU	1[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBTU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HF	hydrofluoric acid
HPLC	High-performance liquid chromatography
Ig	Immunoglobulin
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IUPAC	International Union of Pure and Applied Chemistry
LB	Lysogeny broth
LCMS	High-performance liquid chromatography coupled with a mass spectrometry instrument
Man	Mannose
Mesna	Sodium mercaptoethane sulfonate
NCL	Native chemical ligation

Abbreviations

NEM	<i>N</i> -Ethylmaleimide
Neu5Ac	<i>N</i> -Acetylneuraminic acid
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
POI	Protein of interest
PTM	Posttranslational modification
RNA	Ribonucleic acid
rt	Room temperature
SDS	Sodium dodecyl sulfate
SGP	Sialyl glycopeptide
SPPS	Solid-phase peptide synthesis
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
tp	IgM tailpiece
TUM	Technical University of Munich

Amino acids are depicted in the standard single or triple-letter codes established by the International Union of Pure and Applied Chemistry (IUPAC).

Abbreviations

The following symbols are used in the schematic representation of glycan structures in accordance with the Symbol Nomenclature for Glycans (SNFG)^{1,2}:

 *N*-Acetylneuraminic acid (Neu5Ac)

 Galactose (Gal)

 Mannose (Man)

 *N*-Acetylglucosamine (GlcNAc)

 *Glucose (Glc)*

7 Introduction

7.1 Antibodies

Antibodies, also known as immunoglobulins (Ig), are essential components of the human humoral immune response. They recognize and bind foreign antigens, ideally resulting in the opsonization and destruction of the associated pathogens through other agents of the immune system, such as macrophages or the complement system. They are mainly produced by plasma cells, a type of differentiated B cell³⁻⁵.

The body's natural immune response to an antigen is polyclonal, targeting various epitopes of the same antigen with a variety of distinct antibodies. Köhler and Milstein were the first to describe the controlled production of monoclonal antibodies (mAbs) in 1975⁶, for which they were awarded the Nobel prize in 1984 together with Jerne. mAbs have since been subjected to extensive efforts in protein- and glycoengineering⁷⁻⁹ and antibody drug conjugation^{7,10}. Today, mAbs are widely used to treat cancer and autoimmune diseases such as multiple sclerosis¹¹⁻¹⁵, heavily contributing to a paradigm shift in drug development away from small molecules towards biologics. The FDA has also approved numerous antibody therapeutics for treating viral diseases such as rabies¹⁶, varicella¹⁷, measles¹⁸, hepatitis A¹⁸, hepatitis B¹⁹, and others²⁰.

Each year, hundreds of antibody therapeutics are evaluated in clinical studies. In 2022 nearly 1200 antibody therapeutics were in clinical studies, almost 140 of which were in Phase 2 or 3²¹. More than 160 antibody therapies have been approved by at least one regulatory agency in the world, with more than a hundred monoclonal antibody products approved by the FDA^{22,23}.

The global market for antibody therapies was estimated at 186 billion USD with a compound annual growth rate of above 13%, potentially reaching a size of over 440 billion USD in 2028²². Beyond their tremendous success as therapeutics, antibodies are well-established tools in diagnostics and research. Immunoblots, such as the ubiquitous western blot²⁴, immunoassays, such as the enzyme-linked immunosorbent assay (ELISA²⁵) and lateral flow tests (LFT²⁶), and cell staining applications, such as flow cytometry/FACS^{27,28} and immunoprecipitation²⁹, rely on the specificity of an antibody to its antigen to detect proteins of interest.

7.2 Antibody structure

For historical reasons, antibodies are also divided into an antigen binding- and a crystallizable region (Figure 1). This nomenclature is based on experiments by Rodney R. Porter in which antibodies were digested with papain^{30,31}. This resulted in two fragments still capable of antigen binding (hence called Fragment antigen binding or Fab) and a fragment that could be crystallized (therefore named Fragment crystallizable or Fc)³². Complex mechanisms such as V(D)J recombination³³, somatic hypermutation³⁴, and affinity maturation³⁵ create a vast repertoire of distinct antigen-binding regions. At the same time, a rigorous selection process prevents the formation and release of self-targeting antibodies. Failure of this selection process results in various autoimmune disorders such as systemic lupus erythematosus³⁶, rheumatoid arthritis³⁷, multiple sclerosis³⁸⁻⁴⁰, Graves' disease⁴¹, Hashimoto's thyroiditis^{42,43}, and others^{44,45}.

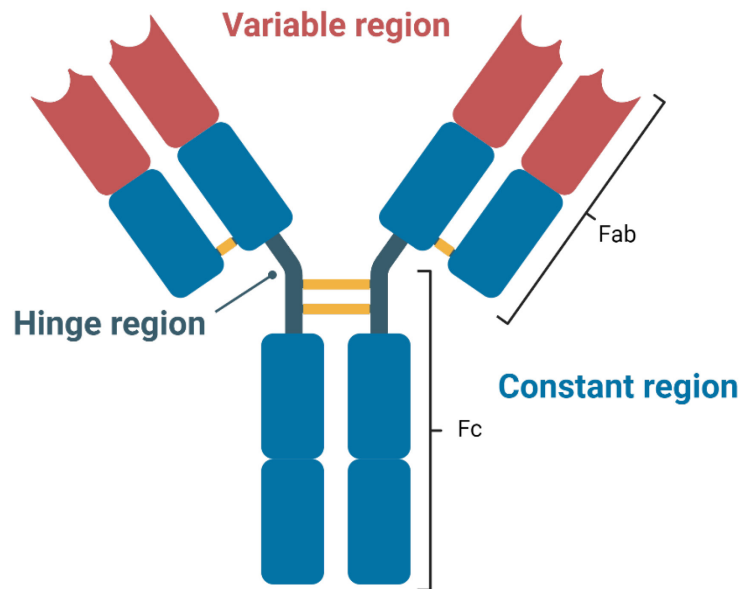


Figure 1: Generalized antibody structure. Variable region (red), constant region (blue). Disulfide bonds are highlighted in yellow.

Antibodies are made of two light (L) and two heavy (H) chains connected *via* disulfide bonds resulting in H_2-L_2 dimers^{46,47}. This H_2-L_2 structure will be referred to as the antibody monomer throughout this thesis.

Both chains can be divided into variable- (V) and constant- (C) domains (Figure 2). Light chains consist of one V_L and one V_C domain. Heavy chains contain one V_H and multiple C_H domains.

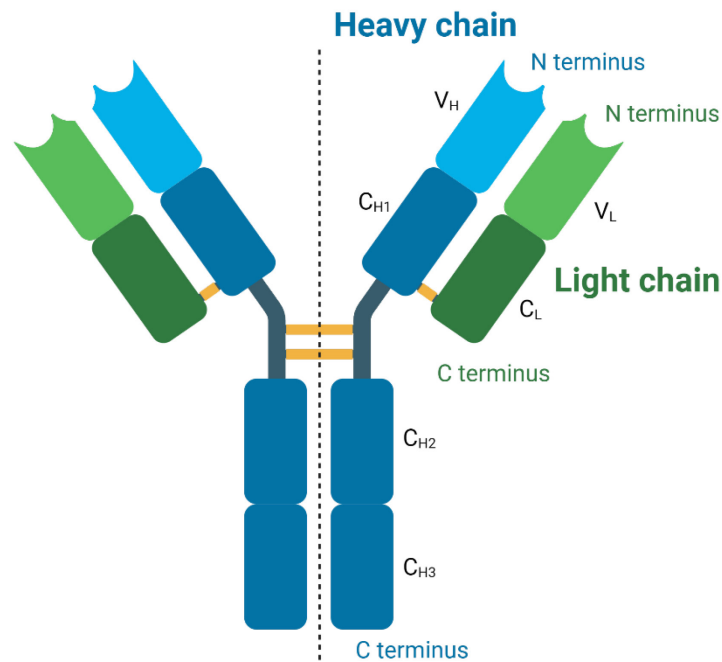


Figure 2: Schematic overview of IgG. Light chains (L) are depicted in green, and heavy chains (H) are depicted in blue. Antigen binding domains are highlighted with lighter shading. Disulfide bonds are highlighted in yellow.

The variable domains of L- and H- chains (V_L and V_H) form the variable region responsible for antigen binding, while the constant domains of both chains (C_L and C_H, respectively) provide structure and effector functions and thus determine the isotype of the antibody, which will be discussed extensively in the next section (7.3). Each of these domains is approximately 110 amino acids long and somewhat similar in structure. These conformations are therefore referred to as the “immunoglobulin fold”. They consist of two antiparallel β -sheets that stack on top of each other. The β -sheets of the constant domains contain four β -strands ($\downarrow A \uparrow B \downarrow E \uparrow D$) and three β -strands ($\downarrow C \uparrow F \downarrow G$). An intra-domain disulfide bond between two cysteine residues in the B and F β -strands covalently links the the two β -sheets, while the β -strands themselves are connected *via* short loops⁴⁸.

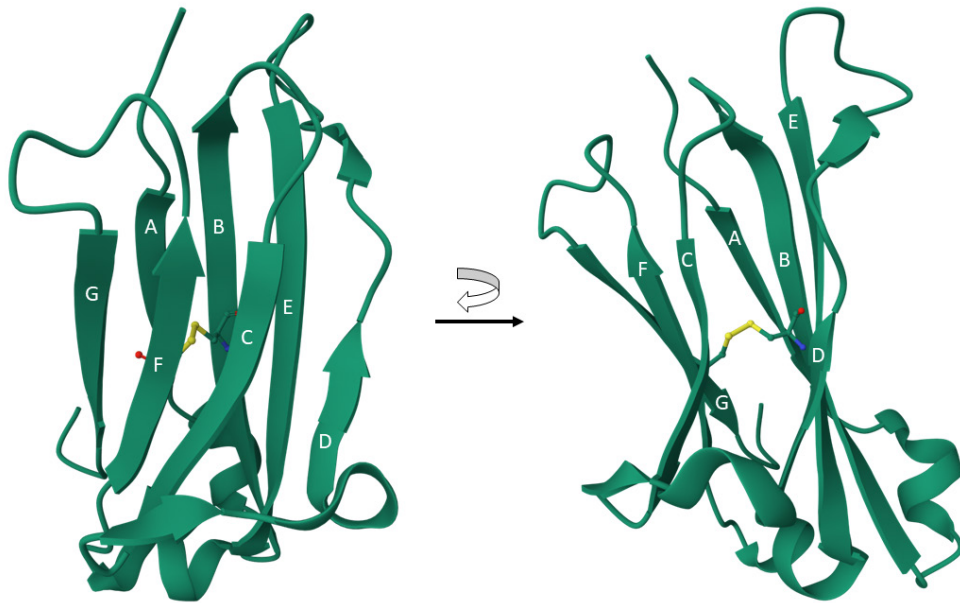


Figure 3: Immunoglobulin fold represented by a mouse IgM C μ 4 domain as cartoon model. Disulfide bond depicted in yellow. Capital letters identify β -strands (PDB ID: 4JVV)⁴⁹⁻⁵¹.

7.3 Antibody isotypes in humans

Antibodies are categorized into different isotypes based on their Fc regions. In humans, the following five antibody isotypes are found: IgA, IgD, IgE, IgG, and IgM⁵². The respective heavy chains are denoted with Greek letters matching their isotype: α for IgA, δ for IgD, ϵ for IgE, γ for IgG, and μ for IgM. The length of these heavy chains varies from ~450 amino acids to ~550 amino acids and three (IgA, IgD, and IgG) to four (IgE and IgM) constant domains. IgD, IgE, and IgG are present as monomers, while IgA and IgM can form oligomers (Figure 4). These isotypes vary in time of production, tissue distribution and efficacy against different types of pathogens. Of these isotypes, IgG is the most prevalent in serum, comprising approximately 75% of all Igs in circulation, followed by IgA (~15%) and IgM (~10%)^{23,53}.

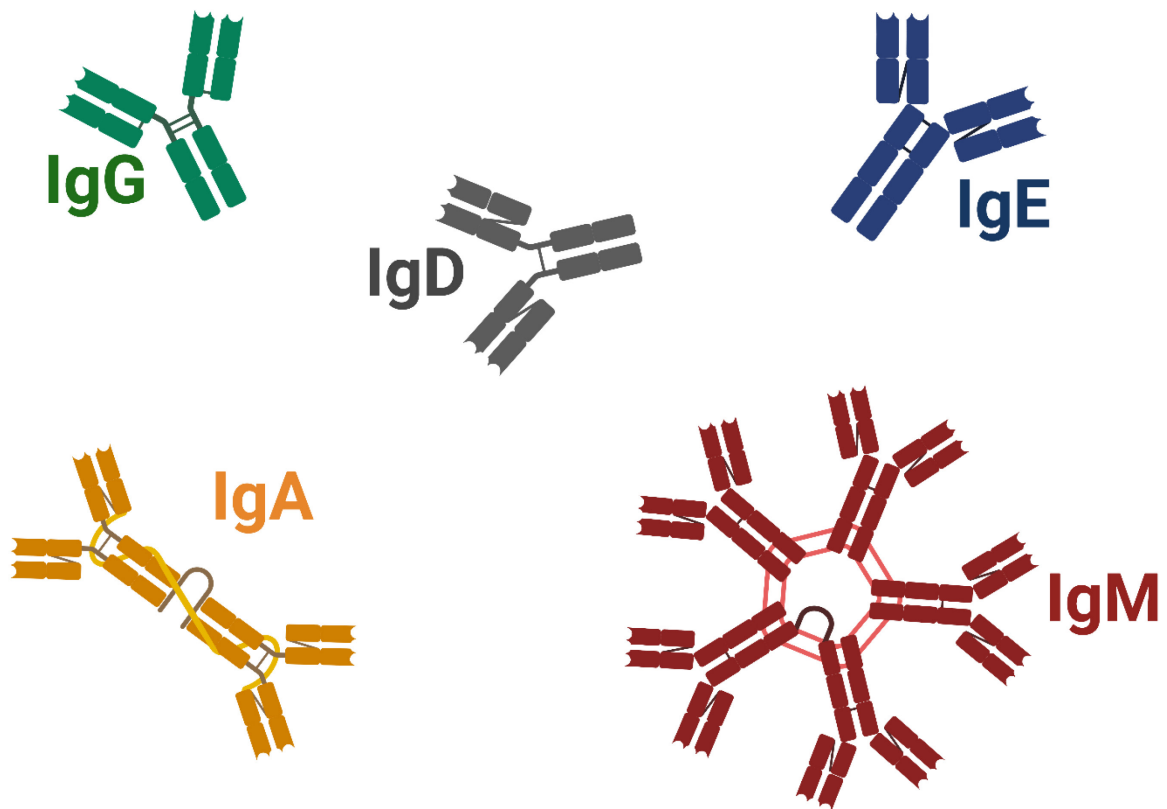


Figure 4: Schematic overview of human antibody isotypes.

7.3.1 IgA

IgA is present in serum primarily as monomer, while dimeric IgA is secreted in the mucosa in a complex with the so-called secretory component, a heavily glycosylated peptide⁵⁴. This dimerization and secretion depends on the presence of a short protein called the J-chain⁵⁵ that is also present in secretory IgM and is discussed further in section 7.4. This unique structure allows IgA to survive in the proteolytic environment of the mucosal surfaces, where it prevents the invasion of pathogens, and influences the metabolism of the local microbiomes⁵⁶. IgA can be divided into two subclasses, IgA1 and IgA2. Both play crucial roles in humoral mucosal immunity⁵² and are the most abundant antibody isotypes in mucosal

secretions⁵⁷. Therapeutic application of plasma-derived antibody preparations containing significant amounts of IgA is very limited and includes IgAbulin, Vennimmun N, Pentaglobin, and Trimodulin⁵⁸. Nonetheless, inhalation of antibodies, including IgA, may be a potential strategy for the treatment of inflammatory lung diseases⁵⁸⁻⁶⁰.

7.3.2 IgD

IgD is present in mammary, nasal, salivary, lacrimal, and bronchoalveolar secretions and is critically important in regulating immune responses in the respiratory tract^{57,61}. It binds to mast cells and basophils, inducing the production of antimicrobial peptides and inflammatory cytokines⁶². IgD likely contributes to mucosal homeostasis. While IgD has been highly conserved for the last 500 million years of evolution⁶³, its role in the immune system remains, as-of-yet, not well defined.

7.3.3 IgE

IgE is an isotype best known for its role in allergic diseases such as allergic asthma, allergic dermatitis, allergic rhinitis, and food allergies⁶⁴. However, it is also involved in the immune response against infections with parasites, such as helminths⁶⁵⁻⁶⁷. It is present in nasal, bronchoalveolar, and intestinal secretions⁵⁷. IgE has the shortest half-life of all antibody isotypes but is stabilized when bound to mast cells, which provides mast cells and basophils with antigen specificity in allergic responses⁶⁸.

7.3.4 IgG

IgG is the most prevalent antibody in humans, accounting for 10-20% of protein content in serum⁵². It is also the primary antibody isotype utilized in protein engineering and monoclonal

therapeutics⁶⁹. There are four subclasses of IgG (IgG1-4) in humans. Although these subclasses display an above 90% similarity in their amino acid sequences, they vary in antigen binding, complement activation, half-life, and placental transport. IgG1 appears to be primarily produced as a response against soluble proteins and membrane proteins, while IgG2 is overrepresented in immune responses against bacterial capsular polysaccharides. IgG3 is a highly pro-inflammatory antibody with a shorter half-life. While it plays an important role in viral infections, it rarely appears to be the dominant IgG subtype. IgG4 antibodies are often formed after repeated exposure to an antigen in a non-infectious setting and may be involved in regulating allergic responses⁷⁰. The subtypes also vary in the flexibility of the hinge region, likely resulting in enhanced antigen and complement activation for IgG1 and IgG3 as opposed to IgG2 and IgG4⁷¹.

7.3.5 IgM

IgM is discussed in detail in the next section, as it is the antibody isotype that this thesis focuses on.

7.4 IgM

IgM was first described by Kabat and Pedersen during the study of antibody sedimentation constants⁷². Due to its large size of ~1,000 kD, this antibody was named γ -macroglobulin. Based on the World Health Organization nomenclature of immunoglobulin isotypes^{72,73}, the macro remains as M in IgM.

IgM is the most highly conserved class of antibody. It can be found in all jawed vertebrates, including mammals, birds, and amphibians, but not in jawless fish. Interestingly, IgM is also

present in vertebrate classes that have not evolved features like class switching or germinal centers, such as bony fish and cartilaginous⁶³.

IgM has two distinct forms in humans: membrane-bound IgM (mIgM)⁷⁴ and secretory IgM (SIgM)⁷⁵. mIgM is initially produced as part of a membrane-bound B cell receptor. During an immune response, IgM is produced by plasma cells and secreted across mucosal surfaces as a soluble pentamer, making sIgM the most structurally complex antibody isotype (7.4.3).

7.4.1 Structure of the IgM monomer

Like other antibody isotypes, IgM forms monomers consisting of two light and two heavy chains (Figure 5). The light chain contains two domains (V_L and V_H). The heavy chain of IgM is the so-called μ chain. It has one variable domain (V_μ) and four constant domains ($C\mu1$ to $C\mu4$). The $C\mu4$ domain carries an 18 amino acid-long C-terminal extension called the tailpiece (tp). Together the $C\mu4$ domain and the tailpiece ($C\mu4tp$) are necessary and sufficient for the formation of IgM oligomers⁷⁶, which are further discussed in section 7.4.4.

The light chain is linked to the $C\mu2$ domain via a disulfide bond in position Cys140. The heavy chains of the IgM monomer are connected by a disulfide bond in position Cys337, located at the end of the $C\mu2$ domain⁷⁷. Intramolecular disulfide bonds in each constant domain of the heavy chain are present between Cys22 and 97, Cys153 and 212, Cys259 and 320, Cys367 and 426, and 474 and 536, respectively⁷⁸.

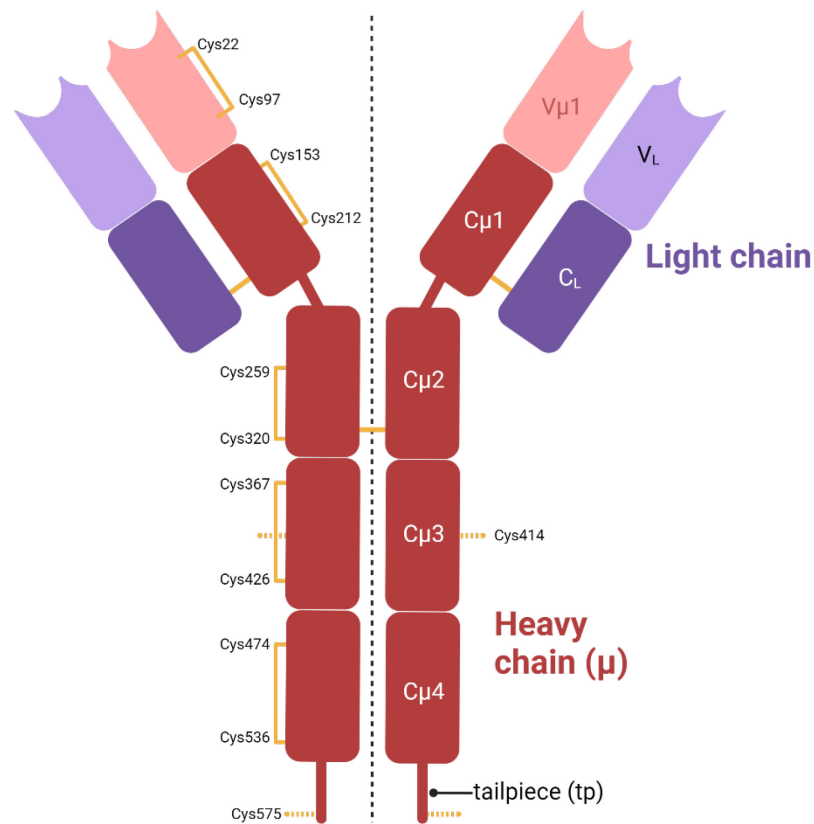


Figure 5: Schematic overview of the IgM monomer. Light chains (L) are depicted in purple, and heavy chains (H) are depicted in red. Antigen binding domains are highlighted with lighter shading. Disulfide bonds are highlighted in yellow. Dashed yellow lines indicate disulfide bonds formed with neighboring IgM monomers in the oligomer.

7.4.2 Membrane-bound IgM and the B cell receptor

B cells, or B lymphocytes, are critical players in the adaptive immune system. Their plasma membrane carries immobilized antibodies as part of the B cell receptor. Upon binding their antigen, they proliferate and differentiate, turning into antibody-secreting plasma cells^{79,80}.

They start as naïve bone marrow-derived B cells expressing both IgM and IgD that develop into germinal center B cells, which undergo hypersomatic mutation during the T-cell-dependent immune response. This hypersomatic mutation dramatically increases the affinity

of the V region of the antibody genes towards their antigen. Selected B cells then undergo class switching, which changes the isotype of the expressed antibody, and are released into the periphery, where they act as long-lived memory B cells. These memory B cells dramatically decrease the response time of the innate immune system upon repeated exposure to the respective antigen⁸¹⁻⁸³. However, not all memory B cells undergo class switching resulting in IgM expressing memory B cells⁸¹.

B cells recognize foreign antigens via their surface-expressed B cell receptor (BCR)⁸⁴. BCRs are composed of two transmembrane proteins, Ig α (CD79A) and Ig β (CD79B), connected to membrane-bound antibodies, such as IgM^{85,86} (Figure 6). The membrane-bound, monomeric antibody retains the canonical Y shape outlined in section 7.2. The Fab parts of the antibody are free to bind their antigen, while the Fc region binds the Ig α -Ig β heterodimer at a 1:1 ratio⁸⁷. The transmembrane helices of the antibody and the Ig α -Ig β heterodimer stick together inside the B cell membrane, interacting with each other through various polar residues in the otherwise hydrophobic membrane environment. Disruption of these interactions *via* mutation severely inhibits antigen presentation downstream of receptor activation⁸⁸. Interestingly, this transmembrane structure resembles the T cell receptor transmembrane structure⁸⁹.

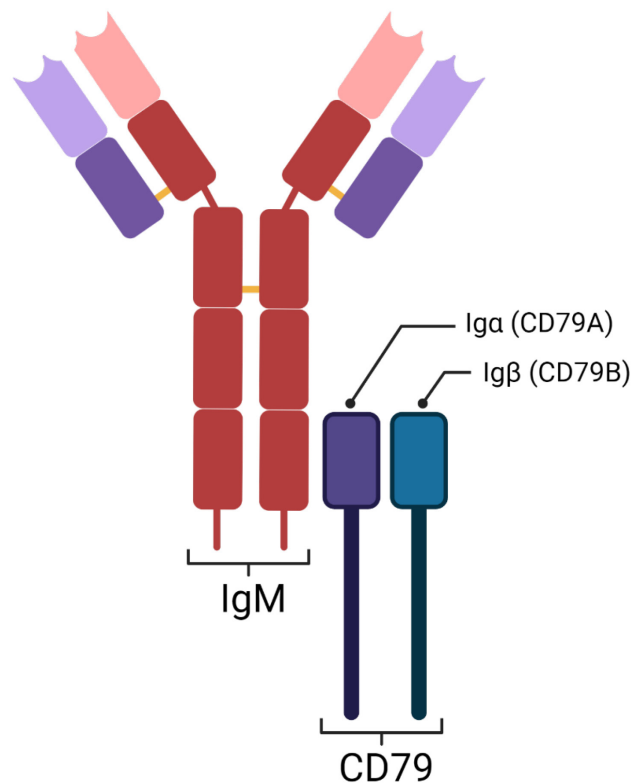


Figure 6: Schematic representation of the BCR structure.

IgM is the first antibody isotype expressed in B cell development⁷⁴ with naïve B cells co-expressing IgD and IgM BCRs with identical specificity^{90,91}. Upon exposure to their antigen, naïve B cells form germinal centers. These transient structures are inside secondary lymphoid organs such as the spleen and lymph nodes. They provide a microenvironment for B cells to undergo antibody affinity maturation. This process is based on additional diversification of variable immunoglobulin genes through somatic hypermutation³⁴. BCRs sense the affinity of their antigen binding, providing B cells with a superior ability to recognize an antigen with a selection advantage, thereby gradually improving the binding capabilities of antibodies to an antigen over the course of an infection resulting in high-affinity antibodies^{92,93}. This ultimately results in the B cell undergoing class switching and differentiating into either antibody-

secreting plasma cells or memory cells, providing long-term immunity⁹⁴. The BCR also regulates essential B cell functions beyond proliferation, such as affinity maturation, and plays a vital role in vaccine-induced immunity⁹⁵.

The proliferation induced by BCR is regulated by secondary signals provided, amongst others, by CD4⁺ helper T cells. Without these secondary signals, BCR signaling induces cell death. The BCR functions as an endocytic receptor, capturing antigens and delivering them to intracellular proteolytic processing, after which the antigens are loaded onto the major histocompatibility complex class II (MHC II), which presents the antigen to antigen-specific helper T cells⁹⁵. T cells also induce class-switching, which changes the IgD and IgM antibody isotypes to IgA, IgE, or IgG³⁴, all of which also form BCR on the B cell surfaces⁹⁵.

7.4.3 Secretory IgM

Secretory IgM is an early-phase antibody most active in the early stages of an immune system response against pathogens. At this stage, antibodies have not yet undergone affinity maturation, resulting in a relatively low affinity⁹⁶. This is counteracted *by* the high avidity displayed by oligomeric IgM⁹⁷⁻⁹⁹ (see section 7.4.4). Interestingly, IgM production is not limited to the initial antigen response. IgM-expressing memory B cells can produce IgM that has undergone somatic hypermutation^{100,101}. These mutated IgM-expressing B cells may account for up to 20% of B cells in peripheral blood⁸¹.

Secretory IgM is the most potent complement activator of all antibody isotypes¹⁰², primarily caused by its unique oligomeric structure (see section 7.4.4). The complement system is a highly conserved, critical component of the innate immune system that can also be activated

by complement-activating antibodies and serves as a first-line defense against pathogens. It encompasses a complex proteolytic cascade that ultimately forms a membrane attack complex (MAC), a large pore in the target cell's membrane, which leads to lysis and cell death¹⁰³.

It is mainly found at mucosal surfaces in the lung, gut, and liver⁷⁵, where it serves as an early defense mechanism to antigen challenge and facilitates the removal of apoptotic cells¹⁰⁴. It can compensate for selective deficiency of IgA, which should also be present at these mucosal surfaces, to a large extent^{105,106}. However, this compensation is less effective in the respiratory tract than in the gastrointestinal system, resulting in an increased susceptibility of IgA-deficient individuals to infections of the airways¹⁰⁷. Selective IgM deficiency displays symptoms similar to allergies, respiratory infections, and autoimmune diseases¹⁰⁸ and significantly changes the distribution of B cell subsets compared to healthy individuals¹⁰⁹. Secretory IgM is transferred through breast milk early in life and plays an essential role in the development of immune tolerance⁷⁵. Interestingly, IgM-secreting B cells may also play a role in the loss of self-tolerance of the immune system in aging⁷⁵.

There are three receptors for IgM: polymeric Ig receptor (pIgR), Fc α / μ R, and Fc μ R. pIgR is expressed in the intestine, lungs, kidneys, and mucosal epithelium. It selectively recognizes pentameric IgM and dimeric IgA containing a J chain and initiates their transcytosis across to the mucosal epithelial layer^{110,111}. This transport leaves behind a peptide fragment from the ectodomain of pIgR, called the secretory component (SC)¹¹⁰⁻¹¹⁵. Fc α / μ R also binds both IgM and IgA and initiates the internalization of antibody coated microorganisms by B cells macrophages¹¹⁶. It is also expressed in lymph nodes, the appendix, and kidneys. Fc μ R is the

only IgM-specific receptor in mammals and recognized IgM to regulate BCR activity¹¹⁷⁻¹²⁰. This receptor is furthermore overexpressed in chronic lymphocytic leukemia B cells and internalized upon IgM binding¹²¹. FcμR appears to play a role in B-cell survival and immune system development^{117,118,122}. However, its precise role is still not well understood¹²³.

7.4.3.1 Natural IgM

A unique subtype of IgM is so called natural IgM^{124,125}. Natural IgM is a part of the innate immune system that is encoded by unmutated germline variable gene segments and is evolutionarily conserved across jawed vertebrates⁶³. It is spontaneously produced by B-1 cells without prior immunization^{126,127}. Together with natural killer cells, dendritic cells, mast cells, and macrophages, it represents a first line of defense against pathogens such as bacteria and viruses. Natural IgM is polyreactive, targeting conserved antigens such as carbohydrates, phospholipids, lipopolysaccharides, DNA fragments, and autoantigens with low affinity¹²⁶. This low affinity is counteracted by the high avidity of the IgM penta- and hexamers⁹⁷⁻⁹⁹ (see section 7.4.4).

Natural IgM also targets altered self-antigens and neo-epitopes of dying cells and marks these for removal. Natural IgM antibodies in mice enhance phagocytic clearance of dead and dying cells and keep innate immune signaling in check to prevent uncontrolled inflammation¹²⁵. Natural IgM also plays a role in controlling B cell development, selection, and induction of central tolerance to prevent autoimmunity¹²⁸ and appears to stimulate T cell-dependent responses and class-switching upon stimulation with antigen during primary response¹²⁹.

7.4.4 The structure of polymeric IgM

IgM is the most structurally complex of the human antibody isotypes. Like IgA, IgM is polymeric and contains an additional 137 amino acid long peptide called the J or joining chain^{130,131}. In the 1970s, electron micrographs first provided evidence for the formation of IgM pentamers with a five-fold symmetry in the presence of a J-chain¹³⁰⁻¹³² and that this J-chain, as well as disulfide exchanging enzymes, are required for IgM oligomerization¹³³. This oligomerization is assumed to counteract the relatively lower binding of IgM in the early stages of infection with the increased avidity of polymeric IgM⁹⁷⁻⁹⁹. This high avidity also increases IgM effector function, such as complement activation¹³⁴.

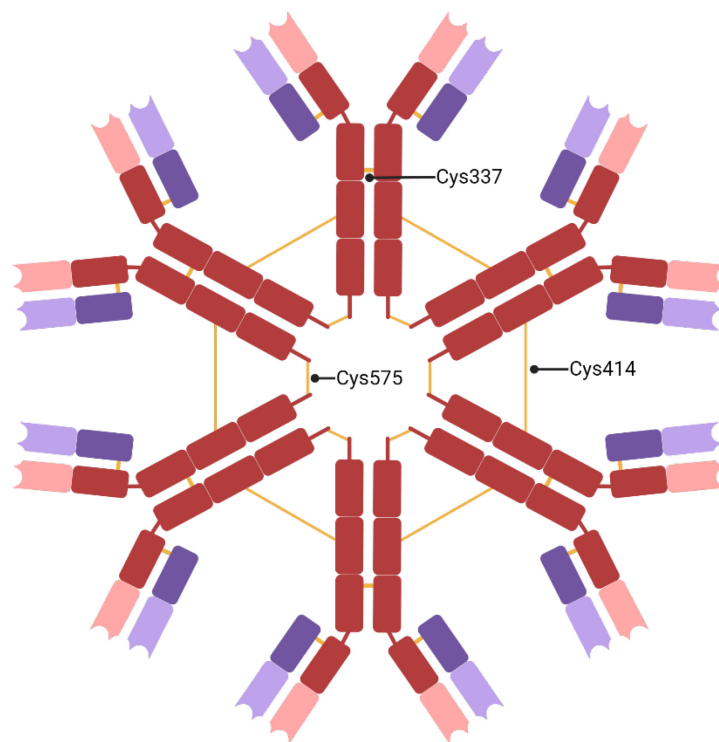


Figure 7: Schematic view of hexameric IgM⁴⁹.

More recent single-particle negative-stain electron microscopy and cryo-electron microscopy (cryo-EM) data suggest an asymmetric pentamer with a 50° to 60° gap occupied by a J-chain^{114,135}. It is possible that an additional (μ_2 -L₂) unit could take the place of the J-chain to form almost perfectly hexagonally symmetric IgM units^{114,136}. This hexameric IgM can also be detected in serum but comprises only 5% of total serum IgM⁹⁹. This results in a complex of over 1.200 kDa, secreted by plasma cells and transcytosed through the mucosal epithelial layer by the polymeric immunoglobulin receptor (pIGR). pIGR binds to J-chain-containing, pentameric IgM at the basolateral surface of epithelial cells and releases it at the apical side *via* proteolysis of its ectodomain^{55,113}. This process leaves a part of the ectodomain, the secretory component (SC), bound to IgM^{115,137}. This secretory component is present in both secreted IgA and IgM as an additional glycoprotein component^{130,138,139}.

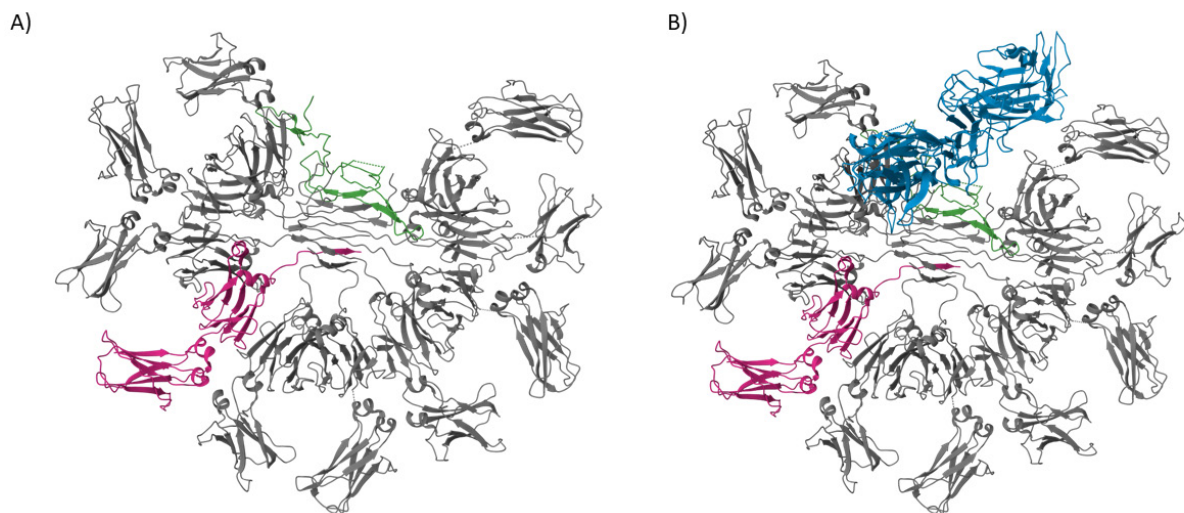


Figure 8: A) Cryo-em structure of IgM pentamer (grey with an individual μ -chain highlighted in red) with J-chain (green, SC hidden to improve clarity) and B) SC (blue) as cartoon model (PDB ID: 6KXS)^{51,114,140}.

Li *et al.* published high-resolution cryo-EM data of the IgM Fc region (Fc μ), J-chain, SC complex¹¹⁴. In the Fc μ dimer, the C μ 3 domains were observed not to be in contact with each other. This contrasts with the pentamer, where the C μ 3 domains were reported to contribute to pentamer formation. This IgM Fc μ pentamer, in conjunction with the J-chain, forms a near-planar structure with a triangular SC docking perpendicular to the Fc μ -J plane ¹¹⁴. Previous negative-stain electron microscopy experiments by Feinstein *et al.* showed that the J chain occupies an approximately 50° gap in this structure¹³². In contrast, Li *et al.* report a 61° gap with the five Fc μ units arranged in almost perfect hexagonal symmetry¹¹⁴.

The Cys414 of the C μ 3 domain and the penultimate Cys575 of the C μ 4tp form disulfide bonds between adjacent Fc μ units, with Cys575 being an essential prerequisite for assembly ^{49,76,78,141}. However, the C μ 4 domain and its C-terminal tailpiece (C μ 4tp, Figure 9) are sufficient for the specific polymerization of C μ 4 dimers into covalently linked hexamers ⁴⁹. Cys575 also connects to the J-chain (for 2 out of 10 monomer subunits) *via* J-chain residues Cys14 and Cys68, respectively ¹¹⁴.

C μ 4–**PTLYNVSLIMSDTGGTCY**–OH

Figure 9: Amino acid sequence of C μ 4 tail piece.

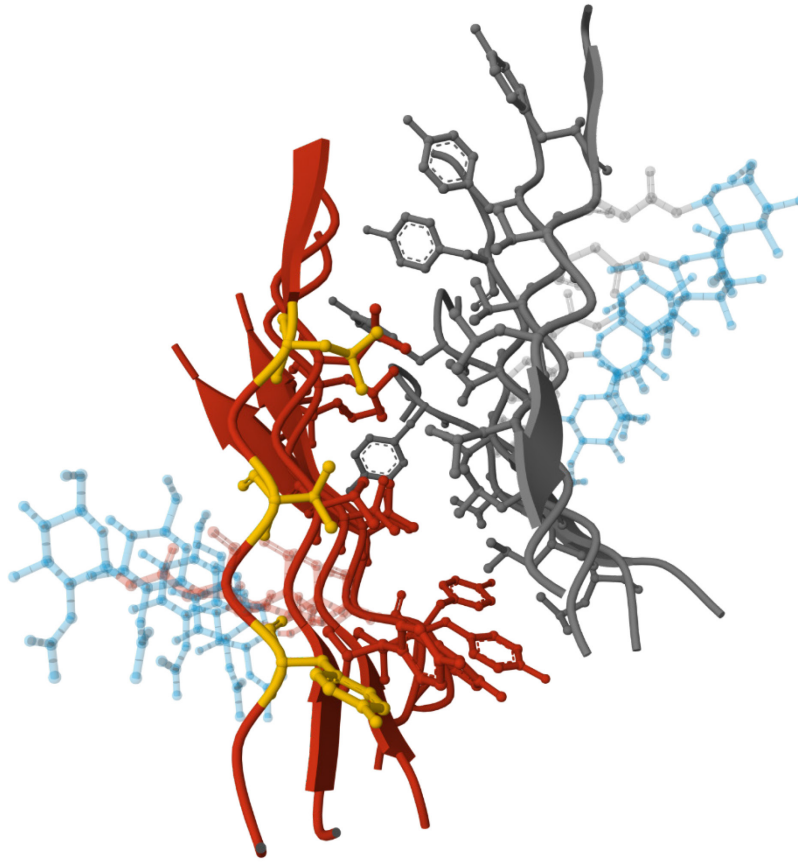


Figure 10: A) Cryo-EM structure of IgM pentamer (grey with an individual μ -chain highlighted in red) with J-chain (green) and B) detailed view of the tailpiece assembly, including N-acetylglucosamine (GlcNAc, depicted in blue) attached to Asn563 (PDB ID: 6KXS)^{51,114,140}.

Pasalic *et al.* proposed a model for C μ 4tp oligomerization, wherein conformational changes induce the formation of Cys575 disulfide bonds, which is followed by the assembly of C μ 4tp oligomers, leading to the formations of hexamers⁷⁶. The residues Tyr562 to Met568 in each tp form beta-strands, resulting in two five-stranded parallel beta sheets that pack onto one another, in an antiparallel fashion, in the IgM pentamer¹¹⁴. This is reminiscent of amyloid proteins and peptides¹⁴². Tyr562, Val564, Leu566, and Val567 residues point inward and

mediate hydrophobic interactions between the two sheets ¹¹⁴. In mice, Y562A, V564, and L566A are exclusively secreted as monomers in mice, which further highlights the importance of hydrophobic interactions between the disrupting the hydrophobic interaction of the C μ 4tp beta sheets for IgM assembly ⁷⁶. Val567 residues in human IgM remain exposed to the surface, stack onto each other, and form extended hydrophobic patches ¹¹⁴, while I567 in mice causes severe disruption of oligomerization when mutated ⁷⁶.

7.4.5 C μ 4tp as hexamerization aid

As described above, IgA also contains a short C-terminal peptide extension with high sequence similarity to the IgM tp, including patches of hydrophobic residues and the presence of a cysteine residue in the penultimate position. Analogous to the function of the tailpiece in IgM, the IgA tp also drives the oligomerization of IgA¹⁴³.

Fusion of the C μ 4 tailpiece to the C-terminus of IgG was reported to promote spontaneous IgG hexamer formation and recruitment of C1q, the first subcomponent of the classical pathway of complement activation. This resulted in increased complement-dependent cytotoxicity and off-target complement activation. Replacement of the penultimate cysteine in position C575 with a serine residue prevented spontaneous complement formation and facilitated reversible, concentration-dependent hexamer formation. The C575S-mutant displayed increased complement activation only after target binding, highlighting the potential uses of tail-piece mutants in fusion constructs and antibody engineering¹⁴⁴.

Another fascinating use of C μ 4tp hexamerization is in the design of fusion constructs for the neutralization of viral proteins. Svilenov *et al.* used this strategy to develop polymeric ACE2-

IgM-Fc fusion constructs for neutralization of SARS-CoV-2 variants, that displayed vastly improved potency compared to monomeric ACE2-IgM-Fc ¹⁴⁵.

7.5 Antibody Glycosylation

7.5.1 Antibody glycosylation

All human immunoglobulins are glycoproteins. Glycans help to improve the stability¹⁴⁶ and solubility of the attached antibody ¹⁴⁷, assist in folding^{76,147} maintaining the structure^{146,148}, improve binding to Fc receptors¹⁴⁹ and lectins¹⁵⁰, transport, secretion, and clearance¹⁵¹. Correct, non-immunogenic Fc glycosylation is essential in producing therapeutic antibodies, with antibodies produced in plant and animal cell culture expression systems suffering from reduced effector functions due to differences in glycan structures¹⁵². Significant effort has been put into the development of technologies for the production of fully humanized antibody therapeutics, including humanized mammalian cell culture systems^{148,152}, yeast expression systems expressing human glycosylation enzymes^{153,154}, and plant-based expression systems with modified trafficking systems aimed at avoiding plant-specific glycosylation in the Golgi^{152,155,156}.

Naturally occurring glycoproteins are generally highly heterogeneous, with distinct glycoforms at the same position of a given protein ¹⁵⁷⁻¹⁵⁹. The controlled production of proteins carrying fully defined glycan structures is an ongoing challenge discussed in chapter 7.6.6.

Natural N-glycosylation of antibodies is initiated in the lumen of the endoplasmic reticulum (ER), where an oligosaccharide transferase co-translationally transfers a preassembled N-glycan structure consisting of fourteen sugar residues (Glc3Man9GlcNAc2, Figure 11) from

dolichol pyrophosphate to the asparagine residue of the N-linked glycosylation consensus sequence Asn-X-Ser/Thr, where X is any amino acid except proline. The tetradeca oligosaccharide is then trimmed through sequential hydrolytic cleavage of the glucose residues by glucosidases I and II in ER, which assists in chaperone-assisted protein folding. Before leaving the ER, the glycan is subjected to ER mannosidase I, which results in a Man8GlcNAc2 precursor. The antibody is then translocated into the Golgi, where it is further trimmed and remodelled^{148,160}.

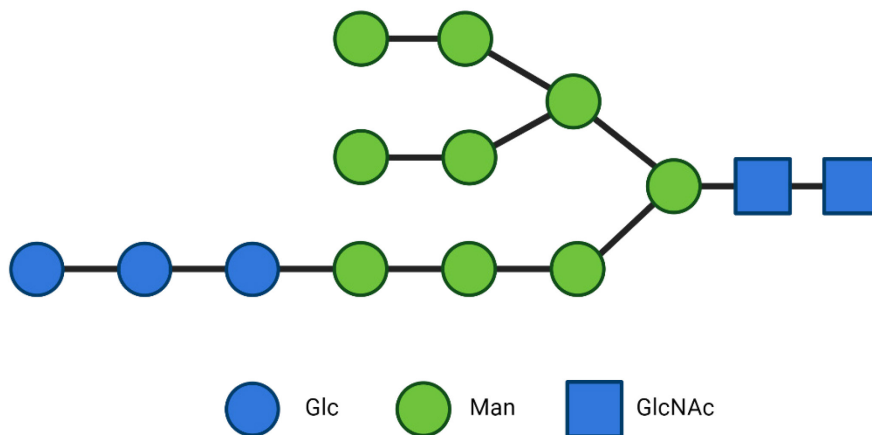


Figure 11: Schematic structure of Glc3Man9GlcNAc2.

The final glycans are classified into three types: mannose-rich, hybrid, and complex. Oligomannose glycans contain between five to nine branching mannose residues that are not further modified. Complex glycans cover a large variety of branched structures that dramatically exceed the high mannose oligosaccharide in complexity. They contain a core structure consisting of GlcNAc2Man3GlcNAc2 and various additional sugar residues such as galactose, fucose, and sialic acid^{148,157,160}.

The glycans on immunoglobulins show significant variability in the amount and position of N-linked glycosylation sites, ranging from two to fourteen percent of the total molecular weight of the heavy chain being contributed by glycosylation depending on the isotype¹⁵². In contrast, light chains do not appear to contain conserved glycosylation sites¹⁵². Oligosaccharides attached to immunoglobulins are generally large, approximately 2 kDa each¹⁵².

7.5.2 IgM glycosylation

IgM contains three N-linked glycosylation sites for complex glycans in positions Asn171 (C μ 1), Asn332 (C μ 2), and Asn395 (C μ 3), as well as two glycosylation sites for oligomannose glycans in positions Asn402 (C μ 3) and Asn563 (tp)^{152,161,162}. The J chain also contains an N-linked glycan at position Asn71^{161,163}. These glycosylation sites are occupied at different frequencies, with Asn402 and Asn563 in human serum being occupied at 100% and 17%, respectively, and only 20% of serum IgM binding to mannan-binding lectin, which binds oligomannose and GlcNAc-terminating structures¹⁶⁴. The glycosylation site formed by Asn563 and Ser565 appears to be essential in preventing IgM aggregations¹¹⁴, as mouse N563A and S565 mutants lacking glycan moieties are secreted as hexamers and higher MW species^{76,162}.

Sialylated N-linked glycans induce internalization of IgM in T cells, which significantly inhibits T cell responses. In contrast, in the absence of sialic acid, T-cell responses are not deterred, indicating significant contributions of IgM to the regulation of T-cell-mediated immune responses¹⁶⁵. This motivated us to explore semisynthetic strategies for the creation of glycosylated IgM C μ 4 variants.

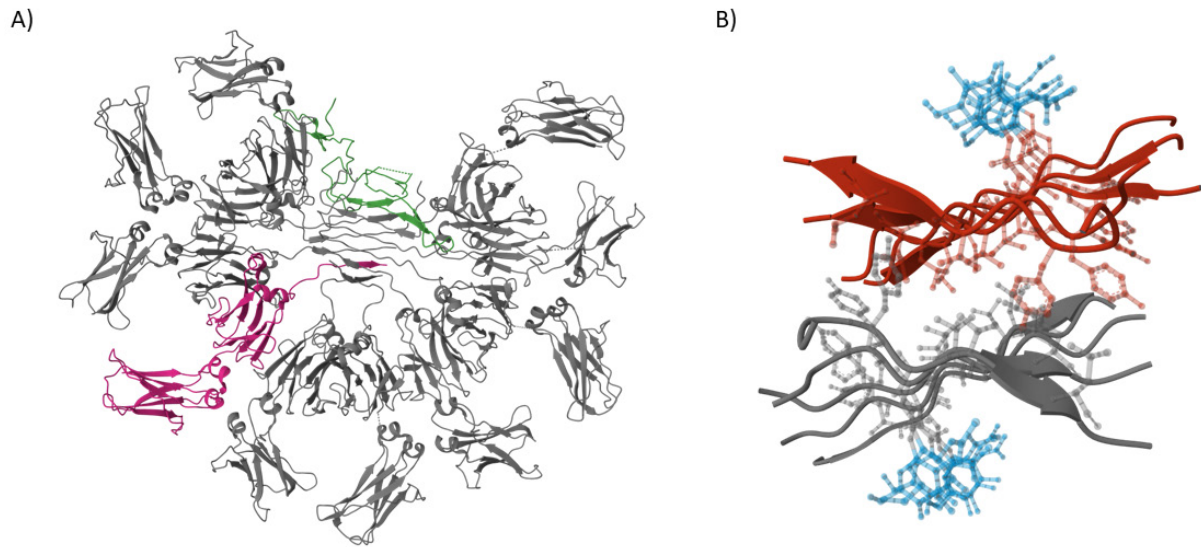


Figure 12: A) Cryo-EM structure of IgM pentamer (grey with an individual μ -chain highlighted in red) with J-chain (green) and B) detailed view of the tailpiece assembly, including N-acetylglucosamine (GlcNAc, depicted in blue) attached to Asn563 (PDB ID: 6KXS)^{51,114,140}.

7.6 Methodological considerations regarding the semisynthesis of C μ 4tp

The study of structure-function relationships in proteins often requires the controlled modification of these proteins. This is especially true for posttranslational modifications (PTM), such as glycosylations, that are highly heterogenic in their naturally occurring forms. These PTMs massively diversify the human proteome beyond what is encoded in the human genome and can have dramatic effects on the behavior of proteins both *in vivo* and *in vitro*¹⁵⁸.

7.6.1 Overview of protein modification strategies

Site-directed mutagenesis is a powerful tool for the modification of proteins and the investigation of structure-function relationships. However, its scope is limited to naturally occurring amino acids, preventing the introduction of chemical functionality not genetically

encoded in nature¹⁶⁶. Extending the genetic code *via* synthetic biology facilitates the incorporation of non-canonical amino acids but must overcome several challenges related to the cellular synthesis machinery¹⁶⁷⁻¹⁶⁹.

Direct chemical modification of cysteine or other reactive amino acid sidechains in proteins has widely been used to add additional functionality. Still, achieving selectivity when using direct chemical modification methods remains challenging¹⁷⁰.

The chemical synthesis of peptides and using naturally-occurring and non-naturally occurring amino acids provides excellent flexibility in modifying side chains and altering the chemistry and chirality of the peptide backbone¹⁷¹. Controlled production of glycosylated proteins requires the use of glycosylated building blocks during synthesis, chemo enzymatic synthesis, modification of the cellular synthesis apparatus in yeast, or enzymatic remodeling of recombinantly expressed glycoproteins¹⁷²⁻¹⁷⁵ and is discussed in detail later in this chapter (sections 7.6.6 and 7.6.7).

7.6.2 Solid phase peptide synthesis – history and limitations

The chemical synthesis of peptides was first performed at the beginning of the 20th century by Emil Fisher¹⁷⁶. However, the scope of this method was very limited until the further development of protection group strategies. In 1932, Bergmann and Zervas introduced the first urethane-protecting group for the α -amino group, allowing for the almost racemization-free synthesis of short peptides *via* solution-phase synthesis⁵⁰. The development of solid-phase peptide synthesis (SPPS) by Robert Bruce Merrifield in 1963 enabled dramatically easier

and faster synthesis of longer peptides¹⁷⁷. Further, iterative improvements to this methodology resulted in the development of automated SPPS¹⁷⁸.

However, the total length of peptides produced *via* SPPS is limited by the reaction yields of each amino acid coupling, the tendency of growing peptide chains to aggregate, and the challenges in purifying the final product from side products accumulated during synthesis¹⁷⁹. Modern methods such as SPPS at elevated temperatures and flow synthesis allow the production of impressively long peptides of over 100 amino acids but are not yet cost-effective at scale¹⁸⁰.

7.6.3 Native chemical ligation

A practical alternative to synthesizing longer and, thus, more challenging peptide sequences is the assembly of large peptides and small proteins from smaller synthetic peptides. Several methods developed for this purpose introduce non-native peptide backbones, which limit the application of the produced proteins. This was overcome by introducing native chemical ligation (NCL) in 1994 by Phil Dawson *et al.*¹⁸¹. Today, NCL is a frequently used method for joining peptide segments. It requires a C-terminal thioester at the N-terminal segment and an N-terminal cysteine on the C-terminal segment. The reaction itself is characterized by a transthioesterification step, in which the thiol moiety of the cysteine attacks the carbonyl group of the thioester, followed by an S→N-acyl shift, resulting in a native amide bond (Figure 13). Typically, NCL is carried out at neutral pH using thiols as ligation mediators.

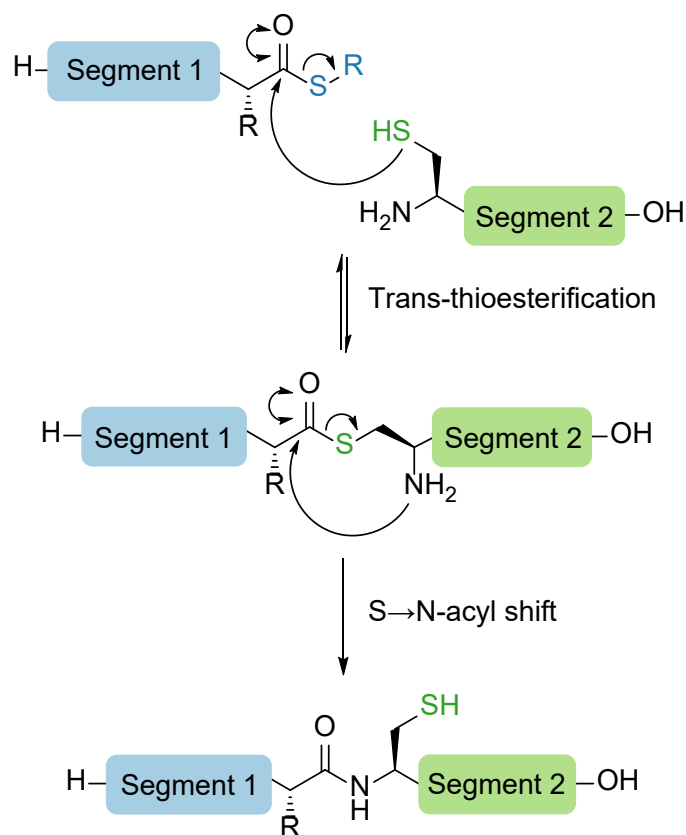


Figure 13: Native chemical ligation (NCL), schematic overview.

Despite the above-described advances in SPPS and impressive achievements in the total synthesis of proteins from SPPS-derived segments *via* NCL, larger proteins requiring multiple consecutive ligation steps are still largely inaccessible *via* this strategy¹⁸²⁻¹⁸⁴. A solution to this problem is protein semisynthesis.

7.6.4 Protein semisynthesis

Protein semisynthesis is a group of strategies and methods that combine the controlled introduction of non-canonical amino acids during SPPS with the advantages in size and scale of the ribosomal protein synthesis machinery^{166,182,185,186}. When NCL is used to ligate a synthetic peptide to an expressed protein, instead of ligating two synthetic segments, it is referred to as

expressed protein ligation (EPL)¹⁸⁷. This requires the production of proteins carrying C-terminal thioesters. This is solved by creating and expressing intein fusion constructs. Inteins are sequences that catalyze their own excision from a peptide sequence. They often serve a posttranslational regulatory function in responding to environmental conditions. The sequences flanking the intein at its N- and C-terminus are referred to as N- and C-extein. During cleavage, a nucleophilic amino acid (commonly a cysteine or serine) at the N-terminal end of the intein attacks the amide bond between itself and the last amino acid of the N-extein, forming a (thio)ester. This thioester is then attacked by a nucleophilic amino acid located at the N-terminus of the C-extein, resulting in the excision of the intein *via* the formation of another (thio)ester. Finally, an S $\rightarrow$ N-acyl shift forms a native amide bond between the N- and C-exteins.

By mutating the intein C-terminus and the nucleophile at the N-terminus of the C-extein to less reactive amino acids, this cleavage can be arrested after the initial thioester formation. The addition of thiols, such as Mesna in an aqueous environment, can then be used to initiate a N-terminal cleavage, leaving behind the protein of interest (as N-extein) carrying a C-terminal thioester, which can be used as starting material for expressed protein ligation (Figure 14)^{182,187,188}.

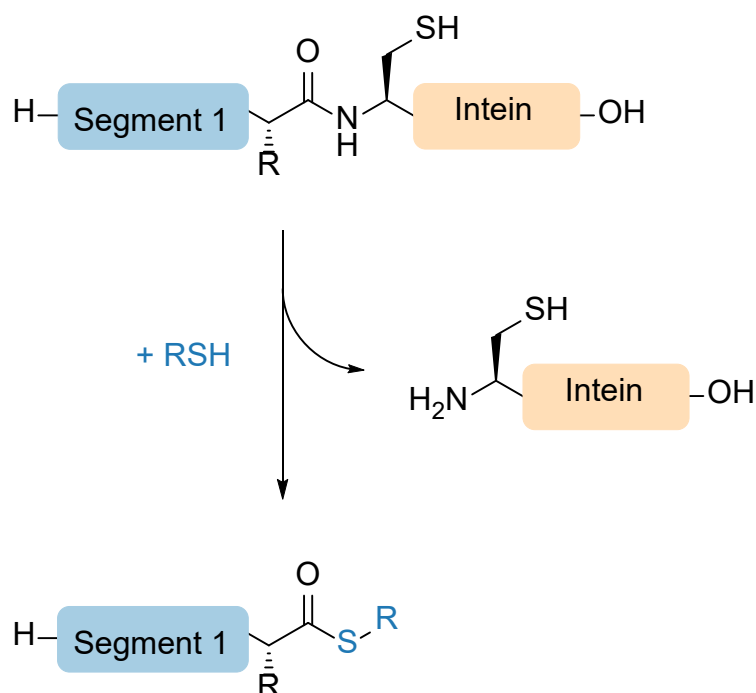


Figure 14: Intein cleavage for expressed protein ligation.

It is also possible to perform EPL with a synthetic N-terminal segment and a recombinantly expressed C-terminal segment. This requires the use of thioester precursors, such as 3,4-diaminobenzoic acid (Dbz)¹⁸⁹ or hydrazide¹⁹⁰ during SPPS¹⁹¹.

7.6.5 Photo-affinity labelling

Photoaffinity labelling (PAL), sometimes referred to as photocrosslinking, is a powerful tool for investigating protein-protein interactions¹⁹². It relies on the ability of photocrosslinking functional groups to form highly reactive species upon photoirradiation. These highly reactive species quickly form covalent bonds with adjacent molecules, thus allowing for the microenvironment mapping around the photocrosslinker in a spacio-temporal manner¹⁹².

PAL was first developed by the research group of Westheimer in 1962¹⁹³. Since then, various photocrosslinkers have been developed based on aryl azide, benzophenone¹⁹⁴, and diazarine¹⁹². We chose p-benzoyl-L-phenylalanine (Bpa) (Figure 15 A) for this work, as it is relatively resistant to ambient light. It can also be incorporated into proteins using codon suppression, as demonstrated by Chin *et al.*¹⁹⁵⁻¹⁹⁷, and is commercially available as a building block for Fmoc SPPS (Figure 15 B).

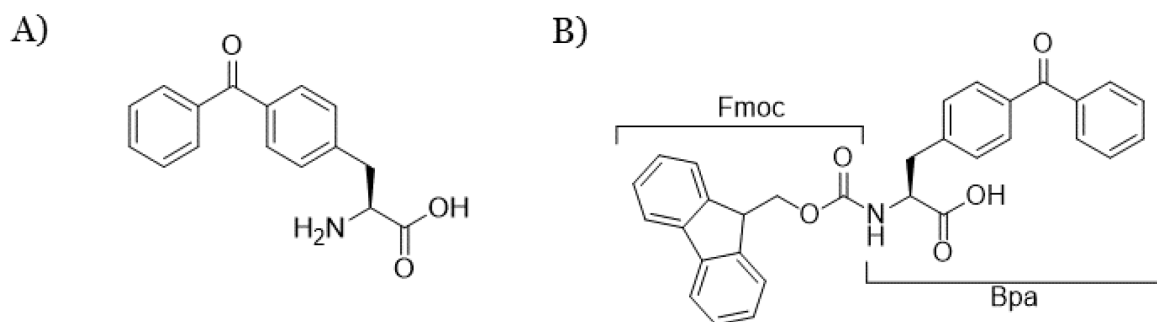


Figure 15: Bpa (p-Benzoyl phenylalanine) (A), and Fmoc-Bpa building block for SPPS (B).

Upon irradiation with UV light at approximately 360 nm, Bpa forms covalent bonds with nearby molecules, such as other proteins, through a radical mechanism depicted in Figure 16.

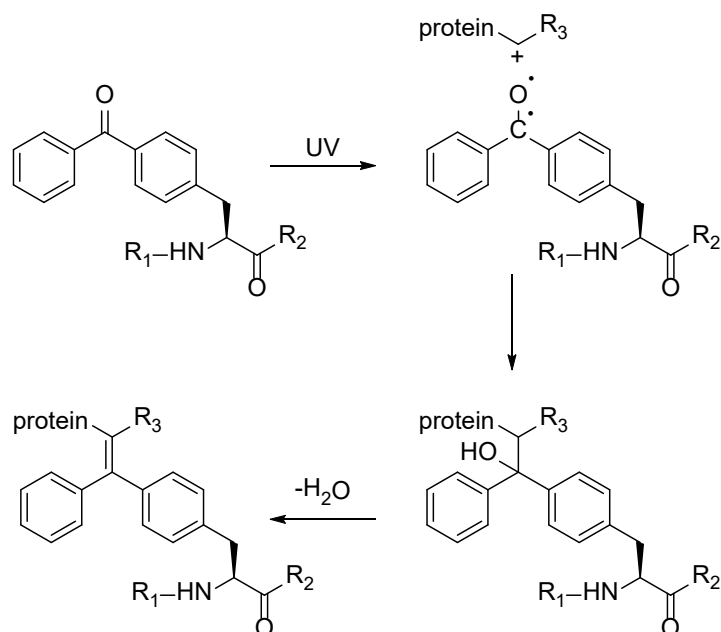


Figure 16: *p*-Benzoylphenylalanine photocrosslinking reaction¹⁹⁸.

7.6.6 Glycosylation in protein semisynthesis

The production of homogenously glycosylated proteins is a significant challenge in academia and industry^{157,174}. Recombinantly expressed glycoproteins are often highly heterogenous in their glycosylation pattern and thus only of limited use when investigating the structure-activity relationship of specific glycans. While chemoselective modification of peptides and proteins with glycans is possible when targeting rarely occurring amino acids¹⁹⁹, these strategies often result in heterogenous products and often have to rely on the microenvironment around specific residues to achieve a higher level of selectivity. Elaborate protection strategies, such as the insertion of non-canonical amino acids or entire motifs and tags for subsequent biorthogonal reactions, have been developed to minimize these side reactions. However, these are often time and resource-intensive^{185,200}. Additionally, when used

for coupling larger oligosaccharides, this strategy suffers from steric hindrance between the glycan and the peptide side chains, resulting in low efficiency^{157,201}.

Semisynthetic strategies are, therefore, often used for the controlled introduction of glycans into proteins. During SPPS, glycosylated amino acids can be used as a building block, some of which, such as Fmoc-L-Asn((Ac)3- β -D-GlcNAc)-OH, are commercially available. These glycosylated segments can be joined *via* conventional NCL or EPL^{157,183,202}.

Alternatively, non-natural amino acids carrying functionalized sidechains can be introduced *via* SPPS or genetic code expansion²⁰³. These then allow for the site-directed introduction of additional modifications, including glycans. A prime example of this are amino acids carrying azides to facilitate click reactions with carbohydrates containing alkynes²⁰⁴. A downside of this otherwise elegant method is the formation of a non-native linkage between the protein and the carbohydrate structure.

A group of strategies to produce homogenously glycosylated proteins that do not require complex protecting group methods and produce native linkage are so-called chemoenzymatic synthesis strategies^{166,174,205-207}. These rely on enzymes such as glycosyltransferases for the transfer of glycans to the target protein²⁰⁸, and endo- and exoglycosidases for the remodeling of existing glycan structures²⁰⁹⁻²¹¹. The latter can also be used to trim heterogenous glycosylation patterns on a protein down to a homogenous core. Enzymatic transfer of predefined glycan structures to a target protein has become a well-established strategy for generating glycosylated proteins *via in vitro* glycoengineering²⁰⁶. These structures can be produced either synthetically or chemo enzymatically^{207,208,212-214}, or they can be isolated from nature.

7.6.7 Sialylglycopeptide as a glycan donor

Sialylglycopeptide (SGP) is a hexapeptide carrying a complex bi-antennary N-glycan. Because of its homogenous glycosylation SGP has seen extensive use as a glycan donor in protein semisynthesis²⁰⁶. SGP can be found in hen's egg yolk²¹⁵. Since 1997, various protocols for the isolation of SGP have been developed^{206,215-218}.

Endo- β -*N*-acetylglucosaminidases catalyze the hydrolysis of *N*-linked oligosaccharides at the β -1,4-glycosidic bond between two GlcNAcs and are sometimes able to efficiently transfer the cleaved glycan structure to an Asn-GlcNAc located in another peptide or protein. A mutant of one of these enzymes from *coprinopsis cinerea*, endo-CC1 N180H, displays high transglycosylation activity of human sialated and not sialated complex human oligosaccharides with significantly reduced hydrolase activity²¹⁹. This makes Endo-CC N180H an excellent glycosyltransferase for use in protein semisynthesis, as the enzyme is capable of transferring glycans from SGP to a broad range of acceptors, including monosaccharides and antibodies containing a 1,3-diol structure (Figure 17)²²⁰.

Introduction

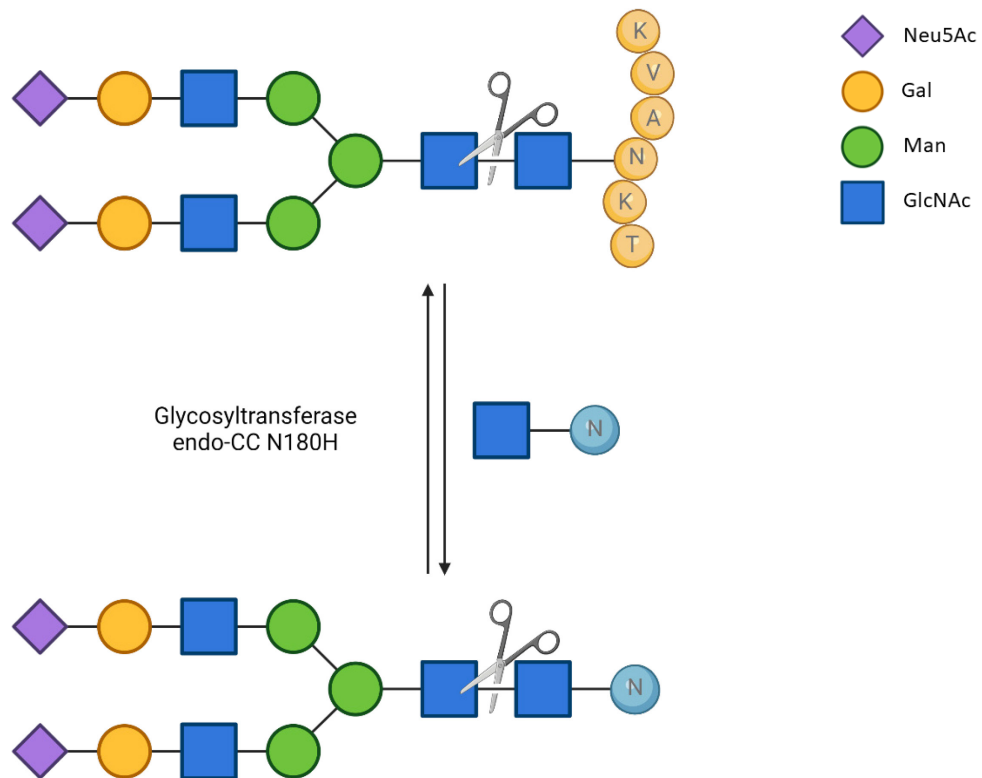


Figure 17: direct transfer of the glycan from sialylglycopeptide to AsnGlcNAc by endo-CC N180H.

7.7 Aim

This thesis aimed to generate new insights into the underlying molecular mechanisms of IgM assembly through design, synthesis, and characterization of novel semisynthetic C μ 4tp variants. To accomplish this goal, we set out to engineer C μ 4tp variants incorporating Bpa substitutions in various positions in the tailpiece. These would then be subjected to refolding and oligomerization experiments by our collaboration partners at the Technical University of Munich, aiming to identify yet unknown interaction sites before, during and after IgM assembly.

Furthermore, we aimed to design and produce semisynthetic C μ 4tp variants containing modifications at the Asn563/S565 N-linked glycosylation consensus motif, including a variant incorporating an Asn-GlcNAc building block. This GlcNAc would then serve as acceptor for the enzymatic transfer of larger glycan structures derived from SGP onto C μ 4tp. Through this we wanted to generate new insights into the impact of C μ 4tp glycosylation on IgM folding and oligomerization and provide a starting point for future semisynthetic synthesis strategies for IgM glycosylation.

Beyond the scope of this thesis, we hope to contribute to opening IgM oligomers for rational-driven design approaches regarding stability, binding properties, oligomer composition, and posttranslational modifications.

8 Materials and Methods

8.1 Production of N-terminal C μ 4tp segments via protein expression

8.1.1 Protein expression

C μ 4-Intein fusion constructs were expressed from a modified pTXB1 Amp^R vector containing a His7 tag between the MxeInteine and the CBD provided by Dr. Alexander Kravchuk. The C μ 4 DNA sequence information for both constructs was provided by Benedikt Weber of the research group of Johannes Buchner at the Technical University of Munich and ordered from EurofinsGenomics. These DNA sequences were incorporated into the pTXB1 vector using a NdeII and a SpeI restriction site. The exact sequence of the constructs is shown in the appendix (section 12.1).

Expressions were performed as follows: A 5 L Erlenmeyer flask containing 2 liters of LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 100 μ g/L ampicillin) was inoculated with an over-night culture of BL21(DE3)Gold carrying pTXB1- C μ 4-1-91-Mxe-H7-CBD or pTXB1- C μ 4-1-108-Mxe-H7-CBD. The flask was incubated at 37°C with agitation. The expression was induced *by* adding IPTG to a final concentration of 1 μ M at an OD₆₀₀ of approximately 0.6. Samples were taken at multiple time points and analyzed *via* SDS-PAGE (see section 8.6.1).

Bacterial cells were harvested *via* centrifugation at 12,000 xg for 10 min at 4°C. The cells in the solubilized pellet were lysed *via* physical disruption using either one of two protocols: Bacterial cell lysis *via* high-pressure disruption (section 8.1.2) or Bacterial cell lysis *via* sonication (section 8.1.3).

8.1.2 Bacterial cell lysis *via* high-pressure disruption

Bacterial cells were disrupted using a Constant Systems CF1 Cell Disruptor system. Before lysis, bacterial cells were homogenized *via* stirring in approximately 50 mL PBS per L of expression media at 4°C. Disruption was performed at 4°C at a pressure of 1.9 kbar. All samples were run through the machine thrice to ensure complete lysis.

8.1.3 Bacterial cell lysis *via* sonication

Bacterial cells were lysed using a Fisherbrand Model 505 Sonic Dismembrator system. Before lysis, bacterial cells were homogenized *via* stirring in approximately 50 mL PBS per L of expression media at 4°C. All samples were sonicated for a total duration of 10 min at 4°C with a pulse length of 15 seconds at 60% amplitude. Pulses were separated by 45 second cooling intervals to prevent the samples from overheating.

8.1.4 Workup of bacterial cells after lysis

Bacterial cells that had undergone lysis, as described in section 8.1.2 and section 8.1.3, were centrifuged at 58,000 xg and 4°C for 45 min. The resulting pellets were initially washed twice with 1% Triton X-100 in PBS and once with PBS. The wash with Triton X-100 was skipped during later expressions, as this resulted in significant contamination of the samples with PEG. The pellets were then homogenized in 10 mL solubilization buffer (6M Gdn.HCl, 100 mM Tris.HCl, 30 mM 2-mercaptoethanol, 20 mM imidazole, 10 mM dithiothreitol, pH 7.5) per L of expression media at room temperature overnight. The suspension was centrifuged at 58,000 xg for 30 min. The supernatant was subjected to either affinity column purification (see

section 8.1.5) followed by dialysis of the expressed protein solutions or directly dialyzed (see section 8.1.6).

8.1.5 Affinity column purification

Samples were purified via Cytiva HisTrap HP NiNTA affinity columns on ÄKTAprime plus systems. Samples were loaded after 1:1 dilution in loading buffer (6 M Gdn.HCl, 100 mM Tris, 20 mM imidazole, pH 7.5). The column was then washed with 10 column volumes of loading buffer before the proteins of interest were eluted using elution buffer (6 M Gdn.HCl, 100 mM Tris, 500 mM imidazole, pH 7.5). Samples were gathered in 5 mL fractions at a 5 mL/min flow rate and analyzed via SDS-PAGE (see section 8.6.1).

8.1.6 Dialysis of expressed protein solutions

Protein solutions from section 8.1.4 and section 8.1.5 were subjected to dialysis. Up to 50 mL of protein solution were dialyzed twice against 1 L of the following buffer: 4 M urea, 2 mM EDTA, 100 mM Tris, pH 7.1 at 4°C for at least 3 h. Precipitation, when occurring, was removed via centrifugation at 4,000 xg for 10 min. All samples then underwent Intein cleavage using Mesna (see section 8.1.7).

8.1.7 Intein cleavage using Mesna

All intein cleavages were performed by adding 1 M aqueous MesNa solution, pH 7.4, to the samples at a 1:1 ratio, resulting in a final concentration of 0.5 M MesNa, 2 M urea, 50 mM Tris.HCl, and 2 mM EDTA. The pH of the solution was then checked and if necessary adjusted to 7.4 *via* the addition of 1M sodium hydroxide solution. The reaction was incubated on a

rotator at room temperature for 24 h. The final product was purified *via* preparative HPLC (see section 8.1.8).

8.1.8 Thioester purification *via* preparative HPLC

Protein thioesters were purified *via* peak collection on preparative HPLC systems (Varian ProStar HPLC System, Waters HPLC System (Waters 2545 Quarternary Gradient Module, Waters 2489 UV-VIS Detector & Waters Fraction Collector III), Shimadzu Prominence (LC-20AP Solvent Delivery Unit, SPD-10A UV-VIS Detector, FRC-10A Fraction Collector)) using a linear gradient of 5-65% eluent A (H₂O, 0.1% TFA) to eluent B (ACN, 0.08% TFA) in 30 min at 60°C and a flow rate of 10 ml/min on a Kromasil C4 columns (300- 10-C4, 250 x 21.2 mm, 10 µm particle size). The collected fractions were analyzed *via* MS direct injection on a Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Samples of sufficient purity containing the expected masses were then pooled and lyophilized. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹ or MassLynx MS Software.

8.1.9 Quality control of synthetic N-terminal C_μ4tp segments

The purity of N-terminal C_μ4tp segments was analyzed *via* analytical HPLC (Thermo Fisher Scientific Dionex and Vanquish systems) using a linear gradient of 5-65% eluent A (H₂O, 0.1% TFA) to eluent B (ACN, 0.08% TFA) in 30 min at a flow rate of 1 ml/min on Kromasil C4 columns (300-5-C4, 150 × 4.6 mm, 5 µm particle size) and MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in

positive ion mode. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics).

8.2 Synthesis of C-terminal C μ 4tp segments

8.2.1 Solid Phase Peptide Synthesis

All peptides were produced using either fluorenylmethyloxycarbonyl (Fmoc) or tert-butyloxycarbonyl (Boc)-based SPPS. Different resins were used to create the synthetic C μ 4tp segments. Variants carrying C-terminal Bpa were synthesized using 2-Chlorotrityl chloride and Wang resins. Variants containing native C-terminal tyrosine were synthesized on preloaded Fmoc-Tyr-Tentagel resin or Boc-Tyr(2-bromo-Z)-PAM 200-40 mesh.

8.2.1.1 Loading of 2-Chlorotrityl chloride resin

The loading of 2-Chlorotrityl chloride resin was performed according to manufacturer guidelines²²². The resin was swollen in dry DCM (10 ml/g resin) for 15 min. Excess DCM was removed from the resin, and a solution of 1.2 eq amino acid and 2.5 eq DIPEA in DCM was added. The mixture was incubated for at least 60 minutes at room temperature with agitation. To cap unreacted remaining trityl groups, 1 ml of a methanol/DIPEA solution (5:1, V:V) were added per gram of resin, and the mixture was agitated for 15 min. The solution was discarded, the resin was washed at least five times with DCM and three times with methanol, and then dried under reduced pressure.

The resin loading was estimated gravimetrically using Equation 1 (as per manufacturers guidelines²²²) or photometrically as follows: Approximately 10 mg of dry resin were deprotected with 1 mL 20% piperidine for 15 minutes. 50 μ L of supernatant was then diluted

with 950 μ L of 20% piperidine, and the absorbance at 301 nm measured using a spectrophotometer (Bormac Spectrophotometer UV-30). The resin loading was calculated with Equation 2.

$$L = \frac{(m2 - m1) \times 1000}{(MW_{Xaa} - 36.46) \times m2}$$

Equation 1: Gravimetric determination of resin loading

$$L \left(\frac{mmol}{g} \right) = \frac{A^{301} \times 20,000}{7800 \times m (mg)}$$

Equation 2: Photometric determination of resin loading

8.2.1.2 Coupling of amino acids via Boc-SPPS

The coupling of amino acids *via* Boc-SPPS was performed as follows: The Boc protecting group was removed by first rinsing the resin twice with TFA. The resin was then incubated twice with five resin volumes of TFA for 1 minute. TFA was removed, and the resin washed ten times with five resin volumes of DMF.

Coupling of the following amino acid was done using 10 equivalents excess of protected amino acids, 9.8 equivalents of HBTU (0.5 M in DMF) and 9.8 equivalents of DIPEA. The reaction was incubated for 15 minutes at room temperature. The resin was then washed ten times with five resin volumes of DMF.

Upon completion of the peptide sequence, the N-terminal Boc protecting group was removed by rinsing the resin twice with TFA. The resin was then incubated twice with five resin volumes

of TFA for 1 minute. TFA was removed, and the resin washed ten times with five resin volumes of DMF and five times with five resin volumes of DCM. Finally, the resin was dried under reduced pressure²²³.

8.2.1.3 Coupling of amino acids *via* Fmoc-SPPS

The coupling of amino acids to the growing peptide *via* Fmoc-SPPS was performed as follows: The Fmoc protecting group was removed by adding five resin volumes of 20% piperidine in DMF. The reaction was incubated at room temperature for 7 minutes. The solution was removed, and the resin was incubated again in five resin volumes of 20% piperidine in DMF for another 3 minutes. The solution was again removed, and the resin washed ten times with five resin volumes of DMF.

Coupling of the next Fmoc-amino acid was performed using five eq. excess of protected amino acids, 4.8 equivalents of HBTU (0.5 M in DMF) and 4.8 equivalents of DIPEA for a duration of 30 min at room temperature. The resin was then washed ten times with five resin volumes DMF.

Upon completion of the peptide sequence, the N-terminal Fmoc protecting group was removed by adding five resin volumes of 20% piperidine in DMF. The reaction was incubated at room temperature for 7 minutes. The solution was removed, and the resin incubated in five resin volumes of 20% piperidine in DMF for another 3 minutes. The resin was washed ten times with five resin volumes of DMF and five times with five resin volumes DCM. Finally, the resin was dried under reduced pressure.

8.2.1.4 Automated Fmoc-SPPS

Automated synthesis of peptides was performed using Liberty Blue and Liberty Prime synthesizers (CEM) according to the manufacturer's guidelines. Fmoc-deprotection was performed using 20% piperidine in DMF at 90°C for 1 minute. Amino acids were activated with 5 equivalents of oxyma and 10 equivalents of DIC and coupled at 90°C for 2 minutes, except for His(trt) coupling on Liberty Blue at 50°C for 10 minutes. The following pseudo-proline dipeptides were used for automated synthesis: Thr-Gly, Arg-Thr, Val-Ser, and Asp Thr. After the final deprotection step, the resin was removed from the synthesizer and washed manually, as described above (section 8.2.1.3).

8.2.1.5 Peptide cleavage from resin

Peptides produced *via* Fmoc SPPS were cleaved off the resin *via* treatment with TFA : H₂O : TIPS : EDT (92.5 : 2.5 : 2.5 : 2.5) or TFA : H₂O : TIPS : DMS (92.5 : 2.5 : 2.5 : 2.5) for at least 3 h²²⁴. Crude peptides were precipitated *via* addition of five volumes of ice-cold Et₂O per volume of cleavage cocktail and centrifuged at 3,000 xg for 10 minutes. The precipitate was washed with ice-cold Et₂O. The remaining Et₂O was evaporated at room temperature before the pellets were dissolved in 50% ACN in H₂O containing 0.1% TFA, and lyophilized.

Peptides produced *via* Boc SPPS were cleaved off the resin following an HF cleavage protocol by the resin manufacturer²²⁵. At this point we would like to emphasize that HF is an extremely toxic, corrosive, and volatile substance and must only be handled with the utmost care following rigorous serious safety precautions. The cleavage was performed *via* treatment of the resin with 10 mL of hydrofluoric acid and 1 mL of p-Cresol per g of resin at 0°C for 1 h.

Afterward, the remaining HF was evaporated under a stream nitrogen. The resin was extracted from the matrix with TFA and precipitated with 10 volumes of ice-cold Et₂O. The remaining Et₂O was evaporated at room temperature before the pellets were dissolved in 50% ACN in H₂O containing 0.1% TFA and lyophilized.

8.2.2 Removal of acetyl protecting groups from GlcNAc modified peptides

Peptides containing an AsnGlcNAc building block had to undergo an additional deprotection protocol. This was necessary, as the acetyl groups protecting the hydroxy groups of the Asn-GlcNAc building block are not removed during TFA cleavage. This Ac deprotection was achieved by dissolving the peptide in MeOH and adding 0.08 M NaOCH₃. The reaction was incubated at room temperature under agitation and quenched by adding 1/6 reaction volume acetic acid. The reaction was then diluted with mixture of 80% solvent A (0.1% TFA in H₂O) and 20% solvent B (0.08% TFA in ACN) and purified *via* HPLC (section 8.2.3).

This protocol was later switched to a simplified version: We incubated the peptide at pH 11 for 2 h at room temperature. The reaction was then diluted with mixture of 80% solvent A (0.1% TFA in H₂O) and 20% solvent B (0.08% TFA in ACN) and purified *via* HPLC (section 8.2.3).

8.2.3 Peptide purification *via* preparative HPLC

All peptides were purified *via* peak collection on preparative HPLC systems (Varian ProStar HPLC System, Waters HPLC System (Waters 2545 Quaternary Gradient Module, Waters 2489 UV-VIS Detector & Waters Fraction Collector III), Shimadzu Prominence (LC-20AP Solvent Delivery Unit, SPD-10A UV-VIS Detector, FRC-10A Fraction Collector)) using, unless stated otherwise, a linear gradient of 5-65% eluent A (H₂O, 0.1% TFA) to eluent B (ACN, 0.08% TFA) in

30 min at a flow rate of 10 ml/min on a Kromasil C4 column (300- 10-C4, 250 x 21.2 mm, 10 μ m particle size). The collected fractions were analyzed *via* MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Samples of sufficient purity containing the expected masses were then pooled and lyophilized. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹ or MassLynx MS software.

8.2.4 Quality control of synthetic C-terminal C μ 4tp segments

The purity of C-terminal C μ 4tp segments was analyzed *via* analytical HPLC (Thermo Fisher Scientific Dionex and Vanquish systems) using a linear gradient of 5-65% eluent A (H₂O, 0.1% TFA) to eluent B (0.08% TFA in ACN) in 30 min at a flow rate of 1 ml/min on Kromasil C4 column (300-5-C4, 150 $\times$ 4.6 mm, 5 μ m particle size) and MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹ or MassLynx MS software.

8.3 Ligation of C μ 4tp segments

8.3.1 Native chemical ligation

Ligations of synthetic C μ 4tp segments (C μ 4tp 92-132, see section 8.2) to expressed C μ 4tp thioester segments (C μ 4tp 1-91, see section 8.1) were performed as described previously

^{181,182,226}.

Gdn.HCl (6 M), MPAA (200 mM), and a small amount of TCEP (0.1 mM) were dissolved in 300 mM Sodium Phosphate buffer, pH 7.2. The solution was degassed using purging with argon for approximately 10 min. The peptide was dissolved in the degassed solution before adding the thioester segment. After addition of the thioester, the pH of the solution was readjusted to 7.2 with 1M NaOH if necessary. Both peptide and thioester were added at a concentration of 10 mg/ml. This ratio of 2.2 mM peptide to 1 mM thioester was chosen to minimize the hydrolysis of the thioester.

The reaction was incubated at room temperature for 24 h under agitation on a rotator and regularly monitored *via* LCMS (section 8.3.2). The reaction samples and the reaction itself were quenched by diluting the reaction in 6 M Gdn.HCl containing 20% 2-mercaptoethanol at a ratio of 1:5.

8.3.2 Reaction monitoring

Progress of the NCL reaction was monitored on a Waters Auto Purification HPLC/MS system (equipped with 3100 Mass Detector, 2545 Binary Gradient Module, 2767 Sample Manager and 2489 UV/Visible Detector) using electrospray ionization in positive ion mode. The reaction samples were quenched by diluting them in 4.8M Gdn.HCl containing 20% 2-mercaptoethanol at a ratio of 1:5. The diluted samples were analyzed using a linear gradient of 5-65% eluent A (0.1% TFA in H₂O) to eluent B (0.08% formic acid in ACN) in 12 min at a flow rate of 1 ml/min on a Kromasil C4 column (300-5-C4, 150 × 4.6 mm, 5 μm particle size). Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹ or MassLynx MS software.

8.3.3 Purification of semisynthetic C_μ4tp variants

C_μ4tp variants were purified immediately after ligation. To remove 2-mercaptoethanol adducts, a small amount of TCEP (0.1 mM) was added to the reaction and incubated for 20 min.

The C_μ4tp variants were purified *via* peak collection on HPLC systems (Varian ProStar HPLC System, Waters Purification HPLC system (Waters 2545 Quaternary Gradient Module, Waters 2489 UV-VIS Detector & Waters Fraction Collector III), Shimadzu Prominence (LC-20AP Solvent Delivery Unit, SPD-10A UV-VIS Detector, FRC-10A Fraction Collector)) using a linear gradient of 25-55% eluent A (0.1% TFA in H₂O) to eluent B (0.08% TFA in ACN) in 60 min at a flow rate of 10 ml/min on Kromasil C4 columns (300-10-C4, 250 x 10 mm, 10 μm particle size) C4 columns. Maximum loading did not exceed 10 mg. Ligations on a larger scale were purified over multiple runs. The collected fractions were analyzed *via* MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹ or MassLynx MS software. Samples of sufficient purity containing the expected masses were then pooled and lyophilized.

8.3.4 Quality control of C_μ4tp variants

The purity of C_μ4tp variants was analyzed *via* analytical HPLC (Thermo Fisher Scientific Dionex and Vanquish systems) using a linear gradient of 5-65% eluent A (0.1% TFA in H₂O) to eluent B (0.08% TFA in ACN) in 30 min at a flow rate of 1 ml/min on Kromasil C4 columns (300-

5-C4, 150 × 4.6 mm, 5 µm particle size) and MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹ or MassLynx MS software.

8.4 Isolation of sialylglycopeptide from egg yolk

Production of SGP as glycan donor was performed and optimized following two different protocols inspired by Sun *et al.*²²⁷ and Alagesan *et al.*²¹⁶.

8.4.1 Initial SGP isolation from whole eggs

Our initial protocol used for SGP isolation was established by Alanca Schmid²²⁸ based on earlier work by Binyang Sun *et al.*²²⁷. It starts with whole eggs and contains multiple column-based purification steps, limiting its throughput. This protocol then proceeds to trim the glycan of SGP. The purification and activation of this isolated glycan require additional time-consuming steps, further limiting the throughput. Hence, changes were made, resulting in the updated SGP isolation protocol described in section 8.4.2.

8.4.1.1 SGP extraction via acetone

Crude SGP extracts from whole eggs were generated as follows: Thirty egg yolks were manually separated from their egg whites, stirred with 100 mL ddH₂O, and lyophilized, creating egg yolk powder. This powder was then generously washed once with 1300 mL and twice with 500 mL diethyl ether over a paper filter. The egg yolk was stirred twice in 300 mL 70% acetone for an hour before removal of the solvent through filtration over a paper filter. The egg yolk was then stirred thrice in 400 mL 40% acetone for 15 min, and the solvent was

collected *via* filtration through a bottle top filter (Steritop MILIPORE Express PLUS, PES membrane 0.2 μm). The volume of the filtrate was reduced on a rotavapor and lyophilized, yielding approximately 2.5 g of crude extract.

8.4.1.2 SGP purification *via* carbon/celite column

Purification of crude SGP after acetone extraction was performed in crude SGP batches of 500 mg each as follows: 10 g activated charcoal and 5 g celite were mixed with ddH₂O and transferred into a 20 mL plastic syringe containing a filter pad. 500 mg of crude SGP were loaded onto the column. The open syringe was placed on top of a vacuum chamber to increase flow rates under suction. The column was washed with increasing concentrations of ACN in ddH₂O containing 0.1% TFA ranging from 0% to 40%.

All fractions were lyophilized and analyzed with analytical HPLC. In brief, ~0.5 mg material was analyzed with a linear gradient of 2-5% eluent A (0.1% TFA in H₂O) to eluent B (0.08% TFA in ACN) in 30 min at a flow rate of 1 ml/min on a Kromasil C18 column (300-5-C18, 150 $\times$ 4.6 mm, 5 μm particle size). Fractions with no peaks after the 15 min mark were pooled and lyophilized.

8.4.1.3 SGP purification *via* size exclusion

The SGP was then further purified *via* size exclusion chromatography (SEC). SGP was loaded onto a PD10 column (GE Healthcare, Germany) using 0.1M acetic acid in H₂O as an eluent. Samples were collected in 0.5 mL fractions and analyzed via analytical HPLC, as described above (section 8.4.1.2). Fractions containing SGP were pooled and lyophilized.

8.4.1.4 SGP trimming and cleavage

The glycan from SGP was enzymatically removed and trimmed down to a tetrasaccharide *via* the following protocol: Three different enzymes were used in sequence: β -Galactosidase and β -N-Acetylglucosaminidase for the trimming of SGP and Endoglycosidase for the cleavages of the glycan from SGP. This yields a GlcNAcMan₃ tetrasaccharide and a hexapeptide carrying a GlcNAc on Asn4.

SGP was dissolved in trimming buffer I (0.1M sodium acetate, 0.5 mg/mL BSA, pH 4.5), and 50 μ l β -Galactosidase (Megazyme) were added. The reaction was incubated at 40°C for 1h under agitation. The product was purified *via* SEC using a PD-10 column as described above and lyophilized.

The SGP was then dissolved in 1.5 ml trimming buffer II (50 mM citric acid, 100 mM sodium chloride, 0.1 mg/ml BSA, pH 5), and 10 μ l β -N-Acetylglucosaminidase from *Canavalia ensiformis* (Sigma Aldrich) were added to the solution. The reaction was incubated overnight at 25°C. Then 400 μ l Endoglycosidase solution (endo A from *arthobacter protophormiae*, 0.0025 μ mol/ μ l) was added, and the reaction was incubated for 2 h at 40°C. The volume was then reduced *via* freeze drying, and the glycan was purified *via* Hydrophilic Interaction Chromatography (HILIC) (section 8.4.1.5).

8.4.1.5 Glycan purification *via* HILIC

Glycans derived from SGP were purified *via* peak collection on an analytical HPLC system (Thermo Fisher Scientific Dionex) using a linear gradient of 2-5% eluent A (95% ACN, 5% H₂O, 10 mM NH₄OAC) to eluent B (10 mM NH₄OAC in H₂O) in 10 min at a flow rate of 1 ml/ min on a

HILIC column (ACCUCORE 150-Amide-HILIC, 2.6 μ m 150X4.6mm). A charged aerosol detector (CAD, Dionex Corona ultra RS) was used for detection of glycans. Maximum loading did not exceed 1 mg. Fractions were collected under the same conditions without use of the CAD, as the CAD requires the system to be under nitrogen which interferes with fraction collection. The collected fractions were analyzed via MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Samples of sufficient purity containing the expected masses were then pooled and lyophilized.

8.4.1.6 Glycan activation and transfer

The tetrasaccharide was activated as follows: 1 mg tetrasaccharide and 12 equivalents of Na_3PO_4 were dissolved in 56 μ L H_2O and cooled to 0°C. 4.2 equivalents of 2-Chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride (CDMBI) were added, and the reaction was stirred for 2 h. The reaction was then centrifuged at 12,000 xg for 10 min.

The supernatant was transferred to a new microcentrifuge tube, and a peptide containing AsnGlcNAc was added at a ratio of 1:5 (peptide to tetrasaccharide). The pH was adjusted to 7.4, and 2 μ L endo-cc N180H was added. The final concentrations of the reaction were 10 mM Tris, 150 mM NaCl, 3.6 mM peptide, and 18 mM tetrasaccharide oxazoline. The reaction was incubated with agitation at 30°C for 2 h. Samples were diluted in eluent A (dd H_2O , 0.1% TFA) and analyzed *via* analytical HPLC (Thermo Fisher Scientific Dionex and Vanquish systems) using a linear gradient of 5-65% eluent A (H_2O , 0.1% TFA (VWR, USA)) to eluent B (acetonitrile (Merck, Germany), 0.08% TFA (VWR, USA)) in 30 min at a flow rate of 10 ml/min on a Kromasil

(300-5-C4, 150 × 4.6 mm, 5 μ m particle size) C4 column and MS direct injection on Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Expected masses were calculated in ChemDraw Professional (PerkinElmer Informatics). Mass spectra were deconvoluted using MagTran²²¹.

8.4.2 Updated SGP isolation protocols

A different approach inspired by Alagesan *et al.*²¹⁶ was chosen to increase the yields of our glycan production. The updated SGP isolation protocol starts with egg yolk powder and relies on both liquid and solid phase extraction methods that are not limited by column capacities. This second protocol underwent several variations before we settled on a finalized version. Further below, you will find more information about these variations.

8.4.2.1 SGP extraction via methanol and chloroform

SGP was isolated from egg yolk powder (Backstars, Germany) as follows: Egg yolk powder was mixed with ddH₂O at a 1:3 ratio (w/v) and stirred for at least 2 h or overnight at RT. Then 2 volumes of methanol (egg yolk powder w/v) were added, and the mixture was stirred for at least 1 h at RT. Then the suspension was centrifuged at 3000 xg at 4°C for 5 min. The supernatant was transferred to a new container. Chloroform was added to the supernatant at a ratio of 25:1 (egg yolk w/v), and the mixture was stirred for 10 min. The mixture was then centrifuged at 3000 xg at 4°C for 10 min. The volume of the aqueous phase was reduced using a rotavapor and lyophilized. The crude SGP extracts were purified *via* carbon/celite solid phase extraction.

8.4.2.2 Purification option: size exclusion

1 g of crude SGP extract was dissolved in 100 mM ammonium acetate, pH 7, and purified *via* a 300 mm long Sephadex G50 (fine) column with 100 mM ammonium acetate as eluent. Fractions were collected in 30-sec intervals and analyzed *via* MS direct injection into a Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems using electrospray ionization in positive ion mode. Fractions containing the expected mass were pooled and lyophilized.

8.4.2.3 Purification option carbon/celite solid phase extraction

We eventually chose the purification of SGP *via* carbon/celite solid phase extraction over purification *via* size exclusion. This extraction was performed as follows: Per g of crude SGP extract, 10 g activated charcoal and 5 g celite were mixed with ddH₂O. SGP was added, and the mixture was stirred for 5 min. All liquid was removed *via* filtration with a bottle top filter (Steritop MILIPORE Express PLUS, PES membrane 0.2 µm) under reduced pressure. The solid phase was extracted with incrementally increasing concentrations of ACN in ddH₂O containing 0.1% TFA ranging from 0% to 40%. The liquid phase was collected *via* filtration with a bottle top filter and analyzed *via* MS direct injection on a Waters 3100 or SQ Detector 2 single quadrupole mass spectrometry systems. Fractions containing the expected mass were combined and lyophilized.

8.5 Transglycosylation of Cp4tp with endo-cc N180H

Direct enzymatic transfers of a glycan from SGP to Cp4tp were performed analogously to Manabe *et al.*²²⁰. Cp4tp variants carrying the N119AsnGlcNAc modification were dissolved in

DMSO at a ~ 0.2 mg/ μ L concentration. 50 to 100 equivalents of SGP were dissolved in 20 mM Tris.HCl, pH 7.5, and added to the C μ 4tp variants in DMSO. 1 μ l endo-cc N180H solution was added per 25 μ l reaction mixture. After 24 hours, the reaction was diluted in 6M Gdn.HCl, purified and analyzed using the same methods outlined in section 8.3.

8.6 Analytical and structural methods

8.6.1 SDS-PAGE

The SDS-PAGE protocol for this thesis was derived from works by Ulrich K. Laemmli²²⁹. Gels were cast in multi-gel casting chambers. The composition of the stacking and separation gels for 10 gels is shown in Table 1.

Table 1: Composition of stacking and separation gels for 15% SDS-PAGE

	stacking gel	separating gel
H ₂ O	12.5 ml	12.5 ml
buffer ST	2.5 ml	-
Buffer SE	-	13.5 ml
30% acrylamide	3.5 ml	26 ml
10% SDS	185 μ l	550 μ l
10% APS	185 μ l	550 μ l
TEMED	18 μ l	18 μ l

SDS-PAGE samples were combined with SDS-PAGE loading dye (500 mM Tris, 6% SDS, 35% glycerin, 3.55% 2-mercaptoethanol, 0.05% bromophenol blue) at a 1:1 ratio and heated to

95°C for approximately 5 min. A small amount of sample was then loaded on a 15% SDS-PAGE gel. The gels were run for 35 min at a constant 230 V and stained *via* Coomassie blue staining solution (0.1% Coomassie R250, 10% acetic acid, 45% methanol) for at least 1 h. Afterward, gels were destained in a destaining solution (10% acetic acid, 40% methanol) for 30 min or ddH₂O water for at least three hours. Images of the gels were captured using a Biorad ChemiDoc MP imaging system.

8.7 Folding of C μ 4tp variants

Folding of C μ 4tp variants was performed by Tuan Nguyen at the Technical University of Munich, who kindly provided the following protocol:

The lyophilized C μ 4 domain was solubilized in PBS with 8 M Urea and 5 mM DTT and incubated for 1 h at RT. Concentration of the protein was 0.5 mg/ml. Refolding of the semisynthetic C μ 4 domain was achieved by dialysis. The sample was filled into a 3.5 kDa MWCO dialysis tube and dialysed against 1 L of the C μ 4 refolding buffer for 3 h at 4 °C. The dialysis tube was then transferred in 1 L of the C μ 4 oxidizing buffer and incubated o/n at RT. On the next day, the dialysis step was repeated o/n, after which the POI was dialysed against 1 L of the C μ 4 native buffer and prepared for SEC. The protein sample was concentrated to a final volume of 500 μ L using an Amicon 3.5 MWCO centrifugation filter (Sigma Aldrich) and injected into a Superdex 200 pg 10/300 GL column (Cytiva), which was preequilibrated in PBS. The presence of the POI was followed through absorbance at 280 nm and checked on a non-reducing 4-20 % SDS gel (SERVA).

Table 2: Refolding and oxidation conditions

C μ 4 native buffer, pH 7.4	PBS	1-fold
	DTT	1 mM
	EDTA	1 mM
C μ 4 oxidizing buffer, pH 7.4:	PBS	1-fold
	Urea	1 M
	GSH	0.5 mM
	GSSG	4 mM
	EDTA	1 mM
C μ 4 refolding buffer, pH 7.4:	PBS	1-fold
	Urea	1 M
	DTT	1 mM
	EDTA	1 mM
PBS 1-fold, pH 7.4:	NaCl	137 mM
	KCl	2.7 mM
	Na ₂ HPO ₄	10 mM
	KH ₂ PO ₄	2 mM

We used HPLC to estimate the size of single proteins or protein/protein complexes. As smaller proteins are more likely to interact with the solid phase of the gel filtration columns, they elute at a later point than larger proteins. HPLC measurements were performed on a Shimadzu Prominence HPLC system, coupled to a SPD-20 A 2 channel UV detector. Samples were injected onto a Superdex 200 pg 10/300 increase column. Absorbance was measured at 230 nm and 280 nm. The flowrate was set to 0.75 ml·min⁻¹.

8.8 Reducing and non-reducing PAGE

Analysis of folded C μ 4tp variants via PAGE was performed by Tuan Nguyen at the Technical University of Munich, who kindly provided the following protocol:

For reducing SDS-PAGE, 40 μ L of the sample were mixed with 10 μ L of 5x reducing Lämmli-Buffer²²⁹ and incubated at 95°C for 5 min. 20 μ L of the sample were loaded onto an SDS gel and run at a constant current. The gels were either stained with Fairbanks A, and destained with Fairbanks D or directly stained with ProteinArk Quick Comassie® Stain (SERVA)²³⁰.

For non-reducing SDS-PAGE, 40 μ L of the sample were mixed with 2 μ L of 500 mM NEM and incubated for 15 min. Afterwards 10 μ L of 5x non-reducing Lämmli and incubated at 95 °C for 5 min. Electrophoresis and staining were performed as described above.

8.9 Circular dichroism spectroscopy and Thermal transition

Circular dichroism spectroscopy and thermal transition measurements of folded C μ 4tp variants was performed by Tuan Ngyen at the Technical University of Munich, who kindly provided the following protocol:

CD spectra were measured at a protein concentration of 0.1 mg $\cdot$ mL⁻¹. The ellipticity was then normalized to the amount of peptide bonds within a protein and the molecular weight. This amount is in most cases the amino acid content minus one. The new value is then defined as the mean residual weight ellipticity²³¹.

$$[\theta_{MRW}] = \frac{\theta \cdot \frac{M}{n-1}}{c \cdot d}$$

Equation 3

θ_{MRW}	mean molar ellipticity, $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$
θ	ellipticity, mdeg
M	molecular weight, $\text{g} \cdot \text{mol}^{-1}$
n	number of residues
c	protein concentration, $\text{mg} \cdot \text{cm}^{-3}$
d	optical path, mm

8.10 Crosslinking

Crosslinking of C μ 4tp variants was performed by Tuan Ngyen and Johannes Buchner at the Technical University of Munich, who kindly provided the following protocol:

For crosslinking, C μ 4 Bpa variants were diluted to a concentration of 1 mg/ml (equivalent to A₂₈₀ = 1.37) with a final volume of 200 μ L. The samples were then transferred into a 96-well plate (Greiner BioOne) and excited with a Blak-Ray 100 W UV lamp at 365 nm (UVP Cambridge) for one hour at a distance of 15 cm. After irradiation, the samples were subjected to ESI/MS analysis on an Orbitrap Fusion Lumos (Thermo Fisher Scientific) and diluted to a concentration of 0.2 mg/mL in 50 μ L.

9 Results and Discussions

9.1 Semisynthetic strategy

As outlined in section 7.7, our main goal for this thesis was to establish access to semisynthetic IgM C μ 4tp and to produce variants carrying photocrosslinkers in the tailpiece region to facilitate further exploration of interaction sites before, during and after IgM assembly, and to produce various glycosylated C μ 4tp variants.

The amino acid sequence of C μ 4tp is shown in Figure 18. Figure 18 depicts a threonine/cysteine ligation site at position 91/92. Amino acids to be substituted with a photocrosslinker are underscored. L117 was selected based on its position just prior to the tailpiece β -sheet assembly. Y118 and V120 were chosen due to their structural relevance in the formation of the tailpiece β -sheets at the core of oligomeric IgM¹¹⁴ and reduced oligomerization behavior during an alanine scan by Pasalic *et al.*⁷⁶. Y132 was selected based on its prominent position at the very end of IgM and to gain further insight into the relevance of the tailpiece region located beyond the β -sheet assembly.

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTLYNV SLIMSDTGGT CY

```

Figure 18: Amino acid sequence of C μ 4tp. Underscored black amino acids are amino acids to be substituted with Bpa. Underscored red amino acid represents a conserved glycosylation site. Expressed segment highlighted in blue, synthetic segment in green.

A schematic representation of our semisynthesis strategy is displayed in Figure 19.

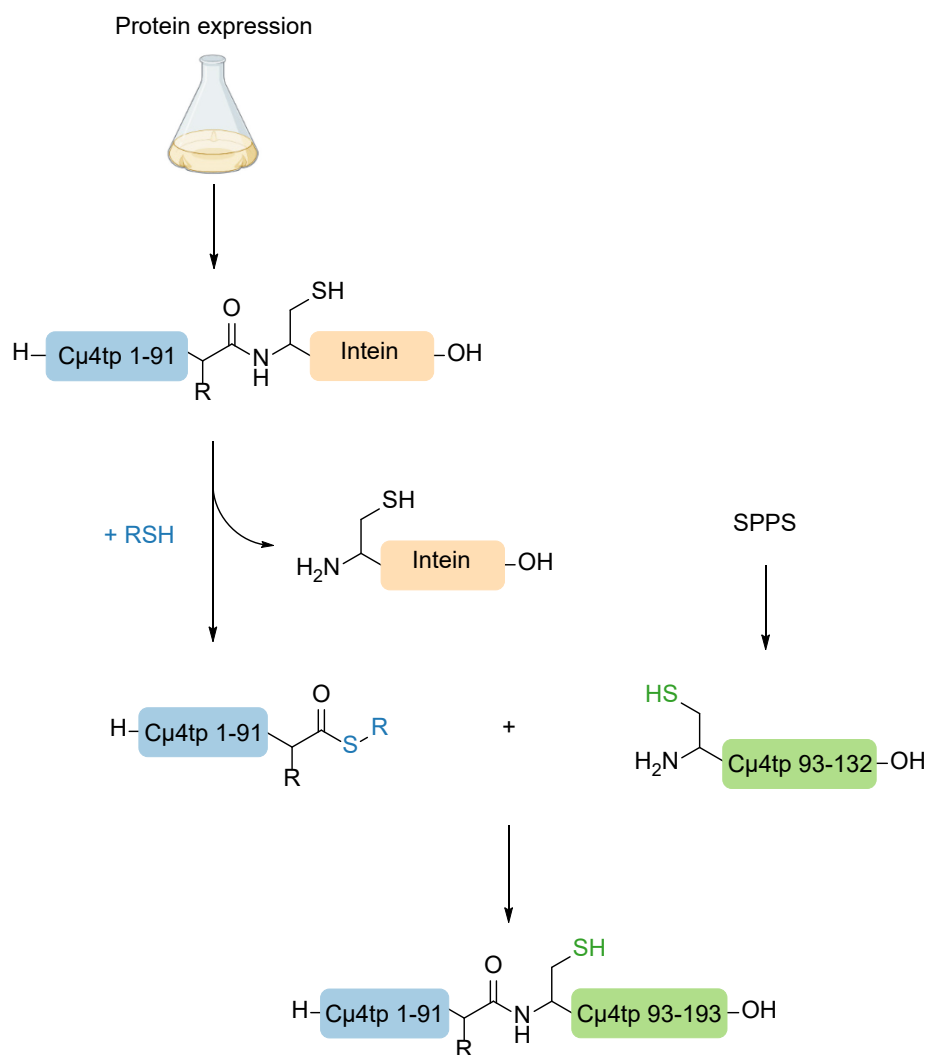


Figure 19: Semisynthesis scheme for $\text{C}\mu 4\text{tp}$.

Our strategy for the production of semisynthetic $\text{C}\mu 4\text{tp}$ variants followed EPL strategies outlined¹⁸². The N-terminal segment (highlighted in blue) was to be expressed in *E. coli* as an intein fusion construct, while the C-terminal segment containing our selected substitution sites (highlighted in green) was to be synthesized *via* SPPS. We decided against the use of

suppression strategies for the introduction of photocrosslinkers, as these are generally less flexible.

We decided to rely on the ligation site located at the native cysteine residue in position 92 highlighted in Figure 18. While this results in a relatively long synthetic segment (40 amino acids), and therefore challenging SPPS procedure, this native cysteine residue allows for simpler workup after ligation. Due to the relatively low abundance of cysteines in naturally occurring proteins (~1.8%), other amino acids have been made available for ligation *via* auxiliaries followed by desulfurization to remove the auxiliary after ligation^{184,232}. For example, the ligation site in Cμ4tp could be moved towards the C-terminus of the domain to position V108/D109, relying on NCL with an N-terminal β-thio Asp on the synthetic segment. A significant advantage of this building block is the ability to selectively remove the auxiliary in the presence of native cysteines using TCEP and DTT at low pH^{233,234}.

9.2 Semisynthesis of C μ 4tp

Before setting out to create our portfolio of modified C μ 4tp variants, we first had to answer the following questions.

- Is the semisynthesis of C μ 4tp based on the strategy outlined in the previous section feasible?
- Does the semisynthetic C μ 4tp domain share the same folding behavior as recombinantly produced C μ 4tp?

To address these questions, we applied our semisynthesis strategy to the synthesis of wild type C μ 4tp. This chapter presents our results related to the recombinant production of N-terminal thioester segment, the solid phase synthesis of the C-terminal segment, and the ligation of these two segments, as well as data comparing the folding behavior of semisynthetic C μ 4tp to recombinantly expressed C μ 4tp. Finally, we discuss these results and how they influenced our consecutive decisions regarding the semisynthesis of C μ 4tp variants.

9.2.1 Production of recombinant C μ 4tp 1-91 thioester

We decided to produce our N-terminal C μ 4tp segments by expressing them as intein fusion constructs and then cleave intein by addition of Mesna to generate C-terminal thioesters of C μ 4tp. While the expression and the intein cleavage themselves worked well, we initially struggled with the purification of the expressed fusion constructs prior to cleavage due to low recovery rates during Ni-NTA affinity chromatography. We eventually opted to skip this purification step, which resulted in far better expression yields.

The respective methods are described in detail in section 8.1.

The full amino acid sequence of the final thioester product is shown in Figure 20.

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYV TSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY T-RSH
  
```

Figure 20: Amino acid sequence of Cμ4tp 1-91 Mesna thioester. Expressed segment highlighted blue. Mesna-thioester highlighted yellow and depicted in cursive. MW_{calc} : 10263.64 Da.

9.2.1.1 Cμ4tp-Intein expression constructs

This work used an Cμ4tp 1-91-MxeIntein-H7-CBD fusion construct created and kindly provided by Alexander Kravchuk. The Mxe GyrA intein derived from *Mycobacterium xenopi* gyrA was chosen due to its long history of successful application in protein semisynthesis^{182,235,236}. H7 represents a poly histidine-tag, that consists of seven consecutive histidine residues and strongly binds to divalent cations such as Ni, Mn, Fe, and Co ions, which allows for efficient purification of the attached protein through the application of Ni-loaded nitrilotriacetic acid columns (Ni-NTA)²³⁷. The chitin-binding domain (CBD) is another tag often used in protein purification²³⁸ but was not relied on for this thesis.

This work relied on the use of a Cμ4tp-1-91-MxeIntein-H7-CBD fusion construct (Figure 21) encoded in a pTXB1 Amp^R plasmid (pTXB1-Cμ4tp-Met-91-MxeIntein-H7-CB) and amplified and expressed in *E. coli* BL21(DE3)Gold. Exact DNA and amino acid sequences of this construct is shown in the appendix (section 12.1).

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCITGDALVA
101    LPEGESVRIA DIVPGARPNS DNAIDLKVLD RHGNPVLADR LFHSGEHPVY
151    TVRTVEGLRV TGTANHPLLC LVDVAGVPTL LWKLIDEIKP GDYAVIQRSA
201    FSVDCAGFAR GKPEFAPTTY TVGVPGLVRF LEAHRDPDA QAIADELTDG
251    RFYYAKVASV TDAGVQPVYS LRVDTADHAF ITNGFVSHAT GLTGIHHHHH
301    HHSGLNSGLT TNPGVSAWQV NTAYTAGQLV TYNGKTYKCL QPHTSLAGWE
351    PSNVPALWQL Q

```

Figure 21: Amino acid sequence of Cμ4tp 1-91-MxeIntein-H7-CBD. Cμ4tp 1-91 highlighted in blue. Intein-H7-CBD highlighted in beige. MW_{calc} : 39197.13 Da.

9.2.1.2 Cμ4tp 1-91-thioester production

Cμ4tp 1-91-MxeIntein fusion constructs were expressed in *E. coli* BL21(DE3)Gold as outlined above (8.1). The fusion protein was primarily detected in inclusion bodies. This allowed for extended expression durations, including overnight expressions, as proteins in inclusion bodies are typically not efficiently degraded by the cellular machinery. A representative expression monitoring is depicted in Figure 22.

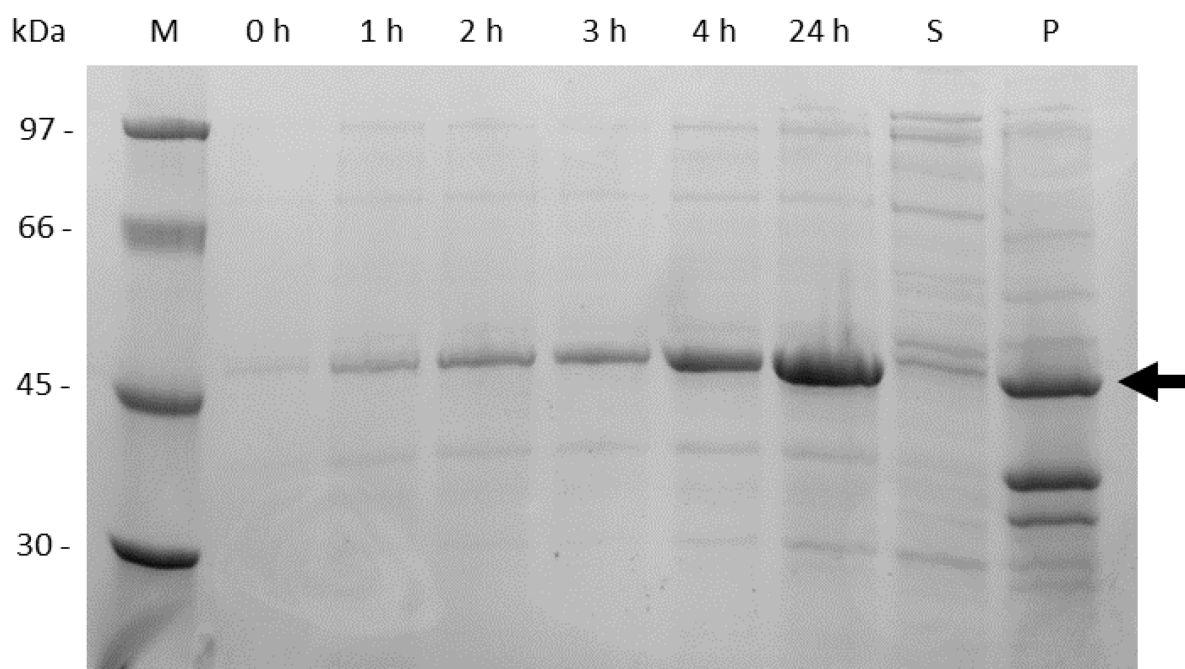


Figure 22: Photography of representative SDS-PAGE monitoring of $C\mu 4tp$ -Intein fusion construct expression in *E. coli*. The arrow points at the protein of interest. Expected mass ~39 kDa; M: protein marker; S: supernatant; P: pellet; Gel stained with Coomassie blue. Timepoints represent hours after induction via addition of IPTG.

The expressed fusion protein was purified *via* cell disruption and affinity chromatography using Ni-NTA columns as outlined in sections 8.1.2 to 8.1.6. Figure 23 depicts a representative analysis of the fractions produced during affinity chromatography.

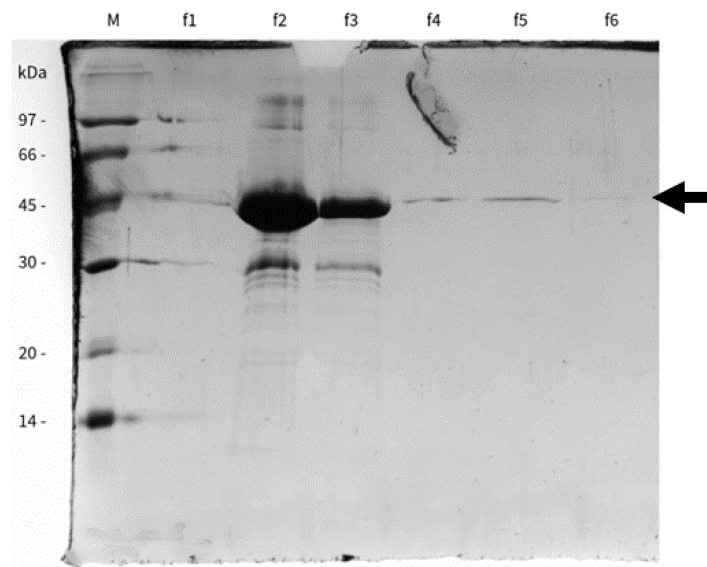


Figure 23: Photography of representative fraction analysis after affinity chromatography using SDS-PAGE. The arrow points at the protein of interest. Expected mass ~39 kDa; Protein marker (M), fractions (f1-6). Gel stained with Coomassie blue.

Ni-NTA improved the purity of our protein, but we also observed low recovery rates after affinity purification, which incentivized us to change this workflow to avoid affinity purification all together. Our changes are described in detail in section 8.1.

Mesna-induced intein cleavage was performed as outlined in section 8.1.7 at a final concentration of 2M urea, as higher concentrations of chaotropic agents prevented Mxe folding and function¹⁸². The progress of the cleavage reaction was monitored *via* SDS-PAGE. Figure 24 displays a representative experiment and shows three distinct bands at different heights. These correspond to the uncleaved fusion construct at ~45 kDa and the intein-H7-CBD protein ~30 kDa. The C μ 4tp Mesna thioester and C μ 4tp, created through hydrolysis of the C μ 4tp Mesna thioester, cannot be distinguished on this gel and both run at ~10 kD.

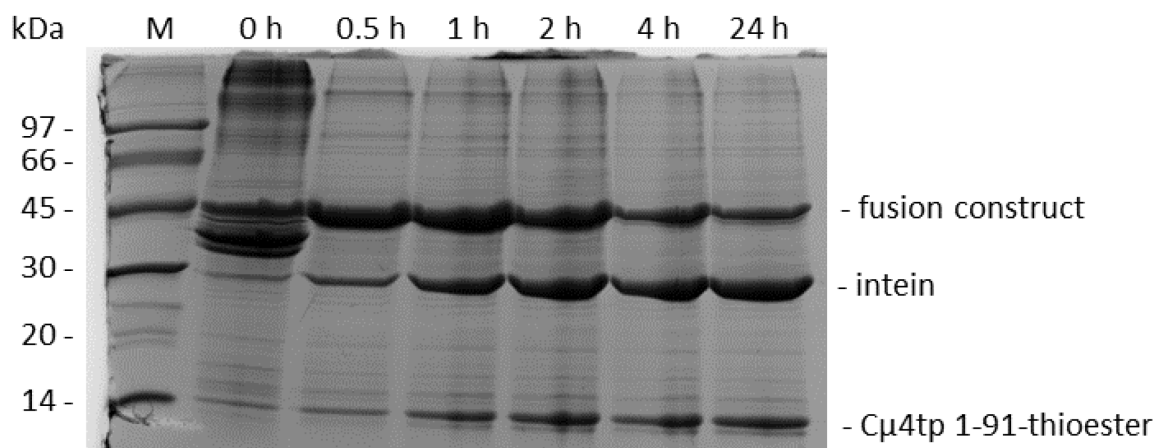


Figure 24: Photography of representative SDS-Page monitoring of Mesna induced intein cleavage of $C\mu 4tp$ -Intein fusion protein. $MW(\text{fusion construct})$: 39 kDa, $MW(\text{intein})$: 39197.23 Da, $MW(C\mu 4tp \text{ thioester})$: 10263.64 Da, $MW(\text{hydrolysis})$: 10139.47 Da, M =protein gel marker (check specifics). Gel stained with Coomassie blue. Timepoints represent hours after addition of Mesna.

These proteins were separated using reversed-phase HPLC systems, as outlined in section 8.1.8. While separation of the intein was trivial, separation of the $C\mu 4tp$ and $C\mu 4tp$ -thioester was not consistently achieved, as both co-eluted at ~45 % buffer B. This is not unexpected, as the polarity of these compounds is almost identical. To our surprise, the uncleaved fusion protein had a very strong tendency to stick the column even at 60°C column temperature and could only effectively be removed *via* Gdn.HCl washes after purification. Fractions containing the desired $C\mu 4tp$ thioester were identified by MS, combined, and analyzed as outlined in the previous chapter (section 8.1.9). Representative results are depicted in Figure 25.

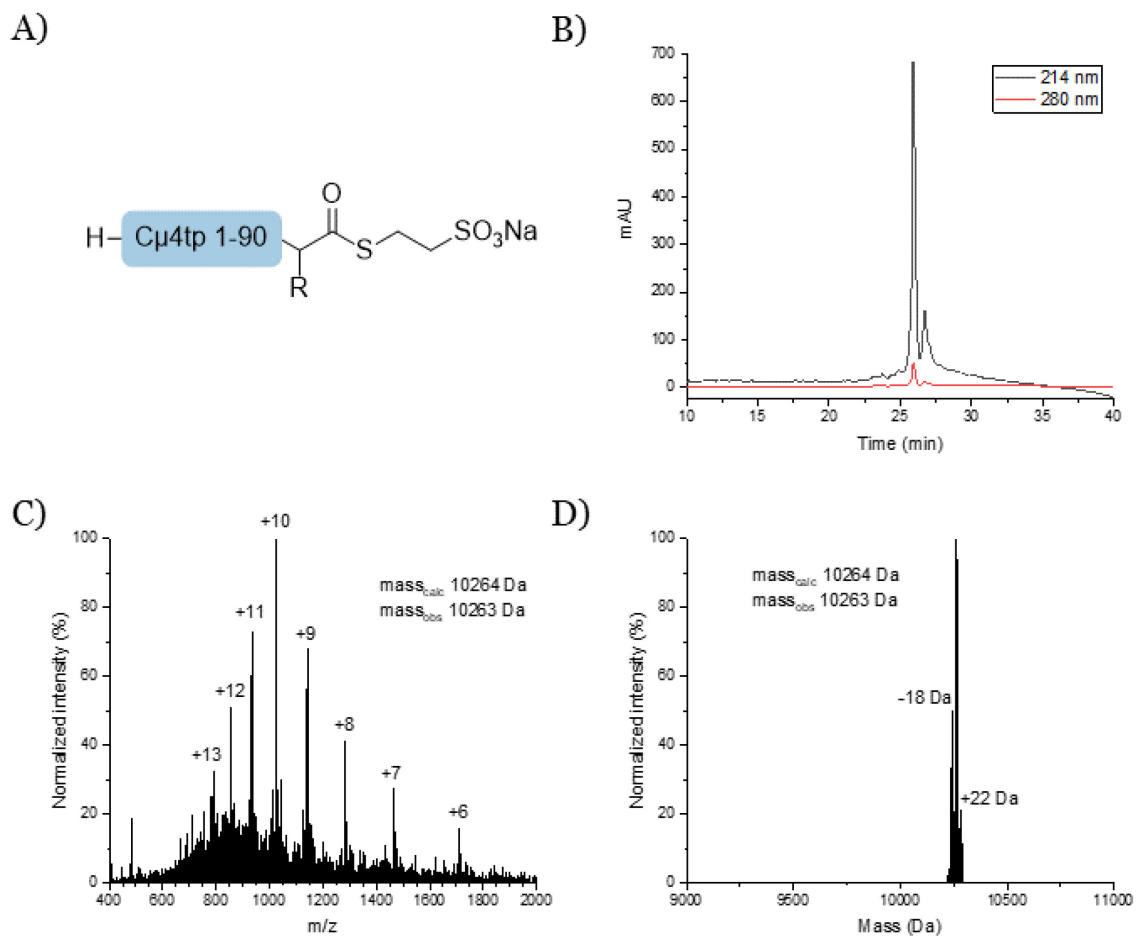


Figure 25: Representative characterization of Cμ4tp 1-91-thioester A) structure of the Cμ4tp 1-91 Mesna thioester B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; Cμ4tp 1-91-thioester MW_{obs} : 10263 Da, MW_{calc} : 10263.64 Da.

During repeated production of Cμ4tp thioester, we noticed weak binding of the Cμ4tp-Intein fusion construct to our Ni-NTA columns resulting in a significantly lower yield per L of expression media. As the fusion protein was already of relatively high purity after cell disruption and solubilization due to compartmentalization in inclusion bodies, we opted to skip the affinity purification steps, directly submitting the fusion construct solubilized in 6M Gdn.HCl to dialysis against 4M urea. This increased the average yield of the protein thioester

production line from ~2 mg to ~8 mg per L of expression media with no detectable differences in purity after preparative HPLC purification. It is unclear what caused the additional signal present in Figure 25 B), as we could not identify additional signals in our MS results.

9.2.2 Synthesis and purification of C μ 4tp 92-132 wt

The efficient production of various C μ 4tp variants required us to produce a range of modified, C-terminal C μ 4tp segments. To achieve this goal, we relied on applying several well-established SPPS strategies, which allow for the incorporation of non-proteinogenic amino acids as described in section 8.2.1. C μ 4tp 92-132 wt was synthesized both manually *via* Boc-SPPS and automated *via* Fmoc-SPPS. Figure 26 shows the full amino acid sequence of the final peptide.



Figure 26: Amino acid sequence of C μ 4tp 92-132 wt. MW 4394.02 Da.

Manual synthesis *via* Boc-SPPS on Boc-Tyr(2-bromo-Z)-PAM resin was performed at a 0.2 mmol scale. This synthesis used HCTU instead of HBTU, due to the latter not being available temporarily owing to long delivery times. The overall yield of the purified peptide was 73 mg or 8.3% theoretical yield.

C μ 4tp 92-132 wt was also synthesized automatically on a Liberty Blue synthesizer on Fmoc-Gly Protide resin at 0.1 mmol scale, to estimate the applicability of this time-saving technology for future C μ 4tp variants. HPLC and MS characterization of the final product prior to ligation is depicted in Figure 27. The overall yield of the purified peptide was 48 mg or

10.9%. The unexpectedly broad signal in panel B) was caused by leakage of the analytical HPLC equipment, that was not detected in time.

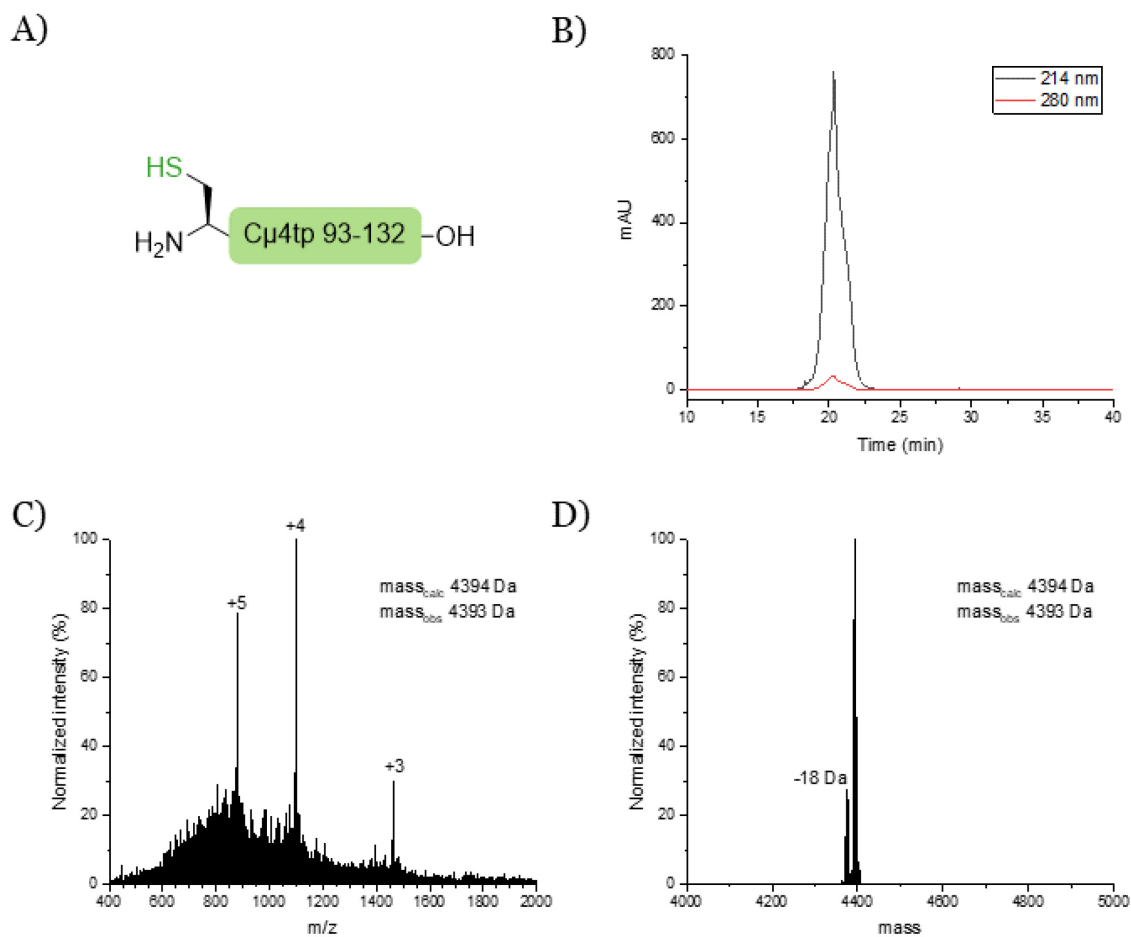


Figure 27: Characterization of Cμ4tp wt A) Cμ4tp 92-132 wt structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 4393 Da, MW_{calc} : 4394.02 Da.

9.2.3 Ligation and purification of Cμ4tp wt

Cμ4tp thioester was ligated to the synthetic segment *via* EPL and purified *via* reverse-phase HPLC, as outlined in section 8.3.

During purification, we faced significant challenges separating the non-reacted Cμ4tp 1-91 thioester (plainly Cμ4tp 1-91) from the ligation products. This hydrolysis is a side reaction during ligation and cannot be prevented entirely. However, we took measures to minimize its occurrence by selecting thioester aliquots with minimal amounts of hydrolyzed Cμ4tp 1-91 as starting materials during ligation, lowering the pH of the ligation reaction, by increasing the overall concentration of the reaction partners during the reaction, and by changing the molar ratio of thioester to synthetic peptide from 1:1 to >1:2. Additionally, we adjusted the gradient of solvent A to solvent B during reversed-phase HPLC from 2% to 0.5% per minute.

The amino acid sequence of Cμ4tp wt is shown in Figure 28.

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTLYNV SLIMSDTGGT CY

```

Figure 28: Amino acid sequence of Cμ4tp wt. Expressed segment highlighted in blue, synthetic segment highlighted in green. Expected mass 114515.48 Da.

In total, the reaction contained 10 mg Cμ4tp 92-132 wt and 10 mg Cμ4tp 1-91-thioester in 2.5 ml ligation buffer and was performed overnight. Figure 29 displays reaction monitoring performed *via* LCMS. Due to the significant overlap of the ligation product with the N-terminal

segment shown in panel A, it is difficult to estimate the reaction progress based on peak height or area. However, deconvoluted MS data (panel B) shows a steady increase in ratio of product to N-terminal segment.

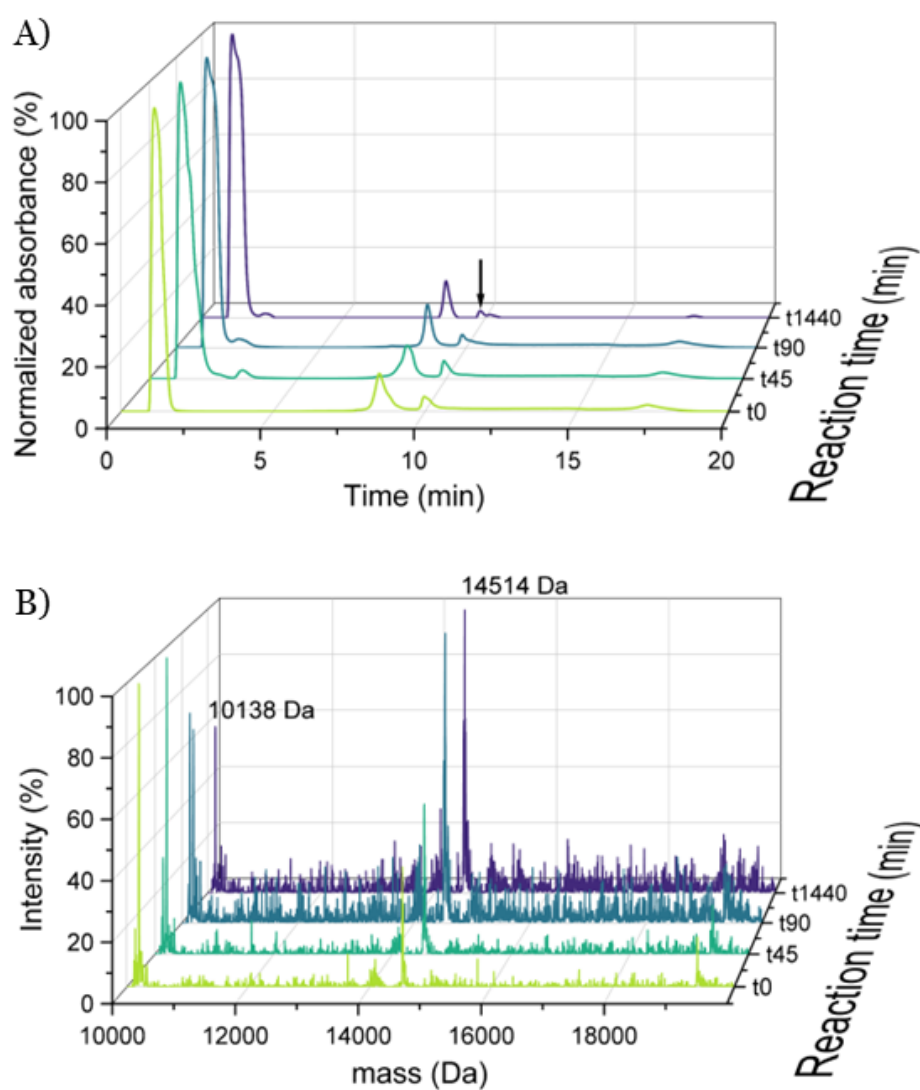


Figure 29: Reaction monitoring of the $C\mu 4tp$ 1-132 wt ligation using LCMS. A) 214 nm trace during LCMS; the arrow indicates the ligation product as well as the overlapping thioester starting material B) deconvoluted MS of the peak indicated by the arrow in panel B. MW_{obs} : 14514 Da, MW_{calc} : 14515.48 Da.

1 mg of ligation product was collected as main pool, and 2 mg were collected as the side pool. In total, 3 mg of material were obtained after purification. This corresponds to a yield of ~21%. The additional +77 Da mass most likely corresponds to a 2-mercaptoethanol disulfide adduct (76 Da) when accounting for a ± 1 Da inaccuracy, that can be removed through the addition of TCEP. We decided to discard fractions containing even minor amounts of hydrolyzed starting material, as our cooperation partners had reported that this segment negatively affected folding.

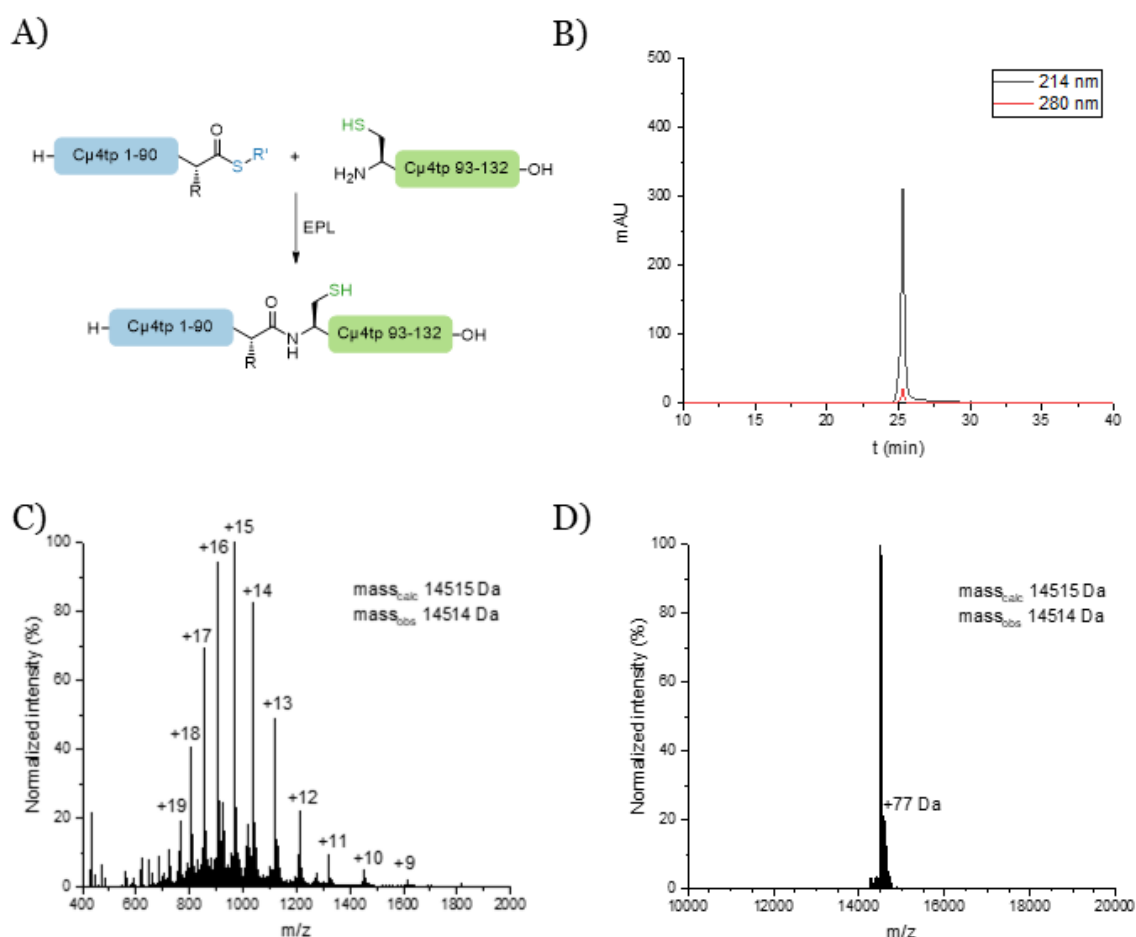


Figure 30: Characterization of Cμ4tp wt main pool A) Cμ4tp 1-132 wt structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs}: 14514 Da, MW_{calc}: 14515.48 Da.

9.2.4 Characterization of C μ 4tp folding behavior

Tuan Nguyen performed the folding and oligomerization analysis of C μ 4tp wt at the Technical University of Munich. C μ 4tp wt was folded *via* a series of buffer exchanges and analyzed using SEC (as outlined in section 8.7), non-reducing PAGE (section 8.8), CD spectroscopy and thermal transition (section 8.9) and compared to recombinantly expressed C μ 4tp wt provided by Johannes Buchner⁷⁶.

Figure 31 shows the distinct elution time of C μ 4tp wt monomer, dimer, and oligomers on SEC, indicating successful folding.

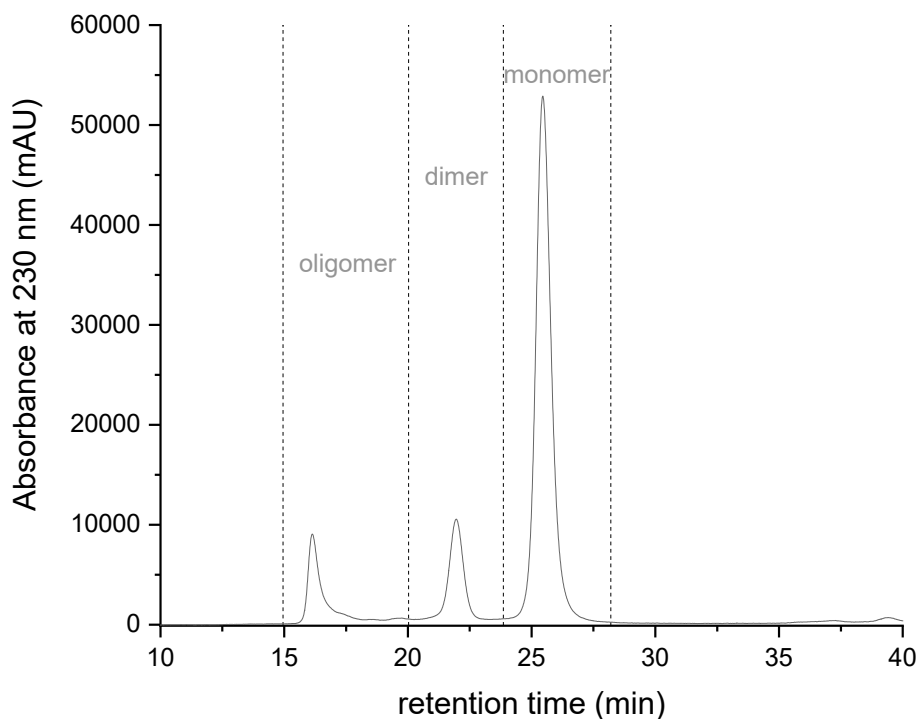


Figure 31: UV-VIS traces of folded C μ 4tp wt during analytical SEC display formation of oligomers and dimers.

We furthermore subjected both recombinantly expressed C μ 4tp and semisynthetic C μ 4tp before and after folding to CD spectroscopy, as shown in Figure 32: CD spectrum of C μ 4tp wt. There is a clear change in structure to the semisynthetic C μ 4tp upon folding, as indicated by the difference in ellipticity. Both semisynthetic C μ 4tp and recombinantly expressed C μ 4tp display very similar CD spectra after oxidation, which further supports our hypothesis, that semisynthetic C μ 4tp behaves similar to recombinantly expressed C μ 4tp provided by the Buchner lab⁷⁶. The broad valley at ~220 nm indicates the formation of β -sheets, which are prominent in the immunoglobulin fold of the C μ 4 domain.

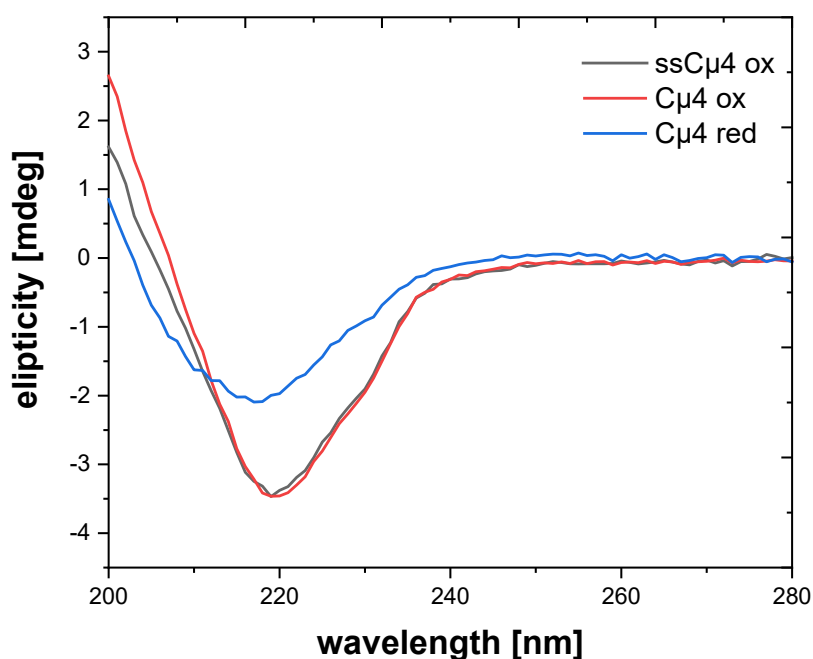


Figure 32: CD spectrum of C μ 4tp wt; ssC4 ox represents folded, semisynthetic C μ 4tp, C μ 4 ox represents folded, recombinantly expressed C μ 4tp, and C μ 4 red represents unfolded, recombinantly expressed C μ 4tp.

Finally, we also recorded thermal transition spectra of recombinantly expressed C μ 4tp and both folded and unfolded semisynthetic C μ 4tp (Figure 33). In accordance with our CD spectroscopy measurements, both semisynthetic and recombinantly produced C μ 4tp display almost identical thermal transition spectra after folding.

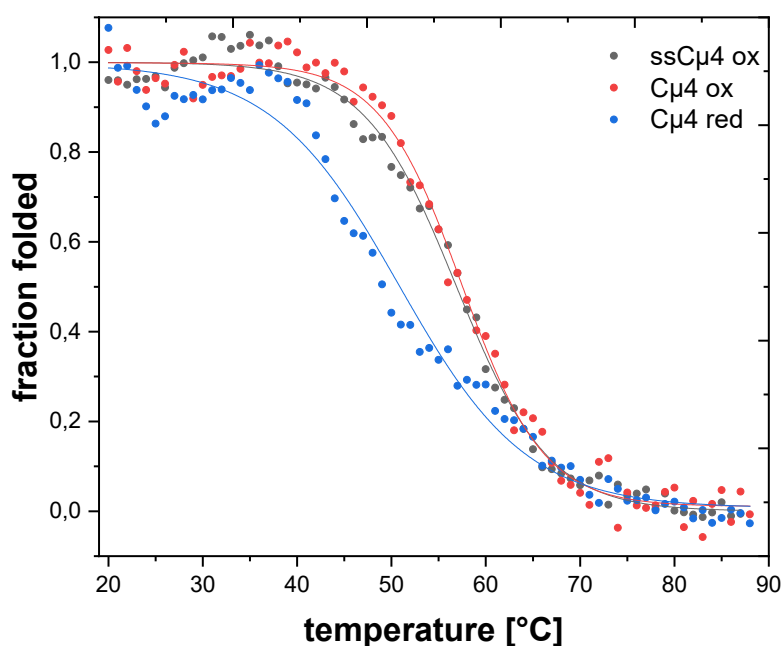


Figure 33: Thermal transition spectrum of C μ 4tp wt, C μ 4 ox represents folded, recombinantly expressed C μ 4tp, and C μ 4 red represents unfolded, recombinantly expressed C μ 4tp.

9.2.5 Discussion of C μ 4tp wt semisynthesis and characterization

Returning to the questions raised at the beginning of this chapter, we successfully demonstrated the feasibility of IgM C μ 4tp semisynthesis and could show similar folding behavior compared to recombinant C μ 4tp.

While obtaining reasonable recovery of the expressed intein-fusion construct from affinity chromatography turned out to be challenging, we were able to obtain C μ 4tp 1-91-thioester at reasonable purity by skipping this affinity purification step entirely. We believe that this is owed to the high percentage of fusion construct located in inclusion bodies during expression.

Both Boc- and Fmoc-SPPS strategies could be applied successfully to obtain C-terminal C μ 4tp segments. The option to use automated Fmoc-SPPS with microwave-heated synthesizers can significantly accelerate the production of additional C-terminal peptide variants. It goes without saying that this is a substantial benefit to future semisynthesis efforts of C μ 4tp.

The ligation of N- and C-terminal segments of C μ 4tp posed two major, intersecting challenges. First, high rates of hydrolysis reduced the overall yield of the reaction. Second, the hydrolyzed thioester, or simply C μ 4tp 1-91, displayed a very similar elution profile to C μ 4tp. Though we did manage to reduce the rate of hydrolysis and improve the separation of the two proteins on RP-HPLC, both steps would benefit from further development. This is of particular importance as C μ 4tp 1-91 contamination of C μ 4tp has negative effects on C μ 4tp folding behavior, thereby dramatically reducing yields during folding. Recovery of the ligation product during RP-HPLC purification should be further improved to push yields beyond 15%. Being able to rely on the naturally occurring cysteine in position 92 for ligation in combination with successful SPPS of the forty amino acid long C-terminal segment, poses a significant advantage over having to perform a ligation at our alternative ligation site which does not

benefit from a naturally occurring cysteine residue. Therefore, this optional strategy was not pursued further.

Folding of C μ 4tp was performed successfully and all subsequent analysis performed by our cooperation partners indicated the same folding behavior as recombinant C μ 4tp. Therefore, we felt confident moving forward with the production and investigation of photocrosslinking C μ 4tp variants.

9.3 The semisynthesis of photocrosslinking C μ 4tp variants

Encouraged by our successful semisynthesis of wild type C μ 4tp, we expanded our efforts to the production of C μ 4tp variants for photocrosslinking. Regarding our choice of photocrosslinker, we ultimately settled on the use of benzophenone-based BPA. This decision was based on the commercial availability of Fmoc-Bpa building blocks for SPPS and ease of handling, as BPA is relatively tolerant of ambient light as opposed to other photocrosslinkers

192 .

Figure 34 shows the position of all attempted C μ 4tp Bpa substitution variants.

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYV TSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTLYNV SLIMSDTGGT CY

```

Figure 34: Amino acid sequence of C μ 4tp. Underscored black amino acids are amino acids to be substituted with Bpa. Underscored red amino acid represents a conserved glycosylation site. Expressed segment highlighted in blue, synthetic segment in green.

9.3.1 Synthesis and purification of C μ 4tp 92-132 Bpa variants

To produce modified, C-terminal C μ 4tp segments, we relied on applying several well-established SPPS strategies, which allow for the incorporation of non-proteinogenic amino acids as described in section 8.2.1.

C μ 4tp 92-132 Y118Bpa was initially produced *via* Boc SPPS (section 8.2.1.2), as good results with this strategy regarding C μ 4tp semisynthesis had previously been accomplished with this

method²²⁶. Boc SPPS was eventually discontinued replaced with a Fmoc-based synthesis strategy (section 8.2.1.3 and 8.2.1.4), which offered the additional advantage of including a cleavage protocol with less organizational overhead, thereby allowing for additional test cleavages during synthesis.

All peptides were purified *via* RP-HPLC as described in section 8.2.3.

9.3.1.1 Synthesis of C μ 4tp 92-132 L117Bpa

C μ 4tp L117Bpa was synthesized *via* Fmoc-SPPS on 2-Chlorotrityl chloride resin preloaded with Fmoc-Tyr at a scale of 0.1 mmol. HPLC and MS characterization of the final product prior to ligation is depicted in Figure 35. Ultimately, we obtained a total yield of 24 mg which corresponds to 6% theoretical yield at a 0.1 mmol scale. An additional 21 mg of material were collected in a side pool (resulting in 11% total theoretical yield). The mass of +99 Da shown in Figure 35 (D) may correspond to a not fully deprotected sidechain carrying a t-Butyloxycarbonyl (tBoc)-protecting group, when accounting for a mass accuracy of ± 1 Da. Additionally, we detected a sodium adduct (+22 Da) that is not clearly visible in the presented figure.

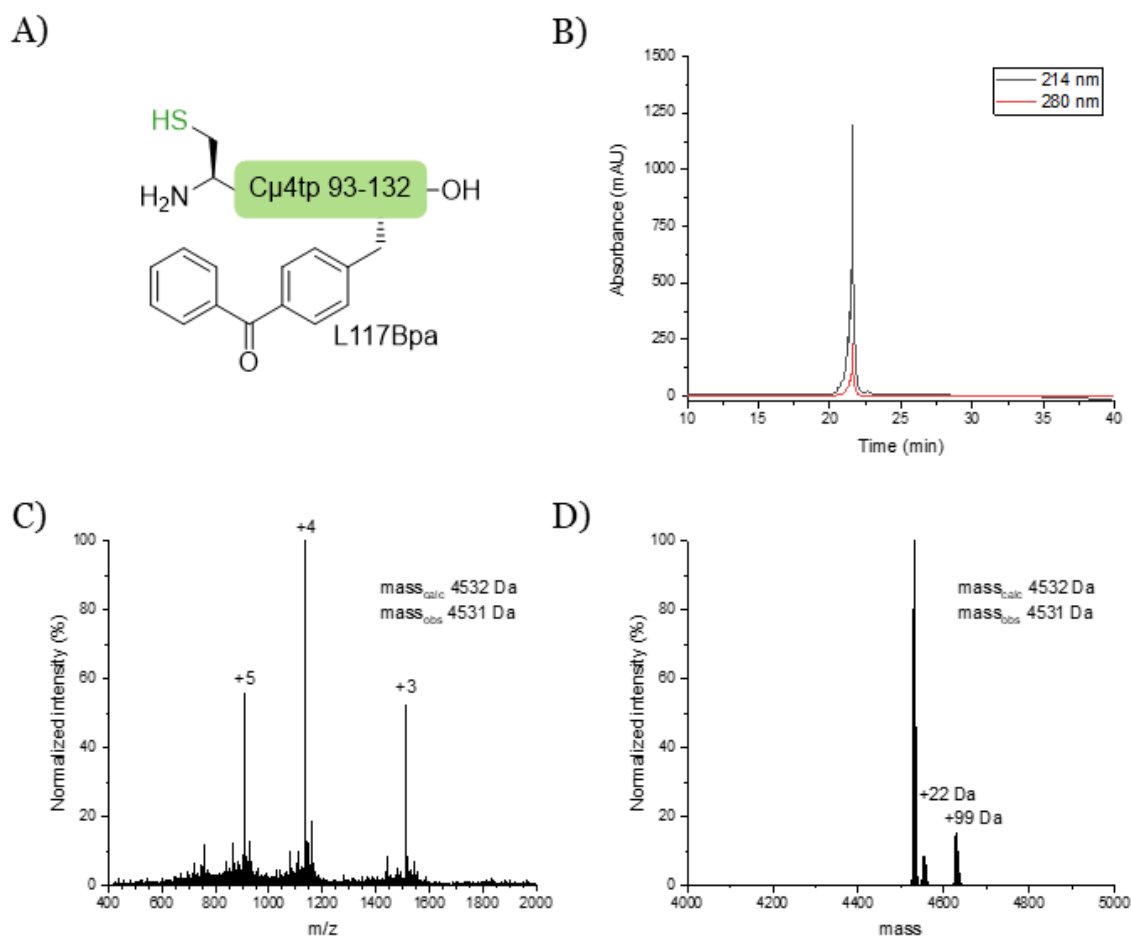


Figure 35: Characterization of C μ 4tp 92-132 L117Bpa via MS and analytical HPLC A) C μ 4tp 92-132 L117Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; C μ 4tp 92-132 L117Bpa MW_{obs}: 4531 Da, MW_{calc}: 4532.15 Da.

9.3.1.2 Synthesis of C μ 4tp 92-132 Y118Bpa

C4tp 92-132 Y118Bpa was synthesized manually *via* Boc-SPPS on Boc-Tyr(2-bromo-Z)-PAM resin (200-400 mesh, 0.45 mmol/g) at a 0.1 mmol scale. HPLC and MS characterization of the final product prior to ligation is depicted in Figure 36. Ultimately, we obtained a total yield of 10 mg which corresponds to 2%. The mass of +22 Da corresponds to a sodium adduct, while

the +98 Da shown in Figure 36 D) may correspond to a not fully removed tBoc protecting group (+100 Da), when accounting for a mass accuracy of ± 2 Da.

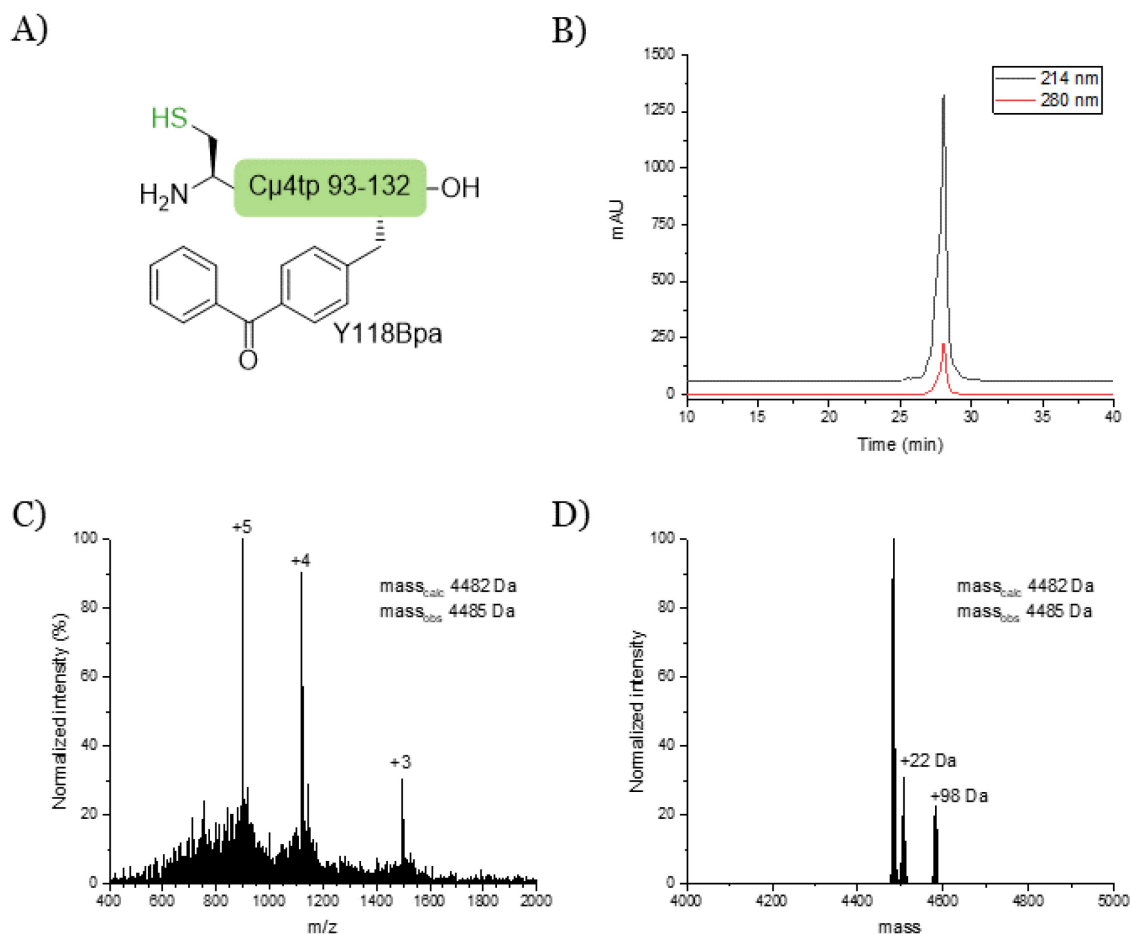


Figure 36: Characterization of C μ 4tp 92-132 Y118 via MS and analytical HPLC A) C μ 4tp 92-132 Y118Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 4485 Da, MW_{calc} : 4482.13 Da.

9.3.1.3 Synthesis of C μ 4tp 92-132 V120Bpa

C μ 4tp V120Bpa was synthesized manually *via* Fmoc-SPPS on 2-Chlorotrityl chloride resin preloaded with Fmoc-Tyr at a scale of 0.1 mmol. HPLC and MS characterization of the final product prior to ligation is depicted in Figure 37. Ultimately, we obtained a total yield of 23 mg which corresponds to 5%. The higher mass found in the deconvoluted spectrum (D) may indicate a not fully removed tBoc protecting group (+100 Da), when accounting for a ± 1 Da mass accuracy. We were unable to identify the source of the signal at ~25 min in panel B).

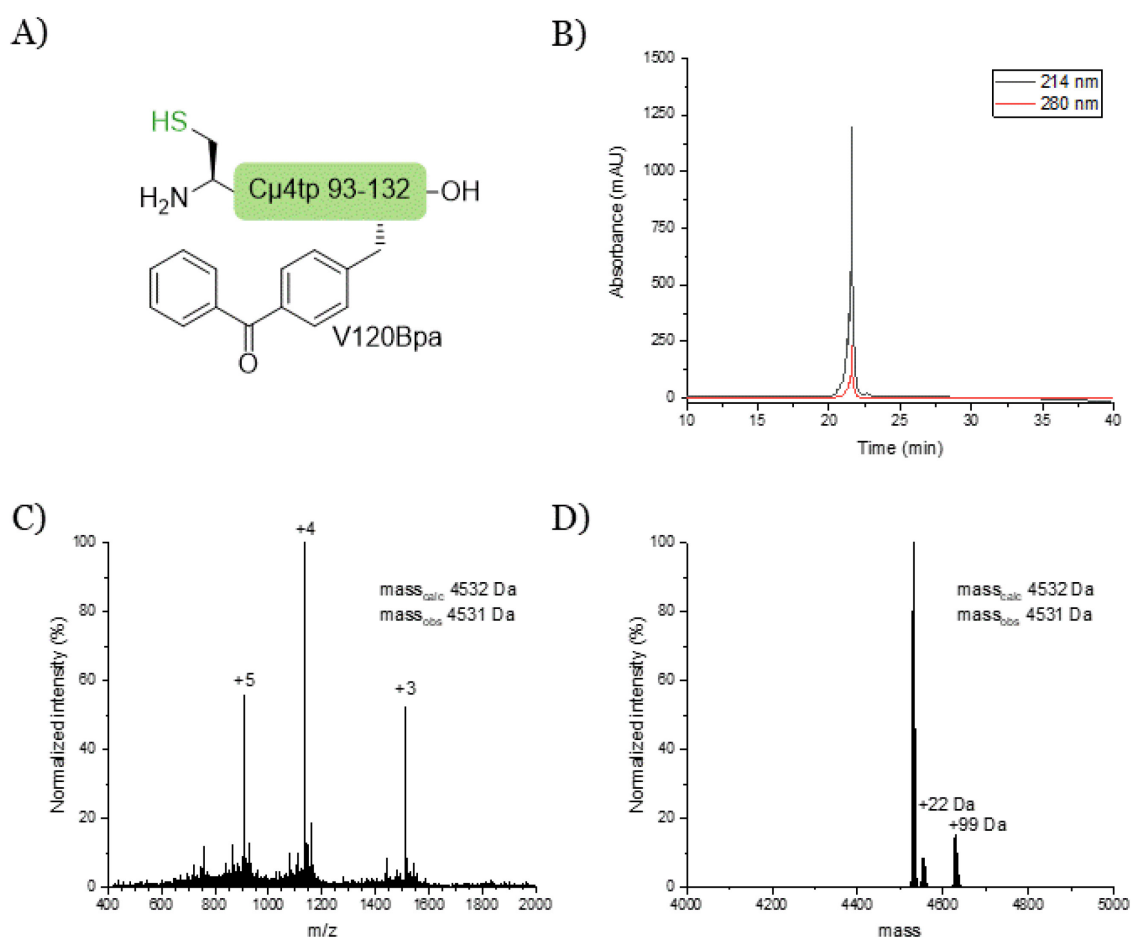


Figure 37: Characterization of C μ 4tp 92-132 V120Bpa via MS and analytical HPLC A) C μ 4tp 92-132 V120Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 4546 Da, MW_{calc} : 4546.17 Da.

9.3.1.4 Synthesis attempts of C μ 4tp 92-132 Y132Bpa

Synthesis of C μ 4tp Y132Bpa was attempted using both Wang and 2-Chlorotrityl chloride resins manually loaded with Fmoc Bpa. Despite our efforts, we were unable to produce the full peptide. While SPPS initially produced a clean product for the first ~9 amino acids (Figure 38, panel C), the quality of the synthesis rapidly deteriorated with the growing peptide chain, ultimately producing a complex mixture of peptides that we were unable to purify or further analyze (panel B and D). Surprisingly, the signals detected in MS in panel D perfectly match the expected Y-series of the peptide up to amino acid fourteen. The profile displayed in analytical HPLC shown in panel B) displays a variety of peaks, indicating the presence of multiple distinct peptides present. It is therefore extremely unlikely that the fragmentation observed in MS is caused during ionization but is clearly present before MS analysis. The clean product visible during test cleavage after the first nine amino acids (panel C) indicates that the observed signals are due to fragmentation of the peptide and are not caused by partial premature termination of the synthesis. It is unclear what caused this behavior. These challenges slowed down our synthesis of IgM C μ 4tp Y132Bpa and were eventually abandoned when we received updates from our collaboration partners at the Technical University of Munich, who had at this point stopped to perform crosslinking experiments with our Bpa constructs, as no significant crosslinking could be observed for any of the variants. This topic is further discussed in detail in section 9.3.5.

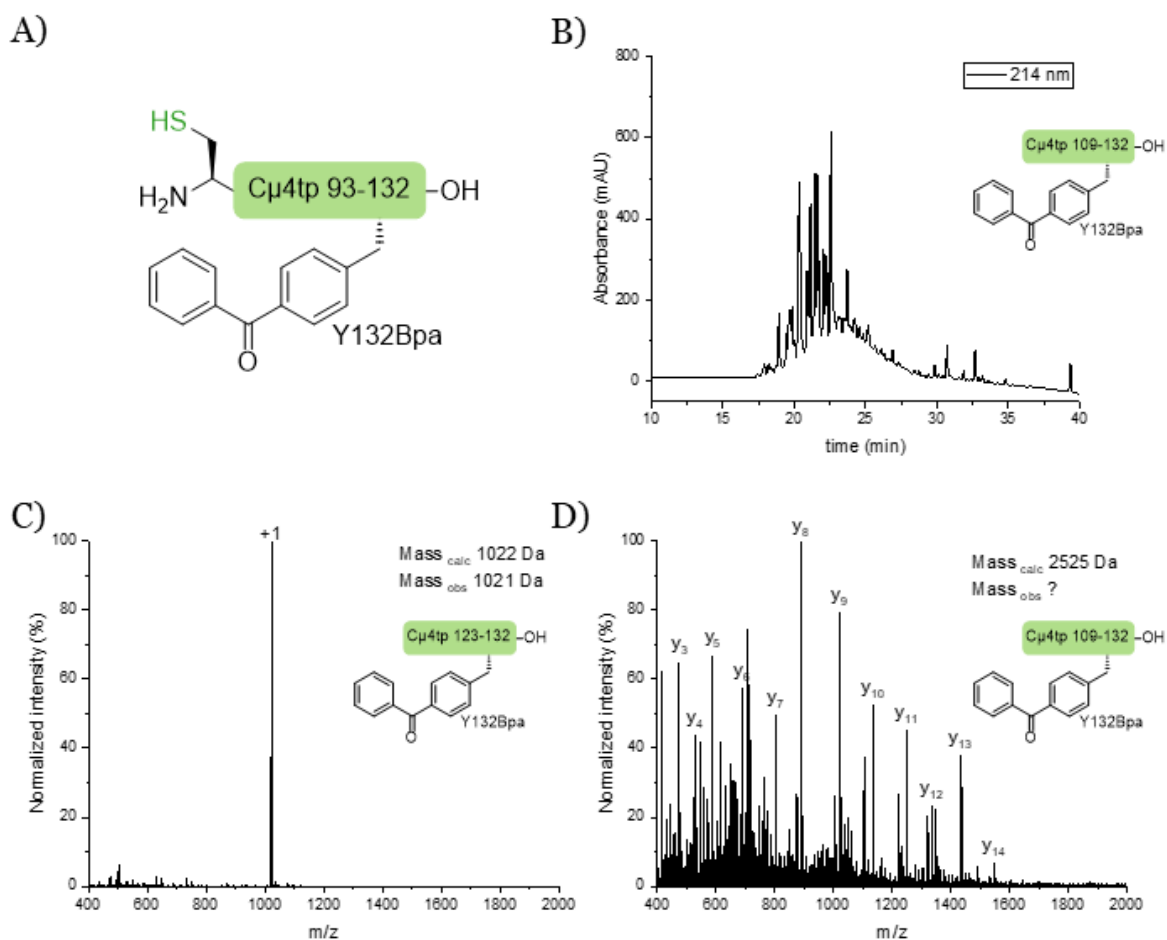


Figure 38: Test cleavage of Cμ4tp 1-132 Y132Bpa via MS A) Cμ4tp 92-132 Y132Bpa structure B) UV-VIS trace of analytical HPLC analysis of Cμ4tp 109-132Bpa after test cleavage C) mass spectrum of Cμ4tp 123-132Bpa after test cleavage D) mass spectrum of Cμ4tp 109-132Bpa after test cleavage; MW_{obs} (Cμ4tp 123-132Bpa): 1021 Da, MW_{calc} (Cμ4tp 123-132Bpa): 1022 Da; MW_{obs} (Cμ4tp 109-132Bpa): n.a., MW_{calc} (Cμ4tp 109-132Bpa): 1022 Da.

9.3.2 Ligation and purification of IgM Cμ4tp Bpa variants

Cμ4tp thioesters were ligated to synthetic segments *via* EPL and purified *via* reversed-phase HPLC, analogous to the previous chapter (section 9.2), and as outlined in Materials and Methods (section 8.3). It is important to note that the ease of separation of the hydrolyzed

thioester and the ligation product is influenced by the variant itself. The following sections show the final analysis of the thus-created semisynthetic C μ 4tp variants.

9.3.2.1 IgM C μ 4tp L117Bpa

Ligation of C μ 4tp L117Bpa was carried out as described in section 8.3. In total, the reaction contained 9 mg C μ 4tp 92-132 L117Bpa and 10 mg C μ 4tp 1-91-thioester in 2 ml ligation buffer and was performed overnight. After purification, we obtained 2 mg C μ 4tp L117Bpa main pool and 4 mg of side pools. In total, this represents a yield of 6 mg or ~40%. The main pool displayed a purity of 93% based on relative peak height at 214 nm in analytical HPLC. Figure 39 depicts our characterization of C μ 4tp L117Bpa main pool.

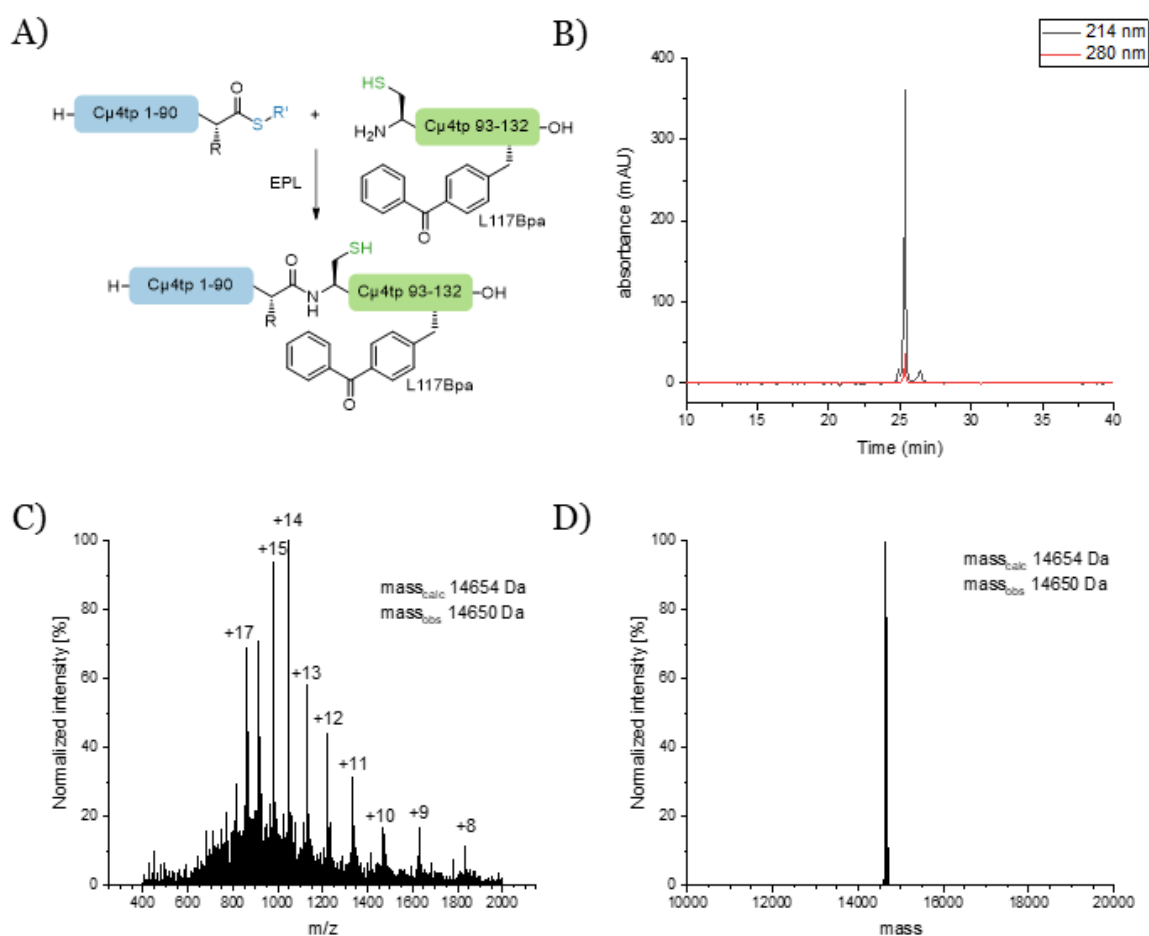


Figure 39 Characterization of Cμ4tp L117Bpa main pool 1 A) Cμ4tp 1-132 L117Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 14650 Da, MW_{calc} : 14653.60 Da.

9.3.2.2 IgM Cμ4tp Y118Bpa

Ligation of Cμ4tp Y118Bpa was carried out as described in section 8.3. In total the reaction contained 5 mg Cμ4tp 92-132 L118Bpa and 5 mg Cμ4tp 1-91-thioester in 0.5 ml ligation buffer and was performed overnight. After purification we obtained two aliquots: one main pool and one side pool. Both contained 1 mg of material which corresponds to a total yield of 2 mg or 25%. The purity of the main pool was estimated *via* relative peak height at 98%, while the purity of the side pool was estimated to be 76%. The characterization of our main pool is displayed in Figure 40. The +97 Da mass corresponds to the formation of a TFA ester.

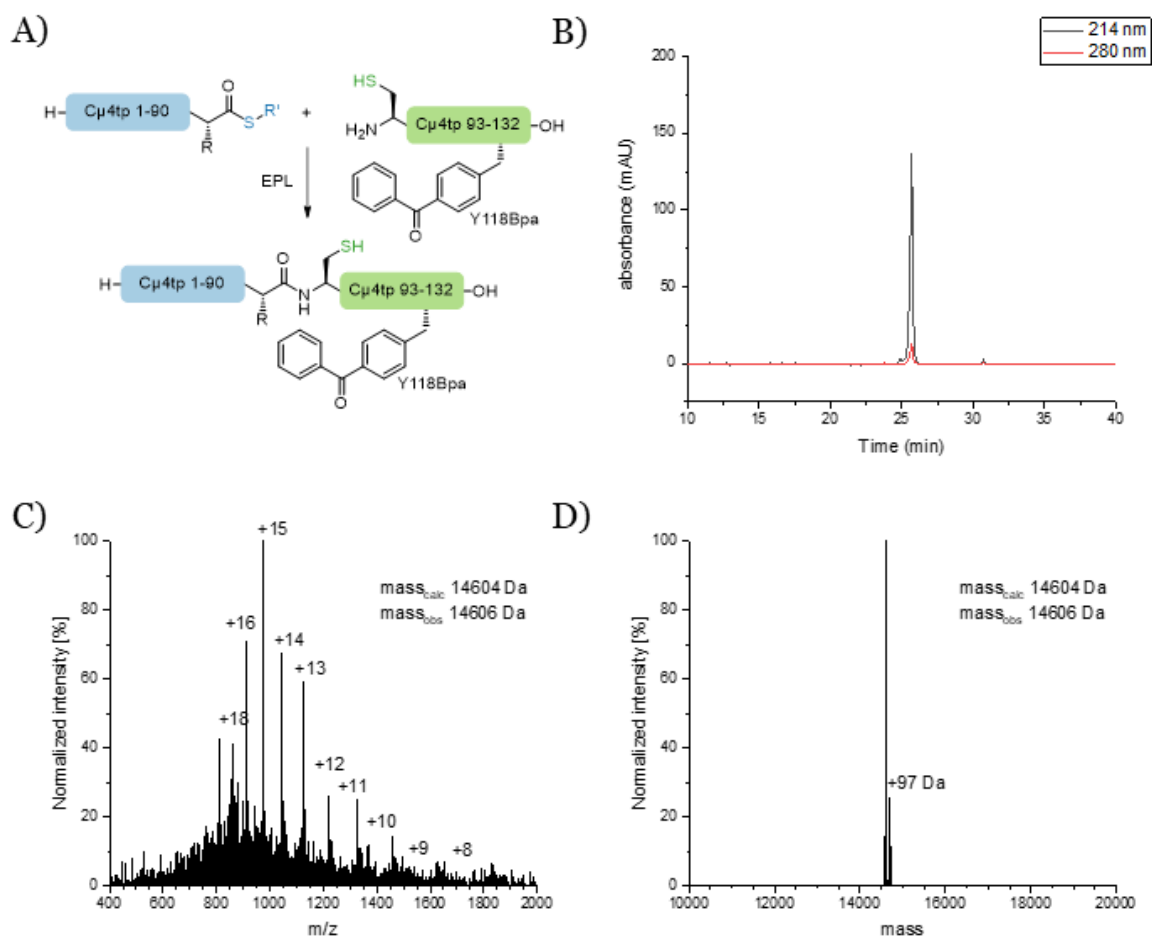


Figure 40: Characterization of Cμ4tp Y118Bpa main pool A) Cμ4tp 1-132 Y118Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 14606 Da, MW_{calc} : 14603.58 Da.

9.3.2.3 IgM Cμ4 V120Bpa

Ligation of Cμ4tp V120Bpa was carried out as described in section 8.3. In total the reaction contained 5 mg Cμ4tp 92-132 V120Bpa and 5 mg Cμ4tp 1-91-thioester in 1 ml ligation buffer and was performed overnight.

After purification, we obtained 2 mg of protein, which is equivalent to 25% theoretical yield.

The purity was estimated to be 94% based on peak height. The +77 Da mass corresponds to a

2-mercaptoethanol adduct, that can be removed *via* addition of TCEP and that disappeared during subsequent folding.

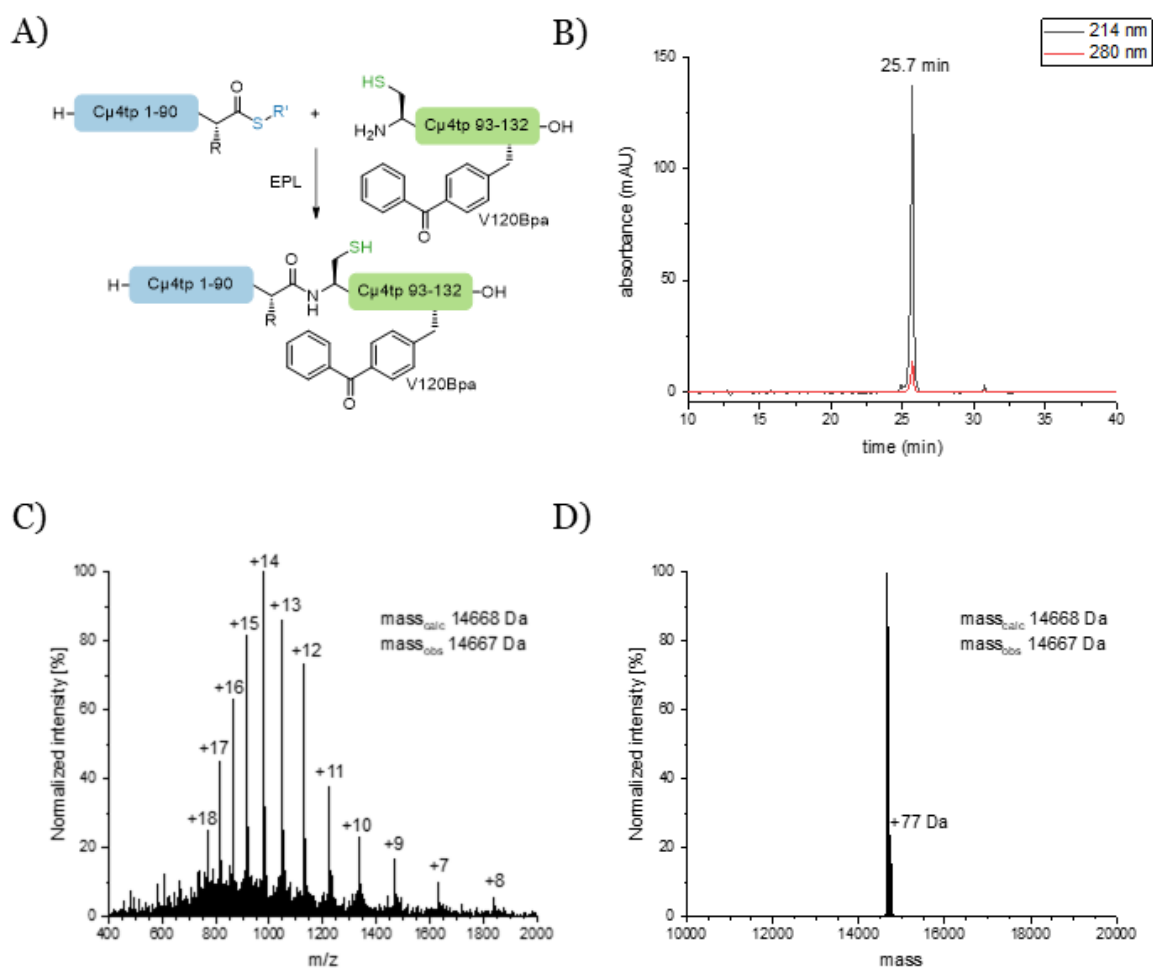


Figure 41: Characterization of Cμ4tp V120Bpa main pool A) Cμ4tp 1-132 V120Bpa structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 14667 Da, MW_{calc} : 14667.63 Da.

9.3.3 Characterization of C μ 4tp Bpa variant folding behaviors

Tuan Nguyen performed the folding and oligomerization analysis of semisynthetic C μ 4tp at the Technical University of Munich. C μ 4tp variants were folded as outlined in section 9.2. These folded variants were investigated using SEC, CD spectroscopy and thermal transition, as detailed in section 8.9.

Figure 42 displays the UV-VIS traces of folded C μ 4tp variants during SEC analysis. The observed peaks can be divided into three categories with increasing elution time: oligomer, dimer, and monomer. All variants display similar folding behavior compared to recombinantly expressed C μ 4tp, although with different ratios in the formation of oligomers and dimers depending on the variant. We observed significantly reduced oligomer formation for the V120Bpa variant compared to wt C μ 4tp. While we hypothesized that our substitutions in the tailpiece could negatively affect the formation of oligomers, through the disruption of hydrophobic interactions between the two β sheets at the center of the C μ 4tp oligomer as described by Li *et al.*¹¹⁴ and in alignment with previous results by Pasalic *et al.*⁷⁶, we did not expect significant differences in the formation of dimers. There is also a shoulder present at the monomer peak, possibly indicating the presence of minor amounts of hydrolyzed N-terminal segment, which may have affected the folding behavior of this variant.

Consequently, we expected similar behavior of the Y118Bpa variant, as the native Tyr118 appears to serve a similar function in the cryo-EM structure presented by Li *et al.* Interestingly, we observed the opposite. Y118Bpa strongly favored the formation of oligomers. It should be noted that the variant also presents two species in the elution window labeled “monomer”. These may correspond to contamination, or incomplete folding.

L117Bpa displays a similar pattern of oligomer / dimer / monomer peaks. However, the data provided on this variant was measured on a different SEC column with a different flow rate (0.5 ml/min as opposed to 0.75 ml/min). It is unclear why this variant displays an additional signal in the oligomer range.

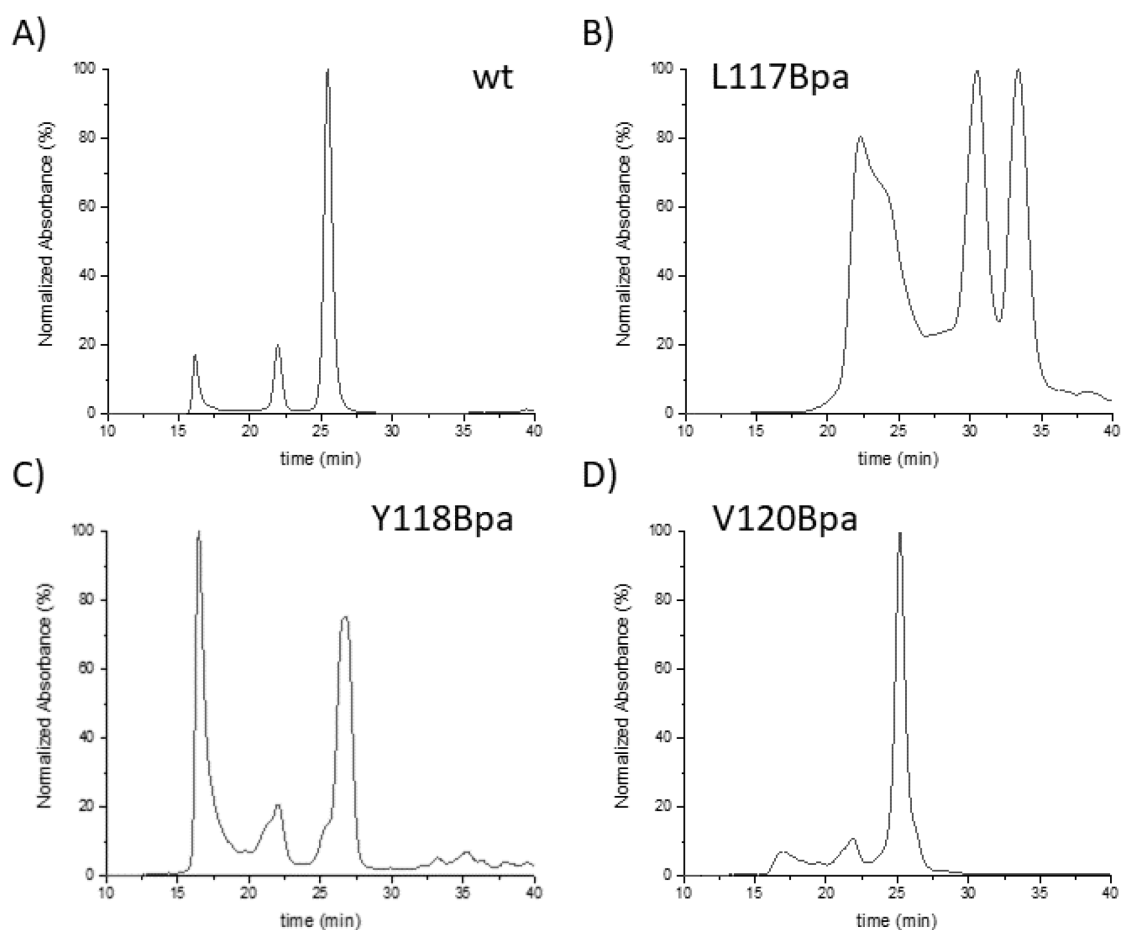


Figure 42: UV-VIS traces of folded $C\mu 4tp$ variants during analytical SEC.

As shown in Figure 43, our Bpa-carrying $C\mu 4tp$ variants display very similar CD spectra to wt $C\mu 4tp$ after folding. The broad valley at ~ 220 nm indicates formation of β -sheets, which are indicative of the immunoglobulin fold of $C\mu 4$. The increased level of noise in the CD spectrum

of E) Y119Bpa can likely be attributed to the low concentration of the protein during measurement.

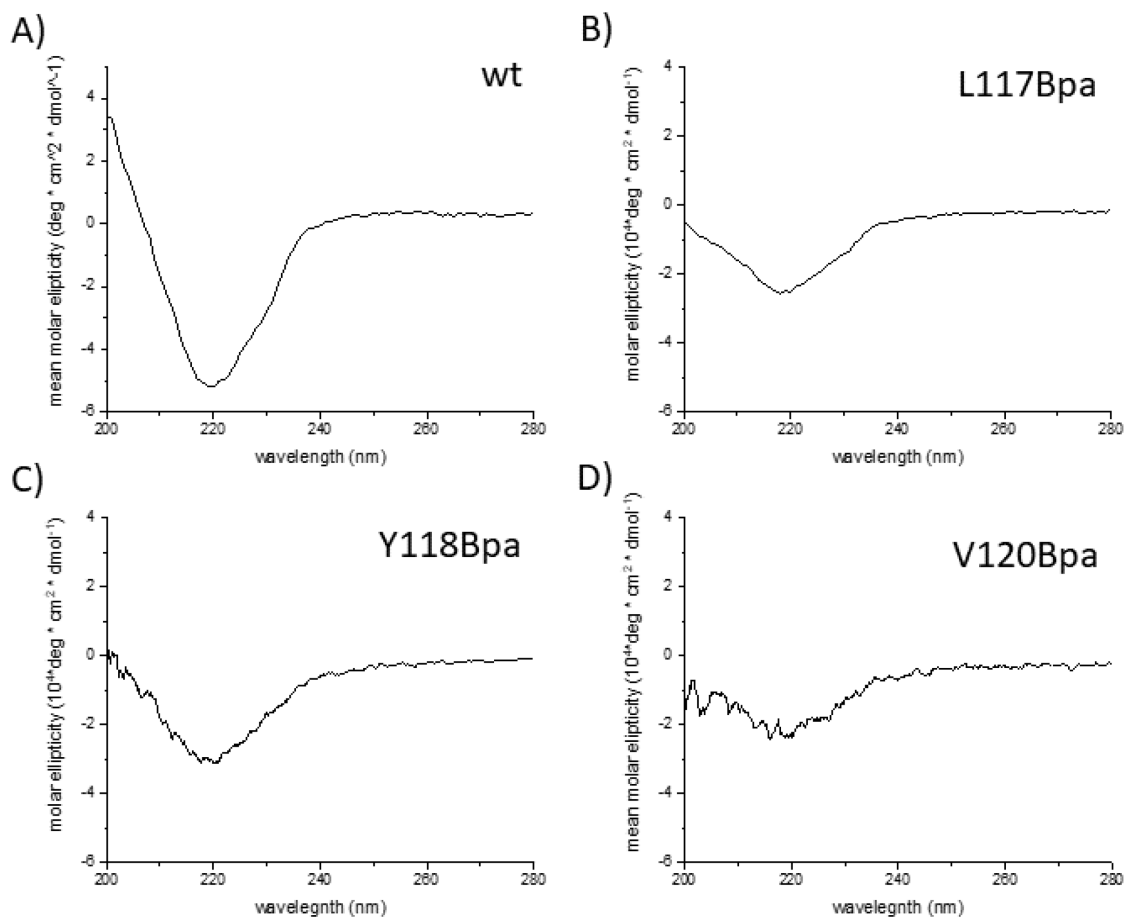


Figure 43: CD spectrum of $C\mu 4tp$ wt and variants. A) wt; B) L117Bpa; C) Y118Bpa; D) V120Bpa.

Figure 44 displays thermal transition data of our $C\mu 4tp$ variants compared to $C\mu 4tp$ wt. V120Bpa D) shows increased levels of noise when compared to $C\mu 4tp$ wt or other variants, which may be a concentration-dependent effect. All $C\mu 4tp$ variants display similar thermal transition behavior, indicating stable folding. We therefore concluded that our $C\mu 4tp$ variants display stable, similar folding compared to $C\mu 4tp$ wt. Both of which are prerequisites for subsequent photocrosslinking experiments.

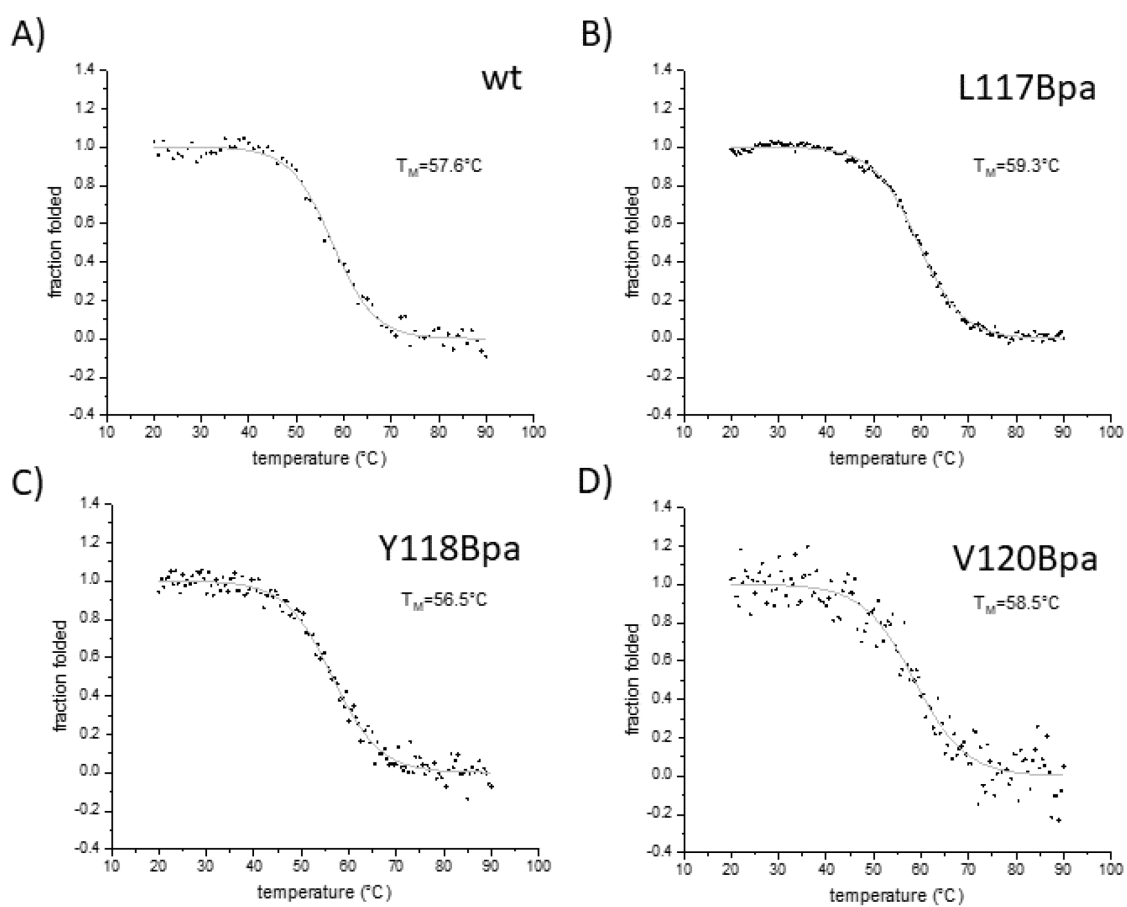


Figure 44: Thermal transition data of $C\mu 4tp$ wt and variants. A) wt; B) L117Bpa; C) Y118Bpa; D) V120Bpa.

9.3.4 Photoaffinity labelling of C μ 4tp Bpa variants

Preliminary photoaffinity labelling experiments of C μ 4tp variants were performed by Tuan Nguyen at the Technical University of Munich as described in section 8.10. The corresponding samples were analyzed *via* SDS-PAGE as shown in Figure 45.

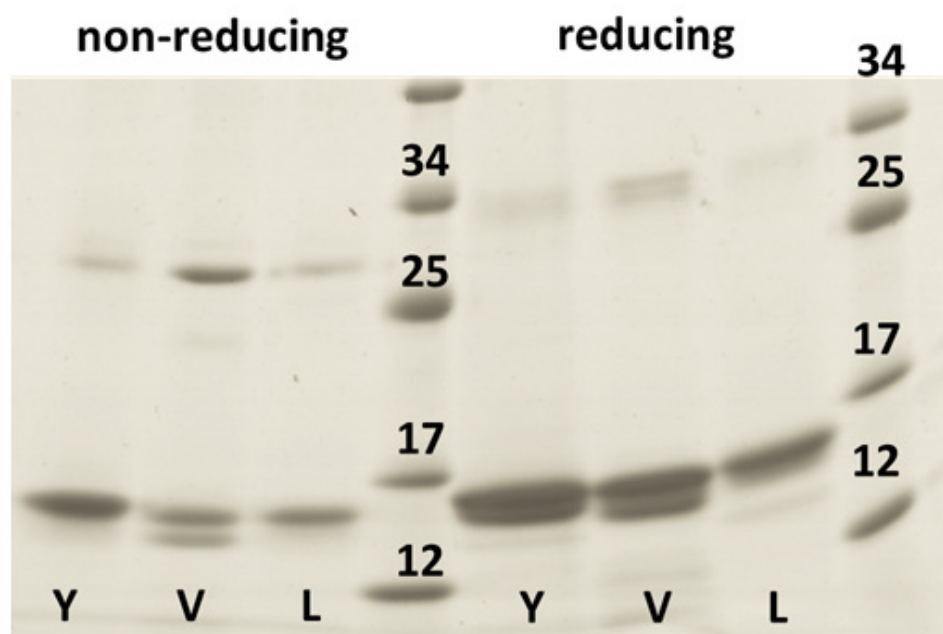


Figure 45: SDS-PAGE analysis of C μ 4tp variants subjected to photocrosslinking.

The non-reducing conditions clearly show the formation of dimers with a MW of ~29 kDa in addition to large amounts of monomer at ~15 kDa. This agrees with the SEC data shown in Figure 42. The nature of the additional bands between 12 and 17 kDa for Y118Bpa and V120Bpa remains unclear. We were able to identify an additional mass of 13437 Da in MS analysis of the V120Bpa variant post crosslinking. We were unable to identify any additional masses during similar MS measurements of the Y118Bpa variant. We believe this observation may have been caused through degradation of the protein.

Under reducing conditions, these dimers are expected to disappear, unless held together by successful crosslinking. While variants L117Bpa and Y118Bpa display only very faint bands, V120Bpa shows two distinct bands in the 25 to 34 kDa range, indicating successful photocrosslinking. The presence of a second band in the monomer region of Y118Bpa and V120Bpa may be the result of incomplete reduction of the internal disulfide bond at the center of the immunoglobulin fold, which is protected by the surrounding β sheets and may thus be difficult to access for reducing agents.

However, it must be kept in mind, that these are preliminary results, and no positive or negative controls were performed during our photocrosslinking experiments limiting the conclusions that can be drawn from these experiments so far. As this is an ongoing investigation, future experiments would significantly benefit from controls such as including C μ 4tp without photocrosslinker, and C μ 4tp with photocrosslinker, but without illumination. Additionally, a positive control carrying Bpa might be desirable.

We also subjected the same samples to MS but were unable to identify signals corresponding to crosslinked dimers. The MS data recorded after 20 min of irradiation provided is depicted in Figure 46 to Figure 48 . Increasing the duration of UV irradiation, to 40 min and up to the point of complete solvent evaporation, did not change these findings. Variant Y118Bpa (Figure 47) shows mass of 14905 Da in the monomer, as opposed to 14604 Da, which may correspond to quantitative glutathionation (+305 Da). This hypothesis is further supported by the absence of this adduct in the dimer, as glutathionation prevents dimerization.

Results and Discussions

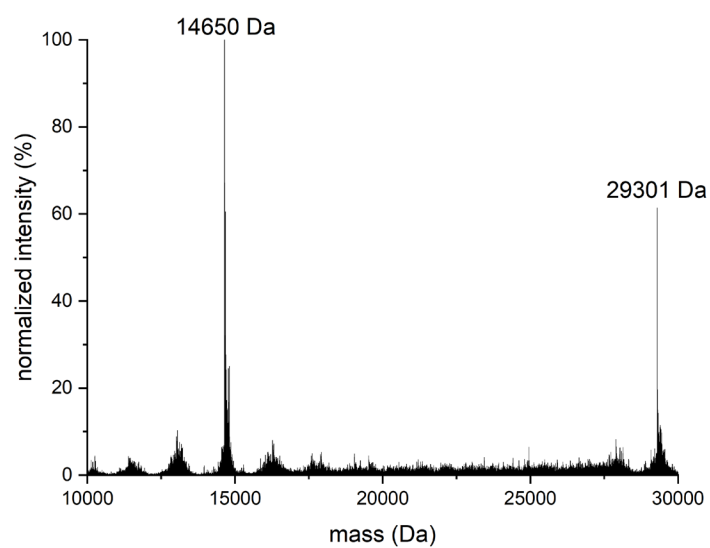


Figure 46: Mass spectrum of variant L117Bpa 20 min after UV irradiation. MW_{obs} : 14650 Da, MW_{calc} : 14654 Da; MW_{obs} : 29301 Da, MW_{calc} : 29298 Da.

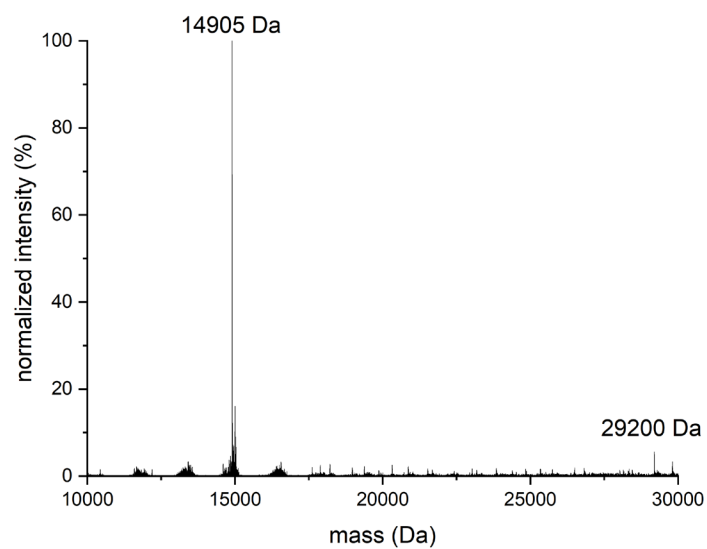


Figure 47: Mass spectrum of variant Y118Bpa 20 min after UV irradiation. MW_{obs} : 14605 Da, MW_{calc} : 14905 Da; MW_{obs} : 29200 Da, MW_{calc} : 29206 Da.

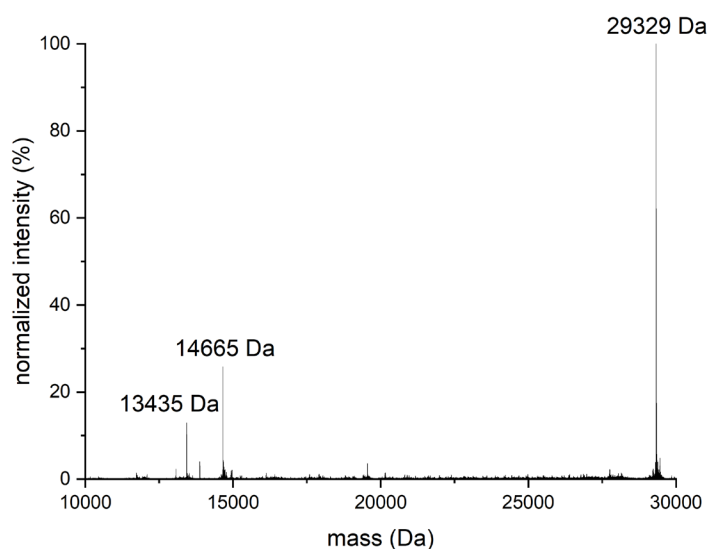


Figure 48: Mass spectrum of variant V120Bpa 20 min after UV irradiation. MW_{obs} : 14665 Da, MW_{calc} : 14668 Da; MW_{obs} : 29329 Da, MW_{calc} : 29333 Da.

9.3.5 Discussion of the semisynthesis and characterization of photocrosslinking C μ 4tp variants

9.3.5.1 Successes and challenges during semisynthesis of C μ 4tp Bpa variants

We successfully produced three of the four planned C μ 4tp variants featuring a Bpa substitution within or near the β -sheet region of the tailpiece. However, synthesis of the C μ 4tp Y132Bpa variant proved challenging and could not be completed over the course of this thesis.

Major complications arose already during SPPS of the C μ 4tp 92-132 Y132Bpa peptide. We assume that these issues resulted from the positioning of the Bpa building block at the C-terminal end of the peptide. Initially, we hypothesized that our procedures for attaching this building block to the resin were inadequate. We therefore performed resin loading and

synthesis on both Wang and 2-Chlorotrityl chloride resins, strictly following manufacturers guidelines. Resin substitution values were determined gravimetrically and photometrically, and aligned well with data from previous, successful SPPS projects within our research group. We further monitored the synthesis progress with regular test cleavages accompanied by LCMS analysis. The results obtained did indeed initially present the desired segment of the peptide with an acceptable purity. Yet, upon continuation of the synthesis, these results changed and started to show very uncharacteristic masses matching fragmentation or cyclization. Analytical HPLC data confirmed the presence of multiple distinct and overlapping peptides, which is a strong indicator against fragmentation in the mass spectrometry system as cause for the observed masses. Contrasting these results with our successful synthesis of other C_μ4tp Bpa variants, we assume that this behavior is not simply caused by the presence of Bpa in the sequence but results from the position of Bpa at the very C-terminus of the sequence.

Overall, both Boc, manual-, and automatic Fmoc synthesis of the remaining variants worked well. While our protocol for Boc synthesis allowed for faster synthesis, changing to Fmoc-SPPS also allowed for the successful synthesizing of the desired variants. As an added benefit, Fmoc-SPPS allowed for efficient test cleavages during synthesis.

Working with photosensitive materials, such as photocrosslinkers, comes with inherent challenges. While all Bpa-containing peptides were synthesized in syringes wrapped in aluminum-foil and kept away from light sources by working under minimal lighting conditions, it is difficult to guarantee the complete exclusion of ambient light, especially when working in busy environments. These procedures did not change between C_μ4tp Bpa variants

and are therefore most likely not to be blamed for our failure to synthesize C μ 4tp 92-132 Y132Bpa. Additionally, we performed literature research on C-terminal Bpa peptides but could not find evidence pointing toward increased susceptibility of C-terminal Bpa to photoirradiation or fragmentation of the attached peptide. Nonetheless, I recommend the use of tinted syringes in addition to the use of aluminum-foil wrapping during SPPS to further decrease the risk of exposure to ambient light when working with photosensitive materials, as well as the use of red light for lighting the work area.

All Bpa variants faced similar challenges during separation of the ligation product from the unreacted N-terminal segment encountered with the wt (detailed in section 9.2). Overall, we successfully designed and established a route for the semisynthesis of the IgM C μ 4tp domain that allows for the selective incorporation of Bpa into the C μ 4 tailpiece.

9.3.5.2 Folding and photocrosslinking of C μ 4tp Bpa variants

Two out of three C μ 4tp Bpa variants mirrored the folding behaviors of semisynthetic and recombinant, wildtype C μ 4tp, with minor differences in the ratio of monomer to dimer to oligomer between the different variants. The exception to this is variant V120Bpa, which displayed reduced formation of oligomers and dimers compared to the other variants and the wt. The residues between Tyr118 and Met124 in each tail piece form beta strands, which pack together into two antiparallel five-stranded beta sheets that stack on top of each other¹¹⁴. Our V120Bpa variant showing reduced oligomer formation is likely due to interference with the hydrophobic interactions between these beta sheets, in agreement with this structure and previous reports by Pasalic *et al.*⁷⁶. We did not observe the same folding behavior for the

Y118Bpa variant, instead this variant displayed a strong tendency to form oligomers. We also observed the formation of oligomers, dimers, and monomers for the L117Bpa variant, however due to different conditions during SEC analysis, we hesitate to offer a detailed comparison with the other variants.

Interestingly, no significant photocrosslinking was observed for any of the variants, except for marginal photocrosslinking of the C μ 4tp V120Bpa dimer in gel electrophoresis, which could however not be confirmed *via* MS data. Additionally, V120Bpa displayed reduced oligomerization and dimerization behavior compared to wt C μ 4tp in SEC analysis. We are therefore hesitant to draw strong conclusions from these observations. This is especially true for variant Y118Bpa, which displayed seemingly quantitative glutathionation of the monomer. It remains unclear why this was only observed for this variant, but we hypothesize this to be caused by extended storage of the protein in presence of folding buffer. It can't be ruled out that this glutathionation also affects folding and crosslinking behavior. We also stress, that these are preliminary results, and future experiments would significantly benefit from including additional controls, such as including C μ 4tp without photocrosslinker, and C μ 4tp with photocrosslinker, but without illumination. Additionally, a positive control carrying Bpa might be desirable.

Overall, these findings do not provide definitive answers, but raise further questions about the structure of IgM and its folding mechanism. Additionally, they highlight the importance for controls, especially in experiments yielding negative results.

9.4 Towards the semisynthesis of C μ 4tp glycovariants

We decided to further expand our semisynthesis efforts of C μ 4tp to encompass modifications of the conserved N-glycosylation motive in position Asn119 with the primary goal of furthering our understanding of the role of this motive in C μ 4tp structure and oligomerization. This glycosylation site formed by Asn119 and Ser121 appears to be essential in preventing IgM aggregations¹¹⁴, as mouse N563A and S565 mutants lacking glycan moieties are secreted as hexamers and higher MW species^{76,162}. First, we wanted to create a minimally glycosylated variant carrying a simple N-acetyl glucosamine on the respective asparagine residue. We concluded that the most straightforward method to achieve this goal would be the incorporation of the GlcNAc modification during SPPS in the form of a commercially available Fmoc-AsnGlcNAc building block.

Additionally, we wanted to explore the transfer of more complex glycan structures onto our semisynthetic C μ 4tp, and to monitor their effects on C μ 4tp folding. Specifically, we wanted to focus on the potential of SGP as glycan donor and our C μ 4tp AsnGlcNAc variant as glycan acceptor for transglycosylation reactions. As previously explained in section 8.5, endo-CC N180H displays strong transglycosylation activity when transferring glycan structures from SGP to other peptides containing N-acetyl glucosamines linked to asparagine and asparagine-like residues^{219,220}. We therefore hoped to use the above-mentioned minimally glycosylated C μ 4tp AsnGlcNAc variant as starting point for our transglycosylation experiments. Previous projects in our research group had had success with isolation, trimming, and chemical activation of the SGP glycan structure prior to enzymatic transfer of the glycan onto new targets²²⁸. For our C μ 4tp semisynthesis project, we asked which benefits

and limitations could be observed when performing direct enzymatic transfers of the SGP glycan onto C μ 4tp AsnGlcNAc.

This chapter presents our results regarding the semisynthesis of C μ 4tp AsnGlcNAc and the challenges encountered during the purification of the final product *via* RP-HPLC. We furthermore show and discuss our progress towards the semisynthesis of more complexly glycosylated C μ 4tp variants. We also present and compare two different strategies for isolating SGP from egg yolk powder. Finally, we put these results into context and discuss implications for future semisynthesis projects involving C μ 4tp and/or SGP glycans.

9.4.1 Synthesis of C μ 4tp 92-132 N119AsnGlcNAc

Synthesis of the C μ 4tp 92-132 N119AsnGlcNAc peptide was performed with both manual and automated Fmoc-SPPS using tentagel resin preloaded with Tyr as outlined in section 8.2, as Boc-SPPS protocols can negatively affect the stability of the GlcNAc building block, by cleaving the glycosidic bond. During automated SPPS the AsnGlcNAc building block itself was also coupled manually as follows: The automated synthesis was stopped at the respective position and then continued via the synthesizer after a manual coupling followed by a test cleavage. We chose not to automatically couple the building block because this would require a significant amount of amino acid to work with the machine. Instead, we opted for a more cost-effective strategy due to the limited availability of the Fmoc-AsnGlcNAc building block.

This building block required an additional deprotection step to remove the acetyl protecting groups on GlcNAc. We initially performed this deprotection

using sodium methoxide in methanol as described in section 8.2.2, but later switched to incubating the protected peptide at pH 11 for 2 h.



Figure 49: Amino acid sequence of C_μ4tp 92-132 N119AsnGlcNAc. X = AsnGlcNAc building block. MW 4597.22 Da.

Overall, both synthesis strategies were successful, yielding 21 mg of peptide for the manual synthesis performed at a 0.05 mmol scale and 49 mg of peptide for the automated synthesis at a scale of 0.05 mmol after Ac-deprotection and RP-HPLC purification.

Figure 50 displays the results of our sodium methoxide deprotection strategy. The desired product was detected with a MW of 4599 Da which matches the expected mass of 4597 Da when accounting for a ± 2 Da accuracy. The major product observed at a mass of 4611 Da, or a mass difference of +14 Da, may correspond to a methylation of the desired product. This could have occurred during the deprotection of the sugar residue with sodium methoxide, though the exact mechanism is unclear. The mass observed at +111 Da when compared to the desired product most likely represents an adduct to the 4611 Da peptide, and therefore an actual mass difference of +98 Da, which may correspond to a not fully removed tBoc protecting group as observed in previous synthesis of C_μ4tp 92-132 peptides.

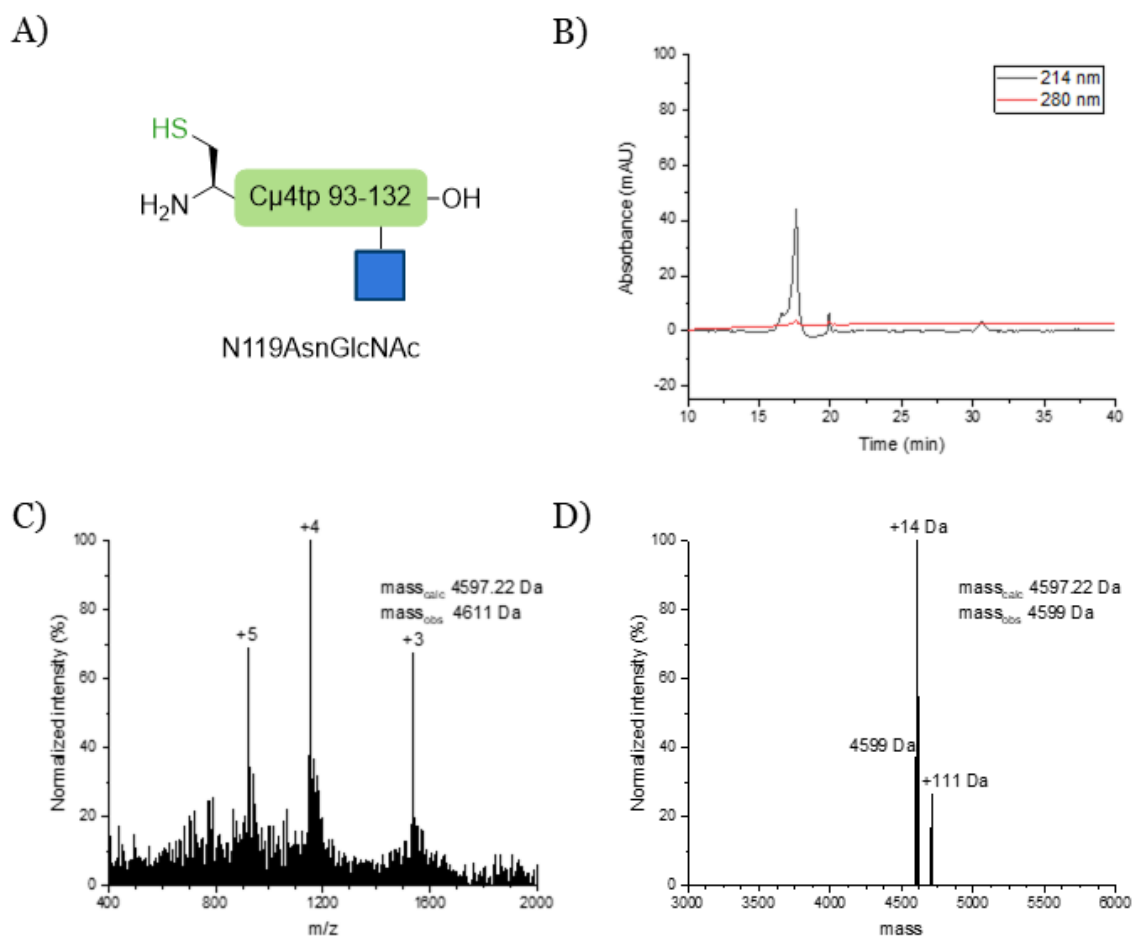


Figure 50: Characterization of manually synthesized N119AsnGlcNAc N119AsnGlcNAc, deprotected with sodium methoxide, via MS and analytical HPLC. A) Cμ4tp 92-132 N119AsnGlcNAc structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 4599 Da and 4611 Da, MW_{calc} : 4597.22 Da.

Figure 51 displays the results of our deprotection strategy without sodium methoxide. The desired product was detected with a MW of 4599 Da which matches the expected mass of 4597 Da when accounting for a ± 2 Da accuracy. Performing Ac-deprotection at pH 11 without sodium methoxide did not result in a +14 mass. The +22 Da mass corresponds to a sodium adduct, coming from the sodium hydroxide used to raise the pH during the reaction. The +98

Da mass may correspond to a not fully removed tBoc protecting group as observed in previous synthesis of C μ 4tp 92-132 peptides.

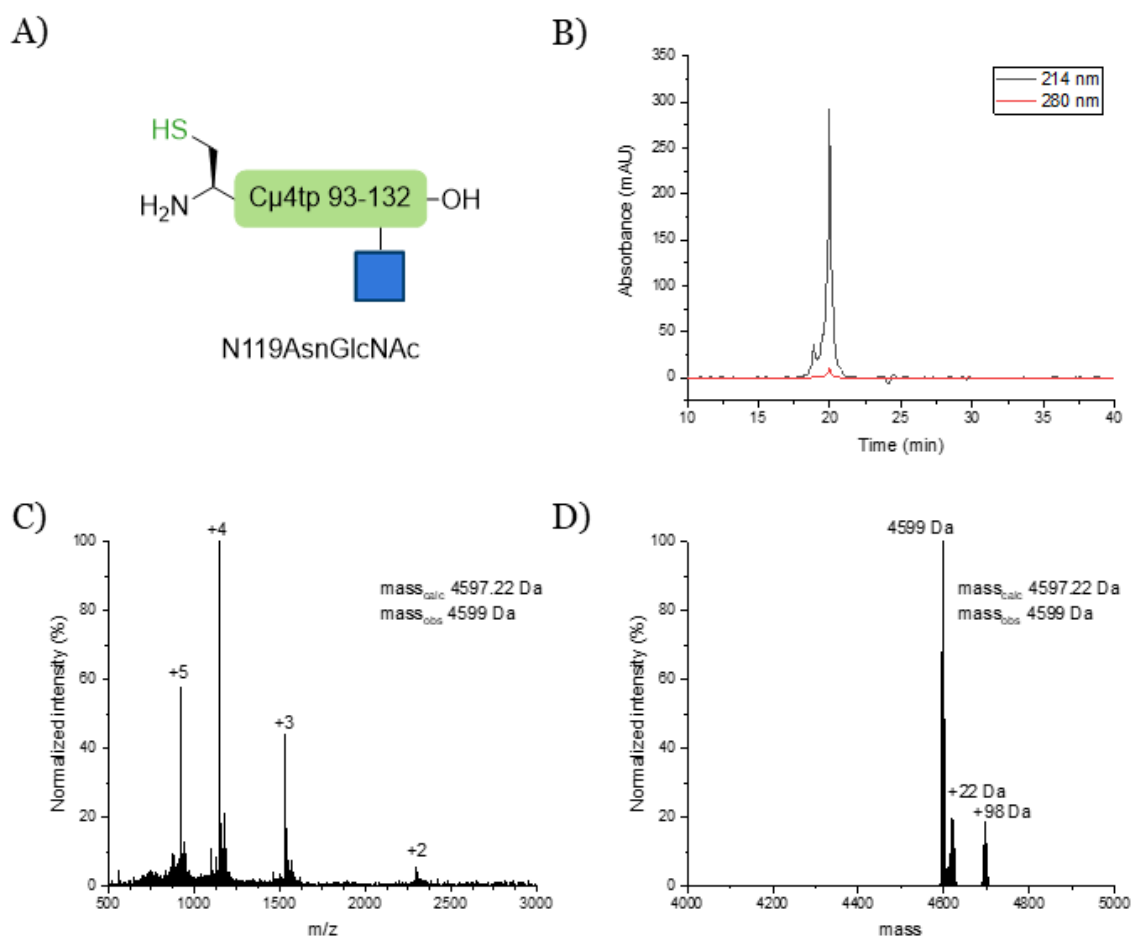


Figure 51: Characterization of automatically synthesized C μ 4tp 92-132 N119AsnGlcNAc, deprotected at high pH, via MS and analytical HPLC. A) C μ 4tp 92-132 N119AsnGlcNAc structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 4599 Da Da, MW_{calc} : 4597.22 Da.

9.4.2 Ligation of Cμ4tp N119AsnGlcNAc

Ligation of Cμ4tp N119AsnGlcNAc has performed analogously to the above-described Bpa variants (section 9.3.2). For further details please refer to section 8.3. The full amino acid sequence of the desired final product is shown in Figure 52.

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYV TSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTXY SLIMSDTG GT CY

```

Figure 52: Amino acid sequence of Cμ4tp N119AsnGlcNAc. X = AsnGlcNAc. Expressed segment highlighted in blue, synthetic segment in green. Expected mass 14718.67 Da.

In total the reaction contained 13 mg Cμ4tp 92-132 N119AsnGlcNAc and 9 mg Cμ4tp 1-91-thioester in 2 ml ligation buffer and was performed overnight. The ligation resulted in 1 mg of main pool (Figure 53) and 1 mg of side pool after purification. This corresponds to a yield of ~2 mg or 15%. Purifying the ligation product *via* RP-HPLC, as outlined in section 8.3.3, proved exceptionally challenging, as the incorporation of GlcNAc reduced the retention time of the ligation product on our columns just enough to cause a significant overlap between the hydrolyzed N-terminal segment and the ligation product.

Figure 53 displays the characterization of the main pool. The desired product was detected with a MW of 14717 Da which matches the expected mass of 14719 Da when accounting for a ± 2 Da accuracy. The +77 Da adduct likely corresponds to a 2-mercaptoethanol adduct when accounting for a ± 2 Da accuracy.

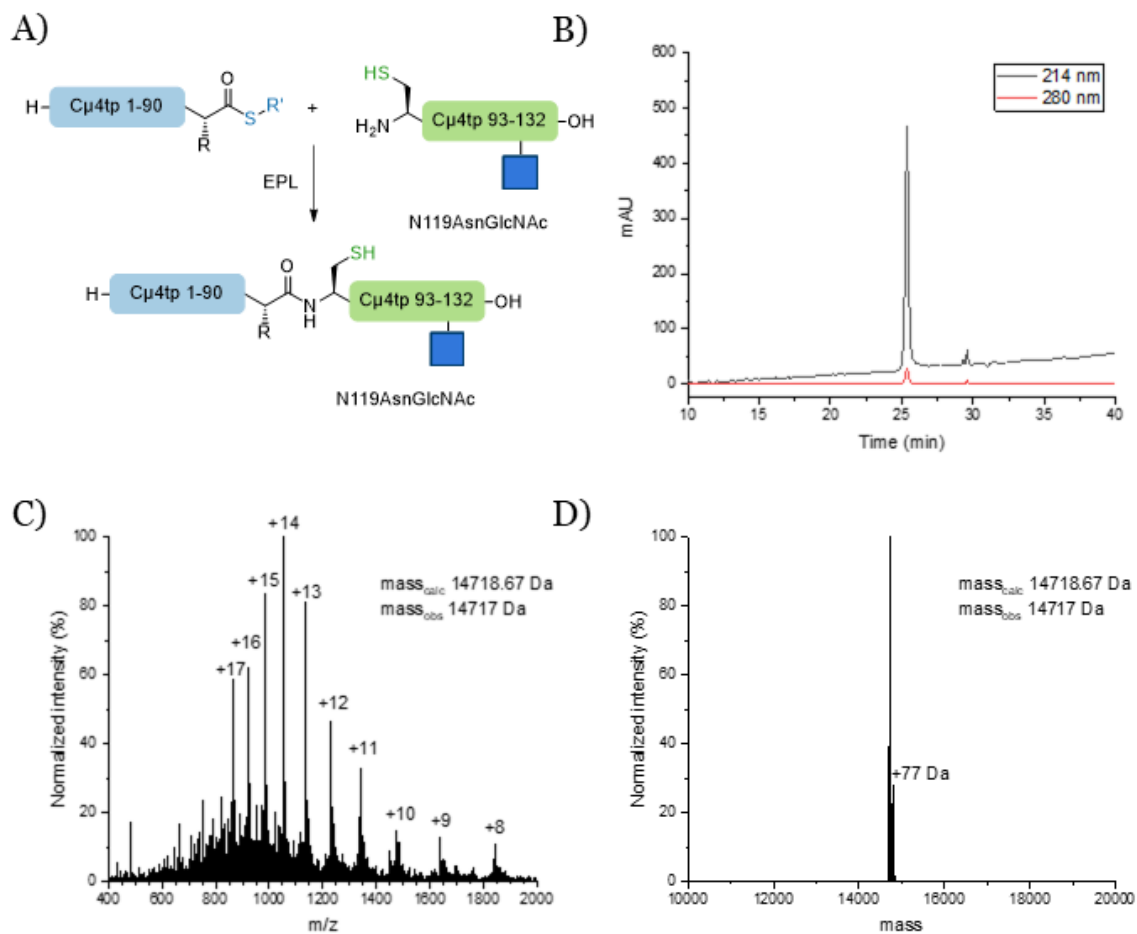


Figure 53: Characterization of Cμ4tp N119AsnGlcNAc main pool 2 via MS and analytical HPLC A) N119AsnGlcNAc structure B) UV-VIS trace of analytical HPLC analysis C) direct injection mass spectrum D) deconvoluted mass spectrum; MW_{obs} : 14717 Da, MW_{calc} : 14718.67 Da.

No folding experiments were performed with these variants, as the project ran out of allocated time. Next, we wanted to investigate the transfer of complex glycan structures on our Cμ4tp 92-132 N119AsnGlcNAc peptide and Cμ4tp 1-132 N119AsnGlcNAc.

9.4.3 Isolation of and transfer of SGP glycan from egg yolk

To support systematic experiments related to the enzymatic transglycosylation of our minimally glycosylated C μ 4tp variant, we needed to obtain significant quantities of SGP. The high cost of commercially available SGP, as well as the low recovery rates of C μ 4tp after ligation of the previously generated C μ 4tp Bpa variants combined with the moderate recovery rates during C μ 4tp folding, motivated us to find efficient and economically viable strategies to isolate SGP from egg yolk on our own.

9.4.3.1 Initial isolation protocol

Initially, we relied on SGP isolation strategies previously established in our research group by Schmid²²⁸ (detailed in section 8.4.1). Briefly, this protocol started with generating egg yolk powder from chicken egg yolks, resulting in approximately 540 g of powder per 30 yolks. This powder was then extracted using diethyl ether and acetone, resulting in about 2.5 g of crude extract. This extract was loaded onto a carbon/celite column and eluted with stepwise increasing concentrations of ACN in H₂O. All fractions were analyzed *via* RP-HPLC. Fractions containing SGP were then fractionated again *via* PD-10 size exclusion chromatography and analyzed *via* RP-HPLC (Figure 54). The additional masses visible in the MS data obtained *via* ESI-MS are not derived from impurities but are caused by fragmentation of the glycan during ionization. Figure 55 depicts the most prevalent of these fragments. For future analysis, we recommend the use of MALDI-MS, which relies on a softer ionization method and thus causes less fragmentation of the glycan, thus simplifying the analysis.

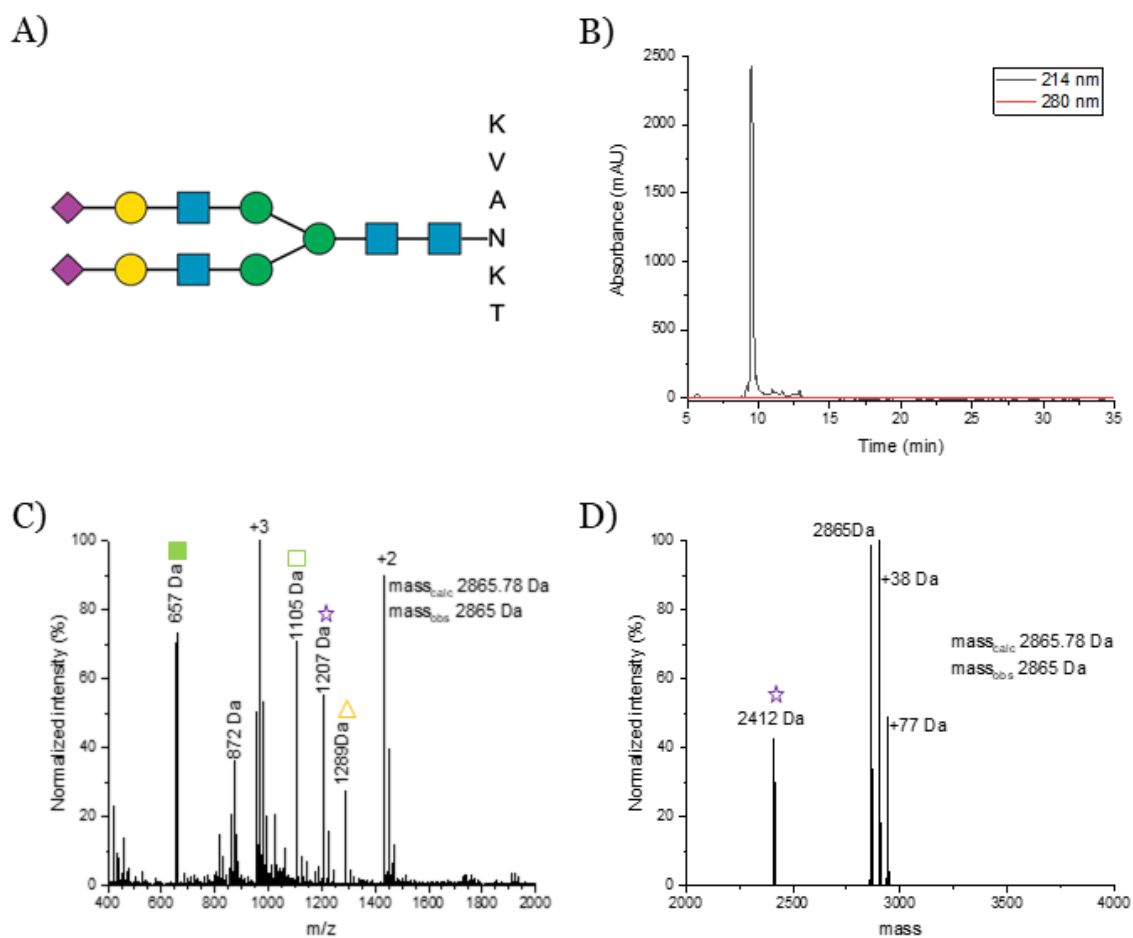


Figure 54: Characterization of SGP isolated with the initial protocol, A) schematic representation of SGP B) UV-VIS trace during analytical RP-HPLC C) MS direct injection D) deconvoluted mass MW_{obs} : 2865 Da, MW_{calc} : 2865.78 Da.

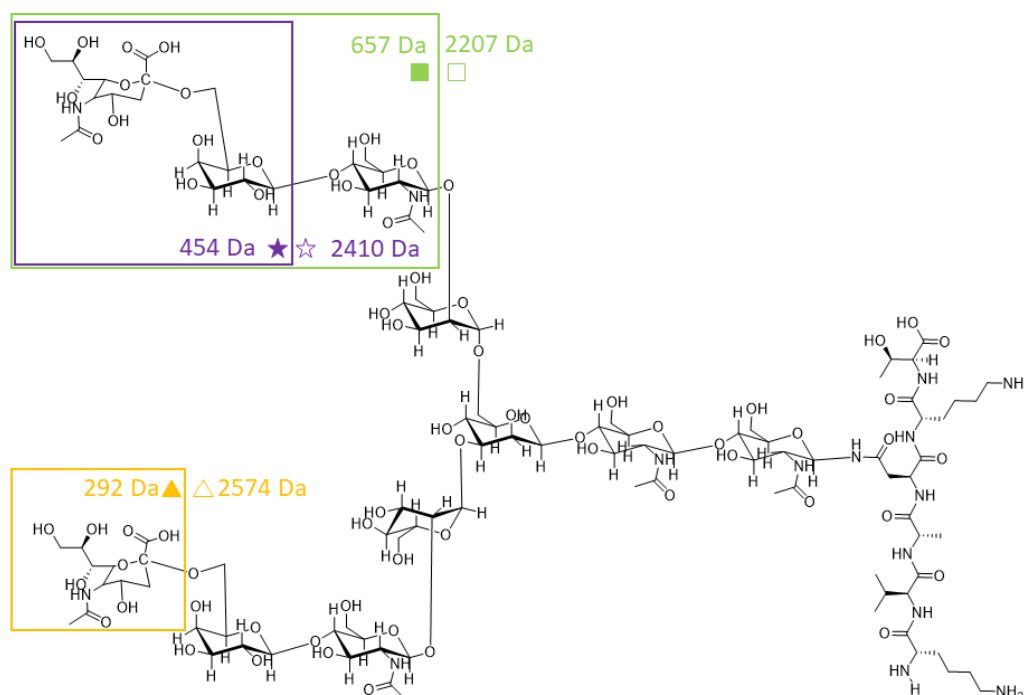


Figure 55: Schematic representation of SGP fragmentation during ESI-MS.

Our initial transfer protocol relied on multiple consecutive enzymatic reactions, to trim the SGP glycan down to a tetra saccharide, and to then perform an oxazoline activation on this tetra saccharide prior to enzymatic transfer onto the target sequence. The exact protocol can be found in sections 8.4.1.4 - 8.4.1.6. Briefly, the glycan was to be trimmed with β -Galactosidase, β -N-Acetylglucosaminidase, and Endoglycosidase, before being separated from the peptide by β -N-Acetylglucosaminidase. This resulted in a GlcNAcMan₃ tetra saccharide to be purified *via* HILIC using a CAD detector on an analytical scale. The tetra saccharide was then to be activated via CDMBI and enzymatically transferred onto the target peptide *via* endo-cc N180H using a large excess of tetra saccharide oxazoline.

9.4.3.2 Discussion of initial SGP isolation protocols

The high cost in time and resources compared to the yield of usable material for enzymatic glycosylation led us to investigate potential alternative strategies. Other publications discussed the reduced hydrolase activity of endo-CC mutant N180H, allowing for efficient transglycosylation reactions using SGP as starting material, thereby circumventing the need for prior enzymatic processing and chemical activation of the SGP glycan structure^{219,220}.

We also concluded that the purity of the isolated SGP was of lesser importance, provided that the contaminating products were highly polar. We based this rationale on our previous results in purifying C μ 4tp variants. While the modification of C μ 4tp with a larger glycan structure may significantly reduce the retention time of the glycosylated protein on a reverse phase HPLC column, we concluded that this shift would not be significant enough to overlap with highly polar compounds left over from our SGP production pipeline. It should be noted, however, that these considerations are only valid if these remaining compounds would not interfere with the transglycosylation performed by endo-CC N180H.

We identified the following bottlenecks in our original SGP isolation strategy:

1. Generating yolk power from whole eggs

Depending on the number of eggs used, this step is quite time-consuming and blocks a significant amount of space on our lyophilizers. While it gives considerable control over the procurement of the eggs in question, such as buying eggs from ethically raised chickens, this step wastes a large amount of egg white. It does not provide a

cost-benefit compared to purchasing egg yolk powder from manufacturers. We therefore opted to rely on commercially available egg yolk powder.

2. Diethyl ether/acetone extraction

This strategy used large amounts of solvents and was performed in a large glass funnel containing a paper filter. This funnel itself proved problematic when working at a larger scale, as increases in the amount of material to filter caused less homogenous flow of solvents through the material. We, therefore, tried to identify strategies with reduced solvent consumption and different separation methods. Ultimately, we settled on a methanol/chloroform solid phase extraction method relying on centrifugation inspired by the work of Alagesan *et al.*²¹⁶ described in section 8.4.2.1.

3. Reliance on consecutive chromatography columns

Our initial protocol relied on the consecutive application of multiple chromatography columns to isolate SGP and corresponding glycan structures. While this approach resulted in high purity of the desired product, it severely limited the throughput and overall yield. This is especially true for the final purification steps after trimming the glycan on analytical HPLC columns, limiting the sample amount to below 1 mg per run. Under the above-outlined assumptions that the efficiency of the transglycosylation reaction by endo-CC N180H as well as the purification of the final product is not severely interfered with by highly polar contaminations left over from SGP isolation, we conclude that such a high purity as provided by our original protocol might not be a necessity.

4. Enzymatic trimming

Opting for a direct transglycosylation reaction from SGP to C μ 4tp does not require prior separation of the glycan structure from SGP. Skipping these enzymatic cleavage steps allows us to reduce the use of precious enzymes and avoid an additional purification step. We also concluded that trimming the glycan structure down to a penta- or tetra saccharide was not necessary to provide proof-of-concept for the enzymatic transglycosylation of C μ 4tp. We thus decided to skip all enzymatic processing of SGP and its glycan structure prior to transglycosylation with endo-CC N180H.

5. Chemical activation

Chemical activation of the glycan was a prerequisite for successfully transferring the tetra saccharide to C μ 4tp. It is not a prerequisite for direct transglycosylation *via* endo-CC N180H from SGP. We'd like to point out that chemical activation also reduces the required excess of glycan during the transfer reaction, as no back transfer from C μ 4tp to SGP can occur if SGP is not present in the reaction to begin with. We hoped to correct for this disadvantage through higher excess of SGP, as demonstrated by Manabe *et al.* ²²⁰.

9.4.3.3 Updated SGP isolation protocol

Based on these considerations, we set out to design an additional protocol for the isolation of SGP in our laboratory. The exact steps are detailed in section 8.4.2.

We directly bought egg yolk powder, thereby reducing our use of lyophilizers and avoiding the disposal of large amounts of egg white. We performed a crude extraction of the powder using methanol/H₂O and chloroform (see section 8.4.2.1). Due to the reliance of large centrifuge buckets and magnetic stirrers, scaling up the process was significantly more convenient than the previous strategy relying on washing the yolk powder in a large funnel. This resulted in approximately 24 mg of brittle, crystalline, yellow material per g of yolk powder. We then compared several methods for further purification of the crude extract, including FPLC , size exclusion using a 300 mm tall Sephadex G50 column under gravitational flow (section 8.4.2.2), preparative HPLC using a C18 column , and carbon/celite solid phase extraction (section 8.4.2.3).

Due to low flow rates of the gravitational SEC, no separation or large-scale purification could be accomplished. Consequently, we decided to switch from gravitational flow to FPLC. However, despite achieving higher flow rates using SEC on FPLC, we were unable to purify larger quantities. This was owed to low loading volume and limited column capacity of our setup, which was designed to primarily purify semisynthetic proteins at lower scales. We also attempted RP-HPLC purification on preparative C18 columns with linearly increasing ACN concentrations in H₂O. While this approach allowed for larger throughput than the previous strategies, we were unsatisfied with the amount of material recovered. We hypothesize that due to the hydrophilic nature of the SGP, a significant percentage of material was lost in the flow through.

Eventually, we settled on fractionating the crude SGP over a series of carbon/celite solid phase extractions using increasing concentrations of ACN in H₂O, followed by analysis of the

fractions *via* MS direct injection. This approach allowed for a larger throughput than the use of carbon/celite columns. On average, this method yielded ~2.3 mg of extract per g of egg yolk powder. We then ran the sample through a PD10 column, resulting in ~1.1 mg of SGP extract per g of egg yolk (Figure 56). Overall, the new protocol resulted in a yield of 0.35 μmol SPG per g of egg yolk powder compared to 0.16 μmol of SGP obtained with our original strategy. Both samples display a high purity of over 90%. The additional masses visible in the MS data obtained *via* ESI-MS are not derived from impurities but are caused by fragmentation of the glycan during ionization. Figure 57 depicts the most prevalent of these fragments. For future analysis, we recommend the use of MALDI-MS, which relies on a softer ionization method and thus causes less fragmentation of the glycan, thus simplifying the analysis.

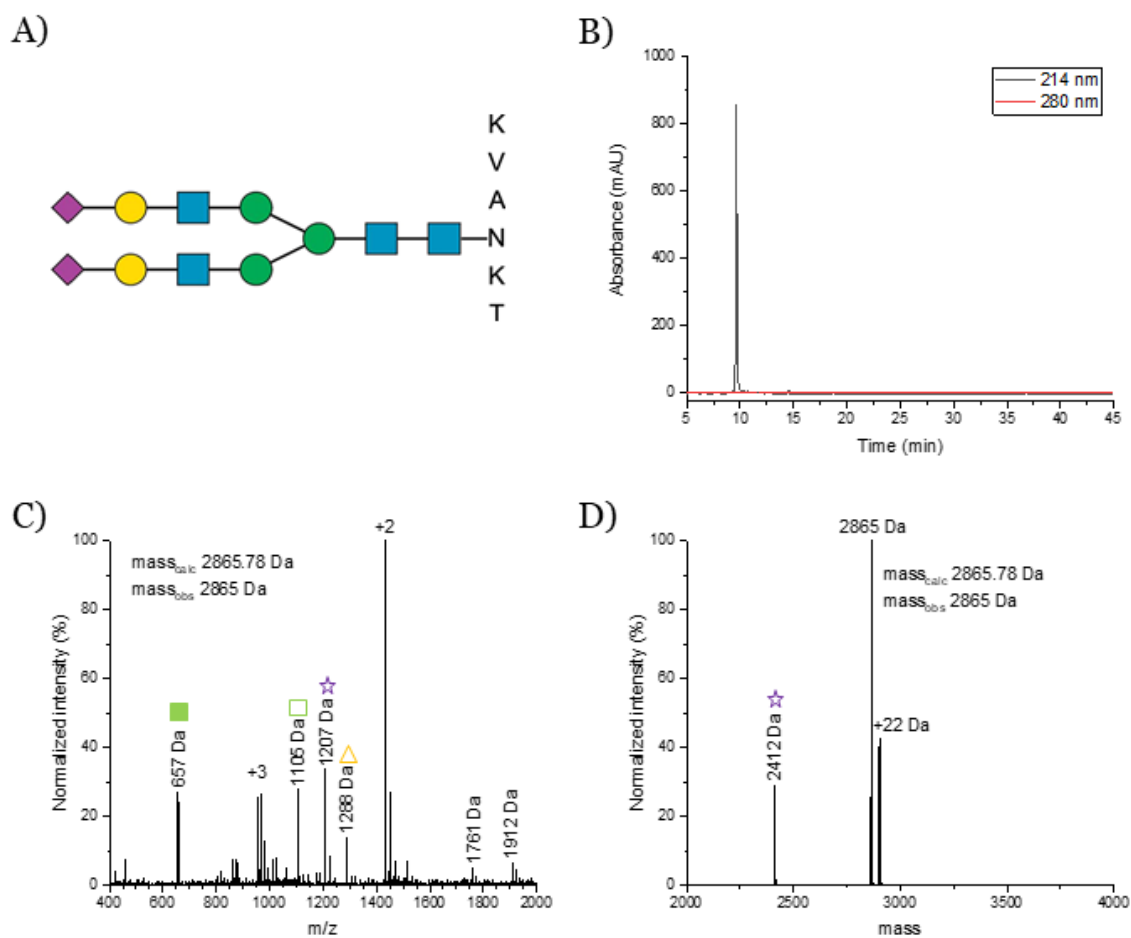


Figure 56: Characterization of SGP isolated with the updated protocol, A) schematic representation of SGP B) UV-VIS trace during analytical RP-HPLC C) MS direct injection D) deconvoluted mass MW_{obs} : 2865 Da, MW_{calc} :2865.78 Da.

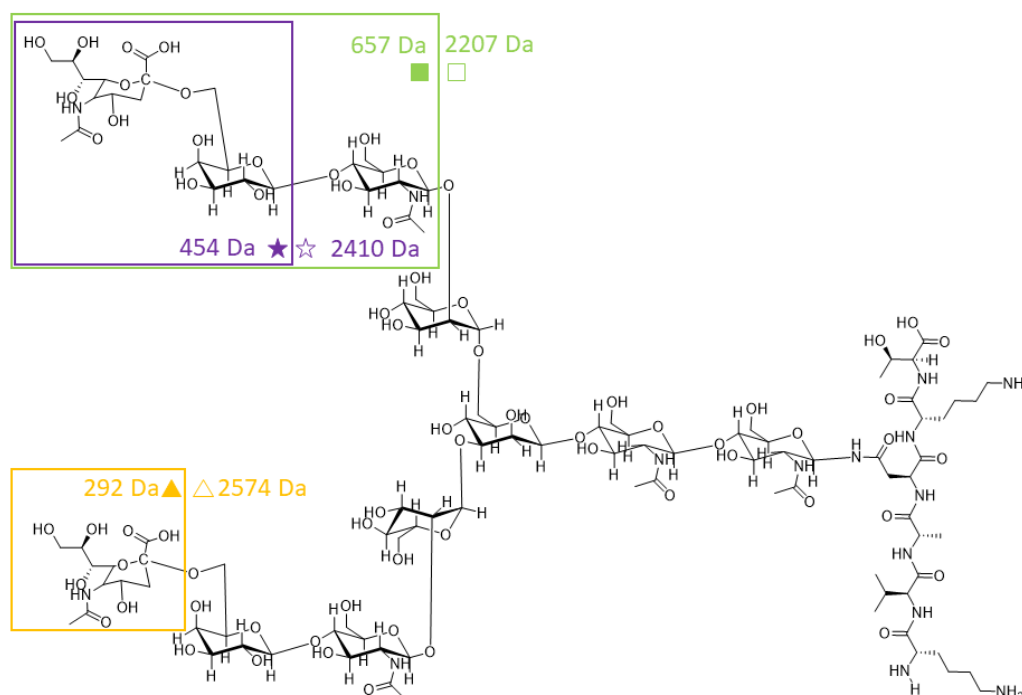


Figure 57: Schematic representation of SGP fragmentation during ESI-MS

9.4.4 Direct, enzymatic transfer of the SGP glycan onto C μ 4tp 1-132 N119NGlcNAc

We attempted to transfer the SGP glycan onto our minimally glycosylated variant C μ 4tp 1-132N119NGlcNAc but were unable to overcome solubility issues of the target protein under the reaction conditions required by the enzyme. Conditions for this enzymatic transfer are described in detail in section 8.5. The reaction was performed with a ratio of C μ 4tp 1-132 N119NGlcNAc to SGP of 1:25. Addition of SGP in reaction buffer caused immediate precipitation of C μ 4tp 1-132 N119NGlcNAc. During the reaction, we could neither observe the starting material nor the desired product in the supernatant *via* MS direct injection.

We therefore decided to attempt the transfer of the glycan onto the C μ 4tp 92-132 N119GlcNac peptide which showed improved solubility compared to the full length C μ 4tp domain.

9.4.5 Direct, enzymatic transfer of the SGP glycan onto C μ 4tp 92-132 N119GlcNac

We transferred the SGP glycan directly onto C μ 4tp 92-132 N199NGlcNa. Conditions for this enzymatic transfer are described in detail in section 8.5. The reaction was performed with a ratio of C μ 4tp 92-132 N199NGlcNac to SGP of 1:25. While the transfer was not complete, we estimate a ratio of 1:1 of the C-terminal segment carrying the glycan structure and not carrying the glycan structure, based on peak size during HPLC purification. While an incomplete glycosylation of the starting material is expected by this strategy, this relatively low conversion rate may also result from precipitation of the starting material under transfer conditions. Purification of the product *via* semi-preparative RP-HPLC proved challenging; we therefore opted to pool all fractions containing the C-terminal segment carrying the glycan structure and not carrying the glycan structure and purify these *via* RP-HPLC on an analytical scale.

While this strategy did increase the purity of our product, characterization of the product *via* MS direct injection revealed the presence of significant amounts of starting material, even though the UV-VIS trace of our analytical HPLC suggested a satisfactory purity (Figure 58). It is also possible that the starting material observed in this sample is caused by fragmentation of the glycan during ESI. The desired product was detected with a MW of 6596 Da, which matches the expected mass of 6600.02 Da when accounting for a ± 3 Da accuracy. Three additional masses were observed at 4594, 4691, and 4731 Da, respectively. These correspond

to the minimally glycosylated peptide (4594 Da), plus 97 Da (minimally glycosylated peptide +tBoc protecting group when accounting for a ± 3 Da accuracy) and 137 Da not previously observed during synthesis. The overall yield of this transfer was approximately 0.4 mg or 27 %. As the correct product could not be separated from the starting material, we did not perform a ligation with this peptide.

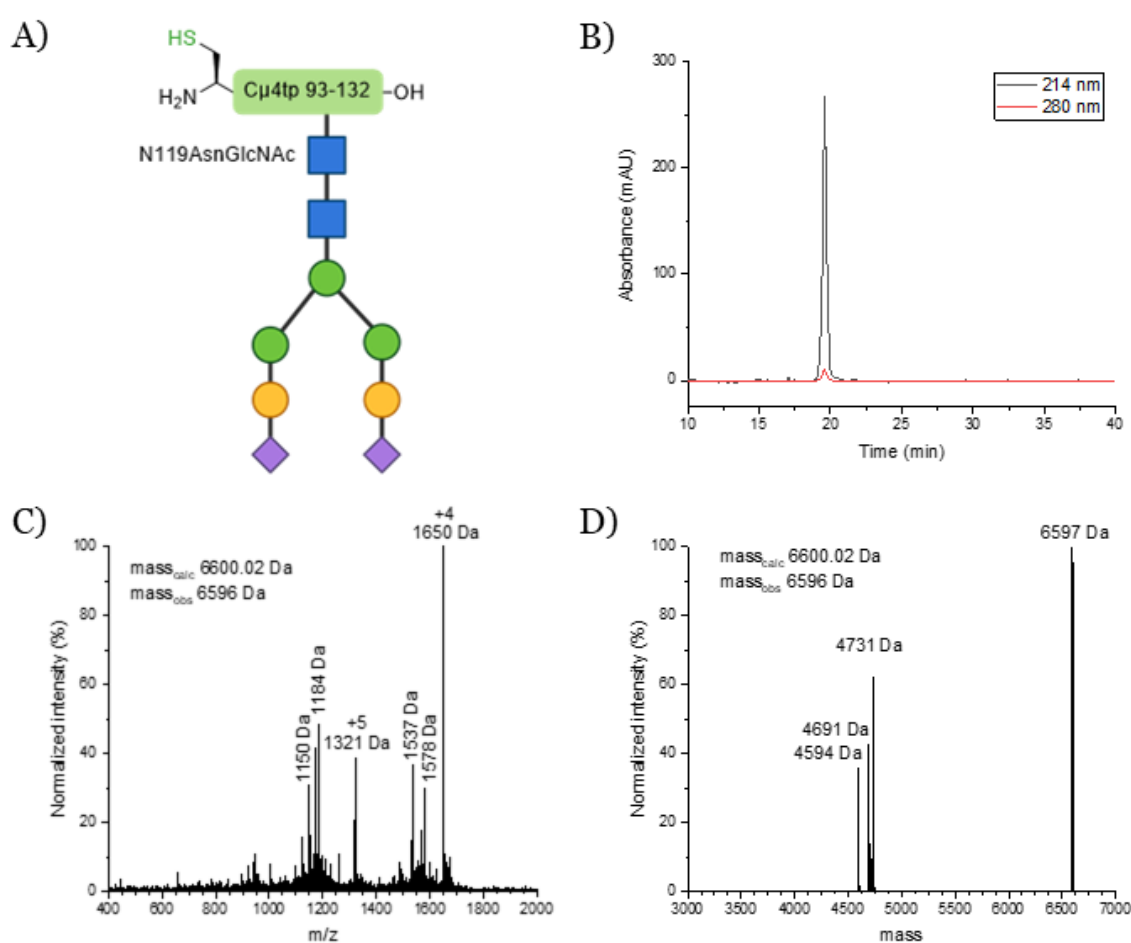


Figure 58: Characterization of the glycosylated Cu4tp 92-132 after direct transfer. A) schematic structure B) UV-VIS trace of analytical RP-HPLC C) direct injection mass spectrum D) deconvoluted mass spectrum MW_{obs} : 6597 Da, MW_{calc} : 6600.02 Da.

9.4.6 Discussion of the semisynthesis of glycosylated C μ 4tp variants

9.4.6.1 Semisynthesis of C μ 4tp N119AsnGlcNAc

We successfully synthesized peptides carrying a GlcNAc modification at the conserved N-linked glycosylation site in position N119 of C μ 4tp. Purification of the ligation product proved once again challenging due to coelution with unreacted N-terminal segment and low recovery rates after RP-HPLC. Further optimization of the ligation reaction to improve the ratio of hydrolyzed starting material to product is generally desirable.

9.4.6.2 SGP isolation

Comparing the resource consumption of the isolation strategy established by Schmid²²⁸ and the glycosylation protocol presented in this thesis, the new protocol resulted in a yield of 0.35 μ mol SPG as opposed to an average of 0.16 μ mol obtained per g of egg yolk through our initial strategy. This aligns well with yields previously reported by Sun²²⁷, Liu²¹⁷, and Alagesan²¹⁶. Both samples display a similarly high purity of over 90% in analytical HPLC. As highlighted by numerous publications on the isolation of SGP^{215-217,227}, there are several ways to reach the desired destination, and choosing the “ideal” strategy depends more on the availability of equipment and resources.

9.4.6.3 Comparing Oxazoline activation and direct transfer of N-glycans

Our research group had previously relied on trimming down the SGP glycan to a pentasaccharide core and then separating it from the SGP hexapeptide *via* enzymatic cleavage of the β 1-4 glycosidic bond between the two N-acetylglucosamines of the N,N'-diacetylchitobiose core of the glycan²²⁸. The resulting tetrasaccharide must be converted into

an oxazoline to allow enzymatic transfer onto the target peptide or protein²²⁸. A disadvantage of this method is the reactivity of the glycosyloxazoline, which can be attacked by reactive amino groups on side chains when reaction conditions are not tightly controlled, forming various side products. This is especially critical when using complex proteins such as antibodies as acceptors and these side products can be challenging to distinguish when using RP-HPLC and MS in quality control^{220,239,240}. Additionally, this strategy also required processing and purification steps during enzymatic trimming.

Therefore, we set out to test a faster method for the transfer of complex glycan structures from SGP onto semisynthetic C μ 4tp. We settled on performing direct enzymatic transglycosylation of the untrimmed SGP glycan onto peptides and proteins containing a single asparagine residue carrying an N-acetylglucosamine using ENGase Endo-cc N180H. This approach was inspired by extensive work on the acceptor range of this enzyme by Manabe *et al.* and their successes with direct transglycosylation of the SGP glycan to an anti CCR4 antibody²²⁰. This strategy allows for simplified and fast access to complexly modified glycosylations on semisynthetic proteins.

In conclusion, it is difficult to strictly recommend one transfer method over the other without consideration for the target of the glycan transfer and the glycan transfer conditions. In our case the solubility of the target protein remains an obstacle to overcome.

9.4.6.4 Semisynthesis of a complexly glycosylated C μ 4tp variant

We tested two distinct strategies for the semisynthesis of a complexly glycosylated C μ 4tp variant. The first strategy encompassed the production of a minimally glycosylated variant

(C μ 4tp 1-132 N119NGlcNAc) before transfer of the SGP glycan onto this variant. However, we were unable to bring this minimally glycosylated starting material into solution under reaction conditions required for enzymatic transfer and were thus forced to rethink our approach. Our second strategy relied on the transfer of the SGP glycan onto a C-terminal peptide (C μ 4tp 92-132 N119NGlcNAc) prior to ligation. This transglycosylation was successful, however the resulting product could not be separated from the educt. Due to time constraints, we did not perform further optimization of the transfer reaction which might allow for improved purity of the starting material, which would be required for a subsequent ligation.

10 Conclusion and Outlook

10.1 Semisynthesis of C μ 4tp and photocrosslinking variants

We set out to generate new insights into the underlying molecular mechanisms of IgM assembly through the design, synthesis, and characterization of semisynthetic C μ 4tp variants containing photocrosslinkers. We successfully designed and established a route for the semisynthesis of the IgM C μ 4tp domain that allows for the selective incorporation of non-canonical amino acids and selective post-translational modifications.

Overall, the IgM C μ 4tp domain is a challenging target for semisynthesis. This is especially true for separating the ligation product from the hydrolyzed thioester starting material when performing NCL using the native Cys92 as a ligation site. Although we could improve the ligation conditions to reduce the rate of hydrolysis and optimize our gradients for better separation during RP-HPLC purification, obtaining a pure product is still a significant challenge, with hydrolyzed starting material occasionally coming up to 40 % of signal height compared to the ligation product based on MS intensity. Additionally, we decided to discard fractions containing more than 5% of hydrolyzed starting material, as our cooperation partners had reported that this segment negatively affected folding, which further reduced yields. Utilizing an alternative valine/ aspartic acid ligation site in position 108/109 might allow for better separation of the starting material from the ligation product²³³. Shortening the synthetic segment would simplify SPPS, potentially increasing yields and purity. Relying on an N-terminal β -thio Asp on the synthetic segment for ligation would allow for selective removal of the auxiliary even in the presence of native cysteine residues.

Incorporation of a photocrosslinker, here Bpa *via* a commercially available Fmoc-Bpa building block, could be achieved without encountering major hurdles for most of our semisynthetic C μ 4tp variants. The only exception is the variant C μ 4tp Y132Bpa containing C-terminal Bpa, which seems to have degraded or reacted during synthesis.

We developed strategies for folding our C μ 4tp variants and demonstrated that semisynthetic wt C μ 4tp behaves identically to recombinantly expressed in CD and thermal transition measurements. Our photocrosslinking variants showed similar behavior, indicating the absence of significant changes in folding due to the presence of Bpa in the tailpiece, with the exception of variant V120Bpa, which displayed reduced formation of oligomers compared to the wt. This may be due to interference with the hydrophobic interactions between β -sheets formed by the tailpieces during oligomerization, in agreement with published IgM cryo-EM structures¹¹⁴ and a previously reported alanine scan⁷⁶.

10.2 Structural implications regarding the C μ 4tp monomer

While our Bpa variants displayed folding in line with our semisynthetic wt, we were unable to observe photocrosslinking of these folded variants in MS analysis. In contrast, during gel electrophoresis of the irradiated variants, V120Bpa appears to have formed stable dimers, which may indicate crosslinking. Contradictory, this variant displayed reduced formation of oligomers and dimers in SEC analysis compared to wt C μ 4tp in SEC analysis. However, it must be noted that these are preliminary results from an ongoing investigation lacking the necessary positive and negative controls to allow for more definitive conclusions. It is possible

to try photocrosslinking using various other photocrosslinkers, including aryl azide or diazirine.

This does not necessarily mean that our attempts to improve our understanding of the structure of monomeric C μ 4tp were entirely unsuccessful. High-resolution cryo-EM data on the IgM oligomer in complex with the J-chain and the SC has become available in recent years, providing extensive detail on the structure of the tail piece in this context¹¹⁴. The residues between Tyr118 and Met124 in each tail piece form β strands, that pack together into two five-stranded β sheets that stack on top of each other in an antiparallel manner. In agreement with this structure and previous reports by Pasalic *et al.*⁷⁶, we observed that our V120Bpa variant displayed reduced oligomer formation, likely through interference with this hydrophobic interactions between these β sheets. We did not observe the same folding behavior for the Y118Bpa variant, instead this variant displayed a strong tendency to form oligomers. It is therefore possible, that this variant suffered from degradation or premature photocrosslinking prior to or during folding, which may have affected the folding behavior of this variant.

10.3 Semisynthesis of glycosylated C μ 4tp variants

We also successfully applied our semisynthesis strategy to the production a minimal glycosylated C μ 4tp variant, carrying an N-linked GlcNAc in position N119. Additionally, we took successful steps toward the semisynthesis of a more complexly glycosylated C μ 4tp variant using direct transfer of the SGP glycan onto the target protein *via* the acetylglucosaminidase mutant endo-CC N180H inspired by work by Manabe *et al.*²²⁰. While we

initially attempted to transfer the glycan from SGP onto our C μ 4tp N119NGlcNAc variant, we were to bring our target protein into solution under transfer conditions. We therefore decided to perform this direct enzymatic transfer of the SGP glycan onto the better soluble C μ 4tp 92-132 N119NGlcNAc. This transfer was performed successfully. However, we could not separate the SGP-carrying product from unreacted starting material *via* RP-HPLC – thus preventing a subsequent ligation attempt. Further work optimizing transfer conditions and separation strategies may allow for successful semisynthesis of this variant in future projects.

We adapted and streamlined existing SGP isolation protocols to isolate SGP from egg yolk. This strategy required us to adapt and streamline existing SGP isolation protocols. When analyzing these methods in the context of the semisynthesis of C μ 4tp, the initial protocol by Schmid²²⁸ was found to be more resource and time intense than the updated protocol inspired by Alagesan *et al.*²¹⁶. Additionally, the updated protocol prioritized scalability without sacrificing purity, as both protocols yielded products with purity above 90%. It should however be noted that ideal isolation strategies depend on the availability of equipment and resources and that other research groups may find other SGP isolation methods more suitable for their needs.

We believe that this direct transfer is a viable strategy for rapid and simple production of complexly glycosylated proteins in semisynthesis, as the solubility- and purification-related challenges encountered in this work derive from the target protein and are not inherent to the glycan transfer method. Further optimization of the transfer conditions such as increasing the DMSO concentration above 25% or using chaotropic agents may overcome these challenges. It should be kept in mind that these must not come at the expense of enzyme activity and

stability. Further modification of the glycan *via* enzymatic trimming should also be possible quantitatively after transfer using established enzymes.

10.4 Outlook

Future semisynthesis projects related to the C μ 4tp domain would benefit from improved ligation and purification protocols for separating the hydrolyzed starting material from the final ligation product. This could include exploration of an alternative valine/ aspartic acid ligation site in position 108/109 and purification *via* SEC.

We remain curious whether the utilization of a different photocrosslinker, such as aryl azide or diazirine, would allow for synthesizing a C-terminal segment with a photocrosslinker in position Y132. We would also like to expand our work on the semisynthesis of a single domain of IgM to include several domains, such as the semisynthesis of the entire heavy chain of IgM. Another interesting target for semisynthesis related to IgM is IgA, which also contains C-terminal tailpiece driving oligomerization¹⁴³. Additionally, we are interested in observing future applications using Cu4tp fusion constructs for controlled protein oligomerization, as presented by Svilenov *et al.*¹⁴⁵.

We're also looking forward to exploring and observing the utilization of enzymatic transglycosylation in the semisynthesis of various proteins. We are also interested in trimming transferred glycans on the target protein. Focusing on IgM and C μ 4tp specifically, we would like to explore additional purification strategies post transfer, including HPLC columns with a higher sensitivity to glycans and other purification strategies.

Conclusion and Outlook

Our results regarding the C μ 4tp structure support existing cryo-EM data of the IgM oligomer and pose new questions about the structure and folding of the Cu4tp monomer. We look forward to further investigating the C μ 4tp driven assembly of IgM and imagine that a deeper understanding of the relevant structure-function relationship could contribute to engineering novel multivalent antibodies and proteins with controlled oligomerization behavior.

11 References

- 1 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).
- 2 Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **29**, 620-624, doi:10.1093/glycob/cwz045 (2019).
- 3 Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M. & Corcoran, L. M. The generation of antibody-secreting plasma cells. *Nature Reviews Immunology* **15**, 160-171, doi:10.1038/nri3795 (2015).
- 4 Khodadadi, L., Cheng, Q., Radbruch, A. & Hiepe, F. The Maintenance of Memory Plasma Cells. *Frontiers in Immunology* **10**, doi:10.3389/fimmu.2019.00721 (2019).
- 5 Shapiro-Shelef, M. & Calame, K. Regulation of plasma-cell development. *Nature Reviews Immunology* **5**, 230-242, doi:10.1038/nri1572 (2005).
- 6 Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497, doi:10.1038/256495a0 (1975).
- 7 García-Alija, M. *et al.* Modulating antibody effector functions by Fc glycoengineering. *Biotechnol Adv*, 108201, doi:10.1016/j.biotechadv.2023.108201 (2023).
- 8 Thakur, A., Huang, M. & Lum, L. G. Bispecific antibody based therapeutics: Strengths and challenges. *Blood reviews* **32**, 339-347, doi:10.1016/j.blre.2018.02.004 (2018).
- 9 Ministro, J., Manuel, A. M. & Goncalves, J. Therapeutic Antibody Engineering and Selection Strategies. *Adv Biochem Eng Biotechnol* **171**, 55-86, doi:10.1007/10_2019_116 (2020).
- 10 Buss, N. A., Henderson, S. J., McFarlane, M., Shenton, J. M. & de Haan, L. Monoclonal antibody therapeutics: history and future. *Curr Opin Pharmacol* **12**, 615-622, doi:10.1016/j.coph.2012.08.001 (2012).
- 11 Baldassari, L. E. & Rose, J. W. Daclizumab: Development, Clinical Trials, and Practical Aspects of Use in Multiple Sclerosis. *Neurotherapeutics* **14**, 842-858, doi:10.1007/s13311-017-0553-8 (2017).
- 12 Craddock, J. & Markovic-Plese, S. Immunomodulatory therapies for relapsing-remitting multiple sclerosis: monoclonal antibodies, currently approved and in testing. *Expert Rev Clin Pharmacol* **8**, 283-296, doi:10.1586/17512433.2015.1036030 (2015).
- 13 Deiss, A., Brecht, I., Haarmann, A. & Buttmann, M. Treating multiple sclerosis with monoclonal antibodies: a 2013 update. *Expert Rev Neurother* **13**, 313-335, doi:10.1586/ern.13.17 (2013).
- 14 Kappos, L. *et al.* Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. *Lancet* **378**, 1779-1787, doi:10.1016/S0140-6736(11)61649-8 (2011).
- 15 Sorensen, P. S. *et al.* Safety and efficacy of ofatumumab in relapsing-remitting multiple sclerosis: a phase 2 study. *Neurology* **82**, 573-581, doi:10.1212/WNL.000000000000125 (2014).
- 16 Grifols USA, L. HYPERRAB - Prescribing Information, <<https://www.accessdata.fda.gov/spl/data/f993778d-01fb-4670-af67-a0e08d6b258b/f993778d-01fb-4670-af67-a0e08d6b258b.xml>> (accessed on 21 June 2023).
- 17 Sol Therapeutics, I. VARIZIG - Prescribing Information, <<https://www.accessdata.fda.gov/spl/data/272379b7-f0e7-4560-8d79-3fd0024c3010/272379b7-f0e7-4560-8d79-3fd0024c3010.xml>> (accessed on 21 June 2023).
- 18 GRIFOLS USA, L. GAMASTAN - Prescribing information, <<https://www.accessdata.fda.gov/spl/data/38a323af-7c25-42d1-9c29-532ef61999b8/38a323af-7c25-42d1-9c29-532ef61999b8.xml>> (accessed: on 21 June 2023).
- 19 GRIFOLS USA, L. HyperHEP B Prescribing Information, <<https://www.accessdata.fda.gov/spl/data/391b2218-8a15-4e5e-8717-aa49efcc2210/391b2218-8a15-4e5e-8717-aa49efcc2210.xml>> (accessed on 21 June 2023).
- 20 Struble, E. B., Rawson, J. M. O., Stantchev, T., Scott, D. & Shapiro, M. A. Uses and Challenges of Antiviral Polyclonal and Monoclonal Antibody Therapies. *Pharmaceutics* **15**, doi:10.3390/pharmaceutics15051538 (2023).
- 21 Kaplon, H., Crescioli, S., Chenoweth, A., Visweswaraiiah, J. & Reichert, J. M. Antibodies to watch in 2023. *MAbs* **15**, 2153410, doi:10.1080/19420862.2022.2153410 (2023).

References

- 22 Lyu, X. *et al.* The global landscape of approved antibody therapies. *Antib Ther* **5**, 233-257, doi:10.1093/abt/tbac021 (2022).
- 23 Mullard, A. FDA approves 100th monoclonal antibody product. *Nat Rev Drug Discov* **20**, 491-495, doi:10.1038/d41573-021-00079-7 (2021).
- 24 Burnette, W. N. "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry* **112**, 195-203, doi:[https://doi.org/10.1016/0003-2697\(81\)90281-5](https://doi.org/10.1016/0003-2697(81)90281-5) (1981).
- 25 Lequin, R. M. Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clinical Chemistry* **51**, 2415-2418, doi:10.1373/clinchem.2005.051532 (2005).
- 26 Di Nardo, F., Chiarello, M., Cavallera, S., Baggiani, C. & Anfossi, L. Ten Years of Lateral Flow Immunoassay Technique Applications: Trends, Challenges and Future Perspectives. *Sensors* **21**, 5185 (2021).
- 27 Herzenberg, L. A., De Rosa, S. C. & Herzenberg, L. A. Monoclonal antibodies and the FACS: complementary tools for immunobiology and medicine. *Immunology Today* **21**, 383-390, doi:10.1016/S0167-5699(00)01678-9 (2000).
- 28 Gross, A. *et al.* Technologies for Single-Cell Isolation. *Int J Mol Sci* **16**, 16897-16919, doi:10.3390/ijms160816897 (2015).
- 29 Buck, M. J. & Lieb, J. D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349-360, doi:<https://doi.org/10.1016/j.ygeno.2003.11.004> (2004).
- 30 Porter, R. R. The hydrolysis of rabbit γ -globulin and antibodies with crystalline papain. *Biochem J* **73**, 119-126, doi:10.1042/bj0730119 (1959).
- 31 Porter, R. R. A chemical study of rabbit antiovalbumin. *Biochemical Journal* **46**, 473-478, doi:10.1042/bj0460473 (1950).
- 32 Deisenhofer, J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry* **20**, 2361-2370, doi:10.1021/bi00512a001 (1981).
- 33 Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology* **11**, 251-263, doi:10.1038/nri2941 (2011).
- 34 Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22, doi:10.1146/annurev.biochem.76.061705.090740 (2007).
- 35 Doria-Rose, N. A. & Joyce, M. G. Strategies to guide the antibody affinity maturation process. *Curr Opin Virol* **11**, 137-147, doi:10.1016/j.coviro.2015.04.002 (2015).
- 36 Arbuckle, M. R. *et al.* Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med* **349**, 1526-1533, doi:10.1056/NEJMoa021933 (2003).
- 37 Song, Y. W. & Kang, E. H. Autoantibodies in rheumatoid arthritis: rheumatoid factors and anticitrullinated protein antibodies. *QJM: An International Journal of Medicine* **103**, 139-146, doi:10.1093/qjmed/hcp165 (2009).
- 38 Minden, S. L. & Schiffer, R. B. Affective disorders in multiple sclerosis. Review and recommendations for clinical research. *Arch Neurol* **47**, 98-104, doi:10.1001/archneur.1990.00530010124031 (1990).
- 39 Dargahi, N. *et al.* Multiple Sclerosis: Immunopathology and Treatment Update. *Brain Sci* **7**, 78, doi:10.3390/brainsci7070078 (2017).
- 40 Gholamzad, M. *et al.* A comprehensive review on the treatment approaches of multiple sclerosis: currently and in the future. *Inflamm Res* **68**, 25-38, doi:10.1007/s00011-018-1185-0 (2019).
- 41 Davies, T. F. *et al.* Graves' disease. *Nature Reviews Disease Primers* **6**, 52, doi:10.1038/s41572-020-0184-y (2020).
- 42 Weetman, A. P. An update on the pathogenesis of Hashimoto's thyroiditis. *J Endocrinol Invest* **44**, 883-890, doi:10.1007/s40618-020-01477-1 (2021).
- 43 Rose, N. R. & Witebsky, E. Studies on organ specificity. V. Changes in the thyroid glands of rabbits following active immunization with rabbit thyroid extracts. *J Immunol* **76**, 417-427 (1956).
- 44 Xiao, Z. X., Miller, J. S. & Zheng, S. G. An updated advance of autoantibodies in autoimmune diseases. *Autoimmunity Reviews* **20**, 102743, doi:<https://doi.org/10.1016/j.autrev.2020.102743> (2021).

References

- 45 Mané-Damas, M. *et al.* Autoimmunity in psychotic disorders. Where we stand, challenges and opportunities. *Autoimmunity reviews*, 102348 (2019).
- 46 Edelman, G. M. DISSOCIATION OF γ -GLOBULIN. *Journal of the American Chemical Society* **81**, 3155-3156, doi:10.1021/ja01521a071 (1959).
- 47 Edelman, G. M. Antibody Structure and Molecular Immunology. *Science* **180**, 830-840, doi:10.1126/science.180.4088.830 (1973).
- 48 Chiu, M. L., Goulet, D. R., Teplyakov, A. & Gilliland, G. L. Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies* **8**, 55 (2019).
- 49 Müller, R. *et al.* High-resolution structures of the IgM Fc domains reveal principles of its hexamer formation. *Proceedings of the National Academy of Sciences* **110**, 10183, doi:10.1073/pnas.1300547110 (2013).
- 50 Bergmann, M. & Zervas, L. Über ein allgemeines Verfahren der Peptid-Synthese. *Berichte der deutschen chemischen Gesellschaft (A and B Series)* **65**, 1192-1201, doi:<https://doi.org/10.1002/cber.19320650722> (1932).
- 51 Sehnaal, D. *et al.* Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research* **49**, W431-W437, doi:10.1093/nar/gkab314 (2021).
- 52 Lu, L. L., Suscovich, T. J., Fortune, S. M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nature Reviews Immunology* **18**, 46-61, doi:10.1038/nri.2017.106 (2018).
- 53 Schroeder, H. W., Jr. & Cavacini, L. Structure and function of immunoglobulins. *J Allergy Clin Immunol* **125**, S41-S52, doi:10.1016/j.jaci.2009.09.046 (2010).
- 54 Macpherson, A. J., McCoy, K. D., Johansen, F. E. & Brandtzaeg, P. The immune geography of IgA induction and function. *Mucosal Immunol* **1**, 11-22, doi:10.1038/mi.2007.6 (2008).
- 55 Brandtzaeg, P. & Prydz, H. Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature* **311**, 71-73, doi:10.1038/311071a0 (1984).
- 56 Hand, T. W. & Riboldi, A. Production and Function of Immunoglobulin A. *Annual Review of Immunology* **39**, 695-718, doi:10.1146/annurev-immunol-102119-074236 (2021).
- 57 Cerutti, A., Chen, K. & Chorny, A. Immunoglobulin responses at the mucosal interface. *Annu Rev Immunol* **29**, 273-293, doi:10.1146/annurev-immunol-031210-101317 (2011).
- 58 Bohländer, F. A new hope? Possibilities of therapeutic IgA antibodies in the treatment of inflammatory lung diseases. *Frontiers in Immunology* **14**, doi:10.3389/fimmu.2023.1127339 (2023).
- 59 Parray, H. A. *et al.* Inhalation monoclonal antibody therapy: a new way to treat and manage respiratory infections. *Appl Microbiol Biotechnol* **105**, 6315-6332, doi:10.1007/s00253-021-11488-4 (2021).
- 60 Sécher, T. *et al.* Therapeutic antibodies: A new era in the treatment of respiratory diseases? *Pharmacol Ther* **189**, 149-172, doi:10.1016/j.pharmthera.2018.05.003 (2018).
- 61 Chen, K., Magri, G., Grasset, E. K. & Cerutti, A. Rethinking mucosal antibody responses: IgM, IgG and IgD join IgA. *Nature Reviews Immunology* **20**, 427-441, doi:10.1038/s41577-019-0261-1 (2020).
- 62 Chen, K. *et al.* Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils. *Nat Immunol* **10**, 889-898, doi:10.1038/ni.1748 (2009).
- 63 Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nature Reviews Immunology* **2**, 688-698, doi:10.1038/nri889 (2002).
- 64 Ling, X. J., Wei, J. F. & Zhu, Y. Aiming to IgE: Drug development in allergic diseases. *Int Immunopharmacol* **121**, 110495, doi:10.1016/j.intimp.2023.110495 (2023).
- 65 Kim, J. Y., Yi, M. H. & Yong, T. S. Allergen-like Molecules from Parasites. *Curr Protein Pept Sci* **21**, 186-202, doi:10.2174/1389203720666190708154300 (2020).
- 66 Wan, T. *et al.* The crystal structure of IgE Fc reveals an asymmetrically bent conformation. *Nat Immunol* **3**, 681-686, doi:10.1038/ni811 (2002).
- 67 Gould, H. J. & Sutton, B. J. IgE in allergy and asthma today. *Nature Reviews Immunology* **8**, 205-217, doi:10.1038/nri2273 (2008).
- 68 Kubo, S., Nakayama, T., Matsuoka, K., Yonekawa, H. & Karasuyama, H. Long term maintenance of IgE-mediated memory in mast cells in the absence of detectable serum IgE. *J Immunol* **170**, 775-780, doi:10.4049/jimmunol.170.2.775 (2003).

References

- 69 Jefferis, R. Isotype and glycoform selection for antibody therapeutics. *Arch Biochem Biophys* **526**, 159-166, doi:10.1016/j.abb.2012.03.021 (2012).
- 70 Vidarsson, G., Dekkers, G. & Rispen, T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Frontiers in Immunology* **5**, doi:10.3389/fimmu.2014.00520 (2014).
- 71 Dall'Acqua, W. F., Cook, K. E., Damschroder, M. M., Woods, R. M. & Wu, H. Modulation of the effector functions of a human IgG1 through engineering of its hinge region. *J Immunol* **177**, 1129-1138, doi:10.4049/jimmunol.177.2.1129 (2006).
- 72 Kabat, E. A. & Pedersen, K. O. THE MOLECULAR WEIGHTS OF ANTIBODIES. *Science* **87**, 372, doi:10.1126/science.87.2260.372 (1938).
- 73 Jin, B. K., Odongo, S., Radwanska, M. & Magez, S. Nanobodies: A Review of Generation, Diagnostics and Therapeutics. *Int J Mol Sci* **24**, doi:10.3390/ijms24065994 (2023).
- 74 Jones, K., Savulescu, A. F., Brombacher, F. & Hadebe, S. Immunoglobulin M in Health and Diseases: How Far Have We Come and What Next? *Front Immunol* **11**, 595535, doi:10.3389/fimmu.2020.595535 (2020).
- 75 Michaud, E., Mastrandrea, C., Rochereau, N. & Paul, S. Human Secretory IgM: An Elusive Player in Mucosal Immunity. *Trends in Immunology* **41**, 141-156, doi:10.1016/j.it.2019.12.005 (2020).
- 76 Pasalic, D. *et al.* A peptide extension dictates IgM assembly. *Proceedings of the National Academy of Sciences* **114**, E8575, doi:10.1073/pnas.1701797114 (2017).
- 77 Davis, A. C., Roux, K. H. & Shulman, M. J. On the structure of polymeric IgM. *Eur J Immunol* **18**, 1001-1008, doi:10.1002/eji.1830180705 (1988).
- 78 Putnam, F. W., Florent, G., Paul, C., Shinoda, T. & Shimizu, A. Complete Amino Acid Sequence of the Mu Heavy Chain of a Human IgM Immunoglobulin. *Science* **182**, 287-291, doi:10.1126/science.182.4109.287 (1973).
- 79 Cooper, M. D. The early history of B cells. *Nature Reviews Immunology* **15**, 191-197, doi:10.1038/nri3801 (2015).
- 80 Wang, Y., Liu, J., Burrows, P. D. & Wang, J. Y. B Cell Development and Maturation. *Adv Exp Med Biol* **1254**, 1-22, doi:10.1007/978-981-15-3532-1_1 (2020).
- 81 Klein, U., Küppers, R. & Rajewsky, K. Evidence for a large compartment of IgM-expressing memory B cells in humans. *Blood* **89**, 1288-1298 (1997).
- 82 Martin, F., Oliver, A. M. & Kearney, J. F. Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens. *Immunity* **14**, 617-629, doi:10.1016/s1074-7613(01)00129-7 (2001).
- 83 Martin, F. & Kearney, J. F. Marginal-zone B cells. *Nat Rev Immunol* **2**, 323-335, doi:10.1038/nri799 (2002).
- 84 Xie, W., Wucherpfennig, K. & Patel, D. J. A structural platform for B cell receptor signaling. *Cell Research* **33**, 95-96, doi:10.1038/s41422-022-00724-9 (2023).
- 85 Ma, X. *et al.* Cryo-EM structures of two human B cell receptor isotypes. *Science* **377**, 880-885, doi:10.1126/science.abo3828 (2022).
- 86 Su, Q. *et al.* Cryo-EM structure of the human IgM B cell receptor. *Science* **377**, 875-880, doi:10.1126/science.abo3923 (2022).
- 87 Schamel, W. W. & Reth, M. Monomeric and oligomeric complexes of the B cell antigen receptor. *Immunity* **13**, 5-14, doi:10.1016/s1074-7613(00)00003-0 (2000).
- 88 Shaw, A. C. *et al.* Mutations of immunoglobulin transmembrane and cytoplasmic domains: effects on intracellular signaling and antigen presentation. *Cell* **63**, 381-392, doi:10.1016/0092-8674(90)90171-a (1990).
- 89 Ramesh, S., Park, S., Im, W., Call, M. J. & Call, M. E. T cell and B cell antigen receptors share a conserved core transmembrane structure. *Proceedings of the National Academy of Sciences* **119**, e2208058119, doi:doi:10.1073/pnas.2208058119 (2022).
- 90 Übelhart, R. *et al.* Responsiveness of B cells is regulated by the hinge region of IgD. *Nature Immunology* **16**, 534-543, doi:10.1038/ni.3141 (2015).
- 91 Noviski, M. *et al.* IgM and IgD B cell receptors differentially respond to endogenous antigens and control B cell fate. *eLife* **7**, e35074, doi:10.7554/eLife.35074 (2018).
- 92 Natkunam, Y. The biology of the germinal center. *Hematology Am Soc Hematol Educ Program*, 210-215, doi:10.1182/asheducation-2007.1.210 (2007).

References

- 93 Victora, G. D. & Nussenzweig, M. C. Germinal Centers. *Annu Rev Immunol* **40**, 413-442, doi:10.1146/annurev-immunol-120419-022408 (2022).
- 94 Harwood, N. E. & Batista, F. D. Early events in B cell activation. *Annu Rev Immunol* **28**, 185-210, doi:10.1146/annurev-immunol-030409-101216 (2010).
- 95 Tolar, P. & Pierce, S. K. Unveiling the B cell receptor structure. *Science* **377**, 819-820, doi:doi:10.1126/science.add8065 (2022).
- 96 Boes, M. Role of natural and immune IgM antibodies in immune responses. *Mol Immunol* **37**, 1141-1149, doi:10.1016/s0161-5890(01)00025-6 (2000).
- 97 Thomas, H. I. & Morgan-Capner, P. The avidity of specific IgM detected in primary rubella and reinfection. *Epidemiol Infect* **104**, 489-497, doi:10.1017/s095026880004749x (1990).
- 98 Cooper, N. R., Nemerow, G. R. & Mayes, J. T. Methods to detect and quantitate complement activation. *Springer Semin Immunopathol* **6**, 195-212, doi:10.1007/bf00205873 (1983).
- 99 Hughey, C. T., Brewer, J. W., Colosia, A. D., Rosse, W. F. & Corley, R. B. Production of IgM hexamers by normal and autoimmune B cells: implications for the physiologic role of hexameric IgM. *J Immunol* **161**, 4091-4097 (1998).
- 100 Krishnamurty, A. T. *et al.* Somatically Hypermutated Plasmodium-Specific IgM(+) Memory B Cells Are Rapid, Plastic, Early Responders upon Malaria Rechallenge. *Immunity* **45**, 402-414, doi:10.1016/j.immuni.2016.06.014 (2016).
- 101 van Es, J. H., Meyling, F. H. & Logtenberg, T. High frequency of somatically mutated IgM molecules in the human adult blood B cell repertoire. *Eur J Immunol* **22**, 2761-2764, doi:10.1002/eji.1830221046 (1992).
- 102 Wibroe, P. P., Helvig, S. Y. & Moein Moghimi, S. The Role of Complement in Antibody Therapy for Infectious Diseases. *Microbiol Spectr* **2**, doi:10.1128/microbiolspec.AID-0015-2014 (2014).
- 103 Freeley, S., Kemper, C. & Le Friec, G. The "ins and outs" of complement-driven immune responses. *Immunol Rev* **274**, 16-32, doi:10.1111/imr.12472 (2016).
- 104 Ehrenstein, M. R. & Notley, C. A. The importance of natural IgM: scavenger, protector and regulator. *Nature Reviews Immunology* **10**, 778-786, doi:10.1038/nri2849 (2010).
- 105 Yazdani, R., Azizi, G., Abolhassani, H. & Aghamohammadi, A. Selective IgA Deficiency: Epidemiology, Pathogenesis, Clinical Phenotype, Diagnosis, Prognosis and Management. *Scand J Immunol* **85**, 3-12, doi:10.1111/sji.12499 (2017).
- 106 Brandtzaeg, P., Nilssen, D. E., Rognum, T. O. & Thrane, P. S. Ontogeny of the mucosal immune system and IgA deficiency. *Gastroenterol Clin North Am* **20**, 397-439 (1991).
- 107 Brandtzaeg, P. & Nilssen, D. E. Mucosal aspects of primary B-cell deficiency and gastrointestinal infections. *Current Opinion in Gastroenterology* **11** (1995).
- 108 Louis, A. G. & Gupta, S. Primary selective IgM deficiency: an ignored immunodeficiency. *Clin Rev Allergy Immunol* **46**, 104-111, doi:10.1007/s12016-013-8375-x (2014).
- 109 Louis, A. G., Agrawal, S. & Gupta, S. Analysis of subsets of B cells, Breg, CD4Treg and CD8Treg cells in adult patients with primary selective IgM deficiency. *Am J Clin Exp Immunol* **5**, 21-32 (2016).
- 110 Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the secretory immunoglobulin A core. *Science* **367**, 1008-1014, doi:10.1126/science.aaz5807 (2020).
- 111 Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the human secretory immunoglobulin M core. *Structure* **29**, 564-571.e563, doi:10.1016/j.str.2021.01.002 (2021).
- 112 Sørensen, V., Sundvold, V., Michaelsen, T. E. & Sandlie, I. Polymerization of IgA and IgM: roles of Cys309/Cys414 and the secretory tailpiece. *J Immunol* **162**, 3448-3455 (1999).
- 113 Johansen, F.-E., Braathen, R. & Brandtzaeg, P. The J Chain Is Essential for Polymeric Ig Receptor-Mediated Epithelial Transport of IgA. *The Journal of Immunology* **167**, 5185, doi:10.4049/jimmunol.167.9.5185 (2001).
- 114 Li, Y. *et al.* Structural insights into immunoglobulin M. *Science* **367**, 1014-1017, doi:10.1126/science.aaz5425 (2020).
- 115 Kaetzel, C. S. The polymeric immunoglobulin receptor: bridging innate and adaptive immune responses at mucosal surfaces. *Immunological Reviews* **206**, 83-99, doi:10.1111/j.0105-2896.2005.00278.x (2005).
- 116 Shibuya, A. *et al.* Fc alpha/mu receptor mediates endocytosis of IgM-coated microbes. *Nat Immunol* **1**, 441-446, doi:10.1038/80886 (2000).

References

- 117 Ouchida, R. *et al.* Critical role of the IgM Fc receptor in IgM homeostasis, B-cell survival, and humoral immune responses. *Proc Natl Acad Sci U S A* **109**, E2699-2706, doi:10.1073/pnas.1210706109 (2012).
- 118 Ouchida, R. *et al.* FcμR Interacts and Cooperates with the B Cell Receptor To Promote B Cell Survival. *The Journal of Immunology* **194**, 3096-3101, doi:10.4049/jimmunol.1402352 (2015).
- 119 Nguyen, T. T. *et al.* The IgM receptor FcμR limits tonic BCR signaling by regulating expression of the IgM BCR. *Nat Immunol* **18**, 321-333, doi:10.1038/ni.3677 (2017).
- 120 Li, Y. *et al.* Immunoglobulin M perception by FcμR. *Nature* **615**, 907-912, doi:10.1038/s41586-023-05835-w (2023).
- 121 Vire, B., David, A. & Wiestner, A. TOSO, the FcμR receptor, is highly expressed on chronic lymphocytic leukemia B cells, internalizes upon IgM binding, shuttles to the lysosome, and is downregulated in response to TLR activation. *J Immunol* **187**, 4040-4050, doi:10.4049/jimmunol.1100532 (2011).
- 122 Choi, S.-C. *et al.* Mouse IgM Fc Receptor, FCμR, Promotes B Cell Development and Modulates Antigen-Driven Immune Responses. *The Journal of Immunology* **190**, 987-996, doi:10.4049/jimmunol.1202227 (2013).
- 123 Sutton, B. J. Multi-faceted immunoglobulin M meets its elusive receptor. *Nature Structural & Molecular Biology* **30**, 866-869, doi:10.1038/s41594-023-01030-7 (2023).
- 124 Grönwall, C. & Silverman, G. J. Natural IgM: beneficial autoantibodies for the control of inflammatory and autoimmune disease. *J Clin Immunol* **34 Suppl 1**, S12-21, doi:10.1007/s10875-014-0025-4 (2014).
- 125 Grönwall, C., Vas, J. & Silverman, G. J. Protective Roles of Natural IgM Antibodies. *Front Immunol* **3**, 66, doi:10.3389/fimmu.2012.00066 (2012).
- 126 Keyt, B. A., Baliga, R., Sinclair, A. M., Carroll, S. F. & Peterson, M. S. Structure, Function, and Therapeutic Use of IgM Antibodies. *Antibodies (Basel)* **9**, doi:10.3390/antib9040053 (2020).
- 127 Griffin, D. O., Holodick, N. E. & Rothstein, T. L. Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+ CD27+ CD43+ CD70. *J Exp Med* **208**, 67-80, doi:10.1084/jem.20101499 (2011).
- 128 Nguyen, T. T., Elsner, R. A. & Baumgarth, N. Natural IgM prevents autoimmunity by enforcing B cell central tolerance induction. *J Immunol* **194**, 1489-1502, doi:10.4049/jimmunol.1401880 (2015).
- 129 Ehrenstein, M. R., O’Keefe, T. L., Davies, S. L. & Neuberger, M. S. Targeted gene disruption reveals a role for natural secretory IgM in the maturation of the primary immune response. *Proceedings of the National Academy of Sciences* **95**, 10089-10093, doi:doi:10.1073/pnas.95.17.10089 (1998).
- 130 Mestecky, J., Zikan, J. & Butler, W. T. Immunoglobulin M and Secretory Immunoglobulin A: Presence of a Common Polypeptide Chain Different from Light Chains. *Science* **171**, 1163, doi:10.1126/science.171.3976.1163 (1971).
- 131 Halpern, M. S. & Koshland, M. E. Novel Subunit in Secretory IgA. *Nature* **228**, 1276-1278, doi:10.1038/2281276a0 (1970).
- 132 Feinstein, A. & Munn, E. A. Conformation of the Free and Antigen-bound IgM Antibody Molecules. *Nature* **224**, 1307-1309, doi:10.1038/2241307a0 (1969).
- 133 Della Corte, E. & Parkhouse, R. M. Biosynthesis of immunoglobulin A (IgA) and immunoglobulin M (IgM). Requirement for J chain and a disulphide-exchanging enzyme for polymerization. *Biochem J* **136**, 597-606, doi:10.1042/bj1360597 (1973).
- 134 Oostindie, S. C., Lazar, G. A., Schuurman, J. & Parren, P. W. H. I. Avidity in antibody effector functions and biotherapeutic drug design. *Nature Reviews Drug Discovery* **21**, 715-735, doi:10.1038/s41573-022-00501-8 (2022).
- 135 Hiramoto, E. *et al.* The IgM pentamer is an asymmetric pentagon with an open groove that binds the AIM protein. *Science Advances* **4**, eaau1199, doi:10.1126/sciadv.aau1199 (2018).
- 136 Sharp, T. H. *et al.* Insights into IgM-mediated complement activation based on in situ structures of IgM-C1-C4b. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 11900-11905, doi:10.1073/pnas.1901841116 (2019).
- 137 Cortini, M. & Sitia, R. ERp44 and ERGIC-53 synergize in coupling efficiency and fidelity of IgM polymerization and secretion. *Traffic* **11**, 651-659, doi:10.1111/j.1600-0854.2010.01043.x (2010).
- 138 Socken, D. J. & Underdown, B. J. Comparison of human, bovine and rabbit secretory component-immunoglobulin interactions. *Immunochemistry* **15**, 499-506, doi:10.1016/0161-5890(78)90080-9 (1978).

References

- Tempio, T. & Anelli, T. The pivotal role of ERp44 in patrolling protein secretion. *Journal of Cell Science* **133**, doi:10.1242/jcs.240366 (2020).
- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).
- Sitia, R. *et al.* Developmental regulation of IgM secretion: The role of the carboxy-terminal cysteine. *Cell* **60**, 781-790, doi:[https://doi.org/10.1016/0092-8674\(90\)90092-S](https://doi.org/10.1016/0092-8674(90)90092-S) (1990).
- Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188-1203, doi:10.1016/j.cell.2012.02.022 (2012).
- Weber, B., Maier, A. & Buchner, J. Peptides in proteins. *Journal of Peptide Science* **26**, e3235, doi:<https://doi.org/10.1002/psc.3235> (2020).
- Sopp, J. M. *et al.* On-target IgG hexamerisation driven by a C-terminal IgM tail-piece fusion variant confers augmented complement activation. *Commun Biol* **4**, 1031, doi:10.1038/s42003-021-02513-3 (2021).
- Svilenov, H. L. *et al.* Multimeric ACE2-IgM fusions as broadly active antivirals that potently neutralize SARS-CoV-2 variants. *Communications Biology* **5**, 1237, doi:10.1038/s42003-022-04193-z (2022).
- Krapp, S., Mimura, Y., Jefferis, R., Huber, R. & Sondermann, P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol* **325**, 979-989, doi:10.1016/s0022-2836(02)01250-0 (2003).
- Mimura, Y. *et al.* The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Mol Immunol* **37**, 697-706, doi:10.1016/s0161-5890(00)00105-x (2000).
- Vattepu, R., Sneed, S. L. & Anthony, R. M. Sialylation as an Important Regulator of Antibody Function. *Frontiers in Immunology* **13**, doi:10.3389/fimmu.2022.818736 (2022).
- Mimura, Y. *et al.* Role of oligosaccharide residues of IgG1-Fc in Fc gamma RIIb binding. *J Biol Chem* **276**, 45539-45547, doi:10.1074/jbc.M107478200 (2001).
- Malhotra, R. *et al.* Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med* **1**, 237-243, doi:10.1038/nm0395-237 (1995).
- Gala, F. A. & Morrison, S. L. The role of constant region carbohydrate in the assembly and secretion of human IgD and IgA1. *J Biol Chem* **277**, 29005-29011, doi:10.1074/jbc.M203258200 (2002).
- Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* **25**, 21-50, doi:10.1146/annurev.immunol.25.022106.141702 (2007).
- Hamilton, S. R. *et al.* Production of complex human glycoproteins in yeast. *Science* **301**, 1244-1246, doi:10.1126/science.1088166 (2003).
- Barone, G. D. *et al.* Industrial Production of Proteins with *Pichia pastoris*-*Komagataella phaffii*. *Biomolecules* **13**, doi:10.3390/biom13030441 (2023).
- Triguero, A. *et al.* Differential N-glycosylation of a monoclonal antibody expressed in tobacco leaves with and without endoplasmic reticulum retention signal apparently induces similar in vivo stability in mice. *Plant Biotechnol J* **9**, 1120-1130, doi:10.1111/j.1467-7652.2011.00638.x (2011).
- Nessa, M. U., Rahman, M. A. & Kabir, Y. Plant-Produced Monoclonal Antibody as Immunotherapy for Cancer. *Biomed Res Int* **2020**, 3038564, doi:10.1155/2020/3038564 (2020).
- Shivatara, S. S., Shivatara, V. S. & Wong, C. H. Glycoconjugates: Synthesis, Functional Studies, and Therapeutic Developments. *Chem Rev* **122**, 15603-15671, doi:10.1021/acs.chemrev.1c01032 (2022).
- Aebersold, R. *et al.* How many human proteoforms are there? *Nat Chem Biol* **14**, 206-214, doi:10.1038/nchembio.2576 (2018).
- Streety, X. S., Obike, J. C. & Townsend, S. D. A Hitchhiker's Guide to Problem Selection in Carbohydrate Synthesis. *ACS Central Science*, doi:10.1021/acscentsci.3c00507 (2023).
- Bieberich, E. Synthesis, Processing, and Function of N-glycans in N-glycoproteins. *Adv Neurobiol* **9**, 47-70, doi:10.1007/978-1-4939-1154-7_3 (2014).
- Moh, E. S. X., Lin, C.-H., Thaysen-Andersen, M. & Packer, N. H. Site-Specific N-Glycosylation of Recombinant Pentameric and Hexameric Human IgM. *Journal of the American Society for Mass Spectrometry* **27**, 1143-1155, doi:10.1007/s13361-016-1378-0 (2016).

References

- 162 de Lalla, C., Fagioli, C., Cessi, F. S., Smilovich, D. & Sitia, R. Biogenesis and function of IgM: the role of the
conserved mu-chain tailpiece glycans. *Mol Immunol* **35**, 837-845, doi:10.1016/s0161-5890(98)00073-x
(1998).
- 163 Muraoka, S. & Shulman, M. J. Structural requirements for IgM assembly and cytolytic activity. Effects of
mutations in the oligosaccharide acceptor site at Asn402. *J Immunol* **142**, 695-701 (1989).
- 164 Arnold, J. N. *et al.* Human serum IgM glycosylation: identification of glycoforms that can bind to
mannan-binding lectin. *The Journal of biological chemistry* **280**, 29080-29087,
doi:10.1074/jbc.m504528200 (2005).
- 165 Colucci, M. *et al.* Sialylation of N-Linked Glycans Influences the Immunomodulatory Effects of IgM on T
Cells. *The Journal of Immunology* **194**, 151-157, doi:10.4049/jimmunol.1402025 (2015).
- 166 Thompson, R. E. & Muir, T. W. Chemoenzymatic Semisynthesis of Proteins. *Chem Rev* **120**, 3051-3126,
doi:10.1021/acs.chemrev.9b00450 (2020).
- 167 Arranz-Gibert, P., Vanderschuren, K. & Isaacs, F. J. Next-generation genetic code expansion. *Curr Opin
Chem Biol* **46**, 203-211, doi:10.1016/j.cbpa.2018.07.020 (2018).
- 168 Chin, J. W. *et al.* An Expanded Eukaryotic Genetic Code. *Science* **301**, 964-967,
doi:doi:10.1126/science.1084772 (2003).
- 169 Xie, J. & Schultz, P. G. A chemical toolkit for proteins — an expanded genetic code. *Nature Reviews
Molecular Cell Biology* **7**, 775-782, doi:10.1038/nrm2005 (2006).
- 170 Boutureira, O. & Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chemical Reviews* **115**,
2174-2195, doi:10.1021/cr500399p (2015).
- 171 Kent, S. B. H. Novel protein science enabled by total chemical synthesis. *Protein Sci* **28**, 313-328,
doi:10.1002/pro.3533 (2019).
- 172 Rich, J. R. & Withers, S. G. Emerging methods for the production of homogeneous human glycoproteins.
Nature Chemical Biology **5**, 206-215, doi:10.1038/nchembio.148 (2009).
- 173 Izumi, M. *et al.* Chemical Synthesis of Intentionally Misfolded Homogeneous Glycoprotein: A Unique
Approach for the Study of Glycoprotein Quality Control. *Journal of the American Chemical Society* **134**,
7238-7241, doi:10.1021/ja3013177 (2012).
- 174 Wang, L. X. & Amin, M. N. Chemical and chemoenzymatic synthesis of glycoproteins for deciphering
functions. *Chem Biol* **21**, 51-66, doi:10.1016/j.chembiol.2014.01.001 (2014).
- 175 Kajihara, Y. *et al.* Prompt chemoenzymatic synthesis of diverse complex-type oligosaccharides and its
application to the solid-phase synthesis of a glycopeptide with Asn-linked sialyl-undeca- and asialo-
nonasaccharides. *Chemistry* **10**, 971-985, doi:10.1002/chem.200305115 (2004).
- 176 Fischer, E. & Fourneau, E. Ueber einige Derivate des Glykocolls. *Berichte der deutschen chemischen
Gesellschaft* **34**, 2868-2877, doi:<https://doi.org/10.1002/cber.190103402249> (1901).
- 177 Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the
American Chemical Society* **85**, 2149-2154, doi:10.1021/ja00897a025 (1963).
- 178 Merrifield, R. B. Solid Phase Synthesis (Nobel Lecture). *Angewandte Chemie International Edition in
English* **24**, 799-810, doi:<https://doi.org/10.1002/anie.198507993> (1985).
- 179 Mueller, L. K., Baumruck, A. C., Zhdanova, H. & Tietze, A. A. Challenges and Perspectives in Chemical
Synthesis of Highly Hydrophobic Peptides. *Frontiers in Bioengineering and Biotechnology* **8**,
doi:10.3389/fbioe.2020.00162 (2020).
- 180 Gates, Z. P. & Hartrampf, N. Flow-based SPPS for protein synthesis: A perspective. *Peptide Science* **112**,
e24198, doi:<https://doi.org/10.1002/pep2.24198> (2020).
- 181 Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation.
Science **266**, 776-779, doi:10.1126/science.7973629 (1994).
- 182 Conibear, A. C., Watson, E. E., Payne, R. J. & Becker, C. F. W. Native chemical ligation in protein synthesis
and semi-synthesis. *Chemical Society Reviews* **47**, 9046-9068, doi:10.1039/C8CS00573G (2018).
- 183 Payne, R. J. & Wong, C.-H. Advances in chemical ligation strategies for the synthesis of glycopeptides
and glycoproteins. *Chemical Communications* **46**, 21-43, doi:10.1039/B913845E (2010).
- 184 Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through
expansion of the native chemical ligation concept. *Nature Reviews Chemistry* **2**, 0122,
doi:10.1038/s41570-018-0122 (2018).

References

- 185 Hackenberger, C. P. & Schwarzer, D. Chemoselective ligation and modification strategies for peptides
and proteins. *Angew Chem Int Ed Engl* **47**, 10030-10074, doi:10.1002/anie.200801313 (2008).
- 186 Durek, T. & Becker, C. F. W. Protein semi-synthesis: New proteins for functional and structural studies.
Biomolecular Engineering **22**, 153-172, doi:<https://doi.org/10.1016/j.bioeng.2005.07.004> (2005).
- 187 Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: a general method for protein engineering.
Proc Natl Acad Sci U S A **95**, 6705-6710, doi:10.1073/pnas.95.12.6705 (1998).
- 188 Machova, Z. & Beck-Sickinger, A. G. Expressed protein ligation for protein semisynthesis and
engineering. *Methods Mol Biol* **298**, 105-130, doi:10.1385/1-59259-877-3:105 (2005).
- 189 Blanco-Canosa, J. B. & Dawson, P. E. An efficient Fmoc-SPPS approach for the generation of thioester
peptide precursors for use in native chemical ligation. *Angew Chem Int Ed Engl* **47**, 6851-6855,
doi:10.1002/anie.200705471 (2008).
- 190 Fang, G.-M. *et al.* Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angewandte Chemie
International Edition* **50**, 7645-7649, doi:<https://doi.org/10.1002/anie.201100996> (2011).
- 191 Bondalapati, S., Jbara, M. & Brik, A. Expanding the chemical toolbox for the synthesis of large and
uniquely modified proteins. *Nature Chemistry* **8**, 407-418, doi:10.1038/nchem.2476 (2016).
- 192 Murale, D. P., Hong, S. C., Haque, M. M. & Lee, J.-S. Photo-affinity labeling (PAL) in chemical proteomics:
a handy tool to investigate protein-protein interactions (PPIs). *Proteome Science* **15**, 14,
doi:10.1186/s12953-017-0123-3 (2017).
- 193 Singh, A., Thornton, E. R. & Westheimer, F. H. The photolysis of diazoacetylchymotrypsin. *J Biol Chem*
237, 3006-3008 (1962).
- 194 Galardy, R. E., Craig, L. C., Jamieson, J. D. & Printz, M. P. Photoaffinity labeling of peptide hormone
binding sites. *J Biol Chem* **249**, 3510-3518 (1974).
- 195 Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid
to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**, 11020-11024,
doi:10.1073/pnas.172226299 (2002).
- 196 Chin, J. W. & Schultz, P. G. In Vivo Photocrosslinking with Unnatural Amino Acid Mutagenesis.
ChemBioChem **3**, 1135-1137, doi:[https://doi.org/10.1002/1439-7633\(20021104\)3:11<1135::AID-
CBIC1135>3.0.CO;2-M](https://doi.org/10.1002/1439-7633(20021104)3:11<1135::AID-CBIC1135>3.0.CO;2-M) (2002).
- 197 Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a
photoreactive amino acid. *Nature Methods* **2**, 201-206, doi:10.1038/nmeth739 (2005).
- 198 Schwarz, R., Tänzler, D., Ihling, C. H. & Sinz, A. Monitoring Solution Structures of Peroxisome Proliferator-
Activated Receptor β/δ upon Ligand Binding. *PLOS ONE* **11**, e0151412,
doi:10.1371/journal.pone.0151412 (2016).
- 199 Wang, P. *et al.* Erythropoietin Derived by Chemical Synthesis. *Science* **342**, 1357-1360,
doi:10.1126/science.1245095 (2013).
- 200 Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L. & Bernardes, G. J. L. Contemporary approaches to site-selective
protein modification. *Nature Reviews Chemistry* **3**, 147-171 (2019).
- 201 Liu, L., Bennett, C. S. & Wong, C.-H. Advances in glycoprotein synthesis. *Chemical Communications*, 21-
33, doi:10.1039/B513165K (2006).
- 202 Shin, Y. *et al.* Fmoc-Based Synthesis of Peptide- α Thioesters: Application to the Total Chemical
Synthesis of a Glycoprotein by Native Chemical Ligation. *Journal of the American Chemical Society* **121**,
11684-11689, doi:10.1021/ja992881j (1999).
- 203 Shandell, M. A., Tan, Z. & Cornish, V. W. Genetic Code Expansion: A Brief History and Perspective.
Biochemistry **60**, 3455-3469, doi:10.1021/acs.biochem.1c00286 (2021).
- 204 Lee, D. J., Cameron, A. J., Wright, T. H., Harris, P. W. R. & Brimble, M. A. A synthetic approach to 'click'
neoglycoprotein analogues of EPO employing one-pot native chemical ligation and CuAAC chemistry.
Chem Sci **10**, 815-828, doi:10.1039/c8sc03409e (2019).
- 205 Li, C. & Wang, L.-X. Chemoenzymatic Methods for the Synthesis of Glycoproteins. *Chemical Reviews* **118**,
8359-8413, doi:10.1021/acs.chemrev.8b00238 (2018).
- 206 Tang, F., Wang, L.-X. & Huang, W. Chemoenzymatic synthesis of glycoengineered IgG antibodies and
glycosite-specific antibody-drug conjugates. *Nature Protocols* **12**, 1702-1721,
doi:10.1038/nprot.2017.058 (2017).

References

- 207 Chao, Q. *et al.* Recent Progress in Chemo-Enzymatic Methods for the Synthesis of N-Glycans. *Frontiers in Chemistry* **8**, doi:10.3389/fchem.2020.00513 (2020).
- 208 Moremen, K. W. & Haltiwanger, R. S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology* **15**, 853-864, doi:10.1038/s41589-019-0350-2 (2019).
- 209 Tong, X., Li, T., Orwenyo, J., Toonstra, C. & Wang, L.-X. One-pot enzymatic glycan remodeling of a therapeutic monoclonal antibody by endoglycosidase S (Endo-S) from *Streptococcus pyogenes*. *Bioorganic & Medicinal Chemistry* **26**, 1347-1355, doi:<https://doi.org/10.1016/j.bmc.2017.07.053> (2018).
- 210 Du, J. J., Klontz, E. H., Guerin, M. E., Trastoy, B. & Sundberg, E. J. Structural insights into the mechanisms and specificities of IgG-active endoglycosidases. *Glycobiology* **30**, 268-279, doi:10.1093/glycob/cwz042 (2020).
- 211 Hsu, Y.-P. *et al.* Successive remodeling of IgG glycans using a solid-phase enzymatic platform. *Communications Biology* **5**, 328, doi:10.1038/s42003-022-03257-4 (2022).
- 212 Esposito, D., Hurevich, M., Castagner, B., Wang, C.-C. & Seeberger, P. H. Automated synthesis of sialylated oligosaccharides. *Beilstein Journal of Organic Chemistry* **8**, 1601-1609, doi:10.3762/bjoc.8.183 (2012).
- 213 Wu, B. *et al.* Synthesis of the fucosylated biantennary N-glycan of erythropoietin. *Tetrahedron Lett* **47**, 5577-5579, doi:<https://doi.org/10.1016/j.tetlet.2006.05.086> (2006).
- 214 Hussnaetter, K. P. *et al.* Strategies for Automated Enzymatic Glycan Synthesis (AEGS). *Biotechnology Advances* **67**, 108208, doi:<https://doi.org/10.1016/j.biotechadv.2023.108208> (2023).
- 215 Seko, A. *et al.* Occurrence of a sialylglycopeptide and free sialylglycans in hen's egg yolk. *Biochim Biophys Acta* **1335**, 23-32, doi:10.1016/s0304-4165(96)00118-3 (1997).
- 216 Alagesan, K. & Kolarich, D. Improved strategy for large scale isolation of sialylglycopeptide (SGP) from egg yolk powder. *MethodsX* **6**, 773-778, doi:10.1016/j.mex.2019.04.007 (2019).
- 217 Liu, L., Prudden, A. R., Bosman, G. P. & Boons, G.-J. Improved isolation and characterization procedure of sialylglycopeptide from egg yolk powder. *Carbohydrate Research* **452**, 122-128, doi:<https://doi.org/10.1016/j.carres.2017.10.001> (2017).
- 218 Zou, Y. *et al.* An Efficient Approach for Large-Scale Production of Sialylglycopeptides from Egg Yolks. *Journal of Carbohydrate Chemistry* **31**, 436-446, doi:10.1080/07328303.2012.666689 (2012).
- 219 Eshima, Y., Higuchi, Y., Kinoshita, T., Nakakita, S. & Takegawa, K. Transglycosylation Activity of Glycosynthase Mutants of Endo- β -N-Acetylglucosaminidase from *Coprinopsis cinerea*. *PLoS One* **10**, e0132859, doi:10.1371/journal.pone.0132859 (2015).
- 220 Manabe, S., Yamaguchi, Y., Abe, J., Matsumoto, K. & Ito, Y. Acceptor range of endo- β -N-acetylglucosaminidase mutant endo-CC N180H: from monosaccharide to antibody. *Royal Society Open Science* **5**, 171521, doi:10.1098/rsos.171521 (2018).
- 221 Zhang, Z. & Marshall, A. G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J Am Soc Mass Spectrom* **9**, 225-233, doi:10.1016/s1044-0305(97)00284-5 (1998).
- 222 Iris Biotech GmbH. 216 (Iris Biotech GmbH, <https://iris-biotech.de/resin-guideline>, 2023).
- 223 Schnölzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S. B. In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* **40**, 180-193, doi:10.1111/j.1399-3011.1992.tb00291.x (1992).
- 224 Merck Biosciences AG. Novabiochem Fmoc resin cleavage guide. 6 (2012). <https://www.emdmillipore.com/Web-US-Site/en_CA/-/USD/ShowDocument-Pronet?id=201503.030>.
- 225 Merck Biosciences AG. Novabiochem Boc resin cleavage and deprotection. (2007).
- 226 Gajsek, O. *Semi-synthesis as a tool for the exploration of protein structure and function* Master of Science thesis, University of Vienna, (2019).
- 227 Sun, B. *et al.* A simplified procedure for gram-scale production of sialylglycopeptide (SGP) from egg yolks and subsequent semi-synthesis of Man3GlcNAc oxazoline. *Carbohydr Res* **396**, 62-69, doi:10.1016/j.carres.2014.07.013 (2014).
- 228 Schmid, A. *Synthesis of homogeneously N-glycosylated soluble Fas ligand variants*, (2022).
- 229 Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680-685, doi:10.1038/227680a0 (1970).

References

- 230 Fairbanks, G., Steck, T. L. & Wallach, D. F. H. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* **10**, 2606-2617, doi:10.1021/bi00789a030 (1971).
- 231 Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim Biophys Acta* **1751**, 119-139, doi:10.1016/j.bbapap.2005.06.005 (2005).
- 232 Chen, J., Wan, Q., Yuan, Y., Zhu, J. & Danishefsky, S. J. Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angewandte Chemie International Edition* **47**, 8521-8524, doi:<https://doi.org/10.1002/anie.200803523> (2008).
- 233 Thompson, R. E., Chan, B., Radom, L., Jolliffe, K. A. & Payne, R. J. Chemoselective Peptide Ligation–Desulfurization at Aspartate. *Angewandte Chemie International Edition* **52**, 9723-9727, doi:<https://doi.org/10.1002/anie.201304793> (2013).
- 234 Araman, C. *et al.* Semisynthetic prion protein (PrP) variants carrying glycan mimics at position 181 and 197 do not form fibrils. *Chem Sci* **8**, 6626-6632, doi:10.1039/c7sc02719b (2017).
- 235 Telenti, A. *et al.* The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol* **179**, 6378-6382, doi:10.1128/jb.179.20.6378-6382.1997 (1997).
- 236 Evans, T. C., Jr., Benner, J. & Xu, M. Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**, 2256-2264, doi:10.1002/pro.5560071103 (1998).
- 237 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**, 1321-1325, doi:10.1038/nbt1188-1321 (1988).
- 238 Akagi, K.-i. *et al.* Identification of the Substrate Interaction Region of the Chitin-Binding Domain of Streptomyces griseus Chitinase C. *The Journal of Biochemistry* **139**, 483-493, doi:10.1093/jb/mvj062 (2006).
- 239 Wang, N. *et al.* Non-enzymatic reaction of glycosyl oxazoline with peptides. *Carbohydrate Research* **436**, 31-35, doi:<https://doi.org/10.1016/j.carres.2016.11.002> (2016).
- 240 Parsons, T. B. *et al.* Optimal Synthetic Glycosylation of a Therapeutic Antibody. *Angew Chem Int Ed Engl* **55**, 2361-2367, doi:10.1002/anie.201508723 (2016).

References

12 Appendix

12.1 Sequences

12.1.1 DNA Sequence of MxeIntein-H7-CBD

>MxeIntein-H7-CBD

```

TGC ATC ACG GGA GAT GCA CTA GTT GCC CTA
CCC GAG GGC GAG TCG GTA CGC ATC GCC GAC
ATC GTG CCG GGT GCG CGG CCC AAC AGT GAC
AAC GCC ATC GAC CTG AAA GTC CTT GAC CGG
CAT GGC AAT CCC GTG CTC GCC GAC CGG CTG
TTC CAC TCC GGC GAG CAT CCG GTG TAC ACG
GTG CGT ACG GTC GAA GGT CTG CGT GTG ACG
GGC ACC GCG AAC CAC CCG TTG TTG TGT TTG
GTC GAC GTC GCC GGG GTG CCG ACC CTG CTG
TGG AAG CTG ATC GAC GAA ATC AAG CCG GGC
GAT TAC GCG GTG ATT CAA CGC AGC GCA TTC
AGC GTC GAC TGT GCA GGT TTT GCC CGC GGG
AAA CCC GAA TTT GCG CCC ACA ACC TAC ACA
GTC GGC GTC CCT GGA CTG GTG CGT TTC TTG
GAA GCA CAC CAC CGA GAC CCG GAC GCC CAA
GCT ATC GCC GAC GAG CTG ACC GAC GGG CGG
TTC TAC TAC GCG AAA GTC GCC AGT GTC ACC
GAC GCC GGC GTG CAG CCG GTG TAT AGC CTT
CGT GTC GAC ACG GCA GAC CAC GCG TTT ATC

```

ACG AAC GGG TTC GTC AGC CAC GCT ACT GGC
 CTC ACC GGA ATT CAC CAC CAC CAC CAC CAC
 CAC TCC GGT CTG AAC TCA GGC CTC ACG ACA
 AAT CCT GGT GTA TCC GCT TGG CAG GTC AAC
 ACA GCT TAT ACT GCG GGA CAA TTG GTC ACA
 TAT AAC GGC AAG ACG TAT AAA TGT TTG CAG
 CCC CAC ACC TCC TTG GCA GGA TGG GAA CCA
 TCC AAC GTT CCT GCC TTG TGG CAG CTT CAA
 TGA

12.1.2 AA Sequence of MxeIntein-H7-CBD

>MxeIntein-H7-CBD

CITGD ALVAL PEGES VRIAD IVPGA RPNSD NAIDL KVLDR HGNPV LADRL
 FHSGE HPVYT VRTVE GLRVT GTANH PLLCL VDVAG VPTLL WKLID EIKPG
 DYAVI QRSFAF SVDCA GFARG KPEFA PTTYT VGVPG LVRFL EAHHR DPDAQ
 AIADE LTDGR FYYAK VASVT DAGVQ PVYSL RVDTA DHAFI TNGFV SHATG
 LTGIH HHHHH HSGLN SGLTT NPGVS AWQVN TAYTA GQLVT YNGKT YKCLQ
 PHTSL AGWEP SNVPA LWQLQ

12.1.3 DNA Sequence of Cμ4tp 1-91-MxeIntein-H7-CBD

>CH4tp 1-91-MxeIntein-H7-CBD

```

ATG GGT GAA GTT CAT AAA CAT CCG CCT GCA
GTT TAT CTG CTG CCT CCG GCA CGT GAA CAG
CTG AAT CTG CGT GAA AGC GCA ACC GTT ACC
TGT CTG GTTs AAA GGT TTT AGT CCG GCA GAT
ATT AGC GTT CAG TGG CTG CAG CGT GGT CAG
CTG CTG CCG CAA GAA AAA TAT GTT ACC AGC
GCA CCG ATG CCG GAA CCG GGT GCT CCG GGT
TTC TAC TTT ACC CAT AGC ATT CTG ACC GTG
ACC GAA GAG GAA TGG AAT AGC GGT GAA ACC
TAT ACC TGC ATC ACG GGA GAT GCA CTA GTT
GCC CTA CCC GAG GGC GAG TCG GTA CGC ATC
GCC GAC ATC GTG CCG GGT GCG CGG CCC AAC
AGT GAC AAC GCC ATC GAC CTG AAA GTC CTT
GAC CGG CAT GGC AAT CCC GTG CTC GCC GAC
CGG CTG TTC CAC TCC GGC GAG CAT CCG GTG
TAC ACG GTG CGT ACG GTC GAA GGT CTG CGT
GTG ACG GGC ACC GCG AAC CAC CCG TTG TTG
TGT TTG GTC GAC GTC GCC GGG GTG CCG ACC
CTG CTG TGG AAG CTG ATC GAC GAA ATC AAG
CCG GGC GAT TAC GCG GTG ATT CAA CGC AGC
GCA TTC AGC GTC GAC TGT GCA GGT TTT GCC
CGC GGG AAA CCC GAA TTT GCG CCC ACA ACC

```

Appendix

TAC ACA GTC GGC GTC CCT GGA CTG GTG CGT
TTC TTG GAA GCA CAC CAC CGA GAC CCG GAC
GCC CAA GCT ATC GCC GAC GAG CTG ACC GAC
GGG CGG TTC TAC TAC GCG AAA GTC GCC AGT
GTC ACC GAC GCC GGC GTG CAG CCG GTG TAT
AGC CTT CGT GTC GAC ACG GCA GAC CAC GCG
TTT ATC ACG AAC GGG TTC GTC AGC CAC GCT
ACT GGC CTC ACC GGA ATT CAC CAC CAC CAC
CAC CAC CAC TCC GGT CTG AAC TCA GGC CTC
ACG ACA AAT CCT GGT GTA TCC GCT TGG CAG
GTC AAC ACA GCT TAT ACT GCG GGA CAA TTG
GTC ACA TAT AAC GGC AAG ACG TAT AAA TGT
TTG CAG CCC CAC ACC TCC TTG GCA GGA TGG
GAA CCA TCC AAC GTT CCT GCC TTG TGG CAG
CTT CAA TGA

12.1.4 AA Sequence of Cμ4tp 1-91-MxeIntein-H7-CBD

>CH4tp 1-91-MxeIntein-H7-CBD

GEVHK HPPAV YLLPP AREQL NLRES ATVTC LVKGF SPADI SVQWL QRGQL
LPQEK YVTSA PMPEP GAPGF YFTHS ILTVT EEEWN SGETY TCITG DALVA
LPEGE SVRIA DIVPG ARPNS DNAID LKVLD RHGNP VLADR LFHSG EHPVY
TVRTV EGLRV TGTAN HPLLC LVDVA GVPTL LWKLI DEIKP GDYAV IQRSA
FSVDC AGFAR GKPEF APTTY TVGVP GLVRF LEAHH RDPDA QAIAD ELTDG
RFYYA KVASV TDAGV QPVYS LRVDI ADHAF ITNGF VSHAT GLTGI HHHHH

HHSGL NSGLT TNPGV SAWQV NTAYT AGQLV TYNGK TYKCL QPHTS LAGWE
PSNVP ALWQL Q

12.1.5 AA Sequence of Cμ4tp 1-132 wt

>CH4tp 1-132 wt

1 **GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL**
51 **LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP**
101 **HLVTERTVDK STGKPTLYNV SLIMSDTGGT CY**

12.1.6 AA Sequence of Cμ4tp 1-132 L117Bpa

>CH4tp 1-132 L117Bpa

1 **GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL**
51 **LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP**
101 **HLVTERTVDK STGKPTXYNV SLIMSDTGGT CY**

X represents Bpa

12.1.7 AA Sequence of Cμ4tp 1-132 Y118Bpa

>CH4tp 1-132 Y118Bpa

1 **GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL**
51 **LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP**
101 **HLVTERTVDK STGKPTLXNV SLIMSDTGGT CY**

X represents Bpa

12.1.8 AA Sequence of C_μ4tp 1-132 V120Bpa

>CH4tp 1-132 V120Bpa

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTLYNX SLIMSDTGGT CY

```

*X*represents Bpa12.1.9 AA Sequence of C_μ4tp 1-132 Y132Bpa

>CH4tp 1-132 Y132Bpa

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTLYNV SLIMSDTGGT CX

```

*X*represents Bpa12.1.10 AA Sequence of C_μ4tp 1-132 N119GlcNAc

>CH4tp 1-132 N119GlcNAc

```

1      GEVHKHPPAV YLLPPAREQL NLRESATVTC LVKGFSPADI SVQWLQRGQL
51     LPQEKYVTSA PMPEPGAPGF YFTHSILTVT EEEWNSGETY TCVVGHEALP
101    HLVTERTVDK STGKPTXYXV SLIMSDTGGT CY

```

*X*represents NGlcNAc

12.2 Additional Photocrosslinking Data

12.2.1 40 min Photocrosslinking

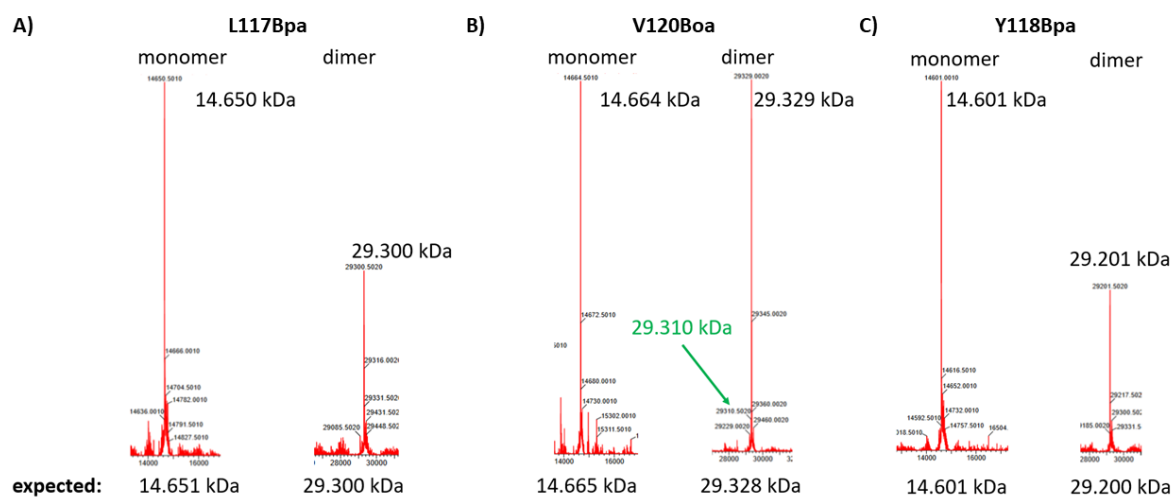


Figure 59: Deconvoluted MS of $C\mu 4tp$ variants after 40 min of irradiation A) L117Bpa B) V120Boa C) Y118Bp