Structural and functional investigation of Glycocin-Glycosyltransferases

Dissertation

zur Erlangung des Grades eines
Doktors (Dr. rer. nat.)

am Fachbereich Biologie, Chemie, Pharmazie

der Freien Universität Berlin

vorgelegt von

Michael Krummhaar

Berlin, April 2023

This work was performed between June 2019 and November 2023 at the Max Planck institute of colloids and interfaces, Potsdam, Germany and the Freie Universität Berlin, Germany under the supervision of Dr. Christian Roth and Prof. Dr. Beate Koksch.

Erstgutachter: Dr. Christian Roth (Max Planck Institute of Colloids and Interfaces, Potsdam)

Zweitgutachter: Prof. Dr. Beate Koksch (Freie Universität Berlin)

Tag der Disputation: 29.04.2024

Acknowledgements

This thesis marks, in a manner of speaking, the end of a long, arduous, and yet mostly pleasant journey. And as with all journeys there are lots of people to feel thankful for once it ends. As it feels unjust to rank all the people who helped me over the years in various ways, I shall thank them chronologically and the order of their appearance is in no way reflective of the amount I appreciate their help.

Firstly are, of course, my parents, who always encouraged my curiosity and without whom I would most definitely be where I am today. Similarly, I would like to thank my siblings for always being there when I needed them. I also wish to thank my very first supervisor in a laboratory, Dr. Jan-Hinrich Guse, whose advice resonates with me to this day. “The most important skill of a scientist is the ability to deal with frustration”, indeed.

For the purpose of this work I would like to firstly thank Dr. Christian Roth, who has been an amazing supervisor during the time I spent working, who supported me whenever I needed help and was willing to entertain even some of my more outlandish ideas I had during my time. Thank you very much.

Secondly, I wish to thank Prof. Dr. Seeberger for the opportunity to work in the Max Planck Institute of colloids and interfaces, as well as Prof. Dr. Beate Koksch for letting me work in her Lab on the peptide part of the project and for being my second supervisor.

Dr. Alexander Langhans and Dr. Valentina Stuhlberg I would like to thank for the great amount of help they were in my efforts to tackle the synthesis of the peptides, for the advice they gave me, as well as them being to listen when I was voicing my frustration about their behaviour. Thank you once again, and sorry for the bother.

I also would like to thank Dr. Bernhard Loll and the kind people joint MX lab in Berlin and the staff of the BESSY beamline, without whom my crystals would have been useless.

I would also like to thank all the people who have worked me in Dahlem, with special thanks to Ruben Ananian and Anika Freitag for many fruitful discussions over the years, not always of a scientific nature.

Finally, I would like you, the reader. Without anyone to read it, the written word is pointless, after all.

Statutory declaration

Hierdurch versichere ich, dass ich meine Dissertation selbstständig verfasst und keine anderen als die von mir angegeben Quellen und Hilfsmittel verwendet habe.

Ich bestätige, dass die elektronische Kopie den gedruckten Kopien gleicht.
Die Dissertation ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

I confirm that the electronic copy is identical to the bound copies of the thesis.
I also declare that I have not submitted the dissertation in this or any other form to any other institution as a dissertation.

Berlin, 30th April 2024 Michael Krummhaar

Contents

Acknowledgements I
Statutory declaration II
Contents III
List of publications VII
Oral presentations VII
Poster presentations VII
Abstract VIII
Zusammenfassung IX
List of abbreviations X
1. Introduction 1
1.1. Bacteriocins and RiPPs 1
1.1.1. Glycocin F-like Glycocins 7
1.1.2. Sublancin-like Glycocins 9
1.1.3. Enterocin 96 11
1.1.4. Other glycocins and glycosylated bacteriocins 13
1.2. Peptide glycosylation 15
1.3. Leloir type glycosyltransferases 20
1.4. Solid phase peptide synthesis 27
2. Aim 36
3. Results 37
3.1. Discovery of putative glycocins and their cognate GTs 37
3.2. Heterologous expression of Glycosyltransferases 40
3.2.1. GccA 40
3.2.2. SacS 41
3.2.3. SunS 43
3.2.4. AciS 44
3.2.5. EntS 46
3.2.6. PltS 48
3.3. Chemical synthesis of the peptide aglycones 51
3.3.1. Synthesis of SacA based peptides 52
3.3.2. Synthesis of SunA derived peptides 56
3.3.3. Synthesis of AciA 56
3.3.4. Synthesis of Ent96 derived peptides 57
3.3.5. Synthesis of PltA and derived peptides 59
3.3.6. Stability of the synthesized peptides 60

3.4. Determination of cosubstrates 61
3.5. Glycosylation of synthetic peptide substrates 63
3.5.1. Glycosylation of SacA 63
3.5.2. Glycosylation of SunA minimal peptide 68
3.5.3. Glycosylation of Enterocin 96 and its minimal recognition sequence 69
3.5.4. Glycosylation of PltA 71
3.6. Acceptor peptide specificity 73
3.7. Reaction kinetics for SacS 76
3.8. An in vivo production method for glycocins 78
3.9. Antimicrobial activity 85
3.10. Crystallization of glycosyltransferases 87
3.11. Crystal structures of Glycocin Glycosyltransferases 95
3.11.1. Overall fold 96
3.11.2. The GT-A fold domain 101
3.11.3. Dimerization domain 102
3.11.4. Donor binding 103
3.11.5. Acceptor peptide binding of EntS 108
3.11.7. Proposed reaction mechanism 113
3.11.9. Comparison between the S-preferential SunS/SacS and O-preferential EntS 116
3.12. Insights into the folding of glycocins 119
4. Discussion 121
4.1. Discovery of glycocins 121
4.2. Synthesis of glycocins 123
4.3. Crystalization and structure determination 127
5. Conclusion and Outlook 135
6. Materials and Methods 137
6.1. Reagents 137
6.2. Growth media 138
6.3. Buffers 139
6.4. Antibiotics 141
6.5. Primers 141
6.6. Bacterial strains 143
6.7. Genome mining 143
6.8. Molecular biology methods 144
6.8.1. Isolation of Genomic DNA 144
6.8.2. Polymerase chain reaction 144

6.8.3. Purification of DNA 145
6.8.4. Ligation 145
6.8.5. Transformation 146
6.8.6. Plasmid purification 147
6.8.7. Site Directed mutagenesis 147
6.9. Biochemical methods 147
6.9.1. Determiation of protein concentration 147
6.9.2. SDS-PAGE 148
6.9.3. Western blotting 148
6.9.4. Isothermal titration calorimetry 149
6.9.5. Differential scanning fluorimetry 149
6.10. Lyophilization 150
6.11. Production and purification of Glycosyltransferases 150
6.11.1. Protein production 150
6.11.2. Lysis of cells 151
6.11.3. Affinity chromatography 151
6.11.4. Size exclusion chromatography 152
6.11.5. Ion exchange chromatography 152
6.11.6. Proteases 152
6.11.7. Purification of GccA 153
6.11.8. Purification of SacS 153
6.11.9. Purification of SunS 154
6.11.10. Purification of AciS 154
6.11.11. Purification of EntS 154
6.11.12. Purification of PltS 155
6.11.13. Storage of proteins 155
6.12. Peptide synthesis 156
6.12.1. Microwave methods and cycles 157
6.12.2. Reaction control 158
6.12.3. Full cleavage and deprotection of peptides 159
6.13. Peptide purification and analysis 159
6.13.1. Analytical HPLC 159
6.13.2. Mass spectrometry 160
6.13.3. Preparative HPLC 161
6.14. Prepared peptides 162
6.14.1. SacA derived peptides 162

6.14.2. SunA derived peptides 164
6.14.3. Enterocin 96 derived peptides 166
6.14.4. PltA derived peptides 168
6.15. Activity tests of GT 170
6.15.1. Qualitative activity test 170
6.15.2. Malachite green assay 170
6.15.3. Enzyme kinetics 171
6.15.4. Hydrolysis assay 171
6.16. Preparation of glycosylated glycocins 172
6.17. In vivo production of glycocins 172
6.17.1. Coexpression of glycocin with its cognate GTase 172
6.17.2. Glycocin purification 173
6.18. Test for antimicrobial activity 174
6.19. Crystallization 175
6.19.1. Crystallization of SacS348 175
6.19.2. Crystalizazion of SunS 175
6.19.3. Crystalizazion of EntS 176
6.19.4. Crystallization of PltS 177
6.20. X-ray diffraction 178
6.20.1. Data collection 178
6.20.2. Structure solution 178
6.21. CD Spectroscopy 179
7. Bibliography 180
8. Annex I 192
8.1. Peptide Data 192
8.1.1. SacA derived peptides 192
8.1.2. SunA derived peptides 194
8.1.3. Enterocin derived peptides 196
8.1.4. PltA derived peptides 198
8.2. Glycosylated Peptides 200
8.2.1. SacA derived peptides 200
8.2.2. SunA derived peptides 201
8.2.3. Enterocin 96 derived peptides 201
8.2.4. PltA derived peptides 203

List of publications

-

Oral presentations

23rd Heart of Europe crystallography meeting – online

24th Heart of Europe crystallography meeting – Lipno, Czech Republic

25th Heart of Europe crystallography meeting – Malchin, Germany

IMPRS of multiscale bio-systems autumn workshop 2022 – Potsdam, Germany

IMPRS of multiscale bio-systems spring workshop 2022 – online

IMPRS of multiscale bio-systems autumn workshop 2021 – online

IMPRS of multiscale bio-systems spring workshop 2021 – online

IMPRS of multiscale bio-systems autumn workshop 2020 – online

IMPRS of multiscale bio-systems spring workshop 2020 – online

Poster presentations

33rd European crystallography meeting – Versailles, France

FEBS course – biomolecules in action 2022 – Hamburg, Germany

Abstract

Antimicrobial peptides are an interesting class on ribosomally and non-ribosomally synthesised peptides found in all kingdoms of life. Among these is the underexplored class of bacteriocins: Short ribosomally synthesized and post-translationally modified peptides that often show high potency, and low toxicity. One subclass requires a glycosylation to show antimicrobial activity and thus have been termed glycoactive bacteriocins, Glycocins.

The glycosylation of glycocins is installed by a specific subfamily of glycosyltransferases, part of the CaZy transferase family 2. Despite being in the same family, these glycosyltransferases show a remarkable diversity in terms of donor specificity, peptide specificity, acceptor selectivity and number of glycosylations that are carried out.

I identified several putative Glycocins and their cognate transferases using bioinformatics. The recombinant production and purification of several glycosyltransferases was established. These were characterised in regard of the metal ion dependency and the carbonucleotide specificity.

Using SPPS several glycocin-peptides were synthesized and their selective glycosylation by the cognate transferase could be proven. A facile and scalable method to obtain the mature, glycosylated, and folded glycocin by heterologous expression of the peptide and its transferase in *E. coli* was established. To investigate the molecular determinants for the observed differences in specificity in terms of preferred sugars and their acceptor selectivity, X-ray crystallography was used as method of choice. For four glycosyltransferases with noticeably different specificities, crystals could be obtained, and the structure could be solved. For the transferase of Enterocin 96 it was possible to obtain snapshots of the enzyme-substrate complex during its reaction.

Zusammenfassung

Antimikrobielle Peptide stellen eine faszinierende Klasse von ribosomal und nicht-ribosomal synthetisierten Peptiden dar, die in allen Lebensreichen vorkommen. Darunter fällt die wenig erforschte Gruppe der Bakteriozine: kurze ribosomal synthetisierte und posttranslationell modifizierte Peptide, die oft eine hohe Wirksamkeit und geringe Toxizität aufweisen. Eine Untergruppe erfordert eine Glykosylierung, um antimikrobielle Aktivität zu zeigen, und wurde daher als glykoaktive Bakteriozine, Glykozine, bezeichnet.

Die Glykosylierung von Glykozinen wird durch eine spezifische Unterfamilie von Glykosyltransferasen eingeführt, zur CaZy-Transferase-Familie 2 gehören. Trotz der Zugehörigkeit zur gleichen Familie zeigen diese Glykosyltransferasen eine bemerkenswerte Vielfalt in Bezug auf Donorspezifität, Peptidspezifität, Akzeptorselektivität und die Anzahl der durchgeführten Glykosylierungen.

Unter zuhilfenahme bioinformatischer Methoden war es möglich mehrere potenzielle Glykozine und ihre kognitiven Transferasen zu identifizieren. Die rekombinante Produktion und Reinigung mehrerer Glykosyltransferasen wurde etabliert. Diese wurden hinsichtlich ihrer Metallionenabhängigkeit und der Nukleotidspezifität charakterisiert.

Mithilfe der Festphasensynthese wurden mehrere Glykozin-Peptide hergestellt, und ihre selektive Glykosylierung durch die kognitiven Transferasen konnte nachgewiesen werden. Es wurde eine einfache und skalierbare Methode zur Gewinnung des reifen, glykosylierten und gefalteten Glykozins durch heterologe Expression des Peptids und seiner Transferase in E. coli etabliert.

Um die molekularen Determinanten für die beobachteten Unterschiede in Bezug auf bevorzugte Zucker und ihre Akzeptorselektivität zu untersuchen, wurde die Röntgenkristallographie als bevorzugte Methode verwendet. Für vier Glykosyltransferasen mit deutlich unterschiedlichen Spezifitäten konnten Kristalle erhalten und die Struktur gelöst werden. Für die Transferase von Enterocin 96 war es möglich, Schnappschüsse des Enzym-Substrat-Komplexes während seiner Reaktion zu erhalten.

List of abbreviations

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophane
Tyr	Tyrosine
Val	Valine
NDP	Nucleotide diphosphate
UDP	Uridine diphosphate
ADP	Adenine diphosphate
CDP	Cytidine diphosphate
GDP	Guanidine diphosphate
Glc	Glucose
GlcNAc	*N*-acetyl glucosamine
Gal	Galactose
Man	Mannose
Rib	Ribose
Xyl	Xylose
APS	Ammonium peroxodisulphate
Boc	*tert*-butyloxycabonyl
CBD	Chitin binding domain
CD	Circular dichroism
CLPC	Cross-linked protein crystal
DCM	Dichloromethane
Dha	Dehydroalanine
DIC	Di-isopropyl carbodiimide
DIPEA	Di-isopropyl ethyl amine
DMF	N,N-dimethyl formamide
DSF	Differential scanning fluorimetry
DTT	Di-thio threitol
Ent 96	Enterocin 96
Fmoc	Fluorenylmethyloxycarbonyl

GccF	Glycocin F
Gdn-Hcl	Guanidinium hydrochloride
GGT	Glycocin glycosyltransferase
GT	Glycosyltransferase
HFIP	Hexafluoroisopropanol
HOBt	1-Hydroxybenzotriazole
HPLC	High performance liquid chromatography
IEX	Ion exchange
IMAC	Immobilized metal ion affinity chromatography
IPTG	isopropyl-β-D-thiogalactoside
ITC	Isothermal titration calorimetry
LB	Lysogeny Broth
MBP	Maltose binding protein
MeCN	Acetonitrile
MIC	minimal inhibitory concentration
MME	monomethyl ether
MS	Mass spectrometry
NMP	N-methyl pyrrolidone
OD	Optical density
OGT	O-GlcNAc glycosyltransferase
OMpe	O-3-methyl-pent-3-yl
ORF	Open reading frame
PEG	Poly(ethylene glycol)
PTM	Post-translational modification
PTS	phosphoenolpyruvate-phosphotransferase system
RiPP	Ribosomally synthesized post-translationally modified peptide
R.m.s.d	Root mean square distance
RP-HPLC	Reverse phase – high performance liquid chromatography
SDS-PAGE	Sodium dodecylsulphate – Polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SOB	Super optimal broth
SPPS	Solid phase peptide synthesis
SUMO	Small ubiquitin-like modifier
Sublancin, SunA	Sublancin 168
TB	Terrific Broth
*t*Bu	*tert*-butyl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TIS	Tri-isopropyl silane
TPR	Tetratricopeptide repeat
Trt	Tritryl

1. Introduction

1.1. Bacteriocins and RiPPs

The discovery of antibiotics is one of the most important discoveries in medicine and has transformed human society by improving the ability to treat infections. However, the wide and often unregulated use of antibiotics lead to a rapid increase of resistant pathogenic bacteria with some having acquired resistance to all clinically available antibiotics. Thus, there is a worrying prospect of an insufficient availability of effective antimicrobial agents in the future, especially since there has been a scarcity of new families of antimicrobial drugs within the last decades[2].

Given that many of the currently used antibiotics have had their origins in natural products[3], further analysis of thus far underexplored natural product species may result in the discovery of new families of antibiotics with new modes of action. One promising class of such natural products are the bacteriocins. Bacteriocins are ribosomally synthesized antimicrobial peptides of bacterial origin, first discovered in 1928[4]. While these peptides often show only a narrow spectrum of bactericidal or bacteriostatic activity[5], there have also been reports of broad range antimicrobial activity[6], making them interesting potential new lead compounds. One of the benefits of bacteriocins is their ribosomal origin and therefore their presence and predictability at the DNA-level. Together with the tendency of proteins involved in their maturation to cluster together, this property allows for the facile prediction and identification of gene clusters *in silico*[7], enables heterologous production[8], and also allows for straightforward approaches to generate engineered variants with improved efficacy or solubility. An example of which is LFF571, an engineered bacteriocin of the thiopeptide family that could show its efficacy in clinical trials[9].

Bacteriocins have been discovered in all bacterial species, which led to the conclusion that almost all bacteria and archaea produce at least one type of bacteriocin[6]. Currently, bacteriocins are mostly classified based on the origin of the producer (gram positive/negative or archea). Though attempts have been made to develop a bacteriocin classification scheme based on specific properties. However, this has been unsuccessful so far, due to their vast structural and functional diversity and partially overlapping features. The first proposed system of Klaenhammer (1993)[10] classified them by the presence of a post-translational modification (PTM) (class I), size < 10 kDa (class II) or >30 kDa (class III) and the presence of a carbohydrate or lipid moiety (class IV). However, it became quickly apparent

that a number of the known bacteriocins would fall into multiple classes. In light of this, the system was revised first by Cotter *et al* (2005)[11], who sorted them generally into class I – modified, and class II – not modified peptides. This was further expanded and revised by Heng *et al* (2007)[12], who introduced a class IV of cyclized peptides, but changed class I to only include lanthipeptides, which leaves more recent additions to the bacteriocins outside of this classification scheme. The most recent classification systems have been proposed by Cotter *et al* (2013)[13], Alvarez-Sieiro *et al* (2016)[14] and Simons *et al* (2020)[15]. In some parts of the literature these classifications are combined, following the classification of Cotter (2013) and Heng (2007) simultaneously[16], creating confusion in the community. In this thesis the classification of Alvarez-Sieiro, which focuses on bacteriocins of lactic acid bacteria (Figure 1) and is well in agreement with the Cotter et al (2013) classification is used. This classification generally distinguishes between modified peptides (class I), unmodified peptides (class II) and large peptides (class III). Class I bacteriocins are therefore part of the ribosomally synthesized post-translationally modified peptides (RiPPs).

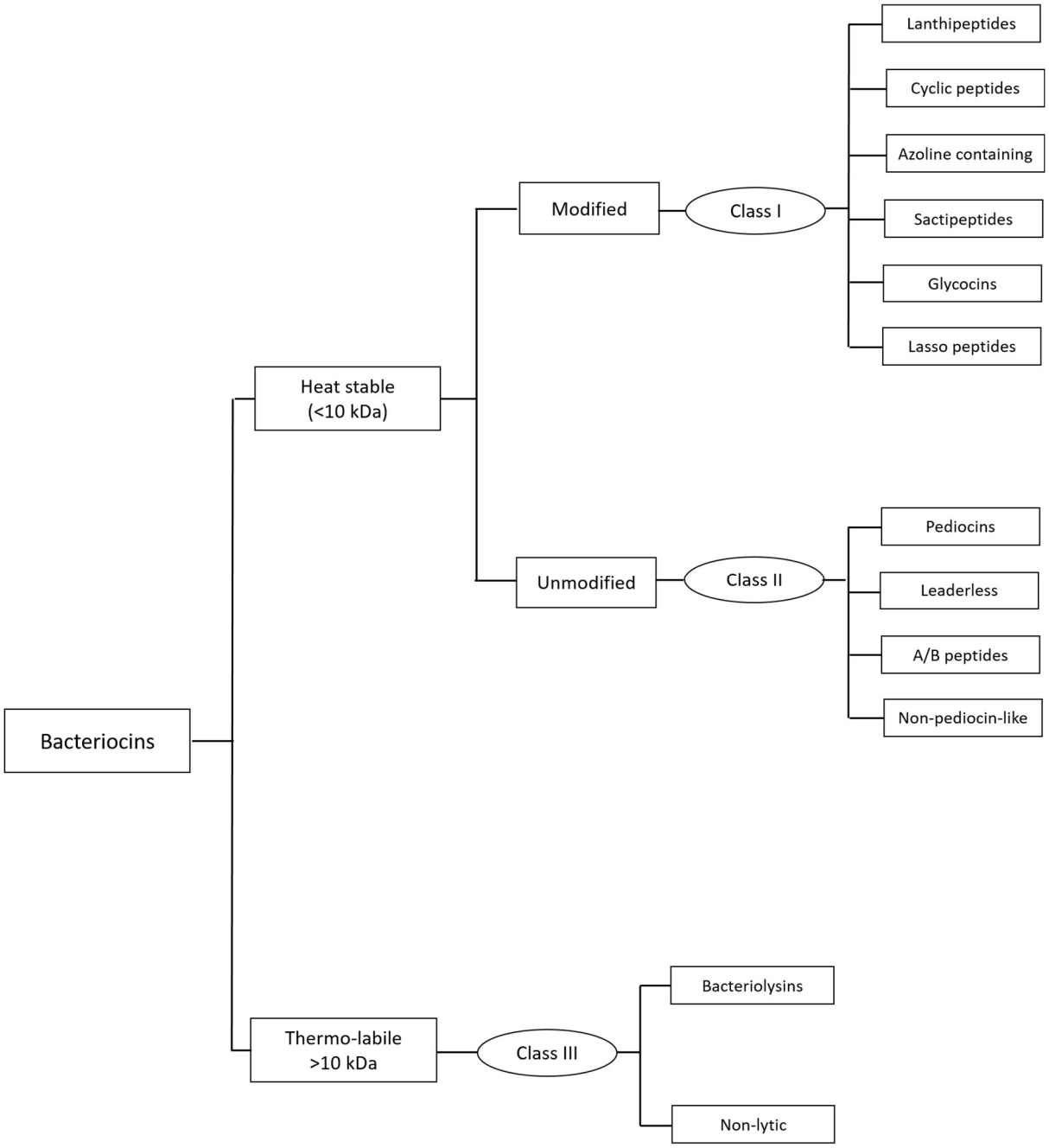


Figure 1 *Classification of bacteriocins produced by lactic acid bacteria according to Alvarez-Sieiro et al. (2016)*[14]

Generally, the RiPPs are produced as a precursor peptide, containing an N-terminal "leader" sequence. This leader sequence can have multiple functions, such as targeting the peptide towards the export mechanism or its PTM introducing machinery[17]. The leader sequence is most often of the so-called "double-glycine" type[18]; a common signal for ABC-type exporters and includes a proteolytic cleavage site between its glycines[19]. A more rarely seen feature of the precursor peptides is a C-terminal follower sequence after the core peptide, which is currently described only for the bottromycin, borosin, amatoxin and pantocin A subclasses[17]. After the ribosomal production and post-translational

modification, the bacteriocin it is exported, by a dedicated exporter cassette. This step is accompanied by the proteolytic cleavage of the leader and follower sequence from the peptide to afford the mature bacteriocin (Figure 2).

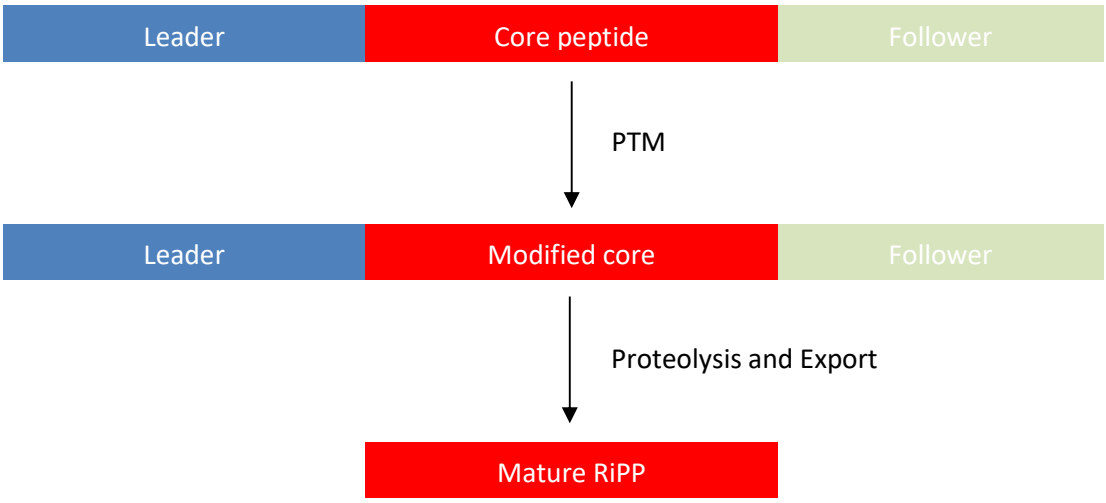


Figure 2 *Schematic representation of the biosynthesis of RiPPs.*

Similar to the diversity of their architecture and sequence, their mode of action is equally diverse. One of the most common mechanisms is the disruption of the cell wall or cell membranes. Siamycin-I, a lasso-peptide, as well as many lanthipeptides for instance primarily target the lipid II molecule of the peptidoglycan[20], impairing the membrane function [21], or using the lipid II as an anchor to form a pore in the cell membrane[22]. Other lanthionine-containing peptides such as duramycin and cinnamycin have been shown to directly interact with the membrane lipids, disrupting the cell wall[23]. Microcin E492, another lasso-peptide on the other hand enters the periplasm of gram-negative cells using the TonB transporter and disrupts the inner membrane of these cells. Interestingly the affinity to the required TonB transporter can be increased by the presence of a C-terminal glycoside[24]. Pediocins, peptides containing a YGXGV-motif called „pediocin box“, act on their target cells by disrupting the membrane potential[25]. In these peptides the N-terminus is vital for their function, whereas the C-terminus determines the specificity for their target strain[26]. Intriguingly, the presence of the Mannose-specific phosphoenolpyruvate-phosphotransferase system (PTS) is required for pediocin-activity, the peptide forcing it into an open conformation[27]. Polytheonamides[28] and the cyclic peptide enterocin AS-48[29] also disrupt the membrane potential by pore formation . Sactipeptides, which are characterized by their head-to-tail cyclization and the presence of multiple nested thioether bonds are thought to display their activity insertion into lipid

bilayer, solubilizing it[30]. Some other peptides, such as Darobactins, inhibit the transport machinery of the target cell, in this case the BamABCDE complex which inserts proteins into the outer membrane[31].

While these membrane or cell wall targeting mechanisms are common there are also plenty of bacteriocins whose target is located in the cytoplasm. The majority of the lasso-peptides target the RNA polymerase, with the notable exception of the two members mentioned above. There they inhibit transcription by blocking the NTP binding site[32]. Other peptides, such as the linear azoline peptide microcin B17, target DNA gyrase[33], inhibiting DNA replication. Most other azoline peptides, as well as bottromycins, thiopeptides and microcin C inhibit the protein production of the target cells at the translation step. Inhibition of the ribosome can occur by blocking the peptide exit channel[34] or by the binding site of either the elongation factor[35] or the aminoacyl-tRNA[36]. Inhibition by direct binding to EF-Tu-GDP is primarily seen for thiopeptides[37] and Microcin C can block the Aspartate-aminoacyl-tRNA-synthetase[38], with engineered derivatives of Microcin C proving capable of inhibiting other tRNA synthetases[39].The peptide Lassomycin appears unique in its mechanism of action, as it blocks the ClpC1 ATPase which is involved in proteolysis as opposed to protein synthesis[40].

For bacteriocins targeting the intracellular machinery, only very little is known about their cell entry. For Microcins it could be determined that they exploit the target cells transport machinery, such as OmpF or FhuA, yet cell entry for other peptides remain elusive[32]. At the same time there are multiple other bacteriocins with an entirely unknown mechanism of action.

Glycocins

Glycocins, **glyco**active bacterio**cins** are a subclass of the bacteriocins, that are characterized by a requirement of a glycosylated residue to display activity[41]. Due to the presence of the carbohydrate moiety they are classified as a type Ie bacteriocin.

Table 1 *Experimentally verified glycocins.*

Glycocin	***Host organism***	***Glycosylation***	***Inhibition***
Sublancin 168[42]	*Bacillus subtilis* 168	S-Glc	Bactericidal
Thurandacin[43]	*Bacillus thuringiensis* serovar andalusiensis	S-Glc O-Glc	Bactericidal
Enterocin F4-9[44]	*Enterococcus faecalis* F4-9	O-GlcNAc O-GlcNAc	Bacteriostatic
Enterocin 96[45]	*Enterococcus faecalis* TX104	O-Glc-Glc	Not determined
Durancin 61A[46]	*Enterococcus durans* 61A	X-Glc X-Ara	Bactericidal
Glycocin F[47]	*Lactobacillus plantarum* KW30	O-GlcNAc S-GlcNAc	Bacteriostatic
ASM1[48]	*Lactobacillus plantarum* A-1	O-GlcNAc S-GlcNAc	Bacteriostatic
Pallidocin/Geocilicin[49, 50]	*Aeribacillus pallidus* sp. 8	S-Glc	Bacteriostatic
Bacillicin CER074[50]	*Bacillus cereus* CER074	S-Glc	Bacteriostatic
Bacillicin BAG2O[50]	*Bacillus cereus* BAG2O	S-Glc	Bacteriostatic
Listeriocytocin[50]	*Listeria monocytogenes* SLCC2540	O-Glc-Glc	Not determined
SvC-neoglycocin[51]	*Streptomyces venezuelae* ATCC15439	S-GlcNAc	Not determined
Cacaoidin[52]	*Streptomyces cacaoi* CA-170360	O-Rha- 6-deoxygulose	Not determined
NAI-112[53]	*Actinoplanes* sp. DSM24059	N-6-deoxyhexose	Not determined

Almost all of the experimentally validated glycocins originate from bacteria of the *Firmicutes* family (Table 1), though there has been a recent genome mining approach that indicates that these peptides are in fact much more common and may even be produced by gram-negative bacteria[16]. Indeed the existence of a glycocins or at least glycocin-like peptide within *Actinomycetia* could be demonstrated[51]. Whether the other gene clusters found through genome mining encode for functional glycocins still remains to be determined.

As RiPPs, glycocins are synthesized by the ribosome with an N-terminal leader peptide that is usually enriched in glutamate and leucine residues. In their biosynthetic operon several other open reading frames (ORFs) can be found, which encode for the cognate glycosyltransferase, responsible for the attachment of the hexose modification, an ABC-type exporter, containing a C39 protease domain and a bipartite oxidoreductase system, which is thought to be responsible for the correct formation of the disulphide connectivity. The folded and glycosylated peptide is thought to be directed towards the ABC-type exporter, where the C39-Protease domain cleaves the leader peptide at a double-glycine cleavage site, releasing the mature glycocin to the extracellular space. Resistance of the producer strain towards their glycocins is thought to be conferred by an immunity protein with the encoding ORF located near, or within, the glycocin-operon. However, only the resistance conferring genes of Glycocin F, termed *gccH*[54, 55], Sublancin 168, termed *sunI*[56] and Enterocin F4-9, termed *enfI*[57] have been experimentally validated. These genes encode for small peptides below 100 amino acids but share little similarity with each other. Indeed, ORFs of similar size were found in most putative glycocin gene clusters, their sequence appears highly variable, making identification based on sequence homology difficult.

1.1.1. Glycocin F-like Glycocins

Glycocin F (GccF) was one of the first discovered glycocin, and initially named plantaricin KW30[58]. It took until 2011 until the identity of its post translational modification (PTM) was determined correctly, which lead to the renaming of this peptide[47]. ASM1, a peptide produced by a different strain of *L. plantarum* and shares high sequence similarity of >80% with GccF, has been discovered later[48, 59]. Both peptides show highly selective activity against bacteria of the *Lactobacillus* genus, primarily targeting other strains of *L. plantarum*[58]. This antimicrobial effect is pronounced, with a described minimal inhibitory concentration (MIC) for GccF in the low nanomolar range[47].

The NMR structure of GccF[60] (Figure 3) shows the conformation of the peptide, which adopts a helix hairpin, stabilized by two disulphide bridges, and a C-terminal, flexible tail. GccF harbours two GlcNAc modifications, one within the loop of its helix hairpin, and one on the C-terminal cysteine.

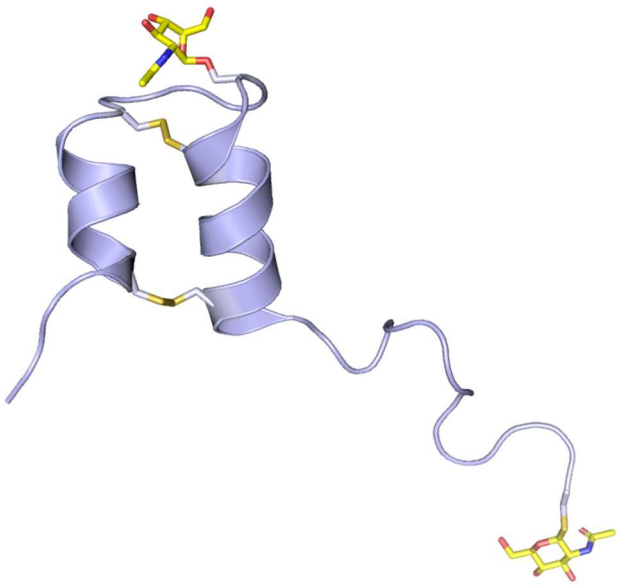


Figure 3 *NMR structure of Glycocin F (PDB: 2KUY). The peptide is shown in ribbon representation with the Hexose moieties and disulphides shown as sticks.*

The natural biosynthetic cluster is activated late in the log phase of growth, with all of the genes in the GccF cluster, with the exception of *gccI*, being under the same transcriptional control[54]. GccF requires both the sugar modification within its loop as well as the correct disulphide connectivity to be active. While the glycosylation of its tail is not strictly necessary for its activity, it improves the activity by almost two orders of magnitude. The nature of this sugar residue is important for its function, with neither mannose nor glucose being able to display improved activity over a mutant lacking this C-terminal modification[61]. The activity of GccF is inhibited by the addition of GlcNAc to the growth medium of the target bacterium[47]. It could be shown that the GlcNAc-specific PTS is necessary for GccF to show its inhibitory effect[55, 62], leading to the assumption that GlcNAc addition may competitively inhibit GccF binding to the PTS. How GccF binding to the PTS system results in a bacteriostatic effect is unclear, however. It is speculated that the PTS may merely anchor GccF on the membrane where it interacts with its true, unknown target[61].

Unfortunately, in part due to difficulties with the recombinant production and total-synthesis approaches, the mechanism of action of this peptide remains elusive.

1.1.2. Sublancin-like Glycocins

Similar to GccF, Sublancin 168 (Sublancin) was first considered to be a lanthipeptide[63], until the correct structure was discovered in 2010[42]. The NMR structure of Sublancin shows large similarity in its architecture compared to GccF[64]. It also adopts a helix-loop-helix hairpin, stabilized by two nested disulphides of C7-C36 and C14-C29, with the glycosylated residue of C22 located within its flexible loop. Contrary to GccF it is lacking the C-terminal tail with a secondary glycosylation site and the identity of the sugar is a glucose instead of an N-acetyl glucosamine.

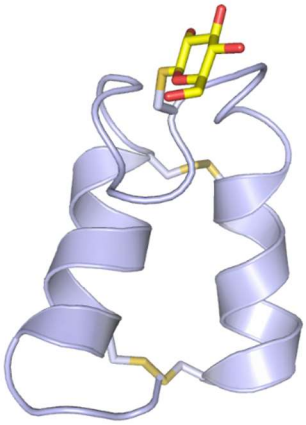


Figure 4 *NMR structure of Sublancin 168 (PDB ID: 2MIJ). The peptide is shown in a ribbon representation and the carbohydrate and disulphide shown as sticks.*

Sublancin is, so far, the best described glycocin with activity reported primarily against other *B. subtilis* strains, such as 6633, which lack the SPβ-prophage that encodes the sublancin gene cluster[65]. Other susceptible strains include *B. cereus*, *B. megaterium*, *Staphylococcus aureus* and *Streptococcus pyogenes*[66].

The mechanism of action of sublancin is still poorly understood. It could be shown that the architecture of the peptide, containing nested disulphides is vital for its function[42]. At the same time, Sublancin requires a carbohydrate moiety to be active against even the most susceptible strain of *B. subtilis*, though the nature of the sugar may vary. Sublancin derivatives decorated with a xylose or galactose proved equally active against a *B. subtilis ΔSP*β indicator strain[67, 68] as the wild type glycovariant. In contrast GlcNAc or Mannose glycosylation still showed activity, albeit at reduced efficacy[67]. Inhibition of the activity of sublancin on the other hand is specific with regards to the added sugar, as only glucose shows inhibitory activity whereas other sugars such as mannose, xylose, galactose or N-acetyl glucosamine failed to protect the indicator strain[67].

Mutagenesis studies indicated that the susceptibility of a strain towards Sublancin is dependent on the presence of the large mechanosensitive channel[69] (MscL), and a glucose-PTS[70], which is involved in glucose import and phosphorylation within *B. subtilis.* Given the viability of ΔPTS knockout strains of *B. subtilis* it appears unlikely that the mechanism of action is caused by the Sublancin-dependent inhibition of the PTS. At the same time a strict dependence of the activity on the Glucose-PTS seems to contradict previous results that has shown that the type of the hexose moiety is not important. A proposed mechanism of action involves Sublancin interacting with the abovementioned transport systems, forcing them into an open configuration and thus dissipating the membrane potential[68]. Surprisingly, Sublancin could be shown to have an immunomodulatory effect as well[71], though not much is known about the mechanism of this effect either.

Immunity to sublancin is encoded by the *SunI* gene, which encodes for a small peptide that is predicted to be membrane associated and co-localizes with the PTS system[68]. This is reminiscent of the immunity mechanism of lactococcicin A, in which the immunity protein binds to the Man-PTS system and thus protects the cell from the effect of lactococcicin A in an unknown manner, possibly by preventing the binding to the PTS system by competitive inhibition. A likewise mechanism may also explain the observed inhibition of Sublancin by the addition of glucose.

Since the discovery of Sublancin, genes with high sequence similarity to those found in the Sublancin operon have been found in multiple bacterial species. However, despite the observed sequence similarities, these display pronounced differences. For instance Thurandacin, a peptide that appears similar to sublancin in both, sequence and activity, is *O*- and *S*- glycosylated at two distinct glycosylation sites[43], by its glycosyltransferase, whereas the cognate glycosyltransferase of Sublancin, SunS, is incapable of *O*-glycosylation and therefore cannot diglycosylate Sublancin in a similar way[42].

Other Sublancin-like peptides that were discovered include Pallidocin (also called Geocillicin), Bacillicin BAG2O and CER074. All three glycocins adopt a helix-loop-helix architecture with two nested disulfides and a Cys decorated with a glucose unit within the loop[50]. Pallidocin exhibited antibacterial activity against several *Bacillus* species and some thermophilic bacteria of the *Geobacillus*, *Parageobacillus* and *Caldibacillus* families[49] in picomolar to nanomolar concentrations. However, some of these activities could not be replicated by a different group that had independently discovered the same peptide, which only described weak activity against *Bacillus cereus*, but not other *Bacilli* like *B. subtilis* or *B. megaterium*[50]. Bacillicin BAG2O and CER074 exhibit a very narrow antimicrobial activity, showing activity only against certain strains of *B. cereus*, with inhibitory concentrations typically in the nanomolar range[50].

1.1.3. Enterocin 96

Enterocin 96 (Ent 96) was discovered in 2009[72], though it took until 2017 to obtain the correct identification of its PTM [45]. Based on its CD spectrum and the signatory disulphide motif it is suggested that Ent96 adopts the same overall architecture observed for Sublancin or GccF. However, so far, no experimental structure to confirm the fold and disulphide pattern is available. It is also of note that, while the other glycocins show a conserved C-X6-C motif for cysteines involved in their nested disulphides[41], Ent 96 shows different distances between its cysteines, making an entirely different type of fold possible. Norris and Patchett (2016)[41] predicted the Ent 96 structure using I-TASSER[73]. This model contained a third helix with α1 composed of residue 6-13, α2 of 16-26 and α3 of 37-47. Helices α1 and α3 were in an almost parallel configuration with a disulfide bond connectivity of Cys6-Cys47 and Cys12-Cys39 (Figure 5). In this configuration Ser33 sticks out from the loop and was therefore proposed to be glycosylated. This glycosylation site was later confirmed by Nagar&Rao[45], supporting the predicted model.

Figure 5 *Predicted structure of Enterocin 96 created with I-TASSER. Figure adapted from Noris & Patchett (2016). The peptide is shown in ribbon conformation with the Ser33 glycosylation site shown as sticks.*

Contrary to the Sublancin- and GccF-like glycocins, Ent 96 contains a disaccharide with a Glc-βGlc linkage attached to Ser33, though it is currently unclear whether these glucose-units are 1-4, 1-3 or 1-6 linked[45]. The glycosylation of Ent 96 appears to occur in a highly site-specific manner. While the position of the glycosylated amino acid within the flexible loop of Sublancin could be changed without a large impact on the glycosylation[74], the Ent 96 peptide contains multiple serines within its loop, of which only a single one is glycosylated[45].

In line with other glycocins, the activity of Ent 96 is highly dependent on its glycosylation, with the monoglycosylated peptide or glycovariants thereof showing no or greatly reduced activity against

target strains. In contrast to both GccF and Sublancin, the disulphides of Ent 96 are not necessary for its function, which is unique amongst glycocins discovered so far[45].

Ent 96 also shows a relatively broad spectrum of activity, primarily targeting gram-positive *Enterococcus* species with strong activity also reported for *Listeria monocytogenes, Staphylococcus aureus* and several *Lactobacilli*[72]. Surprisingly, there have also been reports of activity of Ent 96 against gram-negative strains of *Salmonella enterica* and *Escherichia coli*[72] and even some fungi like *Candida albicans*[75]. The mechanism of action of Ent 96 as well as the immunity-mechanism of its producer strain is poorly understood. While several small peptides akin to SunI or GccH are present within the Ent 96 gene cluster, their protective activity against Ent 96 has so far not been demonstrated.

With Listeriocytocin, a second putative member of the Ent 96-like glycocins was discovered, though it shares little sequence similarity. Listeriocytocin is likewise diglycosylated with two glucose units attached to a Ser within a comparatively small loop[50]. Contrary to Ent 96 this peptide has the C-X_6-C architecture common to most glycocins and forms two nested disulphides as seen in Sublancin- and GccF-like glycocins. It is currently still unclear whether Listeriocytocin has any antibacterial activity and, if so, against which organisms, as this peptide failed to display activity against a variety of tested bacteria[50].

1.1.4. Other glycocins and glycosylated bacteriocins

Two further glycocins with little sequence similarity to the peptides described above have been found in *Enterococcus* strains: Enterocin F4-9 (EntF4-9) and Durancin 61A. EntF4-9, produced by *E. faecalis* F4-9 has been shown to be a bacteriocin with narrow spectrum activity against other *Enterococcus* and some *Bacillus* strains, as well as against the gram-negative *E. coli* JM109[44]. Contrary to other bacteriocins and the glycocins described above, the presence of the leader peptide did not disrupt the function of EntF4-9, but increased its activity instead. Indeed, with the N-terminal leader peptide its spectrum was even expanded towards a *Salmonella enterica* strain[57]. It could be shown that the displayed antimicrobial activity of EntF4-9 is dependent on the presence of two GlcNAc residues attached to Ser37 and Thr46. Similar to the Sublancin- and GccF-like glycocins EntF4-9 also required the presence of two disulphides. Although not experimentally validated, the cysteines are proposed to form a similar nested disulphide pattern[44]. Immunity against EntF4-9 is conferred by a small peptide termed enfI in an unknown manner[57].

Although there is some doubt towards its classification[41], Durancin 61A, expressed by *Enterococcus durans* 61A is a peptide glycosylated with either a glucose, arabinose or both. It possesses a relatively broad spectrum of activity against gram-positive organisms, including several *Listeria*, *Clostridia, Enterococcus* and *Staphylococcus* strains[76]. Furthermore, it displayed activity against the gram-negative bacteria *E. coli* and *Pseudomonas aeruginosa* as well as antifungal activity against *Aspergillus versicolor* and *Penicillium commune*. The mechanism of action of Durancin 61A could be determined to damage the cell wall, causing an ion efflux and thereby death of the cell[46]. Whether the poorly described glycosylation of this peptide is required for its activity is currently unknown.

SvG is the first glycocin or glycocin-like peptide discovered in *S. venezuelae* ATCC 15439, a bacterium of the *Actinomycetales*. Contrary to other glycocins, it only contains only three cysteines and therefore cannot form the nested disulphide bonds. Correspondingly the gene cluster of SvG does not include the oxidoreductases found in other gene clusters. The identified GT modifies the peptide with an *S*-linked GlcNAc within a predicted loop and displays a remarkable promiscuity towards the peptide acceptor. SvG displayed antimicrobial activity only against closely related strains of the producer *S. venezuelae* MTCC 327 and *S. venezuelae* MTCC 4218. However, this activity was independent of its glycosylation. Interestingly, an engineered SvG-SvG dimer displayed activity against *L. monocytogenes* in a GlcNAc dependent manner[51].

Two further potential glycocins are the lanthipeptides Cacaoidin[52] and NAI-112[53]. Both of these peptides are glycosylated and show antimicrobial activity: Cacaoidin against *S. aureus* and *Clostridium difficile*[52] and NAI-112 against various *Staphylococcus* and *Streptococcus* species[53]. The glycosidic residue for both of these peptides is unusual with Cacaoidin containing a β-6-deoxygulopyranosyl-(1-3)-α-rhamnopyranoside O-linked to a tyrosine[52] and NAI-112 containing a glycosylated tryptophan decorated with a 6-deoxyhexose. Both of these peptides display no significant sequence similarity to any other glycocin, while containing multiple unnatural amino acids such as aminobutyric acid and dehydroalanine. Additionally Cacaoidin contains a *N*,*N*-dimethyl lanthionine and a C-terminal S-[(Z)-2-aminovinyl-3-methyl]-d-cysteine amino acid[52], whereas NAI-112 contains two cyclic methyl-labionine rings[53]. All of these unusual amino acid residues are commonly found in lanthipeptides but have so far not been described for glycocins. Since the dependence of their activity on the glycosylation has not been demonstrated yet, it is possible that these peptides should not be classified as glycocins, but as glycosylated lanthipeptides instead.

1.2. Peptide glycosylation

Glycosylation is one of the most abundant post-translational modifications and can be found in all clades of life. Its functions is diverse, ranging from signalling, regulation as well as contribution to folding and facilitating protein-protein interactions.

While peptide glycosylations with other sugars, such as heptoses, pentoses, nonoses and other sugars are found in nature, the most common form of glycosylation is the covalent linkage to an aldohexose. These hexoses naturally occur primarily in their cyclized pyranose-form, although a glycosylation involving the furanose tautomer has also been described[77]. This cyclization of the sugar gives rise to the C1 atom as an anomeric centre with two chemically distinct anomers: the α- and β-anomer (Figure 6). Correspondingly, carbohydrates may be linked to peptides in an α- or β-conformation.

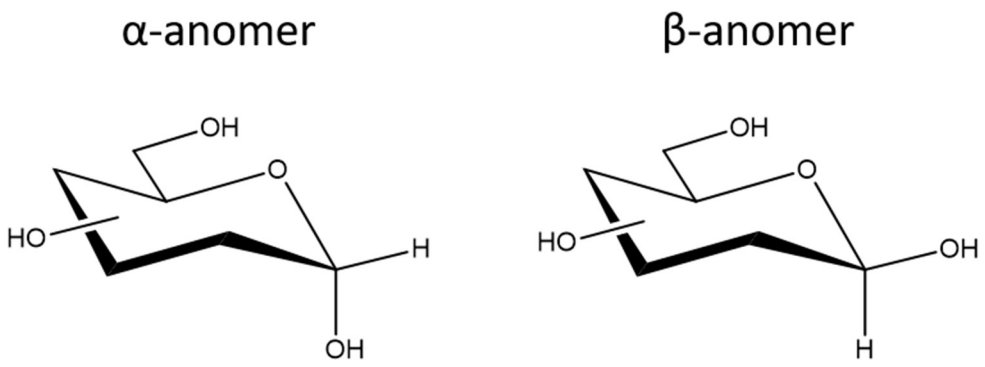


Figure 6 *Anomers of hexoses.*

Further classification of glycoconjugates depends on the atom linked by the formed glycosidic bond. Thus, depending on the amino acid or side chain involved, *N*-, *O*-, *S*- and *C*-glycosylation is observed, which will be briefly introduced in the following sections.

N-glycosylation

N-glycosylation or the formation of a glycosidic bond with a nitrogen of the side chain of either an Asn, Gln, His, Lys, Arg or Trp. No Lys-glycosylation has been discovered to date, and Gln-Glycosylation appears to be a very rare modification. In contrast the linkage to Asn is one of the most widely distributed carbohydrate–peptide bonds and is most commonly found as an attachment point of large Mannose-rich oligosaccharides[78]. The transfer of these glycans to the protein occurs under spatial control on the lumenal side of the endoplasmic reticulum membrane, during or after the translocation of the protein substrate. Here an oligosaccharyltransferase confers the core glycan from the lipid carrier molecule dolichol-phosphate to the Asn of an N-X-S/T (X ≠ P) consensus sequence. This core is subsequently processed by glycosidases and glycosyltransferases in the Golgi apparatus in a highly variable manner, depending on the host organism, cell-type, protein and glycosylation site[78]. Functions of eukaryotic N-glycans are diverse and range from helping the protein folding[79], quality control[80], as well as signaling[81] and cell-cell communication[82].

In prokaryotes the occurrence of *N*-linked glycosylation has been primarily found in proteobacteria. In δ- and ε proteobacteria such *N*-glycans are commonly found in proteins located on the cell surface, such as secretion system subunits or S-layer glycoproteins. These are modified by multiple, usually sulphated oligosaccharides linked to Asn and are part of the outer cell surface where they form a hexagonal, paracrystalline lattice[83]. In archaeal organisms the N-glycosylation is found predominantly in a class of proteins called archaellins, which are involved in cell motility, as well as type IV pilus structures important for cell-adhesion[78, 84]. The transfer of the glycans to the protein occurs *en* bloc, similar to the process in eukaryotes. In contrast to the Mannose-rich glycans of eukaryotes, proteobacterial glycans transferred this way are usually highly enriched in N-acetyl glucosamine instead[78, 84]. Interestingly, in γ-Proteobacteria, such as *Haemophilus influenzae*, *Yersinia enterocolitica*, and *Actinobacillus pleuropneumoniae*, *N*-glycosylation occurs in a processive manner with sugar nucleotides as donors instead of a transfer *en bloc* from a lipid carrier. While this can result in long, complex glycans this glycosylation may also be limited to a single carbohydrate unit[84].

A more infrequent type of N-glycosylation involves the modification of an Arg side chain, which has been found predominantly in gram-negative bacteria, though evidence for this linkage has also been reported in plants[85]. Several strains of gram-negative bacteria rhamnosylate the elongation factor EF-P to rescue stalled ribosomes[86]. *Salmonella enterica* and *Escherichia coli* add a GlcNAc to an Arg of death-domain proteins of an infected cell such as TRADD, RIPK1, or TNFR1, deactivating these as part of the infection cycle[87].

Furthermore, there has also been evidence towards N-linked glycosylation of histidines. These were found almost exclusively in the cyclic theopalaumide-pepides of *Theonella* sponges that contain a hexose-modification of the N^{π}-Atom of an N^{τ}-Histidinoalanine residue[88]. A further example of such a modification is the N^{π}-mannosylation of a modified His (α-amino-β-[4'-(2'-iminoimidazolidinyl)]-β-hydroxypropionic acid) found in mannopeptimycin-E of *Streptomyces hygroscopicus*[89].

Lastly, N-glycosylation of the N^{in} of a tryptophane has been observed for the Chorion peroxidase of *Aedes aegypti*[90] and the lanthipeptide NAI-112 of *Actinoplanes* sp.[53]

O-glycosylation

Similar to the N-glycosylation, O-glycosylation may result either in an O-linked single sugar, a small oligosaccharide, or a large, branched polysaccharide, though the latter typically display a lower degree of complexity than observed for N-glycans[78].

In eukaryotes the longer O-glycans usually consist of N-acetyl glucosamine and lactosamine repeats. Contrary to N-glycans, these sugars are usually attached in a processive manner and not transferred *en bloc*. No consensus sequence for the attachment of these glycans has been identified so far, though they are often found in flexible regions that are rich in serine, threonine and proline[78]. *O*-glycosylation is most commonly found on surface exposed or exported polypeptides. Depending on the density of the glycosylation, their function can be varied, with more densely packed glycans being reported to confer structural rigidity, protection from proteolysis as well as an increased propensity of gel formation in mucin-type peptides. More sparsely packed glycans are predominantly seen in cell adhesion and cell-cell communication, such as the non-self recognition by Siglec-Sialic acid interaction, or the AB0-glood group antigen[91].

Tyr O-glycosylation appears to be rare in eukaryotes. It is found in the glycosidic linkage to Glycogenin, a self-glycosylating enzyme involved in glycogen biosynthesis[92], as well as sialylation of amyloid precursor protein[93] and HexNAcylation of three mitochondria-associated proteins in mice[94].

Long, and branched O-glycans are also found in prokaryotes, commonly attached to S-layer glycoproteins. Especially gram-positive bacteria, where no N-glycosylation of these proteins has been observed, display these kinds of O-linked glycans. In contrast to the eukaryotic O-glycans, these are not only linked Ser and Thr residues, but Tyr O-glycosylation appears to be a common feature[84, 95]. Intriguingly, these long O-glycans are synthesized on a lipid carrier, commonly undecaprenyl

phosphate, and attached to the acceptor peptide *en bloc* by an oligosaccharyltransferase. Similar to the eukaryotic O-glycosylation sites, no consensus sequence has been found for prokaryotes. Indeed, in some proteins the glycosylation sites varied substantially even between individual molecules of the same peptide species[96]. Apart from the S-layer proteins, bacterial O-glycans can also be found in the adhesins and type IV plilins of ε-proteobacteria, which are commonly poly-mannosylated or arabinosylated[84].

In addition to the comparatively large O-glycans, there are also abundant and diverse simple glycosylations of Ser/Thr, which are produced within the cytoplasm. The best characterized of which are eukaryotic O-GlcNAc modifications of nuclear and cytosolic proteins. Additionally, glycosaminoglycan xylosylation, cadherin mannosylation or collagen galactosylation has been observed[91]. Similarly, prokaryotic species produce a vast array of simple Ser/Thr-linked *O*-glycoconjugates, often making use of modified or exotic sugars[77]. Likewise, simple Tyr-glycosylations within prokaryotes can be found in several antimicrobial peptides such as Vancomycin[97], mirabamides of *Siliquariaspongia mirabilis*[98] or Cacaoicidin of *S. cacaoi*[52] (see chapter 1.1.4) and may involve unusual carbohydrates.

C-glycosylation

C-glycosylation is a rare protein modification and thus far the only carbohydrate that has been discovered in such a linkage is a single mannose residue. This modification was first found and identified in human RNase 2 which contains an α-mannose attached to the $C2^{in}$ atom of the Trp via a C–C bond. Since then, several C-mannosylated proteins have been reported[99].

Their glycosylation site is usually found within a $[WxxWxx]_n$ consensus sequence, though the glycosylation site within that sequence is dependent on the mannosyltransferase[100]. C-mannosylation appears to be installed within the ER, using dolichol phosphate as the lipid carrier of a single mannose residue. To date, C-mannosylation of tryptophanes has only been found in higher organisms, such as mammals, avians, amphibians and nematodes, but not in fungi or prokaryotes[99].

S-glycosylation

While first evidence of cysteine S-glycosylation has already been reported in the 1970s in a S-galactosylated urinary peptide[101] and an S-glucosylated membrane peptide of erythrocytes[102], the field of protein S-glycosylation is still largely underexplored. Only few S-glycosylated peptides were

investigated in any detail, most of which belong to the glycocins (see chapter 1.1), with sublancin[42], thurandacin[43] and GccF[47] being the best characterized S-linked glycopeptides.

In 2016 a report by Maynard *et al*. [103] implied that the S-glycosylation is indeed not a rare modification, but that S-linked GlcNAc is a common motif of mammalian proteins and may be conferred by the *O*-GlcNAc transferase OGT. Furthermore, it was shown that these resulting S-linked sugars show an increased stability as compared to the *O*-glycosylated equivalent, especially with regards to cleavage by glycosylhydrolases[103].

1.3. Leloir type glycosyltransferases

Glycosylation and thereby the formation of the glycosidic linkage is catalysed by the enzyme family of glycosyltransferases (GTs). At time of writing (August 2023) the CaZy database[104] currently lists over 1,140,000 glycosyltransferase-sequences, classified into 116 families based on their sequence[105]. There are further, overlapping, yet substantially different methods of classifying GTs, based on their enzymatic properties: By acceptor, by donor, and by fold[106].

GTs in general have thus far been shown to adopt one of only five different folds, GT-A, GT-B, GT-C, GT-D and GT-E (Figure 7). While GT-C and GT-E type folds are predominantly seen in enzymes using lipid phosphate sugars or alkyl-linked glycosyl diphosphates. Enzymes of the GT-A, GT-B and GT-D fold typically use sugar nucleotides as their donor (Figure 8)[106, 107]. GTs using these sugar nucleotides as donors are termed Leloir-type glycosyltransferases, in honour of Lous Frederico Leloir for his discovery of these sugar nucleotides[108].

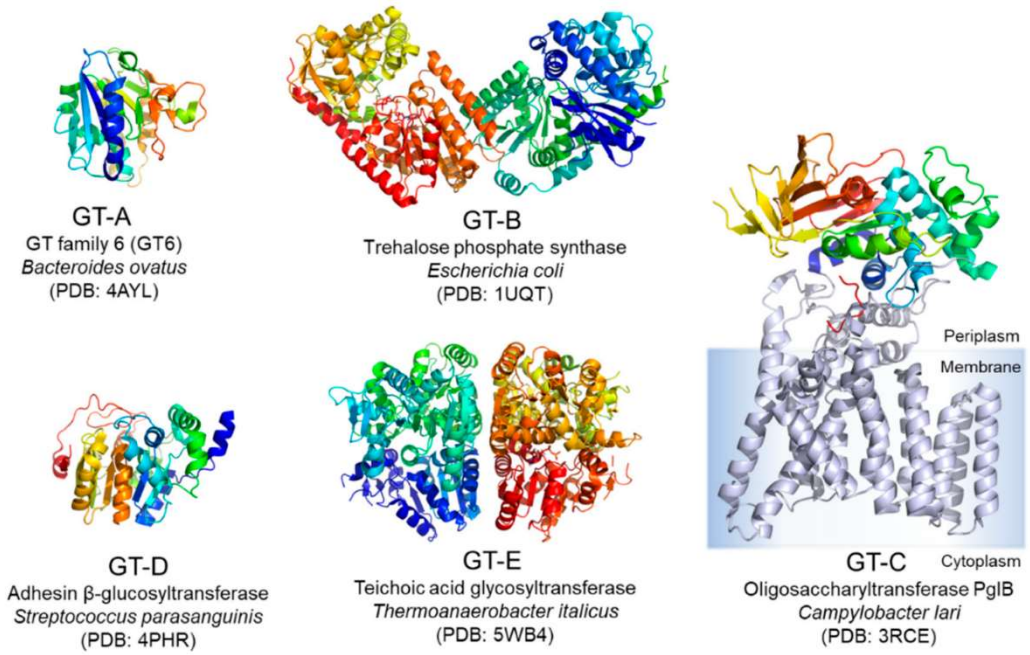


Figure 7 *Common folds of glycosyltransferases, figure adapted from Mestrom et al (2019)[107].*

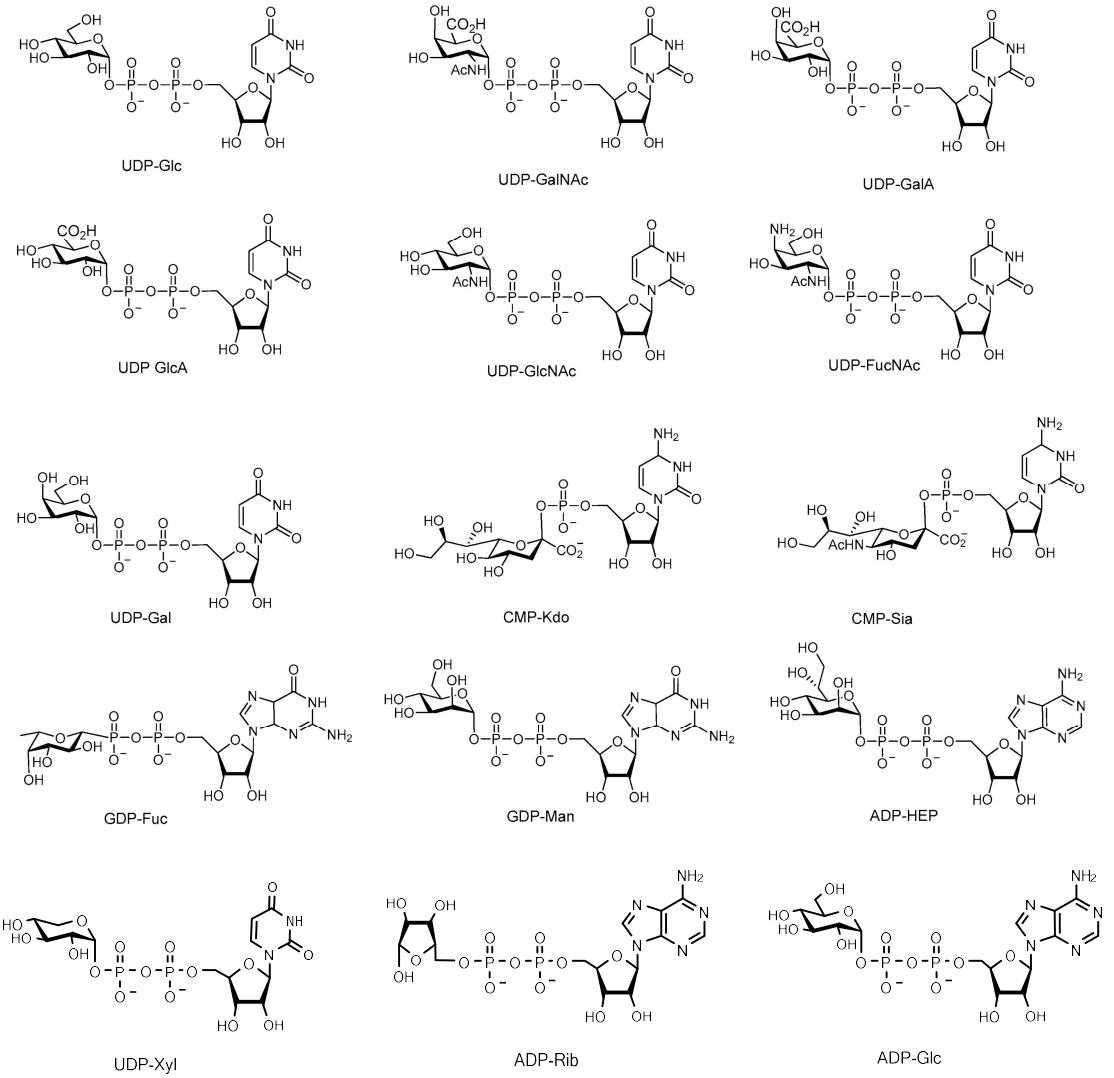


Figure 8 *Sugar nucleotides commonly found in nature. Adapted from Mestrom et al (2019)[107].*

With the exception of CMP-Neu5Ac and ADP-Rib, all of these sugar nucleotides are composed of the α-anomer of their sugar. ADP-Rib, in contrast, has its sugar attached to the nucleotide with its non-reducing end and Neu5Ac is attached in a β-conformation to a monophosphate instead of the more common diphosphate[109].

As mentioned previously, hexoses can be present as either an α- or β-anomer. Thus, GTs can also be divided based on the stereochemical outcome of the glycosylation reaction. If the anomeric configuration of the substrate, in general α, is retained in the product the enzyme is called retaining and if the configuration changes, it is called inverting. As such there are both inverting, as well as retaining type GTs present within each of the GT-A and GT-B folds (Table 2).

Table 2 *Glycosyltransferase fold and their GT-Families according to the CaZy database[104]. GTs-families whose fold and mechanism is not experimentally verified but are predicted by Taujale et al. (2021)[110] are shown in bold.*

Fold	*Mechanism*	*GT-Family*
GT-A	Inverting	2, 7, **12**, 13, 14, 15, 16, **17**, **21**, **25**, 29, 31, **40**, 42, 43, 48, 49, **54**, **60**, **67**, **69**, 75, **82**, **84**, 92, 109, 111
	Retaining	6, 8, 15, 24, 27, 32, 34, 44, 45, 55, 62, 64, 71, 77, 78, 81, 88, **95**, **116**
GT-B	Inverting	1, 9, 10, **11**, 18, 19, 23, 28, 30, **33**, 37, 38, 41, 47, 52, **56**, 61, 63, 65, 68, 70, 80, 90, **102**, **103**, 104, 112, 115
	Retaining	3, 4, 5, 20, 35, 72, **93**, **94**, 99, 107, 113
GT-C	Inverting	**22**, 39, 50, 53, 57, **58**, **59**, 66, 76, 83, 85, **87**, **89**, 98, **105**
GT-D	Inverting	101
GT-E	Inverting	26
Lysozyme-like		51
Unknown		73, 74, 91, 96, 97, 100, 106, 108, 110, 114

Since GTs involved in glycocin biosynthesis are all predicted to belong to the GT-family 2, the following sections about common folds and mechanisms of GTs will be restricted to those of inverting GTs of a GT-A type fold.

The GT-A fold

The canonical GT-A fold displays a central 7-membered β-sheet in a 3214657-topology sandwiched between α-helices (Figure 9), in a Rossmann fold. This type of fold is a motif often found in nucleotide binding proteins[106].

Within the loop L4, usually connecting two β-strands termed β4 and β4', the typical GT-A fold displays a highly conserved D-X-D motif (Figure 9). This motif is essential for the coordination of a mechanistically important, divalent metal ion, typically Mn^{2+} or Mg^{2+} [106, 111, 112]. Interestingly, in inverting GTs, it has been shown that only one of these aspartates, usually the latter, is actually involved in metal-ion coordination, while the other may be involved in binding of the ribose moiety or serve a different function entirely[113]. To satisfy the octahedral coordination sphere of the metal ion, water molecules and/or a nearby histidine are involved in metal complexation[114]. The C-terminal region of the GT-A fold protein is usually variable and is involved in acceptor recognition[113]. This region also contains the catalytic base, typically located within the helix α5.

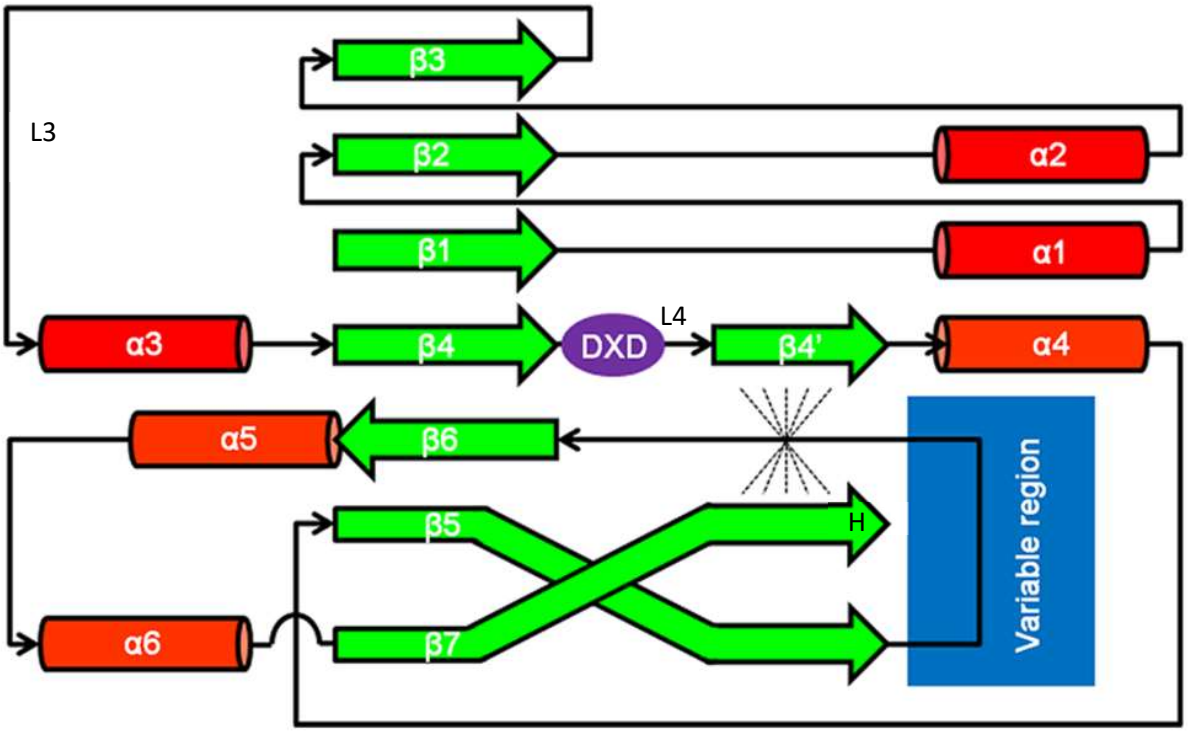


Figure 9 *Consensus topology of the GT-A type fold. Figure adapted from Romero-Garcia et al. (2013)[1]. Dashed lines represent interactions between the β4' and β7 strands, forming a second β-sheet.*

Nucleotide and sugar recognition of GT-A folds

Irrespective of the fold of the GT, recognition of the nucleobase occurs predominantly by π-stacking interactions[111], though some hydrophobic stacking has also been reported[115]. Nucleotide specificity is often imparted by hydrogen bonding with a residue located at the C-terminus of strand β1. While consensus sequences for sugar binding have been observed in other GT folds, GT-A type folds lack any such conserved specificity imparting motifs[111]. In general, interactions with the sugar moiety appear to be minimal, with the change of single amino acids within the binding pockets able to confer pronounced changes in the observed specificities. Additionally, many described enzymes exhibit high promiscuity with regards to the accepted sugar donors[114].

Binding of the sugar-nucleotide is usually accompanied by pronounced conformational changes. In GT-A type folds this is performed by loop L3, which changes from a random coil to adopt an ordered conformation[111, 116]. Oftentimes this loop includes an aromatic residue that shifts towards the nucleotide, interacting with it by π-stacking[114]. The movement of the L3 loop often creates a "lid" over the bound sugar-nucleotide, which is thought to protect it from hydrolysis[113]. This conformational change may also cause a movement of the loop L7 at the C-terminus of the Rossmann fold, adopting an α-helical conformation and assisting with acceptor binding[114, 117]. Additionally for some GTs it has been reported that these conformational changes bring a His residue at the C-term of the β7 strand closer to the metal ion, replacing a coordinating water molecule[114].

Common reaction mechanisms

While some GTs have been reported to follow a random order of substrate binding[118], or the acceptor promoting the binding of the donor substrate[119], most display a cooperative, sequential binding mode with the bound donor promoting the binding of the glycosyl acceptor: The sugar-donor binding to the catalytic pocket, is accompanied by the abovementioned conformational change, which allows for the formation of the acceptor binding site[106, 111, 118]. After the acceptor and donor are bound within the catalytic pocket, the reaction occurs either with retention or inversion of the anomeric configuration.

In inverting type GTs the reaction typically involves a deprotonation of the glycosyl-acceptor by a catalytic base, often a conserved Asp or Glu for GT-A fold proteins though catalytic His has been reported as well[106]. However, some GTs lack suitable side chains within their catalytic pocket. It is speculated that water chains or the α-phosphate of the donor act as catalytic base (OGT, GT-B fold) in such cases[120]. The deprotonated nucleophile subsequently attacks the anomeric carbon of the sugar, resulting in an oxocarbenium ion transition state (Figure 10). According to QM/MM-calculations the

O5-C1 bond of the oxocarbenium ion has double bond character and the glucopyranose ring adopts the usually unstable $^4H_3$ conformation. The cleavage of the C1−O1 glycosidic bond is accompanied by a 17° rotation of the β-phosphate oxygen while going from the ES to TS complex[121]. Stabilisation of the developing negative charge at the phosphate occurs in metal ion dependent GTs primarily via the associated divalent metal ion. In metal-ion independent GTs, salt bridges to positively charged side-chain residues, such as Arg or Lys stabilize the charge instead[106, 109].

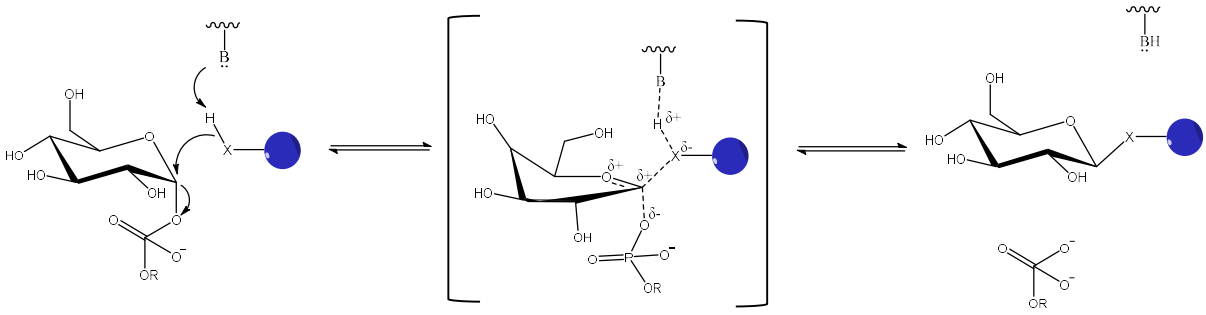


Figure 10 *Reaction mechanism of inverting glycosyltransferases. X = NH, S, O. R = Nucleotide phosphate. The glycosyl acceptor is represented by a blue sphere.*

Independent of the stereochemistry of the transferred sugar, the positioning of the sugar with respect to the nucleotide helps facilitate the reaction. Especially in retaining enzymes, the sugar donor is forced into a conformation in which the sugar situated above the pyrophosphate. This allows the O2 or, in the case of N-acetyl sugars, the acetamide to stabilize the developing negative charge on the phosphate group by hydrogen bonding[122]. Additionally this shape elongates and thereby weakens the anomeric bond by torsion around the Π-angle[113]. Interestingly, a difference in energies released by breaking the NDP-sugar bond can be observed in a nucleotide-specific manner. For instance, the hydrolysis of ADP-Glc is thermodynamically more favourable that GDP-Man[107].

Though GTs are highly efficient enzymes that catalyse the formation of a very stable glycosidic bonds with high specificity and selectivity, they have a few drawbacks, which makes them not the first choice as catalysts for the synthesis of carbohydrate structures. Firstly, same time the availability of the nucleotide sugars is often limited and, in some cases, prohibitively expensive. Additionally due to the high energy of the nucleotide-sugar bond, the sugar donors of Leloir GTs are prone to hydrolytic degradation, especially in the presence of divalent metal ions[123]. This problem is further exacerbated if the sugar nucleotide is bound to the donor binding site in the absence of a glycosyl acceptor. This is referred to as “error hydrolysis”[124]. Moreover, the binding of the sugar donor is facilitated primarily by the nucleotide diphosphate, an already hydrolysed nucleotide may also bind to the GT, causing

product inhibition. Interestingly there seems to be a correlation between the specificity of the GT and the severity of the product inhibition it experiences, with more promiscuous enzymes being less prone to product inhibition. Furthermore, it could be shown that the reaction of a GT is generally reversible[125]. While the equilibrium typically favours the formation of the glycoside with a K_{eq}>10, there are instances where the equilibrium constant is close to 1[118].

1.4. Solid phase peptide synthesis

Ever since Merrifield and coworkers developed solid phase peptide synthesis (SPPS), this methodology has become a widely used system for the quick and facile generation of peptide sequences of an ever-increasing variety. Using SPPS combined with ligation approaches, even proteins of considerable sizes >90 kDa have become accessible to chemical synthesis[126].

In general, the method of SPPS can be divided into Boc- and Fmoc-based strategies, each named after the protecting group of the backbone nitrogen of the respective amino acid building blocks. As the Fmoc strategy generally involves milder conditions, has improved scalability, as well as advantages of the Fmoc-group, such as its inherent fluorescence, this strategy has generally found broader use[127]. A scheme of an Fmoc-SPPS-cycle can be seen in Figure 11.

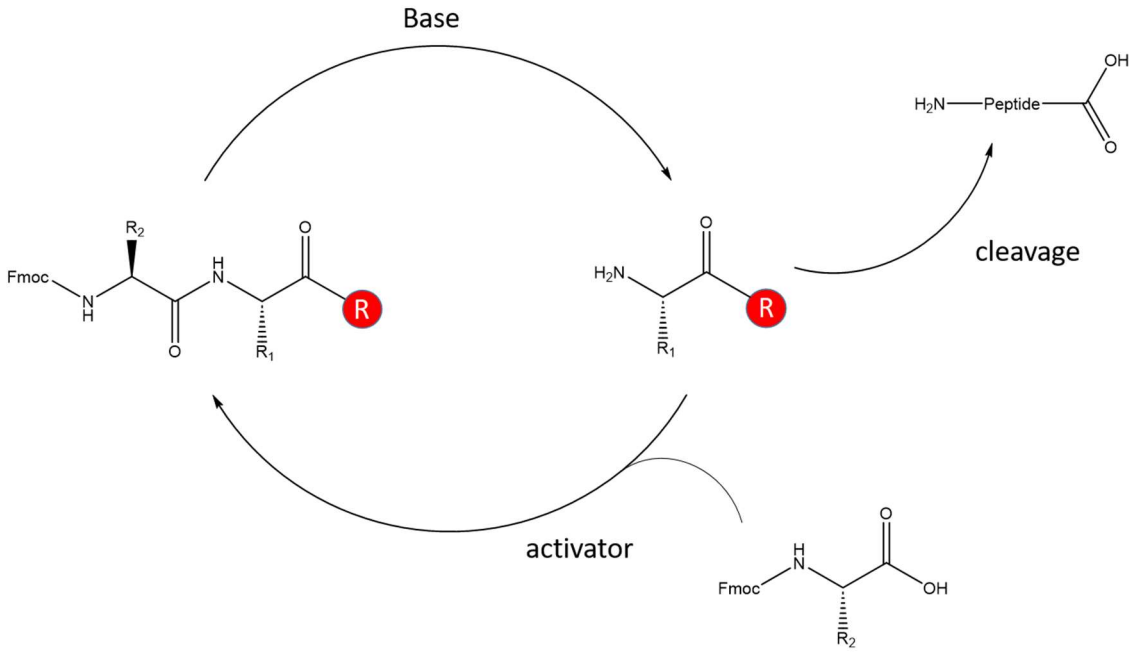


Figure 11 *Schematic representation of the Fmoc SPPS strategy. An Fmoc protected amino acid attached to a peptide chain or resin (red sphere) is deprotected via base treatment. The free amine reacts with a second, Fmoc-protected amino acid in the presence of an activator, extending the chain. This cycle is repeated until the peptide chain is cleaved from the resin after Fmoc-deprotection.*

Given the large number of individual steps involved in SPPS approaches, the minimization of side reactions per step is vital. At the same time the vast diversity and density of functional or reactive groups within peptide sequences makes them prone to side reactions during the deprotection and purification steps, which can decrease the final yield significantly. While a comprehensive description of such side reactions is far outside the scope of this thesis, a selection of these that are commonly

described for sequence motifs common in glycocin-peptides, as well as methods how to reduce their occurrence are briefly described in the following sections.

Cyclization reactions

One of the most well described side reactions during peptide synthesis is a cyclization reaction of aspartates to the corresponding aspartimide[128]. Less well known are the cyclization reactions of asparagine[129], glutamates[130], and glutamines[131], which may undergo a mechanistically similar side reaction. While these reactions are particularly prevalent under the basic conditions of Fmoc abstraction, they have also been described to occur in acidic condition[129], or even neutral conditions upon prolonged storage[132]. The mechanism of the base catalysed cyclization reaction is shown in Figure 12. A base, usually piperidine or piperazine during Fmoc removal, abstract the proton of a backbone amide, which may then form a five- or six-membered ring upon nucleophilic attack on the carboxyl or amide group in an S_N2-type reaction[128]. The resulting cyclical product can be easily detected via mass spectrometry, owing to the decreased mass by -18 or -17 amu, respectively.

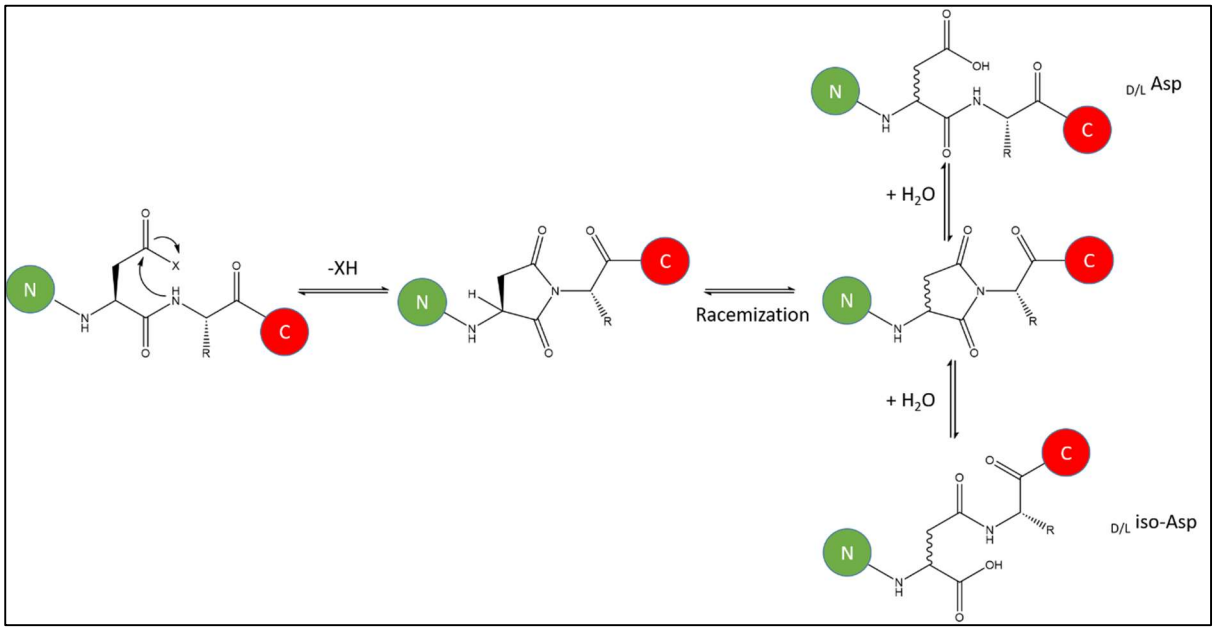


Figure 12 *Mechanism of Aspartimide formation. N- and C-terminal peptide sequences are represented by green and red spheres, respectively.*

However, the aspartimide and glutarimide moieties are highly prone to racemization[133] and hydrolytically unstable. Hydrolysis in a ring-opening reaction may occur at either of the two carbonyls, resulting in a mixture of L/D-Asp and L/D-iso-Asp or L/D-Glu and L/D-iso-Glu, respectively. These

compounds are difficult to remove chromatographically, and cannot be detected via mass spectrometry, unless the originating amino acid was an Asn or Gln derivative, where this reaction results in an effective deamination with a corresponding change in mass of -1 amu[128].

As mentioned previously, this side reaction is a comparatively common issue during Fmoc-SPPS, especially in the Fmoc abstraction step. Additionally, microwave irradiation[134] and the accompanying high temperatures[135] exacerbate this issue further. Many other factors may also influence the occurrence of this cyclisation reaction during peptide synthesis. For example, the identity of the amino acid whose N^{α} is performing the nucleophilic attack is an important variable, though the extent of this effect may vary from peptide to peptide and is impossible to predict. Asparagine, phenylalanine and histidine, which appear to be especially sensitive to this type of reaction. Serine and threonines have also been found to increase the occurrence of aspartimide formation, albeit primarily in their unprotected form[128, 136].

Much work has been done to alleviate the issue of cyclization reactions, especially aspartimide formation during SPPS. The choice of base for Fmoc-removal proved particularly helpful during microwave assisted synthesis with the use of piperazine instead of piperidine and addition of HOBt, or OxymaPure®, reducing the basicity in the Fmoc deprotection cycle[137, 138]. Two further commonly employed strategies involve the protection of the backbone N^{α} and the choice in protecting group of the aspartate. In this regard the OMpe and Odie protecting groups have shown improvement over the standard OtBu (Figure 13), most likely due to their increased bulkiness[135].

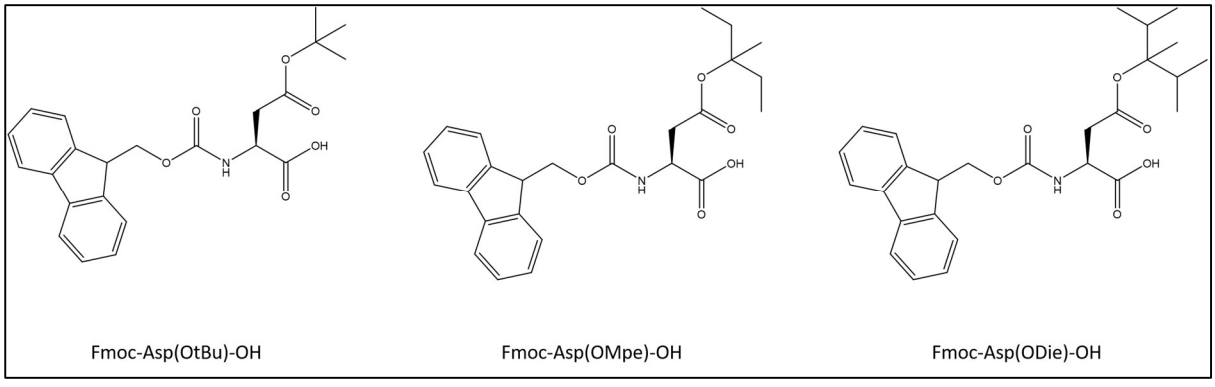


Figure 13 *Chemical structure of aspartic acid with common side chain protecting groups, OtBu, Ompe, and Odie. The N-term is protected with an Fmoc-group.*

Racemisation

Most glycocins discovered so far, include two of the most notorious amino acids with regards to racemisation reactions: Histidine and cysteine. This side reaction is exacerbated during microwave assisted peptide synthesis compared to traditional approaches, most likely due to the increased reaction temperatures[139].

Histidine racemisation is thought to occur via two different mechanisms[140]. One being the direct H^{α} abstraction, which may occur in an intramolecular fashion with the N^{π} functioning as the respective base. Reprotonation of this intermediate would then lead to a racemic mixture of His (Figure 14 A). The other proposed mechanism involves the intramolecular, nucleophilic attack of N^{π} on the His-carboxylate, forming an imidazoline moiety (Figure 14 B). Ring opening reactions of this compound then result in a D- or L-conformation of the His.

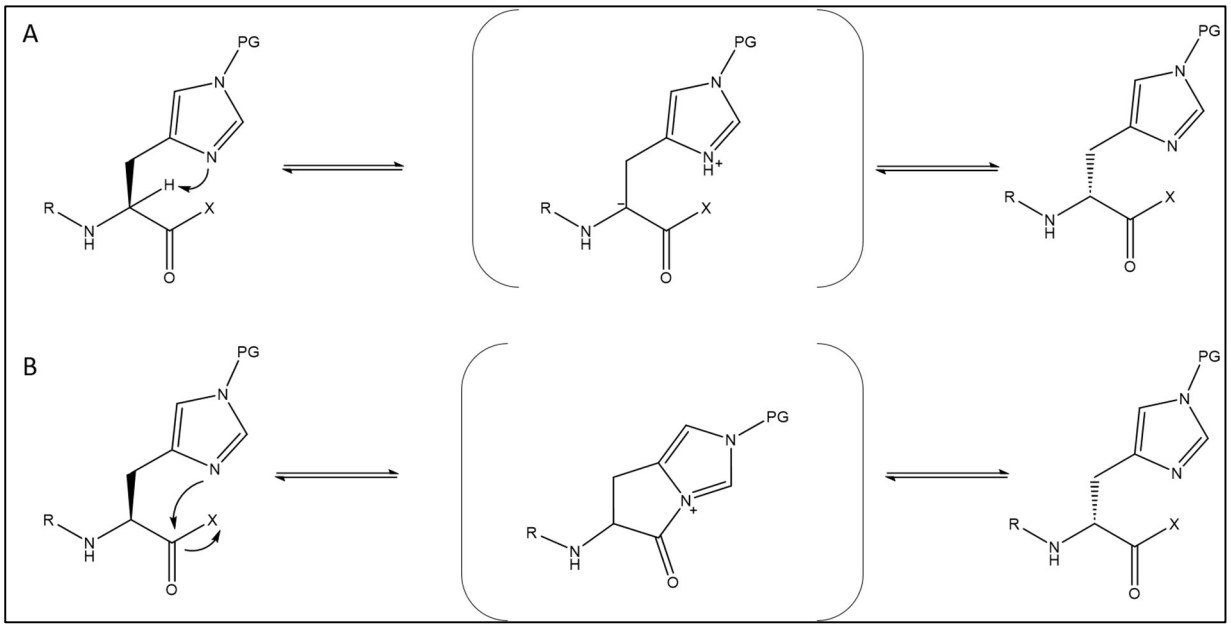


Figure 14 *Mechanism of His-racemization by internal proton abstraction (A) and imidazoline formation (B).*

Given that both of these mechanisms involve the N^{π}, protection of this moiety appears the most logical choice to reduce this side reaction. However, the commonly used protecting groups for N^{π} give rise to several side reactions with other nucleophilic amino acids, and deprotection of these is often difficult. Protection of N^{τ} with an electron withdrawing group such as Boc (Figure 15) instead has been shown to reduce this autocatalytic effect to acceptable levels as well[141].

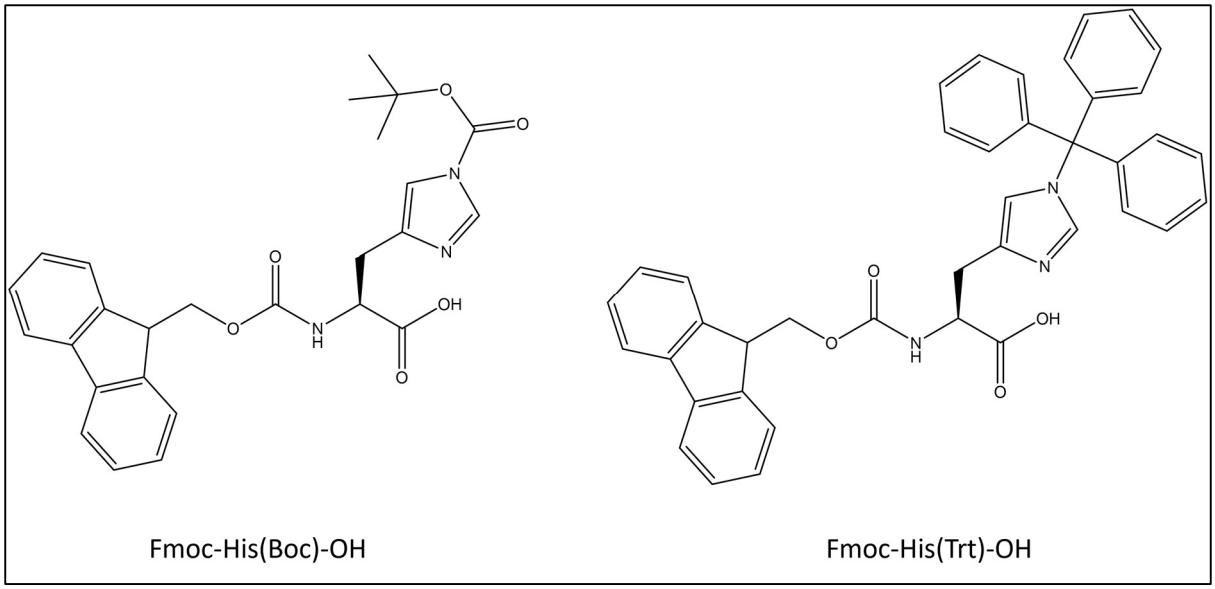


Figure 15 *Fmoc-His(Boc)-OH and Fmoc-His(Trt)-OH*

Racemization of cysteine has been a long-known problem. The most likely mechanism appears to be a direct H^{α} abstraction in the presence of a base[142]. Especially the immobilization reaction of cysteine to the resin linker appears to be prone to racemization reactions, while other positions have been reported to be more resistant towards these side reactions[143]. Indeed, with the use of Trt or S-S*t*Bu in combination with the proper choice of activator base during peptide coupling, such as HOBt, racemisations of internal cysteines may be reduced to acceptable levels[141].

β-elimination of cysteine sulfhydryl-group

Similar to the aforementioned racemisation reaction, H^{α} abstraction may also lead to the elimination of the sulfhydryl group, leading to the formation of dehydroalanine (Dha). The formed Dha can in turn be attacked a nucleophile in a Michael-type addition. This reaction appears to be highly dependent on the Cys microenvironment, as it is primarily described as a problem for C-terminal residues, with the choice of resin displaying a pronounced effect[144], whereas β-elimination of cysteines distant from the C-terminus usually involves strongly basic conditions[145].
Choice of the correct side-chain protecting group may also help alleviate this issue, as Cys(Trt) appeared much less prone to β-elimination than Cys(Acm)[144].

Oxidation of methionine or cysteine

Oxidations of cysteine sulphur atoms are generally well known, given that the arising disulphide bridge is a very common and biologically relevant motif. It is easily reversible by addition of a reducing agent. The formation of the corresponding disulphide should still be considered during peptide purification as it can lead to further side reactions, such as the β-elimination of the sulphur atom. This has been reported primarily for basic milieu[146] or strong heating[147].

Oxidations beyond the oxidation state S(-I) may also occur, leading to the formation of the corresponding sulfinic or sulphonic acid (Figure 16 A), which are irreversible reactions.

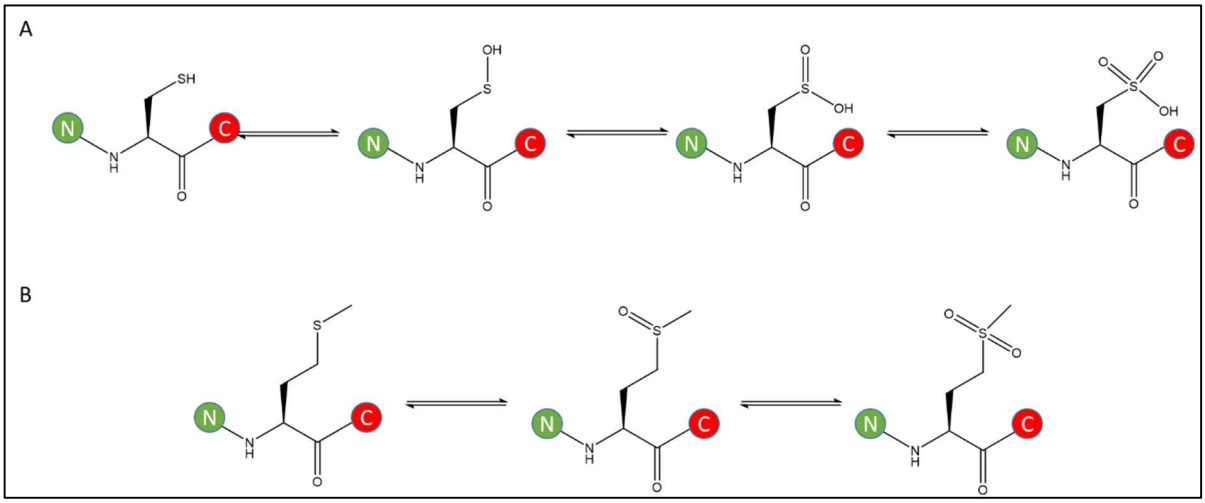


Figure 16 *Oxidation of Cysteine (A) and Methionine (B). N- and C-terminal peptide sequences are represented by green and red spheres, respectively.*

Similarly, the sulphur of methionine is prone to oxidation (Figure 16 B) under mildly oxidative conditions such as contact with air; especially under acidic conditions[148]. The usual route to protect against such undesired side reaction or even recover methionine from oxidized species, is the addition of NH_4I and dimethyl sulphide (Me_2S) to the cleavage cocktail[149]. However Trp residues were shown to be unstable under these conditions[150], and the arising dimethyl sulfoxide (DMSO) and Iodine (I_2) could cause the oxidation of cysteines in more complex peptides[127]. An alternative that is recommended for deprotection strategies of methionine containing compounds is the use of proper scavengers[127]. The most commonly recommended scavengers for methionine containing peptides are Reagent K (TFA-phenol-H2O-thioanisole-EDT (82.5:5:5:5:2.5)[151] and Reagent R (TFA-thioanisole-EDT-anisole (90:5:3:2)[152], both of which contain thioanisole and EDT as main scavengers. However, the use of these scavengers may introduce its own problems as well. Under concentrated TFA conditions, as is common for peptide deprotection strategies, EDT has been shown

to form adducts with tryptophanes at elevated temperatures[153]. Thioanisole may cause several different side reactions, primarily alkylations[154] , which are discussed in the following section.

Hence the use of other reducing agents is recommended for longer incubation times, such as dithio threitol (DTT) or Tris-(2-carboxyethyl)-phosphine (TCEP). Considering the former displays its reducing effect under alkaline conditions, the latter is usually much more suitable for the protection of cysteines during peptide purification.

Alkylation

Owing to the release of multiple electrophilic species arising from common protecting groups, alkylations occur primarily during deprotection and cleavage of the peptide from the resin, making the use of proper scavengers vital. EDT and trialkylsilanes stand out as particularly potent, though water can also function as an effective scavenger for *tert*-butyl cations. The correct choice of silane is critical as, for instance, triethyl silane is able to reduce the indole ring of unprotected Trp[155].

In general, three side chains stand out as being particularly vulnerable to alkylation: the electron-rich indole ring of Trp, as well as Cys and Met, owing to the nucleophilic sulphur atom.

In the case of Trp alkylation, this issue can largely be circumvented to a by use of the N^{In}-Boc protecting group. While the initial cleavage of the Boc-group occurs in a rapid manner, the resulting carbamate is much more stable[156]. Due to the electron-withdrawing effect of such a group, it effectively protects the tryptophan from undesired alkylation reactions[157]. This resulting carbamate is slowly cleaved off, releasing carbon dioxide, under even mildly acidic conditions, allowing for essentially a two-step deprotection of the peptide (Figure 17).

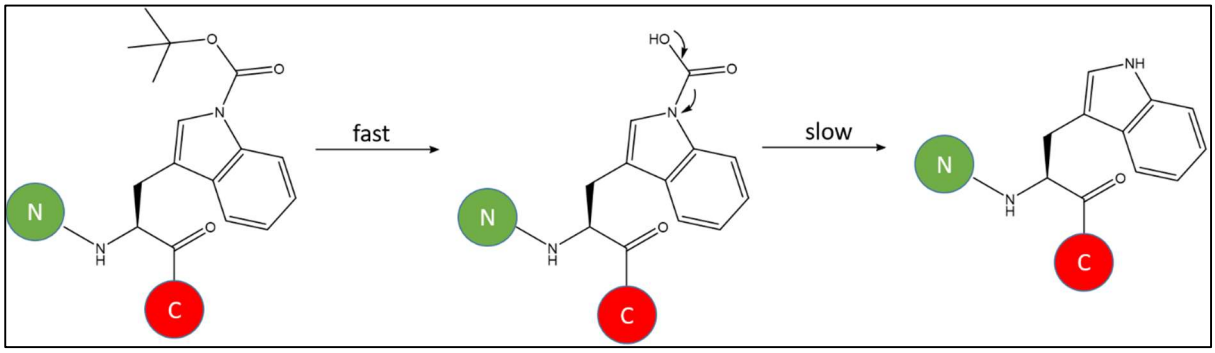


Figure 17 *Tryptophan deprotection occurs in a 2-step process removal of the tBu occurs fast, whereas the decarboxylation of the carbamate occurs much slower. N- and C-terminal peptide sequences are represented by green and red spheres, respectively.*

This may also serve to protect the Trp from reactions with an acid labile resin or the cationic linker species released upon peptide cleavage, as these modifications also are reported to primarily affect the 2-position of the indole. These side reactions have been reported to occur concomitant to cleavage from Wang-resins[158], among others. A different method to effectively protect against these cationic species is the use of correct scavenger. While in some cases TIS and EDT have been shown to be insufficiently protective, addition of thioanisole helped to prevent alkylation of Trp by spacer residues[152].

Similar to Trp, Cys and Met can form alkylation adducts under standard cleavage conditions. Mainly responsible for these side reactions are electrophilic *tert*butyl carbocations arising from *t*Bu and Boc-protecting groups[159], as well as the hydroxybenzyl-linker of Wang-resins[160]. These alkylations result in irreversible modifications. Additionally, the Trt group, if improperly quenched, may get attacked by the freed sulfhydryl group, reversing the original deprotection. As mentioned previously, for suppression of these side reactions the use of proper scavengers like EDT and TIS for *tert*butyl- and Trt-cations, respectively, is crucial.

Asp-Pro cleavage

The sequence motif Asp-Pro is a motif that has been shown to be prone to hydrolysis under acidic conditions even at moderate temperatures[161]. The proposed mechanism involves a nucleophilic attack of the Pro nitrogen on a protonated carboxylic acid of aspartate, forming a bicyclic intermediate. Hydrolysis of this intermediate then gives rise to either the original compound or results in the scission of the peptide chain.

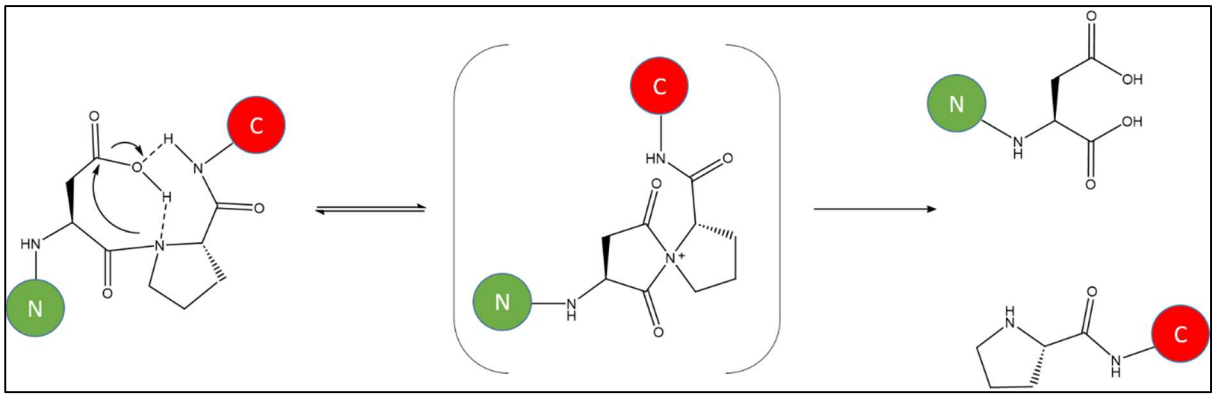


Figure 18 *Reaction mechanism of Asp-Pro cleavage. N- and C-terminal peptide sequences are represented by green and red spheres, respectively.*

Given that most side chain protecting groups are designed to be removed under acidic conditions in Fmoc-based peptide synthesis strategies, this problem may be unavoidable for peptides harbouring this motif.

Aggregation

A “stickiness” or tendency to self-aggregate has often been reported to be a problem for efficient production of bacteriocins and similar peptides. Indeed, for glycocin peptides this issue has already been reported for the solid-phase synthesis of sublancin 186[162] and GccF[163]. Here it was solved by the synthesis of two smaller fragments, combining these via native chemical ligation, or by drastically reducing the loading of the linker.

Other potential solutions to the problem that are usually recommended are the addition of fluorinated solvents such as trifluoroethanol (TFE)[164], use of microwave radiation[137, 165] or the introduction of pseudoprolines[127].
Microwave radiation appeared to be the most practical solution to this problem, though it may exacerbate issues such as racemization and cyclization reactions.

2. Aim

This work was aimed to identify and characterise members of the glycocin family to improve understanding of glycocins, especially with regards to the generation and functional importance of the unusual *S*-glycosidic linkage. To that end, genome mining will be used to identify potential new members of the glycocin family and identify their sequence as well as their putative glycosyltransferases. A method for heterologous expression and purification of the glycocins and the glycocin-glycosyltransferases in *Escherichia coli* will be established and glycocin peptide derivatives and authentic, defined standards will be synthesized using solid-phase peptide synthesis. The glycocins will be characterised in respect of their activity, selectivity, structural characteristics and type of glycosylation. The respective transferases will be characterised in respect of their catalytic mechanism, structure and substrate specificities.

3. Results

3.1. Discovery of putative glycocins and their cognate GTs

To date, only few glycocin-glycosyltransferase pair have been characterized beyond genome annotation. To expand the number of putative glycocin sequences, the NCBI database was searched for peptides or GTs with high sequence similarity to the experimentally validated glycocin or their cognate GTs respectively. The observed, highly similar sequences were investigated with regards to the other sequences in their gene cluster. A few examples for the organization of glycocin biosynthesis clusters can be found in Figure 19.

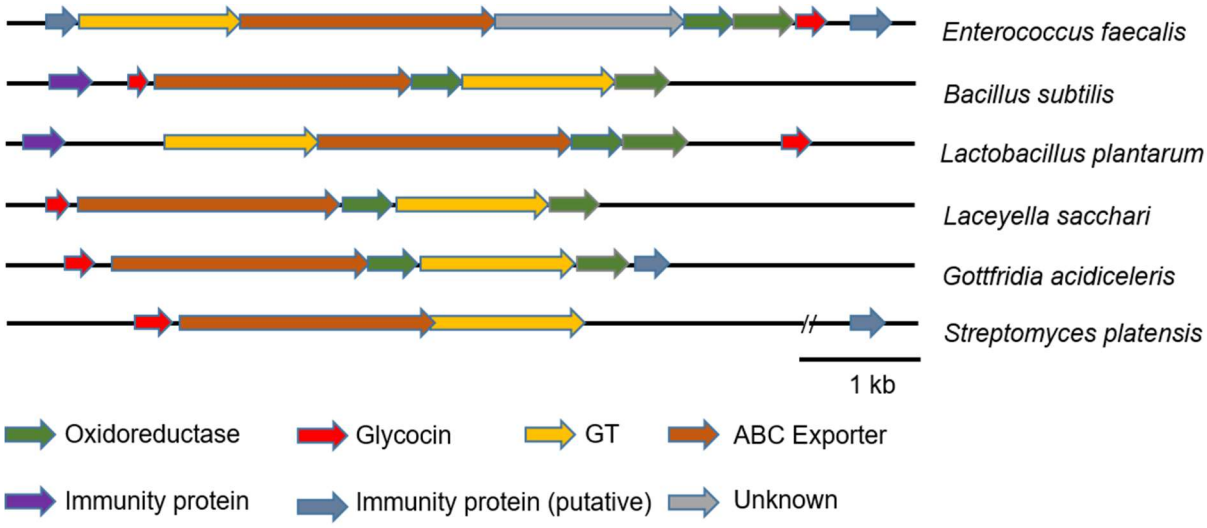


Figure 19 *Typical organization of gene clusters of glycocin producing strains. Genes encoding the glycocin in red, GT in yellow, ABC-type exporter in brown, thioredoxin-type proteins in green. Experimentally validated immunity proteins are shown in purple for and in blue for predicted peptides.*

With the sequences of the found, putative glycocins in hand, a phylogenetic tree was constructed (Figure 20). While the generally low bootstrap values imply a fair amount in uncertainty about the clustering, several distinct groups can be seen.

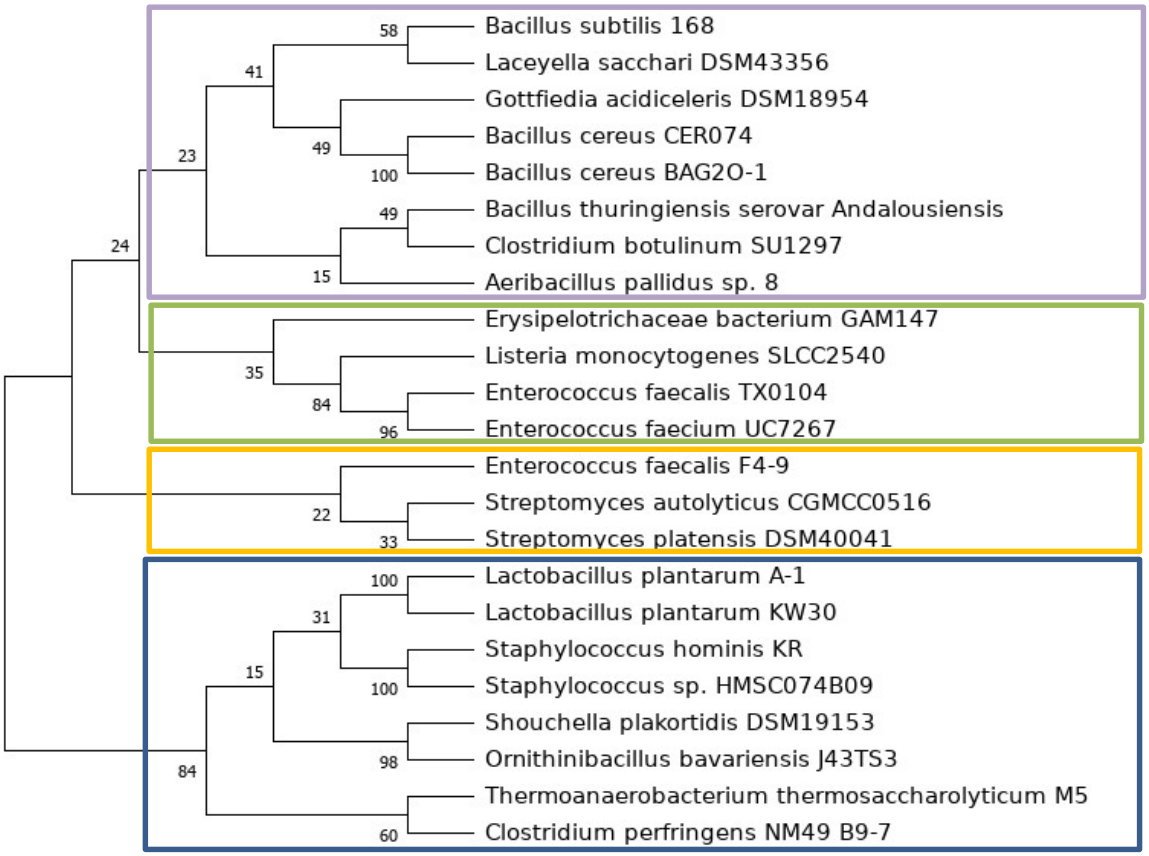


Figure 20 *Maximum likelihood phylogenetic tree of a small selection of glycocins. Numbers next to the nodes represent bootstrap values. These can be approximately clustered into several groups based on their sequence similarity: Sublancin-like (purple), Enterocin 96-like (green), Enterocin 4-9-like (yellow) and Glycocin F-like (blue).*

The Sublancin-like peptides, named after Sublancin 186 (SunA) of *Bacillus subtilis*, which may include the peptides of *Bacillus thuringiensis* and *Clostridium botulinum* as well. Further reliable inclusions into this group include *Laceyella sacchari*, *Bacillus cereus* and *Gottfriedia acidiceleris*. Pallidocin may constitute its own group but has been included into the Sublancin-like for its high sequence similarity to both SunA and ThuA.

In a similar manner the second, much smaller group of Enterocin 96-like peptides was determined. This group only includes the related strains of *Enterococcus faecalis* and *Enterococcus faecium*, which express essentially the same peptide sequence. A sequence found in various bacteria belonging to the *Eryspielotrichaceae* family is also included in this peptide-family.

A third family is found in the Enterocin 4-9 group, which, while also originating from *E. faecalis*, has very limited sequence similarity to Enterocin 96, and shares little of its PTM architecture. While the Streptomyces-based sequences of *S. platensis* and *S. autolyticus* appear to fall into the same family,

they may as well be distinct, given the capability of Enterocin 4-9 to form multiple disulphide bonds and the presence of a thiol-disulfide isomerase within the biosynthetic cluster[44].

The final family is the family of Glycocin F (GccF) of *Lactobacillus plantarum*, the first glycocin to be discovered. These peptides are distinct from the other glycocin families, by their C-terminal disordered, yet glycosylated tail. While the peptides of ASM1, and staphylococcal peptides can be clustered into this group with some degree of certainty, the group including *Thermoanaerobacter thermosaccharolyticum*, *Clostridium perfringens* and an *Ornithinibacillus* may also constitute a fifth group with no currently experimentally validated member.

To investigate a diverse number of different identified glycocin families I chose several members of the different families for investigation. Enterocin 96 (Ent96), Sublancin 186 (SunA) and Glycocin F (GccF) with their respective GTs, EntS, SunS and GccA, were chosen from the experimentally validated glycocins to be investigated further. Of the newly discovered sequences, the putative glycocins of *Laceyella sacchari* and *Gottfriedia acidiceleris* of the Sublancin-like family, and *Streptomyces platensis* of the Streptomyces/Enterocin 4-9-family were chosen. These new, putative glycocins were termed, according to the nomenclature scheme[166]: Saccharicin (SacA, accession code WP_132219678.1 with its cognate GT SacS, accession code TCW40592.1), Acidicin (AciA, accession code WP_088013405.1 with the cognate GT AciS, accession code WP_088013374) and Platicin (PltA, accession code OSY44892.1 with the cognate GT PltS, accession code OSY44890.1), respectively. Bases on sequence similarity all of these GTs belong to the GT-family 2 and should therefore adopt a GT-A type fold.

3.2. Heterologous expression of Glycosyltransferases

To determine the function of the glycocin glycosyltransferases and use them for the chemoenzymatic synthesis of glycocins, first a protocol to obtain suitable amounts of protein had to be developed. As every protein is unique in its own way, these protocols in part substantially differ from each other, and are discussed in the following sections.

3.2.1. GccA

The DNA sequence for GccA was cloned into a pET-3CLIC vector and expression was attempted at first in *E. coli* BL21 (DE3) Gold. While a vast amount of protein was produced under standard expression conditions, the GT was present exclusively in the insoluble phase, indicating severe misfolding. This problem has since also been reported elsewhere[167]. Several attempts to obtain soluble protein were made, such as truncation of predicted domains, expression in Shuffle T7 or Arctic Express cells, different expression temperatures as well as periplasmic expression via a pelB-signal peptide. All attempts to produce soluble active GccA in sufficient amounts for further experiments proved unsuccessful.

Expression with an MBP-tag, however, allowed for the production of a GccA fusion construct in a soluble manner in *E. coli* Arctic Express (DE3) at 11°C. Startlingly, this fusion protein eluted from the affinity chromatography column as a strongly opalescent solution. Cleavage of the tag resulted in the immediate precipitation of the entirety of the GccA-protein of interest. Attempts to refold this protein via rapid dilution from a 6 M guanidinium-hydrochloride (Gdn-Hcl) lead to up to 100 µg per litre of culture when refolding into a buffer of 50 mM MES pH 5.5, 10% Glycerol. This soluble protein was highly unstable however, precipitating within 4 hours. A desalting step was performed directly after refolding in 50 mM MES pH 5.5, 10% Glycerol to remove trace amounts of Gdn-HCl. While this dramatically increased the stability of the protein, attempts to concentrate the protein using centrifugal concentrators failed. Taking all observed problems into account it was decided to not continue with GccA.

3.2.2. SacS

To characterise the putative GT, SacS from *L. sacchari*, the gene was amplified from genomic DNA and cloned into pET-3CLIC.

The created construct expressed well in *E. coli* BL21 (DE3) Gold and the produced recombinant protein was captured using immobilized metal ion affinity chromatography (IMAC). Removal of the imidazole via size exclusion chromatography (SEC) against a standard buffer (Tris pH 7.5, 150 mM NaCl) lead to pure homogeneous protein, which started to precipitate at concentrations above 10 mg/ml. This issue could be alleviated by increasing the pH to 8.0 and the addition of 5% (v/v) glycerol to the buffer. With the improved conditions, concentrations above 20 mg/ml could be reached, though precipitation still occurred with a half-time of the protein $t_{1/2}$ of approx. 24 hours. The stability of this protein could be dramatically increased by removal of the N-terminal His-tag prior to size exclusion. Even after 1 week at room temperature, no precipitation was observed and concentrations of >50 mg/ml could be reached.

The produced and purified wild-type SacS failed to yield diffraction quality crystals, as will be discussed in detail in a later chapter (chapter 3.10). Modifications were designed to facilitate the structure solution: A truncated GT SacS348, comprising the residues 1 to 348, and the SacSA-fusion construct, with a C-terminal extension comprising a short linker sequence followed by the mature peptide sequence.
The former was designed based on the structure of SunS[168] and the generated alphafold2[169] model of SacS. This deletion construct lacks the predicted α-helical dimerization domain. Additionally, SacS348noSt was created due to a reverse primer lacking the Stop-codon. This extended the C-term of the recombinant protein by 44 amino acids with the following sequence: RAFSSHMASMTGGQQMGRGSEFELRRQACGRTRAPPPPPLRSGC.
SacSA on the other hand extended the C-terminus of the GT by a short GSSG-linker, followed by the sequence of the SacA4Ser peptide (see chapter 3.3). The 4Ser sequence was chosen, as it was speculated that the disulphides of the wild-type peptide would prevent the peptide from binding to the GT active site as well as hinder the folding process by the formation of wrong disulphide connections.

Furthermore, a mutant of SacS, SacS(H200A) in which the putative, catalytic histidine is mutated to an alanine was created by site directed mutagenesis. All of these SacS-variants could be easily produced

in *E. coli* BL21 (DE3) Gold using the abovementioned protocol, established for the wild-type SacS. (Figure 21).

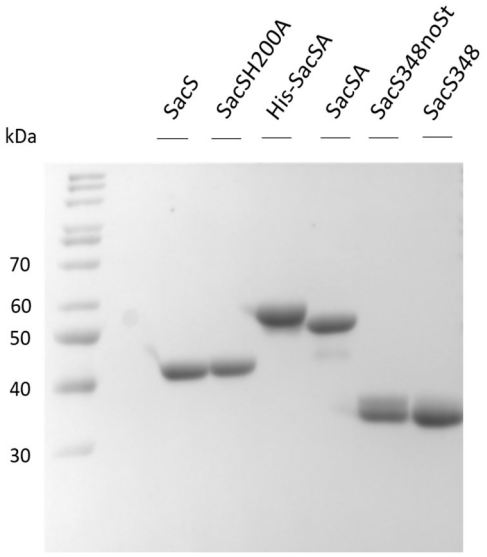


Figure 21 *SDS-Page of pure SacS and its derivatives.*

Interestingly, the fusion construct of SacS and SacA eluted later than the wild-type protein during SEC, despite being a marginally larger protein. Estimating the hydrodynamic radius with the reported separation properties of the used column[170] reveals the SacS to migrate akin to a protein of 100 kDa, matching the theoretical value of the dimer, whereas SacSA migrates akin to a protein of 40-50 kDa, indicating that this protein is present in a monomeric form.

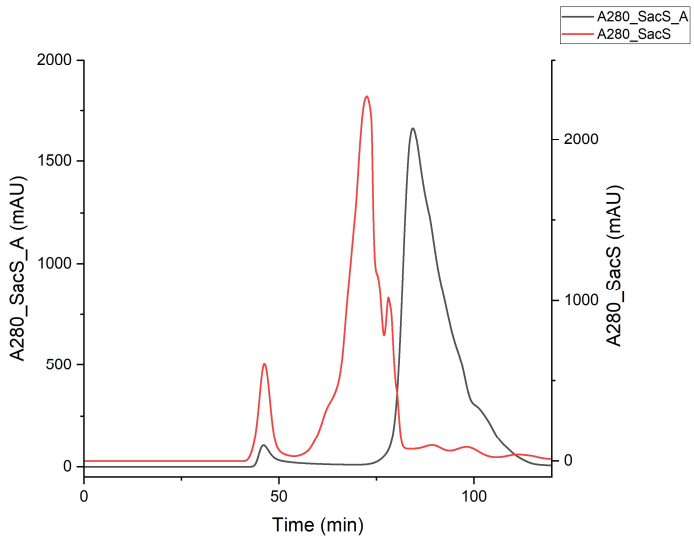


Figure 22 *Size exclusion chromatogram of SacS and SacS-A fusion construct. Absorbance was detected at 280 nm. A major shift in retention time implies the presence of SacS-A as a monomer as opposed to the wild-type dimer.*

3.2.3. SunS

A method to obtain SunS via heterologous expression in *E. coli* has already been described by Oman *et al.* (2011)[42]. Initial attempts to obtain pure GT were performed by cloning the *SunS*-gene into a pET-3CLIC vector and purifying the enzyme, following the published protocol[42].

While SunS could be successfully produced that way, much of the protein seemed to be present in the insoluble fraction of the lysed bacteria, indicating partial misfolding. Given that the small ubiquitin-like modifier (SUMO) protein has often been reported to assist with expression, solubility and folding[171], it was attempted to increase the yield of correctly folded protein by adding an N-terminal SUMO-tag, linked via a small GSSG linker. Furthermore, sequence analysis of SunS revealed the presence of rare codons. Hence, this His-Sumo-SunS construct was produced in *E. coli* Bl21 (DE3) Rosetta. After initial purification using IMAC, the tag was removed by addition of the Ulp1 protease. Similar to the method of Oman *et al*.[42], the usual reverse HisTrap step commonly used to remove the His-tag and the His-tagged protease, was skipped for SunS, as this protein bound to the immobilized nickel even in the absence of the N-terminal hexahistidine tag (Figure 23A). Therefore, the reaction mixture of the tag removal was purified by size exclusion only. Fortuitously, this method afforded sufficiently pure SunS (Figure 23B) in a yield of 30 mg/L culture. That amounts to an approx. tenfold increase, compared to the initial attempt using the published protocol[42].

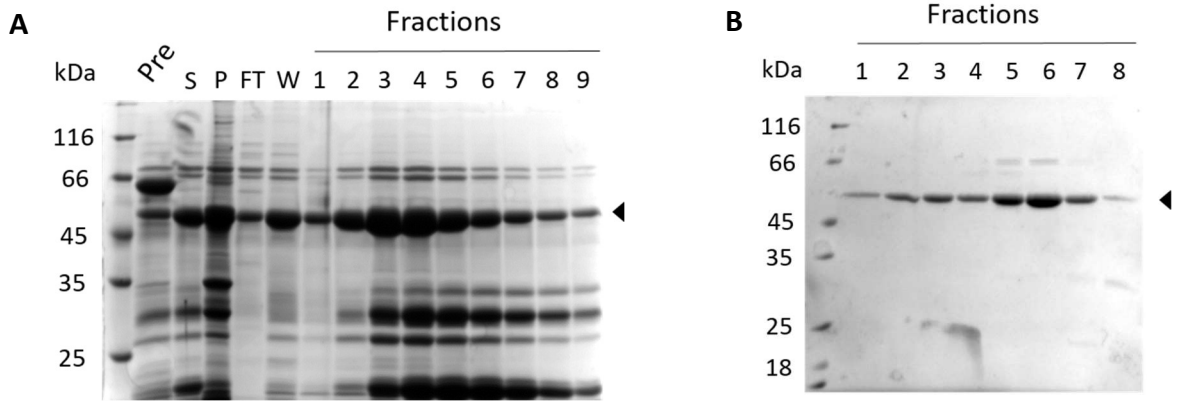


Figure 23 *A) SDS-PAGE of SunS after sumo-tag removal showing the pre-digest mixture (Pre), supernatant post digest (S), precipitate (P) as well as fractions of reverse IMAC corresponding to the flow-through (FT), wash (W) and elution fractions (numbered). A band corresponding to SunS can be found in all fractions. B) SDS-PAGE analysis of fractions collected from SEC purification of the proteolytic cleavage mixture of Sumo-SunS. The bands corresponding to the protein of interest are indicated with an arrow.*

3.2.4. AciS

Similar to the other GTs described above, the first approach to express AciS was in form of a construct with an N-terminal His-tag by cloning the gene in the pET-3CLIC vector. However, despite several attempts, ligation of the amplified DNA into the vector failed. Thus, AciS was cloned into a pET-YSBL vector instead. This vector is identical to the pET-3CLIC with the only difference being the absence of the 3CLIC cleavage-site. Here the ligation using standard methods proved successful. Given the vast number of rare codons in AciS, this protein was expressed in *E. coli* Bl21 (DE3) Rosetta, and a moderate amount of AciS could be detected after induction with IPTG.

However, while the protein was present in the lysate, retention on the His-trap column was poor and the protein precipitated immediately upon elution. The instability could be rectified by reduction of the pH of the purification buffers from 8.0 to 7.5 and the addition of 5% glycerol. However, this did not improve the problem of low retention on the affinity column.

I speculated that the issue may be caused by an inaccessible His-tag. Therefore, I re-cloned the *AciS* into the pET-YSBL vector making use of the option of a C-terminal His-Tag instead of the N-terminal one via frameshift mutation. Indeed, after expression of this construct, its ability to bind to the column was greatly improved. Since the SacS GT was greatly destabilized by the presence of a His-tag, the inability to remove it for AciS was an issue. AciS-3C-His was therefore created via extension PCR followed by intramolecular homologous recombination to introduce a 3C cleavage site between the C-terminus of AciS and the His-tag.

This construct could be expressed equally well and retained its ability to bind to the His-trap column. Up to a four-fold increase in yield (based on absorbance at 280 nm after HisTrap purification) could be obtained by swapping the expression media to Autoinduction medium (described in section 3.8). AciS displayed an increased 260/280 nm absorbance ratio, indicating the presence of nucleotides. Thus, an additional step for the removal of protease, residual tag and the contaminating nucleotide was achieved via ion exchange chromatography (Figure 24 A). AciS was present in almost all elution fractions, though these contained a large amount of impurities. Contrary, the flow-through contained almost pure AciS under these conditions (Figure 24 B). Therefore, the flow-through was collected, concentrated and further purified using SEC (Figure 25 A). Fractions, containing AciS were collected and concentrated to afford AciS in good yield (10 mg per litre of culture) and sufficient purity (Figure 25 B).

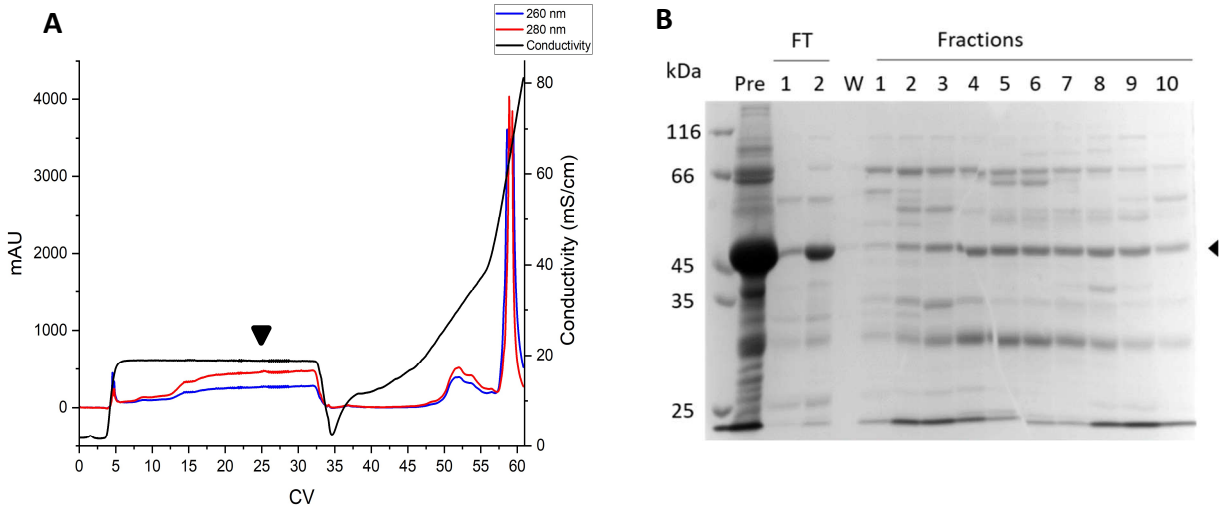


Figure 24 *Chromatogram of AciS during ion exchange chromatography (A). AciS eluted primarily during the flow-through at CV 10-35. A large peak of impurity can be seen eluting at CV 58. Samples of the mixture prior to separation (Pre), the flow-through (FT) and fractions were analysed by SDS-PAGE (B). Bands corresponding to the protein of interest are indicated by an arrow.*

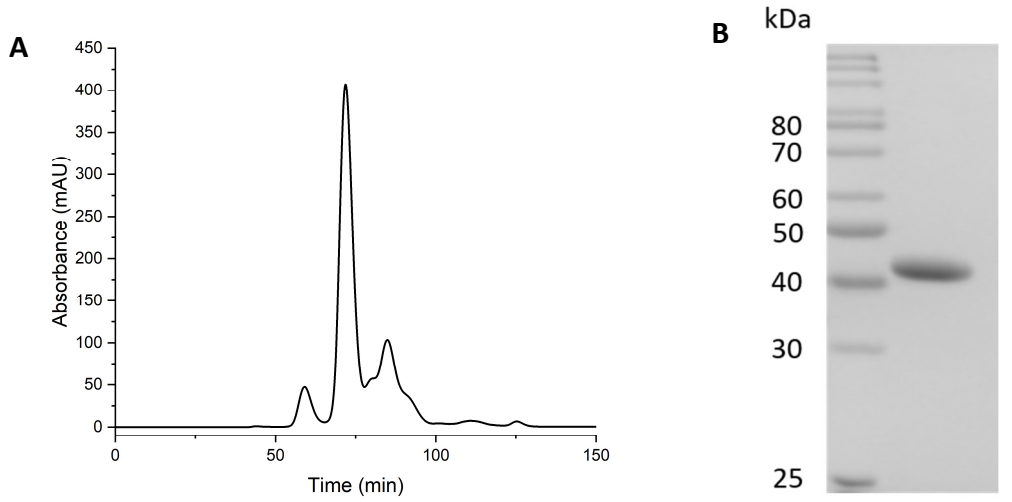


Figure 25 *A) Size exclusion chromatogram of AciS. B) Pooled fractions of AciS analysed by SDS-PAGE*

3.2.5. EntS

Similar to the SacS GT discussed above, the *EntS* gene was cloned from genomic DNA of *E. faecalis* into a pET-3CLIC vector. As an approach to obtain pure EntS by heterologous expression has already been published[172], I first attempted to adapt this protocol. However, in my hands, production levels were poor, and the protein could only be obtained in low yield. Analysing the DNA sequence of the *EntS* gene revealed many rare codons that could hamper the expression. Therefore, I attempted to express the gene in *E. coli* Bl21(DE3) Rosetta. Indeed, improved EntS production was observed in this strain and yields of up to 160 mg of pure EntS per litre of culture could be obtained.

The standard purification protocol for EntS consisted of a His-trap affinity chromatography, followed by tag-removal using 3C protease. Protease, His-tagged EntS and further impurities could be removed by using a reverse His-trap approach, collecting the flow-through. In a subsequent polishing step using SEC EntS eluted with a noticeable shoulder (Figure 26 A). SDS-PAGE analysis confirmed that both the shoulder as well as the main peak contained highly pure EntS. Based on elution time the main peak corresponds to the dimer, whereas the shoulder is likely to correspond to a higher oligomeric state. Fractions containing the dimeric EntS were pooled, concentrated and re-applied to the SEC to obtain dimeric EntS in high purity and yield.

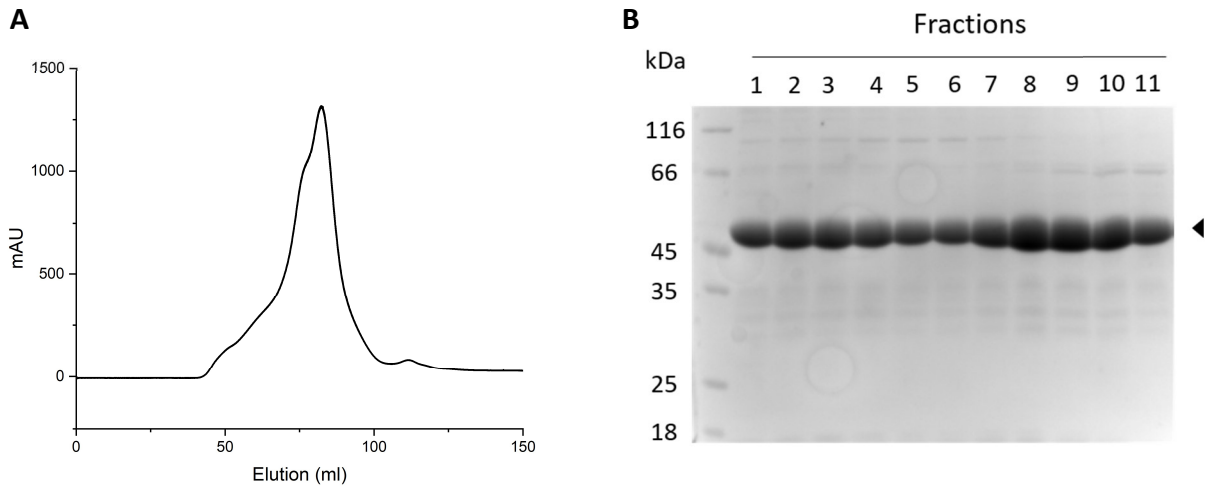


Figure 26 *A) Size exclusion chromatogram of EntS. Absorbance was measured at 280 nm. EntS appears to elute at min 70-80 in a peak with pronounced shoulder. Fractions from that area were analysed by SDS-Page (B) revealing a band corresponding to EntS for all fractions.*

In an effort to elucidate the reaction mechanism of EntS, a mutant was generated replacing the putative catalytic base to an alanine (see chapter 3.11.7). This EntSH214A was purified in a similar manner as the wild-type enzyme described above. However, this mutant unexpectedly displayed an increased ratio of absorbance at 260/280 nm, a property usually indicative of nucleotide impurities. Thus, an additional purification step was introduced before the final gel filtration. Dialyzing EntSH214A against a low salt buffer, followed by ion-exchange chromatography proved an effective method to remove said impurity from the enzyme. It is unclear why this mutant appeared to co-elute with nucleotide-based impurities, compared to its functional wild-type counterpart.

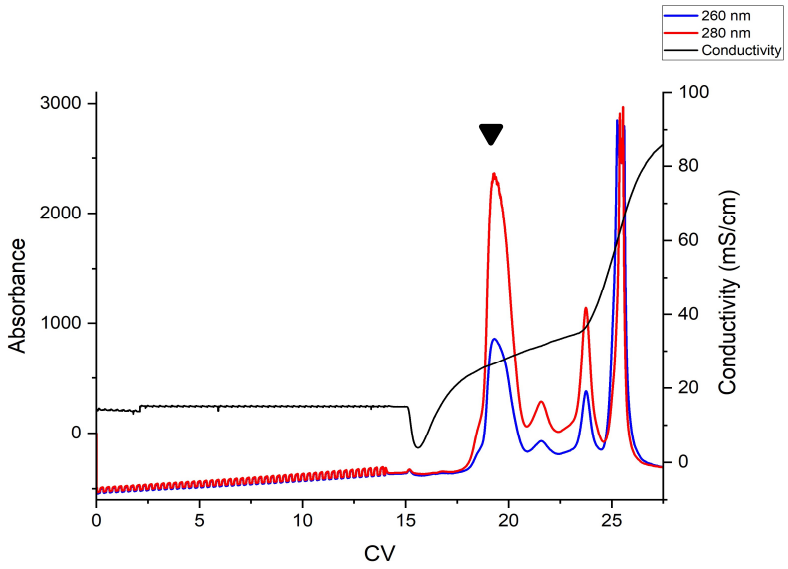


Figure 27 *Chromatogram of EntSH214A during ion-exchange chromatography. The peak corresponding to EntS is marked with an arrow. A strongly UV-active peak elutes at approx. 25 CV, likely corresponding to a DNA-based contaminant.*

3.2.6. PltS

Initial attempts to obtain PltS were performed with a construct in pET-3CLIC, introducing an N-terminal His-tag, in *E. coli* BL21 (DE3) Gold. However, while the protein was produced, the purity of the resulting protein was too low to be suitable for subsequent experiments, even after additional steps using IEX and SEC (Figure 28).

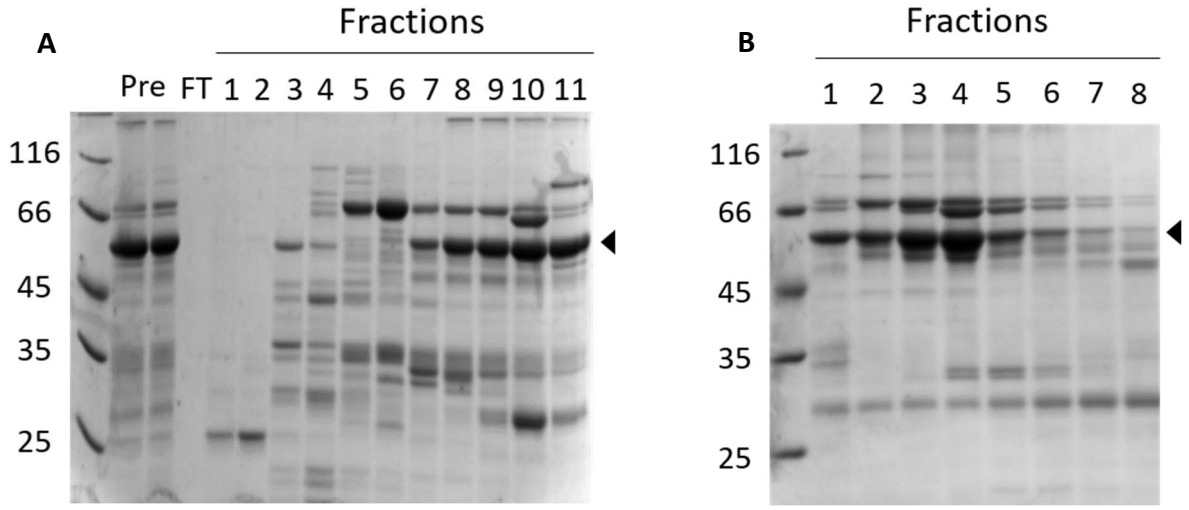


Figure 28 *SDS-PAGE of IEX (A) and SEC (B) fractions of PltS. The protein of interest is indicated with an arrow.*

Moving the His-tag to the C-terminus helped to remove the impurities but revealed 3 bands of similar size after IMAC. Western blotting followed by staining with an α-His antibody revealed the identity of all three of these bands to be PltS with intact C-terminus (Figure 29). Attempts to remove these impurities *via* IEX or SEC were unsuccessful.

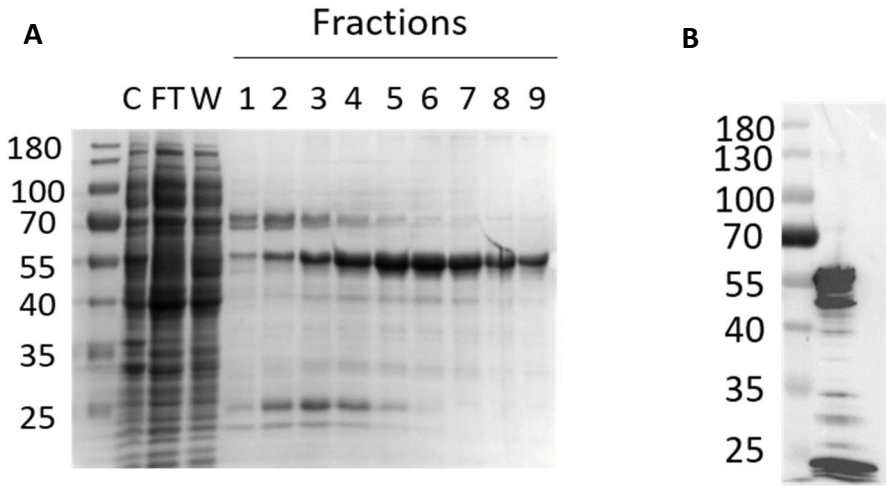


Figure 29 *SDS-PAGE of the affinity chromatography (A) of PltS-His and western blot of the pooled fractions (B) stained with His antibody reveals N-terminal truncations in the purified protein. The size of the full-length PltS-His is indicated with an arrow.*

An attempt to obtain a homogeneous protein sample was made by estimating the size of the truncation based on the SDS-PAGE, and pre-emptively remove the unstable section at the DNA level. To this end PltSΔ1-44 was created, lacking the first 44 amino acids. However, this construct proved to be unsuitably unstable and poorly folded in *E. coli.*

Given the apparent difficulties with the His-PltS purification a different approach was investigated, using another affinity-tag. The Strep-Tag-II was chosen, due to its small size, superior affinity and specificity, especially when compared to a His-Tag. The His-tag was replaced via an extension PCR of the pET-3CLIC-PltS plasmid, followed by intramolecular Gibson assembly. Expression of this construct occurred in *E. coli* BL21 (DE3) Gold, as described for the His-tagged protein. Surprisingly, the eluting fractions of the StrepTrap affinity chromatography were heavily contaminated with a 70 kDa protein, most likely DnaK (Figure 30A). As this protein acts as a chaperone in *E. coli*, it was attempted to reduce this contamination by expressing Strep-tagged PltS in *E. coli* Bl21(DE3) Arctic Express instead, to help with protein folding. This expression strain contains two additional chaperones that allow for expression at greatly reduced temperatures, lowering the chances of the protein misfolding that would recruit DnaK. Indeed, greatly reduced occurrence of the contamination at 70 kDa could be observed for PltS expressed in Arctic Express cells (Figure 30 B). However, this protein appeared to be extremely sensitive to divalent metal ions, usually an essential cofactor for Leloir type transferases, with up to 90% of it precipitating upon the addition, reducing the overall yield to a mere 100 µg per litre of culture.

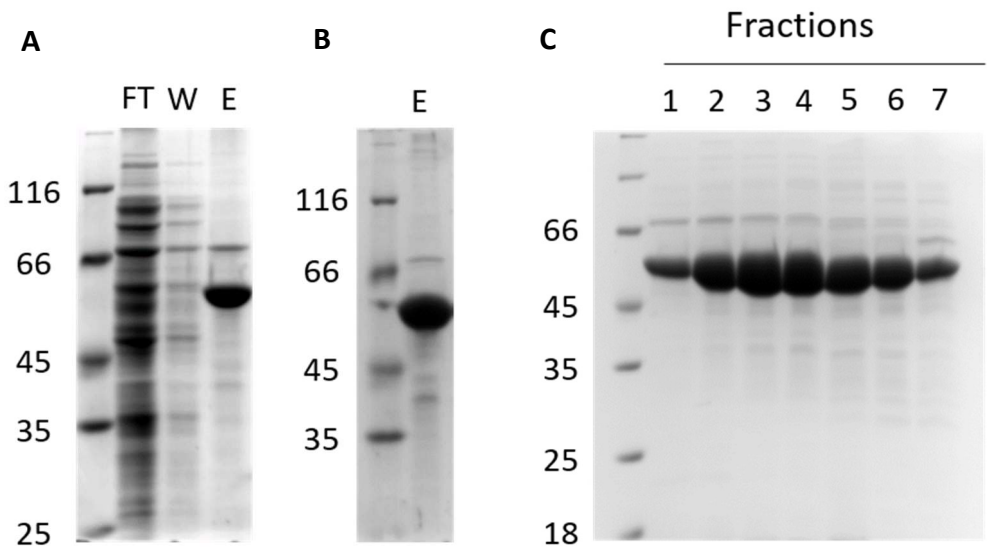


Figure 30 *Strep-PltS expressed in BL21 Gold (A) or Arctic Express (B) after affinity chromatography. FT corresponds to the flow-through fraction, W to the was fraction and E to the pooled desthiobiotin elution C) Fractions of the final SEC of PltS expressed as a Sumo-tagged construct.*

As this amount was insufficient for the planned structural studies, as well as the characterization of its mode of action, an improved method of PltS production and purification was required. As the SUMO-tag had proven successful in the SunS and AciS proteins mentioned above, PltS, including the Strep-tag-II and 3C cleavage site was cloned into a Champion pET-SUMO vector. While expression levels of this construct were fairly low in TB-medium, usage of the autoinduction medium, similar to AciS, increased the expression levels greatly.

Surprisingly, the initial affinity chromatography using immobilized nickel or cobalt already resulted in nearly pure protein, the latter showing a more beneficial elution profile, requiring lower concentrations of imidazole. Gratifyingly, the resulting GT fraction proved resistant to the addition of divalent metal ions such as $MgCl_2$ or $MnCl_2$, even at concentrations exceeding 5 mM. The His-, Sumo- and Strep-tag were removed from PltS by incubation with 3C protease followed by loading the cleavage mixture onto a His-Trap column to remove the cleaved Tag and the 3C protease. The flow-through, containing PltS was concentrated and loaded onto a size exclusion column, affording pure PltS in yields of 20 mg/L of culture (Figure 30 C).

3.3. Chemical synthesis of the peptide aglycones

To determine the enzymatic properties of the GTs, it was necessary to synthesize sufficient amounts of the acceptor peptide substrate first. Earlier work on SunS and EntS, have shown that the leader-peptide of glycocins is not necessary for its function or glycosylation[42, 45]. Therefore, I decided to synthesize only the core peptide and variants thereof.

A list of all peptides synthesized in this study can be found in Table 3. Analytical data for the pure peptides can be found in Annex I.

Table 3 *Peptides synthesized in this study. Changes from the natural sequence are marked in red.*

Peptide	***Sequence***
Enterocin96	MSKRDCNLMKACCAGQAVTYAIHSLLNRLGGDSSDPAGCNDIVRKYCK
Ecm	AVTYAIHSLLNRLGGDSSDPAG
EC(Cys)m	AVTYAIHSLLNRLGGD**C**SDPAG
EC(Ala)m	AVTYAIHSLLNRLGGD**A**SDPAG
SunAm	GKAQCAALWLQCASGGTIGCGGGAV
SunAmAllSer	GKAQ**S**AALWLQ**S**ASGGTIG**S**GGGAV
SunAm2Ser	GKAQ**S**AALWLQ**S**ASGGTIGCGGGAV
PltA	GMSKAECTYLYNLITTGATSSHGCVPSSNYLDLYRSNCKGKGPKL
PltAallSer	GMSKAE**S**TYLYNLITTGATSSHG**S**VPSSNYLDLYRSN**S**KGKGPKL
PltA2Ser	GMSKAE**S**TYLYNLITTGATSSHGCVPSSNYLDLYRSN**S**KGKGPKL
Leader-TEV-SacA	MDQLFKELKLEELE**NLYFQ**GFTAAQCAAFFVQCASGGTIGCGGMWHGRPAACDLYD QYCK
SacA	GFTAAQCAAFFVQCASGGTIGCGGMWHGRPAACDLYDQYCK
SacAallSer	GFTAAQ**S**AAFFVQ**S**ASGGTIG**S**GGMWHGRPAA**S**DLYDQY**S**K
SacA4Ser	GFTAAQ**S**AAFFVQ**S**ASGGTIGCGGMWHGRPAA**S**DLYDQY**S**K

The glycocin sequences were first evaluated *in silico* in order to identify sequence motifs known to be problematic in solid phase peptide synthesis. Several potential motifs prone side reactions could be identified. That includes typical reactions, known to occur during synthesis in an Fmoc-based strategy: Cyclization reactions such as aspartimide or glutarimide formation, racemization of His and Cys, oxidation or alkylation of Cys, Met and Trp, Asp-Pro cleavage and aggregation.

In order to prevent, or at least supress the occurrence of these side reactions several protective measures were used during the synthesis: Deprotection of the N-terminal Fmoc was performed using a piperazine-oxyma or piperazine-HOBt mixture. Given that both Histidine and Cysteine racemization are exacerbated by the use of high temperatures, His coupling was performed at lowered temperatures with similar approaches investigated for Cys. Boc-protected His and Ompe protected Asp were used as required. Similarly, thioanisole was added to the deprotection cocktail as necessary.

Concentration of the peptides was be performed at RT or lowered temperatures to avoid acid-catalyzed side reactions.

3.3.1. Synthesis of SacA based peptides.

Initial synthesis and purification of the wild-type SacA peptide using only standard couplings for the first 20 amino acids and double couplings for all subsequent reactions, were successful in generating pure peptide, albeit only after challenging chromatographic separation and only in low yield (1.8%). Furthermore, wild-type SacA showed only low solubility in aqueous media. Hence two approaches were explored to create a more soluble peptide, while retaining the wild-type sequence. A minimal peptide of SacA was prepared, based on published data for the highly similar Sublancin 186[74], including only the N-terminal helix and the loop until the +3 position from the glycosylation site. In SunA this is sufficient for recognition and decoration with a hexose. However, contrary to the Sublancin minimal peptide, the designed SacA minimal peptide displayed insufficient solubility for chromatographic purification.

The second method explored was to extend the sequence with the natural, N-terminal leader peptide. This sequence is highly enriched in glutamates and the resulting charge density should therefore confer greatly increased solubility in aqueous media at a basic pH. Given that the glycosylation happens before the transport through the membrane, the leader peptide should not influence the reaction with the substrate. Nevertheless, an ability to cleave off the leader peptide would be desirable. An analysis of the sequence showed that the leader sequence shows similarity to the TEV protease cleavage site (Figure 31). Thus, the peptide Leader-TEV-SacA was designed, which displays the TEV recognition sequence in lieu of the natural double-glycine cleavage site.

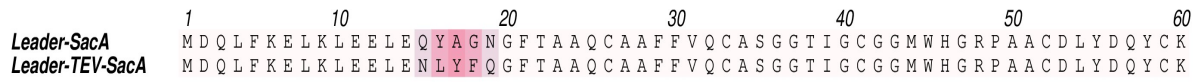


Figure 31 *Alignment of the wild-type Leader-SacA peptide with the designed degree of conservation is shown in a gradient from most conserved (white) to not conserved (red).*

As cysteine containing peptides were heavily prone to aggregation and sensitive to oxygen, several peptides with Cys-to-Ser substitutions were prepared as well.

SacAallSer replaces all cysteines for serines, to afford a highly soluble and stable peptide that would, due to its inability to form disulphide bonds, retain its extended, linear form. However this peptide would also lack the cysteine for the *S*-glycosylation reaction. Therefore, I also prepared SacA4Ser, which replaces only the Cys involved in disulphides with serines to retain its linear shape, while the Cys in the interhelix loop would still be available as acceptor in the glycosylation reaction.

In an attempt to improve the yield of the final product, different deprotection cocktails were explored, composed of either 94% TFA, 2.5% EDT, 2.5% H_2O, 1% TIS or 94% TFA, 2.5% EDT, 2.5% H_2O, 1% TIS, 1% thioanisole. The addition of thioanisole effectively prevented the occurrence of a large amount of side product (Figure 32). Usage of the thioanisole containing cleavage cocktail effectively doubled the yield of the prepared SacA peptides.

Another approach to optimization was the use of Fmoc-His(τ-Boc) and Fmoc-Asp(Ompe) during the synthesis to reduce the amount of histidine racemization and aspartimide formation, respectively. Use of these protecting groups as opposed to the standard Fmoc-His(Trt) and Fmoc-Asp(OtBu) did not result in a detectable increase in purity of the crude peptide.

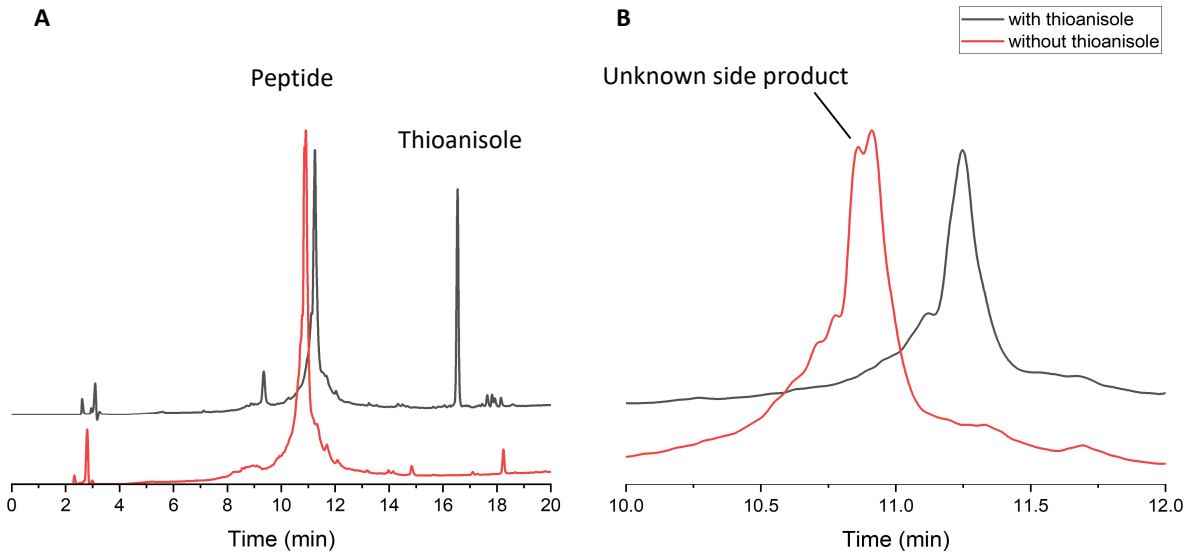


Figure 32 *Deprotection of SacA4Ser in the presence (black) or absence of thioanisole in the deprotection cocktail (A). Residual thioanisole elutes at min 16.4. A zoomed in view of the peptide peak (B) reveals a large side reaction inhibited by the addition of thioanisole.*

Further optimization experiments for the wild-type sequence revealed that one major side product was located within the first 10 amino acids, in particular the C-terminal Cys-Lys pair did not react very well. And attempt to improve the reaction rate by performing a double coupling step at 90°C, led to several side-product species instead, indicating that the Cys coupling may be temperature sensitive. While many different temperature-sensitive side reactions of Cys have been reported, as mentioned in chapter 1.4, these usually affect the very C-terminus of the peptide, with other positions generally considered to be less at risk. Nevertheless, coupling this Cys39 at 50°C for 10 min instead of 90°C noticeably improved the purity of the peptide product (Figure 33), further corroborating the temperature sensitivity of Cys to a Lys-Wang type resin.

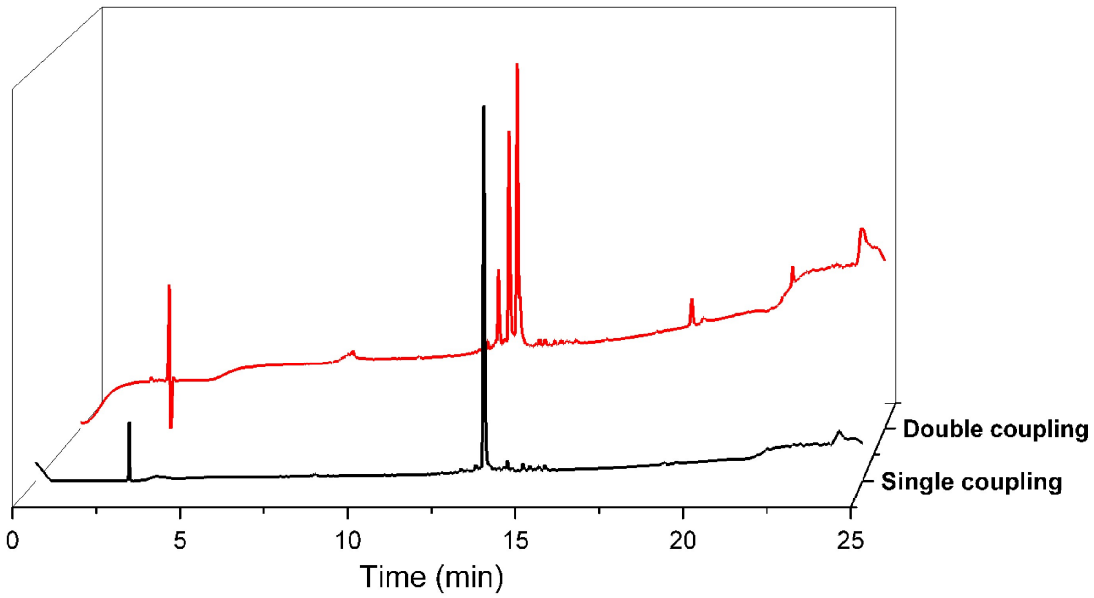


Figure 33 *Chromatogram of the crude SacA30-40 partial peptide with Cys(39) either single (black) or double coupled (red). Absorbance was measured at 220 nm and intensities were normalized. The use of double coupling for cysteines resulted in the emergence of three distinct side products.*

Other Cys couplings within the SacA sequence also seemed to suffer from incomplete couplings. Since the resulting Cys-deletion products proved difficult to separate from the main product, I chose to introduce an acetylation step with acetic anhydride after each Cys coupling. Attempts to couple all cysteines in a similar manner to Cys39 at 50°C with subsequent capping resulted in the emergence of an increased amount of secondary product. Analysis of these products with mass spectrometry revealed masses of 2963 and 2316 amu (Figure 34). These masses correspond to a failure of cysteine

coupling and subsequent acetylation at Ala15 and Gly23, respectively. The reason for the inefficiency of the Cys-Ala, and Cys-Gly couplings is currently unknown but may have been caused by aggregation of the peptide chain. Increasing the temperature of these Cys couplings back to 90°C and performing a double coupling step greatly increased coupling efficiency without showing the detrimental side effects observed for the Cys39-Lys40 pair.

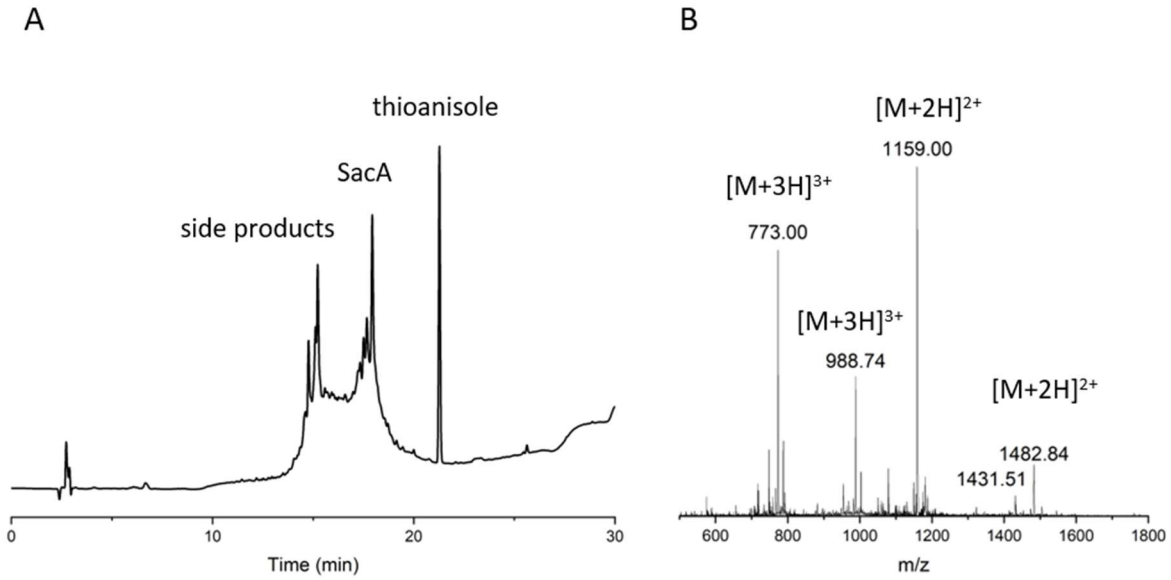


Figure 34 *Chromatogram of crude SacA-peptide synthesized via cysteine single couplings at 50°C (A). The side product peak at min 15 was collected and analysed using mass spectrometry (B). The mass indicates a mixture of truncated and acetylated peptide with deconvoluted mass of 2316 and 2963 amu, respectively.*

All of the synthesized and deprotected SacA-peptides, however, were still heavily prone to aggregation as a crude peptide, a problem that was only aggravated for the pure species. As mentioned in 1.4, the addition of fluorinated solvents, such as TFE or hexafluoroisopropanol (HFIP) is often recommended in such cases. Surprisingly, addition of HFIP to a solution of SacA caused immediate precipitation of the peptide as a white powder. This implied that these peptides interacted with fluorinated solvents to form insoluble aggregates or salts. Indeed, removal of residual TFA after purification increased the solubility of SacA peptides by several orders of magnitude. This effect was less pronounced for the Leader-TEV-SacA peptide, which remained largely insoluble in water. Taken together all variants of SacA could be synthesised under optimised conditions with yields ranging from 3.8 to 10.0 % final product.

3.3.2. Synthesis of SunA derived peptides

The previously mentioned minimal sequence of Sublancin 186[74] was prepared. Using standard methods, I obtained the desired peptide in low yield (0.9%), primarily due to the propensity of the crude peptide to form aggregates that proved resistant to treatment with reducing agent or fluorinated solvents such as HFIP.

Similar to the SacA peptide, two mutants were prepared to increase the solubility and stability of the peptide towards oxidation: SunAm2Ser and SunAmallSer. The former replacing Cys5 and Cys12 with serines, the latter replacing all cysteines that way. These peptides were much less prone towards aggregation as compared to the wild-type and could be prepared in moderate yield (8% and 15%, respectively).

3.3.3. Synthesis of AciA

The predicted glycocin of *G. acidiceleris*, was synthesized using a standard coupling strategy, afforded, after deprotection, a crude peptide mixture that resisted all attempts at solubilisation in organic or aqueous solvents. Thus, work with this peptide was discontinued.

3.3.4. Synthesis of Ent96 derived peptides

Since extensive work of Nagar&Rao[45] has already identified the minimal recognition motif of the EntS protein, first attempts were to synthesize the published sequence. However, the use of standard building blocks and methods proved unsuccessful with large amounts of side reactions occurring (Figure 35 A).

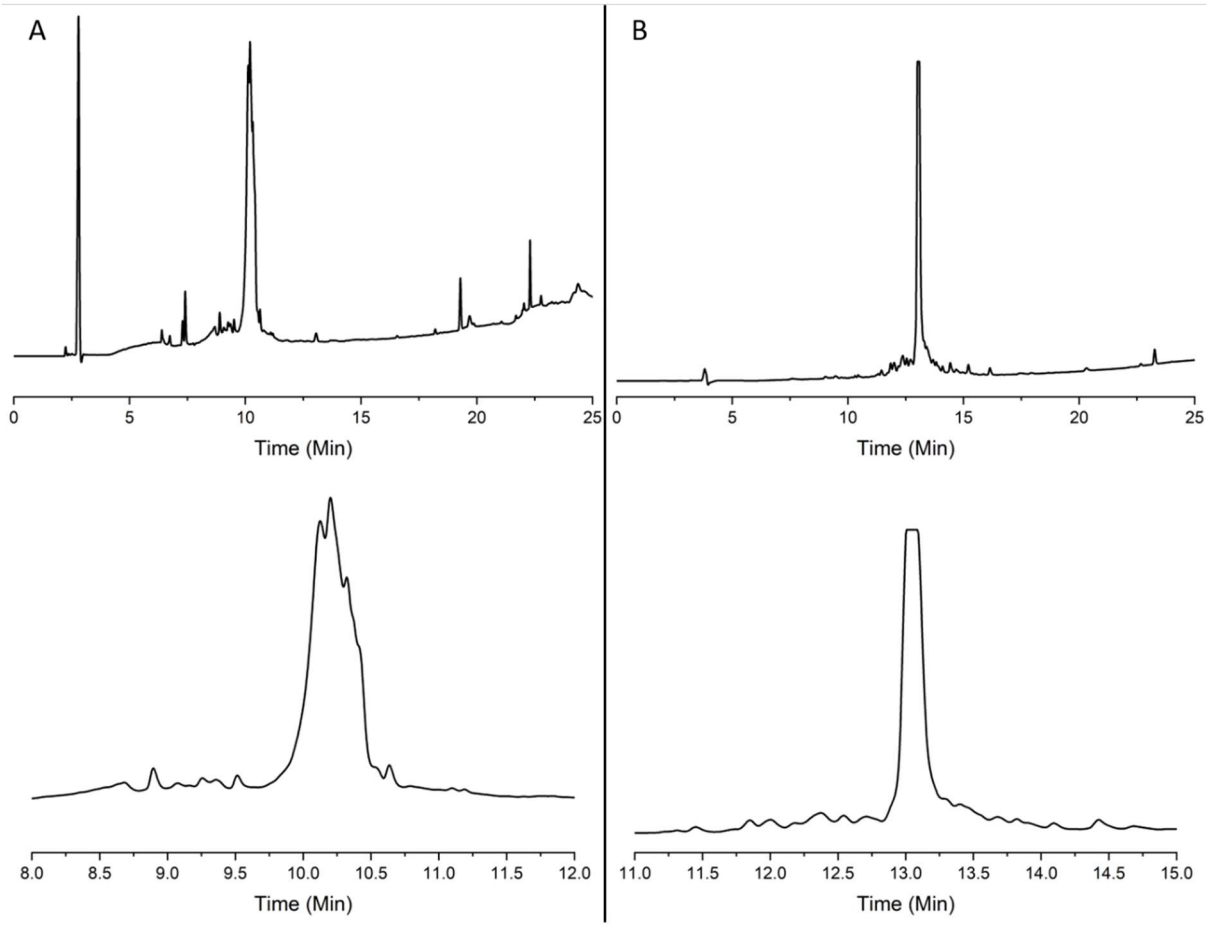


Figure 35 *Chromatogram of crude Ecm peptide prior to optimization efforts during synthesis (A) and after optimization of the synthesis procedure (B). For clarity a zoomed in view is shown on the bottom panels.*

For ease of synthesis, the minimal peptide of Ent96, was extended by a C-terminal Ala-Gly motif, as C-terminal Pro residues have been shown to be problematic, leading, to increased diketopiperazine formation and loss of the peptide chain[173]. Additionally-His(Boc) and Fmoc-Asp(Ompe) were used to reduce the occurrence of the aforementioned side reactions of racemization and cyclization, respectively. The combination of these measures allowed for the successful synthesis (Figure 35 B) and pure Ecm could be isolated in good yields (27%). An analogue to Ecm was designed with an S17C

mutation in the glycosylation site to observe differences between the *O*- and *S*-glycosylation of EntS. Using the same optimized method, synthesis of this mutant was successful, albeit with a much lower yield (14%). A final mutant of this peptide, EC(Ala)m, harbouring a S17A mutation was designed that would lack the nucleophile and therefore be an incompetent acceptor for an EntS-catalysed glycosylation. Synthesis and purification were performed congruent to the other Enterocin 96 minimal peptides leading to a good yield (29%).

The optimized protocol for Ecm together with the optimizations, developed for SacA was used to synthesise full-length Enterocin 96. Here, cysteines were coupled at 50°C and followed by acetylation of free N-termini. Fmoc-His(τ-Boc)-OH was used to reduce racemization reactions and Asp32 was introduced with an Ompe side-chain protecting group. Deprotection was performed similar to SacA in the presence of thioanisole and the peptide was subsequently concentrated at room temperature. This way the 48-mer peptide could be prepared in moderate yield (3%) and high purity.

3.3.5. Synthesis of PltA and derived peptides

Initial efforts to synthesize the identified PltA were performed by splitting the peptide into three blocks with test cleavages in between to analyses the synthesis progress. Already during reaction control of the first block of 20 amino acids, PltA(25-45), a large amount of undesired side-product could be seen (Figure 36).

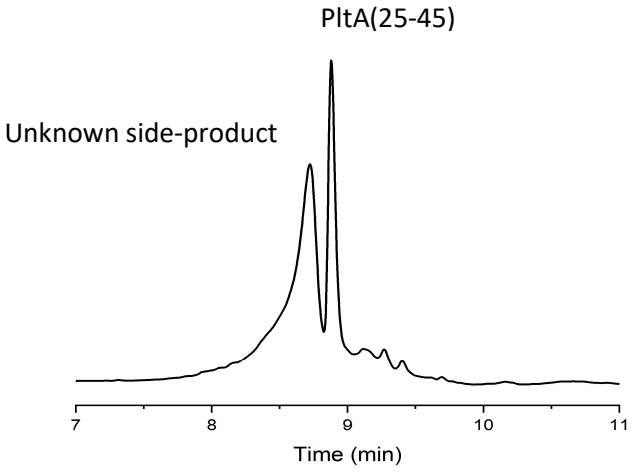


Figure 36 *Chromatogram of the crude PltA(25-45) partial peptide synthesized in a single block without additional washing steps.*

In an attempt to identify the problematic site within the peptide, the size of contiguously synthesized blocks was reduced to 10. Surprisingly, even though the coupling methods, amino acids and protecting groups remained the same between these synthesis attempts, the aforementioned problem was not present in the peptide synthesized in smaller blocks. As each reaction control is accompanied by a thorough wash with N,N-dimethylformamide (DMF) and dichloromethane (DCM), I wondered, if the impurity may stem from side reactions occurring due to incomplete removal of coupling reagent or a thus far unknown side reaction product during the microwave cycles. Thus, an additional wash step was introduced following each coupling step. This method proved successful at obtaining good purities and acceptable yields (3 %) even at contiguous blocks of 25 amino acids.

As with SacA and SunAm, two different Cys-to-Ser mutants of PltA were created: PltA2Ser (C7S, C38S) and PltAallSer (C7S, C24S, C38S). For PltA2Ser the remaining cysteine was coupled at 50°C and followed by a capping step to reduce potential side reactions. Contrary to SacA, no cysteine deletion sequence was detected at that point and PltA2Ser could be isolated in acceptable yields (4.6 %).

3.3.6. Stability of the synthesized peptides

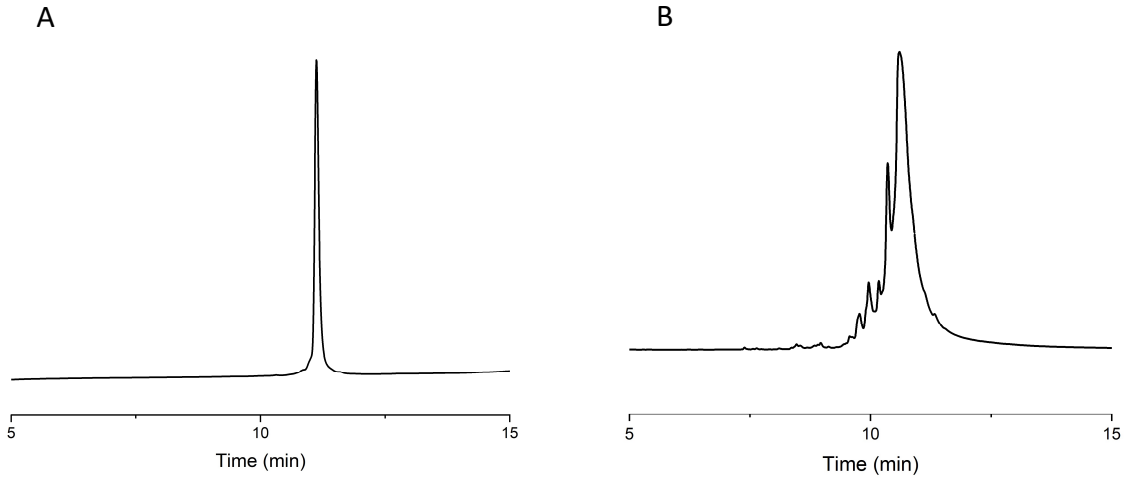


Figure 37 *Chromatogram of pooled fractions of SacA4Ser (A) and after one month of storage at -80°C (B). Absorbance was detected at 220 nm.*

While the minimal sequences of enterocin, Ecm and EcmCys proved to be fairly stable, all other synthesized peptides proved to be highly prone to ageing. For the wild-type SacA peptide, storage at -80°C for one week was sufficient to introduce a variety of additional species that could be separated via HPLC (Figure 37). However, these additional species proved to be hardly detectable via ESI-MS. A potential side product may correspond to a cis/trans isomerization reaction of proline, which has been reported to be in a dynamic equilibrium, even in cyclic disulphide-bridged peptides[174].

For the degradation of SacA4Ser a single side-product mass of -18 amu could be identified after prolonged storage (Figure 39). This is indicative of aspartimide or glutarimide formation and constitutes a common side reaction for peptides, as described further in 1.4.

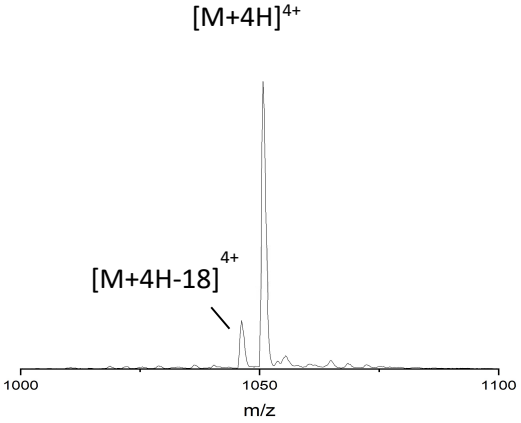


Figure 38 *Mass spectrum of SacA4Ser after extended storage at -80°C reveals a second peak at an m/z of 1046.23, corresponding to a +4 charged SacA4Ser species with a -18 amu modification.*

3.4. Determination of cosubstrates

In an attempt to identify the natural sugar donors of the respective GTs, I used isothermal titration calorimetry (ITC) to determine the K_d values for the nucleotide and the respective nucleotide sugar.

An initial attempt with UDP, because UDP-Glc as the most commonly found sugar donor for Leloir-type GTs, failed. As mentioned previously, the GTs of involved in the biosynthesis of glycocins were predicted to belong to the GT-family 2, a Leloir-type glycosyltransferase where the nucleotide binding is dependent on a metal cofactor (see chapter 1.3). Hence, I speculated that the metal cofactor may have been lost during purification. Indeed, binding to UDP could be detected in the presence of 1 mM magnesium-ions in the buffer. An increase in affinity of approx. 50-fold could be observed when the magnesium was substituted for manganese at identical concentrations (Figure 39). Considering this observation, all further affinity experiments were performed in the presence of 1 mM $MnCl_2$.

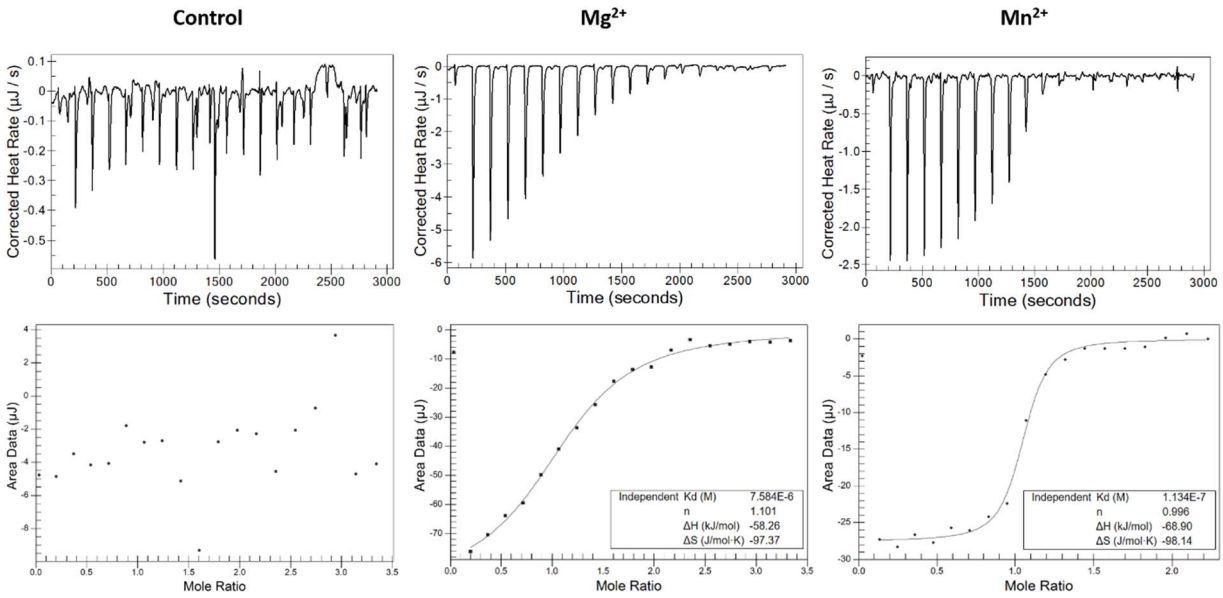


Figure 39 *Baseline subtracted data (top) and integrated peak area with fit assuming an independent set of sites (bottom) of an ITC measurement titrating UDP into SacS. Measurements were performed in the absence divalent metal salts or in the presence of 1 mM $MgCl_2$ or 1 mM $MnCl_2$. No binding could be detected in the absence of divalent metal ions. Binding parameters are shown in the bottom right corner of the fit.*

I performed the binding experiments for all four commonly used nucleotides and a number of potential sugar donors for four out of five GGT's at a standard temperature of 25°C (Table 4). The binding of UDP-GlcNAc to EntS had to be performed at 37°C, as the enthalpy change ΔH at 25°C was insufficient for reliable curve fitting. For PltS no values could be obtained, due to the instability of the protein in the ITC experimental setup. In all cases, glucose seems to be the preferred sugar to be transferred. Interestingly, SacS has an unusual high affinity for ADP-Glc, outperforming UDP-Glc by nearly an order of magnitude. This is surprising, given that ADP on its own binds with lower affinity than UDP. Thus, I also tried ADP-Ribose as substrate. Here a much higher K_d value was determined, pointing to a strong influence of the carbohydrate moiety on the donor affinity in the case of SacS. Due to the prohibitive costs, ADP-heptoses could unfortunately not be investigated. Overall, it was observed that most GGT's are quite specific for their donor substrates, with only one or two donors to be accepted with nanomolar affinities (Table 4). SacS proved to be the most promiscuous GGT, being able to bind all nucleotides and most sugar-nucleotides with reasonable affinities. In contrast AciS showed surprisingly high selectivity with regards to the sugar moiety, given its sequence similarity to both SunS and SacS. EntS, as member of a different glycocin-family showed different affinities, with a much higher selectivity with regards to both nucleotide as well as carbohydrate moiety.

Table 4 *Affinities of GTs towards several nucleotide and nucleotide sugars in the presence of 1 mM $MnCl_2$*

Ligand	***K_d [nM] SacS***	***K_d [nM] SunS***	***K_d [nM] AciS***	***K_d [nM] EntS***
UDP	110 ± 30	240 ± 100	420 ± 50	2500 ± 1100
ADP	260 ± 60	1600 ± 400	1500 ± 220	87700 ± 16400
GDP	1100 ± 200	6200 ± 1700	8850 ± 970	77400 ± 8000
CDP	1300 ± 300	5300 ± 300	4900 ± 1800	182000 ± 60000
UDP-Glc	240 ± 40	200 ± 15	140 ± 80	380 ± 60
UDP-GlcNAc	2500 ± 950	2000 ± 230	26300 ± 4500	4700 ± 130*
UDP-Gal	2200 ± 400	1800 ± 150	16200 ± 740	8700 ± 1300
UDP-Xyl	6400 ± 1500	14000 ± 2500	24600 ± 500	12400 ± 3300
GDP-Man	1200 ± 350	1600 ± 980	11400 ± 450	23000 ± 12000
ADP-Rib	1300 ± 760	44000 ± 5600	26200 ± 2700	51000 ± 33800
ADP-Glc	35 ± 25	720 ± 230	1300 ± 230	8500 ± 4800

* Experiment performed at 37°C

3.5. Glycosylation of synthetic peptide substrates

After successful production of both the GT (chapter 3.2) as well as their proposed glycocin acceptor peptides (chapter 3.3), the glycosylation pattern facilitated by the GTs could be investigated.

3.5.1. Glycosylation of SacA

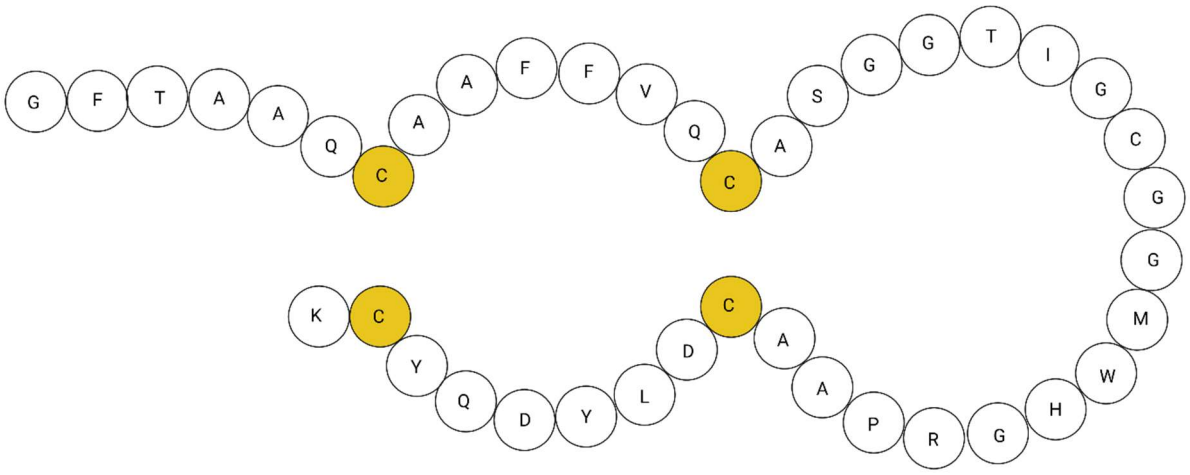


Figure 40 *Schematic representation of the SacA peptide. Cysteines forming the predicted disulphides are shown in yellow.*

To confirm that SacS is able to glycosylate its predicted glycocin, SacA, the wild-type peptide was dissolved in aqueous buffer and incubated together with SacS, $MnCl_2$ and UDP-Glc as the most likely glycosyl-donor, given its excellent affinity to the GT (chapter 3.4) and availability in most organisms[109]. As the predicted glycosylation site required the presence of a free sulfhydryl group, this reaction was carried out under reducing conditions (10 mM TCEP). A pH of 8.0 was chosen, as SacS had shown increased stability under these conditions (chapter 3.2.2). Unfortunately, the wild-type SacA peptide proved highly unstable at this pH, precipitating within 15 min. In order for efficient glycosylation, it was necessary to outspeed the aggregation of the peptide. Therefore, the reaction was performed at a comparatively large concentration of SacS (5 µM, 0.05 eq.), to that of SacA (100 µM). Nevertheless, precipitation of the peptide occurred during the time course of the reaction. While some of this precipitate could be re-dissolved upon acidification to pH 3, the vast majority of it remained insoluble. Investigation of the acidified supernatant of this reaction with RP-HPLC revealed a shift in retention time from 13.09 min to 12.63 min for the peptide (Figure 41). Mass spectrometry revealed the

expected mass of the glycosylated peptide, lacking any disulphide bridges, likely caused by the high concentration of reducing agent.

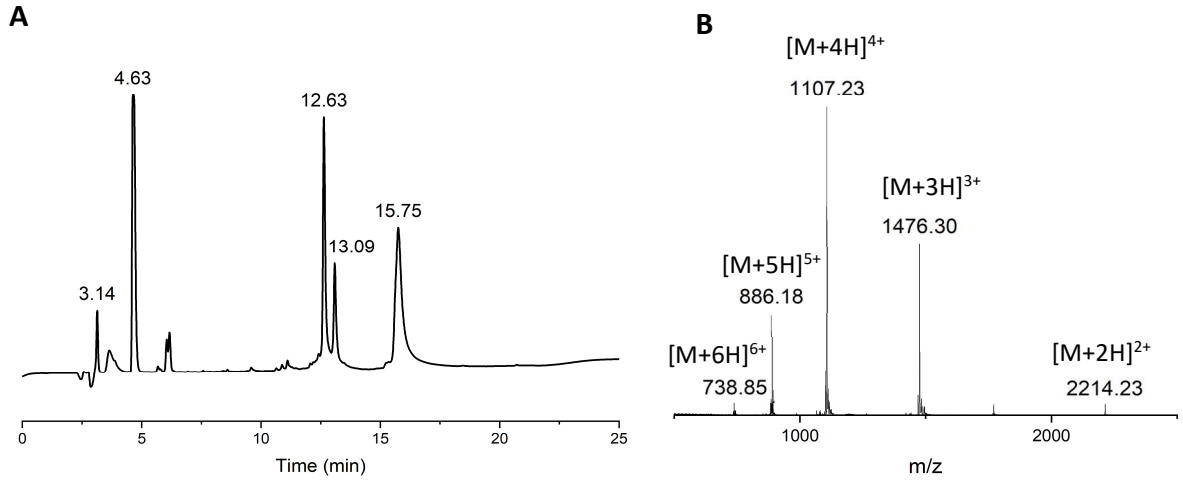


Figure 41 *Chromatogram of the SacA glycosylation reaction with SacS (A). The peak at 12.63 min was collected and analysed via MS (B), showing a mass/charge distribution series indicative of* $[SacA+162+XH]^{X+}$.

Interestingly two further peaks were obtained during the analytical HPLC of the reaction mixture, eluting at min 4.63 min and 15.75 min, respectively. Collection of these peaks and analysis via mass spectrometry did not result in any discernible signal, despite the pronounced UV-activity. Given that these peaks are dependent on the presence of the peptide (Data not shown), it is presumed that these peaks correspond to aggregated forms of SacA which, owing to their size, may be difficult to ionize.

With proof in hand that SacS is indeed active towards the SacA peptide, I proceeded to find the optimal reaction conditions for subsequent experiments. While the presence of salt negatively impacted the solubility of the peptide and especially the emergent, glycosylated species (Data not shown), the choice of pH was, at first, difficult to estimate, given that the SacS enzyme remained stable and in solution in a pH range of 2-12. Thus, a similar reaction as described above was performed in different buffer systems ranging from pH 5.0 to 9.0. To avoid the inherent instability of the SacA-peptide described above, I made use of the SacAallSer peptide that showed greatly improved solubility, especially at basic pH. A change in retention time compared to a control barring any GT was considered indicative of glycosylation. To estimate the efficiency, the glycosylation was quenched after 1h. No glycosylation could be observed at acidic pH, with the concentration of glycosylated species increasing with higher pH (Figure 42). The reaction at pH 9, however, displayed a third peak, indicative

of an unknown side reaction. Given that many different base catalyzed reactions can occur for this peptide, this behaviour is not entirely surprising. Thus pH 8.0 was considered to be the optimal pH for SacS activity measurements.

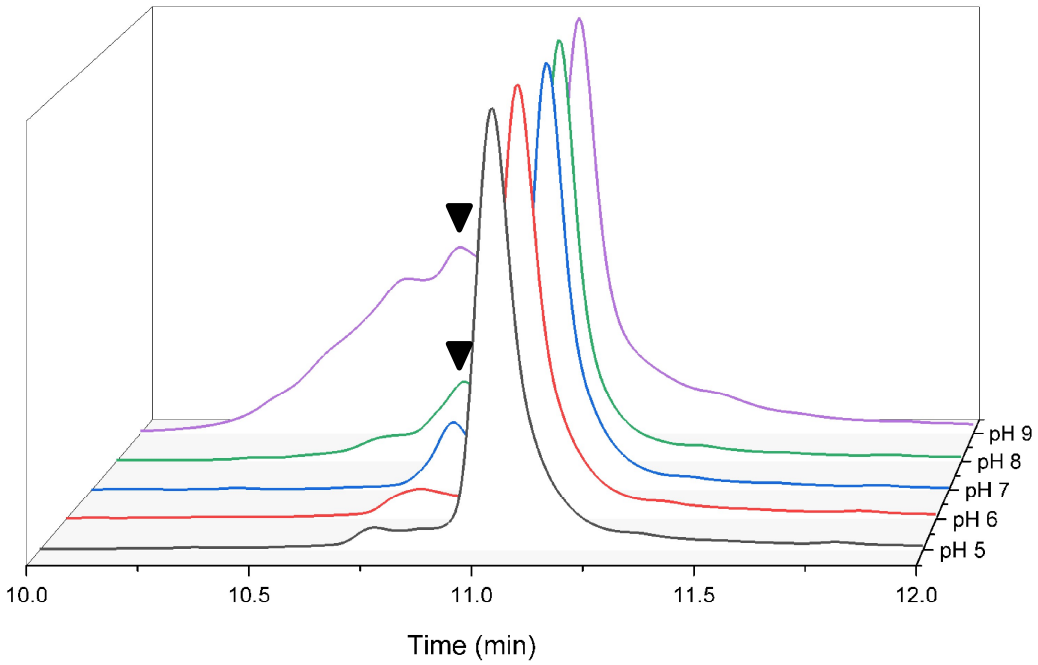


Figure 42 *Reaction of SacAallSer with UDP-Glc within 1h at different pH. Absorbance was measured at 220 nm and intensities were normalized. The product peak is labelled with an arrow. A peak corresponding to an unknown side product arises at pH>8.0.*

Given that SacS displayed extraordinary affinities to most sugar nucleotides, compared to the other investigated GTs (see 3.4), I wondered if this promiscuity would translate to glycosylation activity. Similar to the determination of optimal pH, several reactions of SacS and SacA were set up with different sugar-nucleotides as donor, the reaction mixture lacking any sugar-nucleotide was used as a control. After 1h, shifts in retention time were observed for ADP-Glc, UDP-Glc and UDP-GlcNAc, indicating that these are the preferred sugar-donors of the SacS enzyme (Figure 43A). Unfortunately, a peak with similar retention time to the glycosylated peptide can be observed for all peptides irrespective of the sugar donor or presence of GT. This peak is likely caused by peptide aging during storage (see chapter 3.3.6) and made the evaluation of less efficient sugar donors problematic. Therefore, the glycosylation efficiency of SacS using UDP-Gal, UDP-Xyl, ADP-Rib and GDP-Man was evaluated after 24 hours reaction time. This resulted in glycosylation with UDP-Gal and UDP-Xyl, but not GDP-Man or ADP-Rib (Figure 43 B).

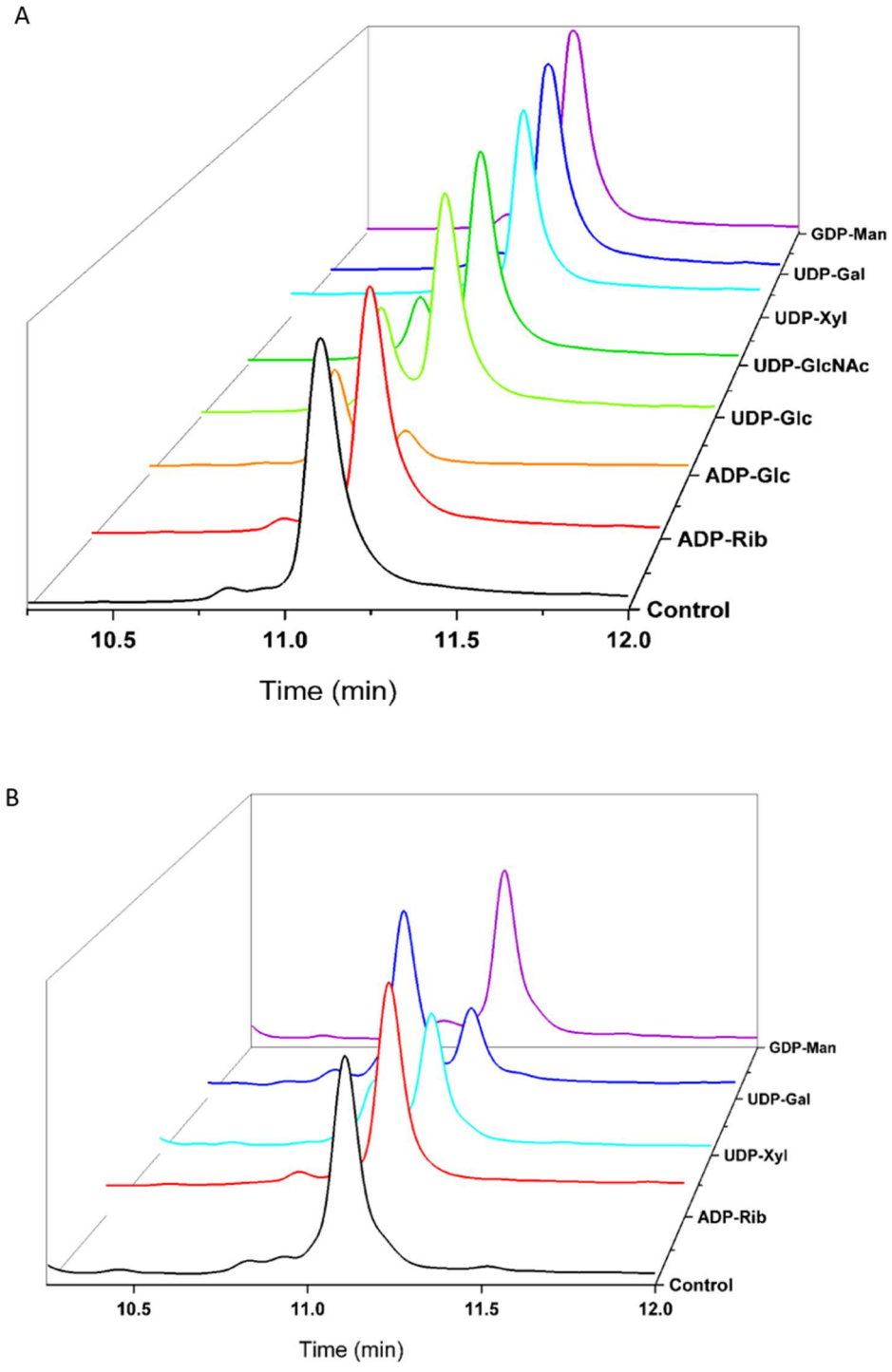


Figure 43 *Glycosylation of SacAallSer with different nucleotides within 1h (A) and 24 h (B). While only UDP-Glc (orange) and UDP-GlcNAc (light green) show a shift in retention time within 1h, this shift can also be detected for UDP-Xyl (light blue) and UDP-Gal (dark blue) after 24 h.*

Surprisingly, when the glycosylation was performed using SacS348 mutants in lieu of the full-length protein, residual glycosylation could be observed (Figure 44). The peak corresponding to the aglycone is larger than that of the glycosylated peptide indicating a much slower speed of the reaction, compared to the wild-type enzyme, most likely due to the lowered affinity towards the acceptor peptide.

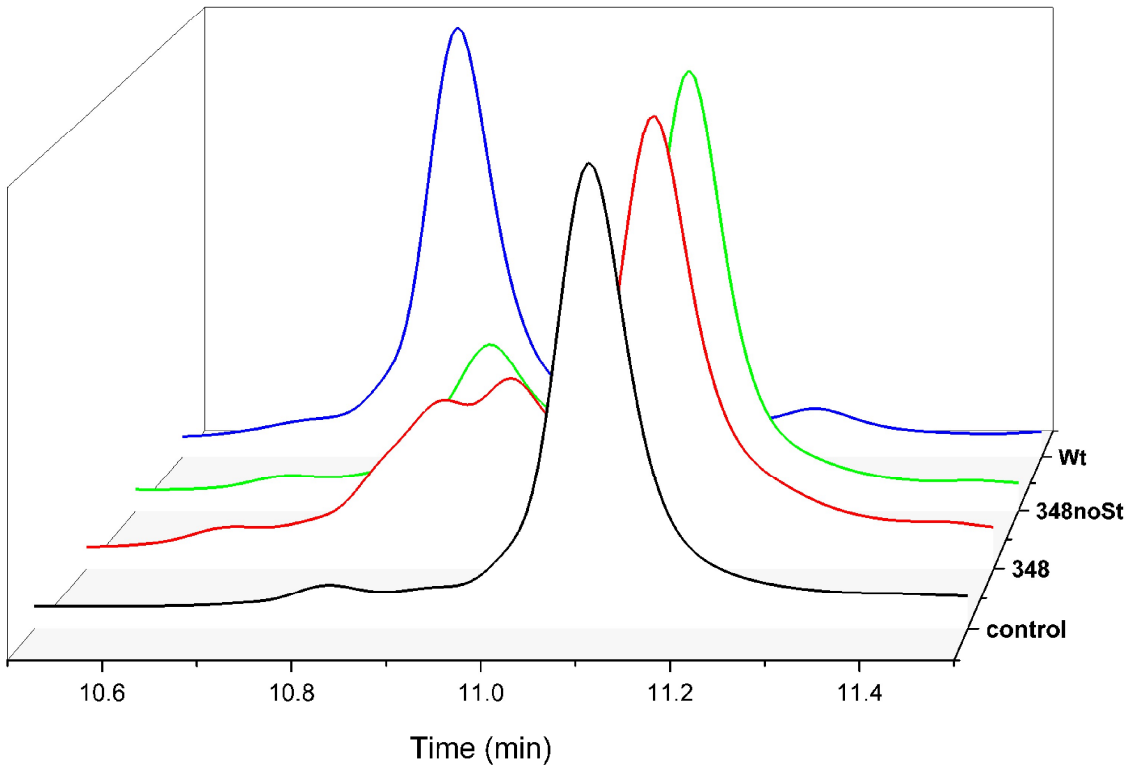


Figure 44 *Chromatogram of the glycosylation of SacA4Ser with UDP-Glc after 24 hours using either SacS348 (red), SacS348noSt (green) or wild-type SacS (blue) as catalyst. A corresponding mixture lacking any enzyme (black) was used as control.*

3.5.2. Glycosylation of SunA minimal peptide

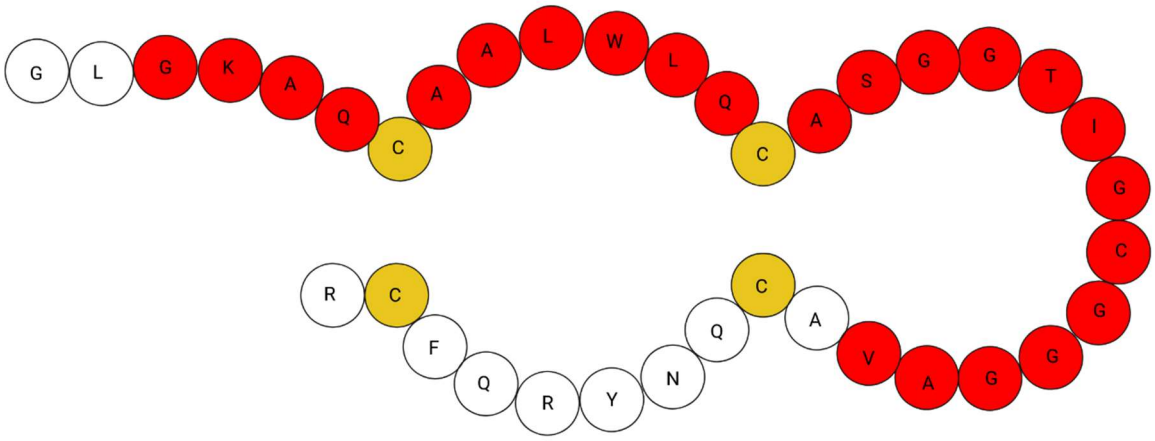


Figure 45 *Schematic representation of the Sublancin peptide. Cysteines forming disulphides are shown in yellow. The reported minimal peptide is shown in red.*

The properties of the SunS GT have been already investigated[42, 74]. Using the reported procedure, *S*-glycosylated SunAm peptide could be obtained. In a similar manner, the SunAm2Ser peptide could be glycosylated as well (Figure 46), indicating that the cysteines involved in disulphide formation are not critical for peptide recognition. Consistent with previous reports[42], SunS failed to glycosylate the SunAmallSer peptide. No change in either mass nor retention time could be observed, even at high amounts of sugar-donor and transferase.

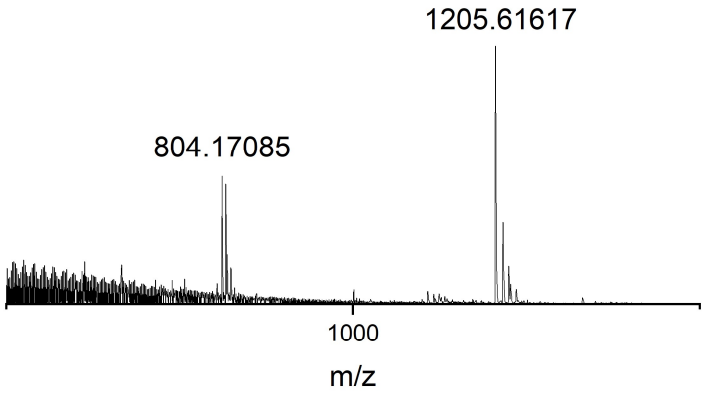


Figure 46 *Mass spectrum of SunAm2Ser-Glc, deconvoluted mass: 2409.218, calculated mass: 2409.517. Additional peaks correspond to +Na peaks.*

3.5.3.Glycosylation of Enterocin 96 and its minimal recognition sequence

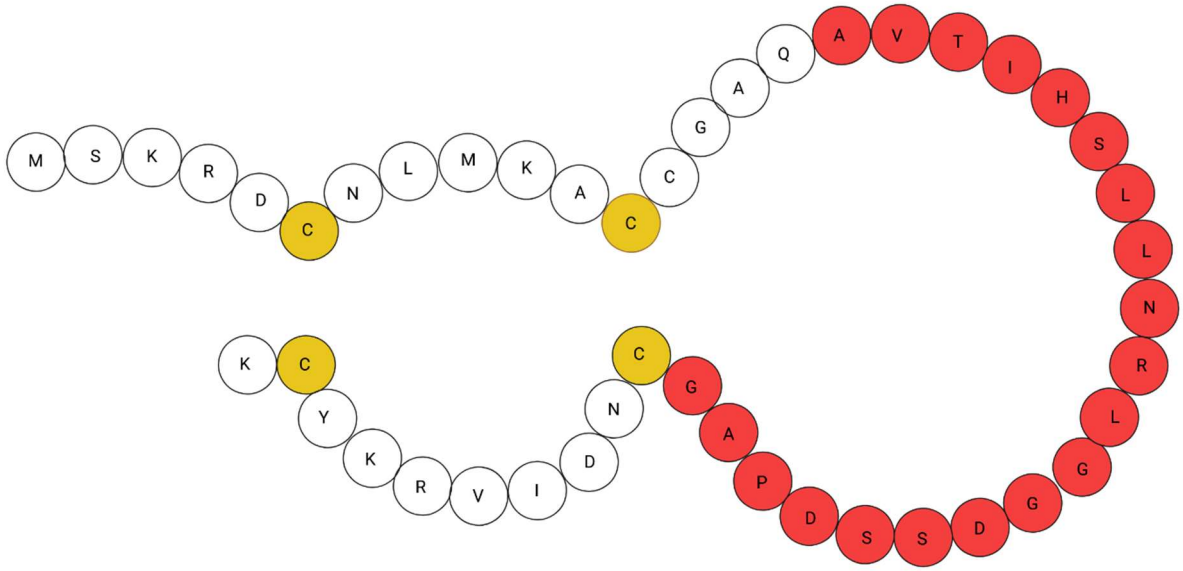


Figure 47 *Schematic representation of Enterocin 96. Cysteines predicted to be involved in disulphide formation are shown in yellow. The Ecm minimal peptide sequence is shown in red.*

The glycosylation of Enterocin 96 and peptides derived from it, has already been reported elsewhere[172]. Given earlier observations about the metal-depended affinity, manganese was substituted for the magnesium compared to the published procedure. Using this protocol, the minimal peptide Ecm could effectively be glycosylated. Surprisingly, when the glycosylation reaction of Ecm was stopped early by the addition of HCl, it could be observed that the monoglycosylated Ecm peptide was present in much lower concentration than both the diglycosylated and the unglycosylated peptide (Figure 48). This may indicate that the second glycosylation step is much more efficient than the first. This is in stark contrast to the data reported for full-length Enterocin 96, where the monoglycosylation noticeably preceded the extension of the carbohydrate. A possible explanation for this discrepancy may be a lowered binding affinity of the used minimal peptide towards the enzyme, which may noticeably improve with the addition of the glucose moiety. Unfortunately, the affinities of the Ecm peptide were too low, and the heat of dilution too high too high to reliably determine the affinities using ITC.

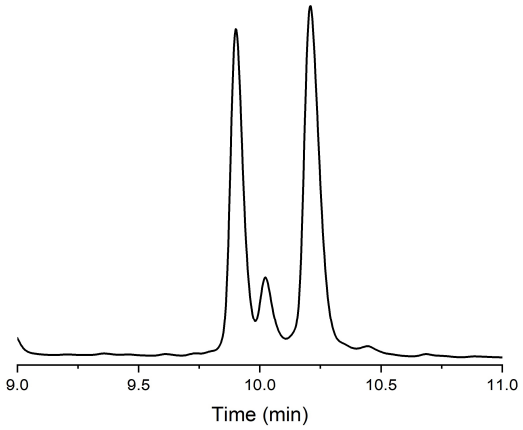


Figure 48 *Glycosylation of the Ecm minimal recognition sequence after 1 hour. Zoomed in on the area of Ecm elution. Glycovariants of Ecm elute at 9.9 min (diglycosylated), 10.02 min (monoglycosylated) and 10.21 min (nonglycosylated).*

Consistent with the results published by Nagar&Rao[45], EntS was also capable to glycosylate the S33C mutant of Ecm, though the reaction had to be performed in the presence of 1 mM TCEP to prevent peptide aggregation. This reaction occurred noticeably slower, with no glycosylated product observable within 1 hour. Extending the time to 24 hours allowed for the efficient production of EC(Cys)m-GlcGlc, however.

Similar to the Cys-mutant of Ecm, the glycosylation of the full-length peptide Ent 96-aglycone was performed under reducing conditions to avoid aberrant formation of disulphides and the mature, diglycosylated Ent 96 glycocin could be isolated by HPLC in a similar manner.

3.5.4. Glycosylation of PltA

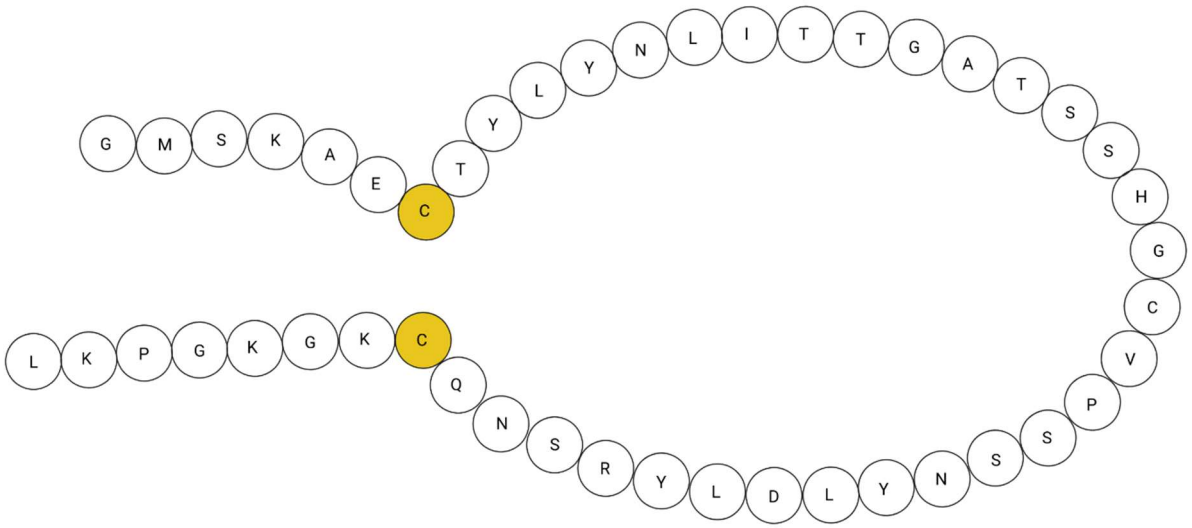


Figure 49 *Schematic representation of the PltA peptide. Cysteines predicted to form disulphides are shown in yellow.*

Given that PltS proved to be unstable under the conditions of ITC, no previous data were available to predict the natural co-substrate of the enzyme. Thus, a screen of several sugar nucleotides was performed, equivalent to the one described above for the SacA peptide.

Interestingly, PltS proved much more selective in its choice of sugar nucleotide, showing no reaction with any sugar donors except for UDP-Glc, ADP-Glc and UDP-GlcNAc within 16 hours (Figure 50). Both the Glc and GlcNAc decorated glycoforms of PltA could be produced *in vitro* in this manner for all of its mutants. Given that the *in vivo* produced PltA (described in detail later, see 3.8) was exclusively decorated with a HexNAc, the natural sugar donor is most likely UDP-GlcNAc.

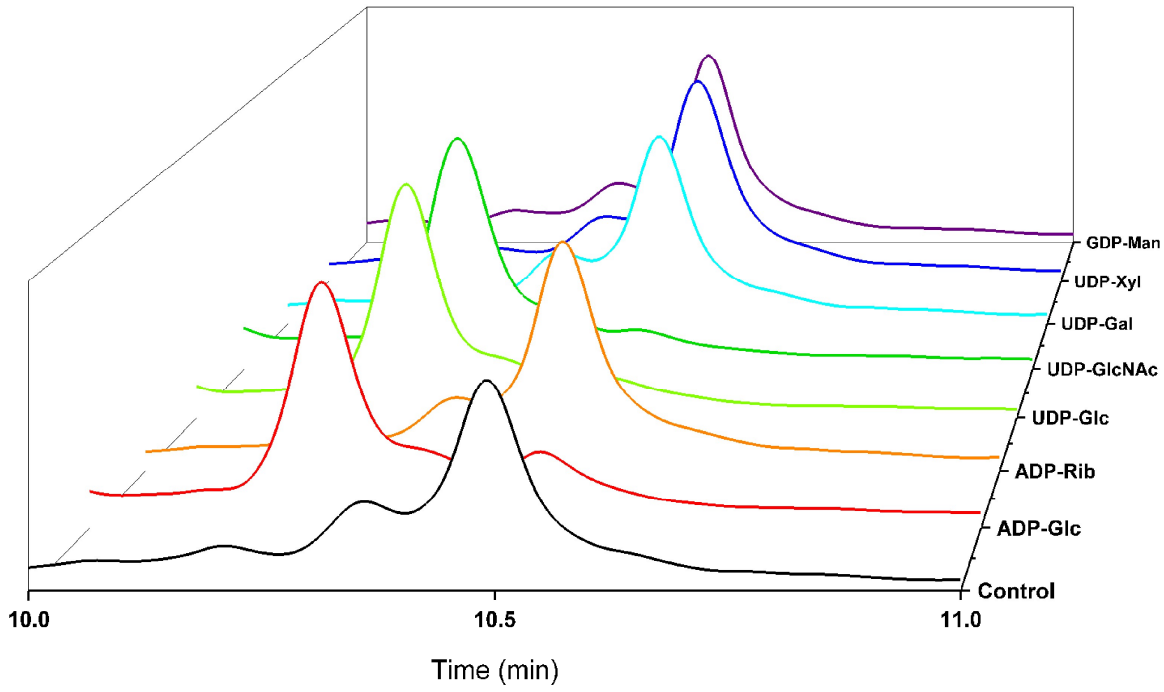


Figure 50 *Chromatograms of PltA incubated with PltS and different sugar-nucleotides. Zoomed in between min 9.5-11 for better visibility of shifts in retention time. Only ADP-Glc (red), UDP-Glc (light green) and UDP-GlcNAc (dark green) showed noticeable shifts in retention time indicative of glycosylation. Incubation with PltS lacking any sugar donor was used as control (black).*

3.6. Acceptor peptide specificity

For GGTs there have been reports of extraordinary promiscuity towards the acceptor peptide sequence. For example, ThuS is reported to accept the sublancin peptide as well as peptides containing sequences originating from γ-synuclein or cyclic AMP-response binding protein, with little similarity to the natural substrate beyond an N-terminal helix[43]. Thus, I have investigated the ability of the GTs described in this work to glycosylate each other's peptides (Table 5). The acceptor peptide was incubated with 0.1 eq of GT and 20 eq. of the preferred sugar donor. After incubation for 24 h the reaction was quenched with hydrochloric acid and the mixture analysed with RP-HPLC. Given the inherent instability of several of the wild-type peptides, their much more stable mutants SacA4Ser, PltA2Ser and SunAmallSer were used in these experiments.

Table 5 *Glycosylation of peptides by different GTs. ✓ indicates glycosylation, (✓) indicates weak glycosylation, X indicates no observable reaction. ND is not determined.*

	SacS	SunS	AciS	EntS	PltS
SacA4Ser	✓	✓	(✓)	X	X
PltA2Ser	X	X	X	X	✓
Ent96	X	X	X	✓	✓
SunAmallSer	X	X	X	X	✓
SunAm2Ser	(✓)	✓	X	X	ND
SunAm	X	✓	X	ND	ND

Given the unique sequence of PltA, it is unsurprising that none of the other GTs were able glycosylate this peptide, whereas all SunS-like GTs were able to decorate SacA4Ser, albeit only to a minor extent in the case of AciS. Similarly, in agreement with published data, no SunS-like glycosyltransferase was capable of glycosylating Ent96. However, surprisingly, PltS was capable of catalysing a reaction for both Ent96, as well as SunAmallSer. Interestingly, despite the pronounced sequence similarity between SunA and SacA within the minimal recognition sequence of the former, SacS was unable to glycosylate SunAmallSer. That is, despite evidence that SacA is, in fact capable of *O*-glycosylation in case of SacAallSer (Figure 43). Changing the nucleophilic species back to a sulphur atom in SunAm2Ser allowed both SunS and SacS to glycosylate this peptide (Figure 52 A). EntS and AciS were still unable to glycosylate this SunA-derived peptide.

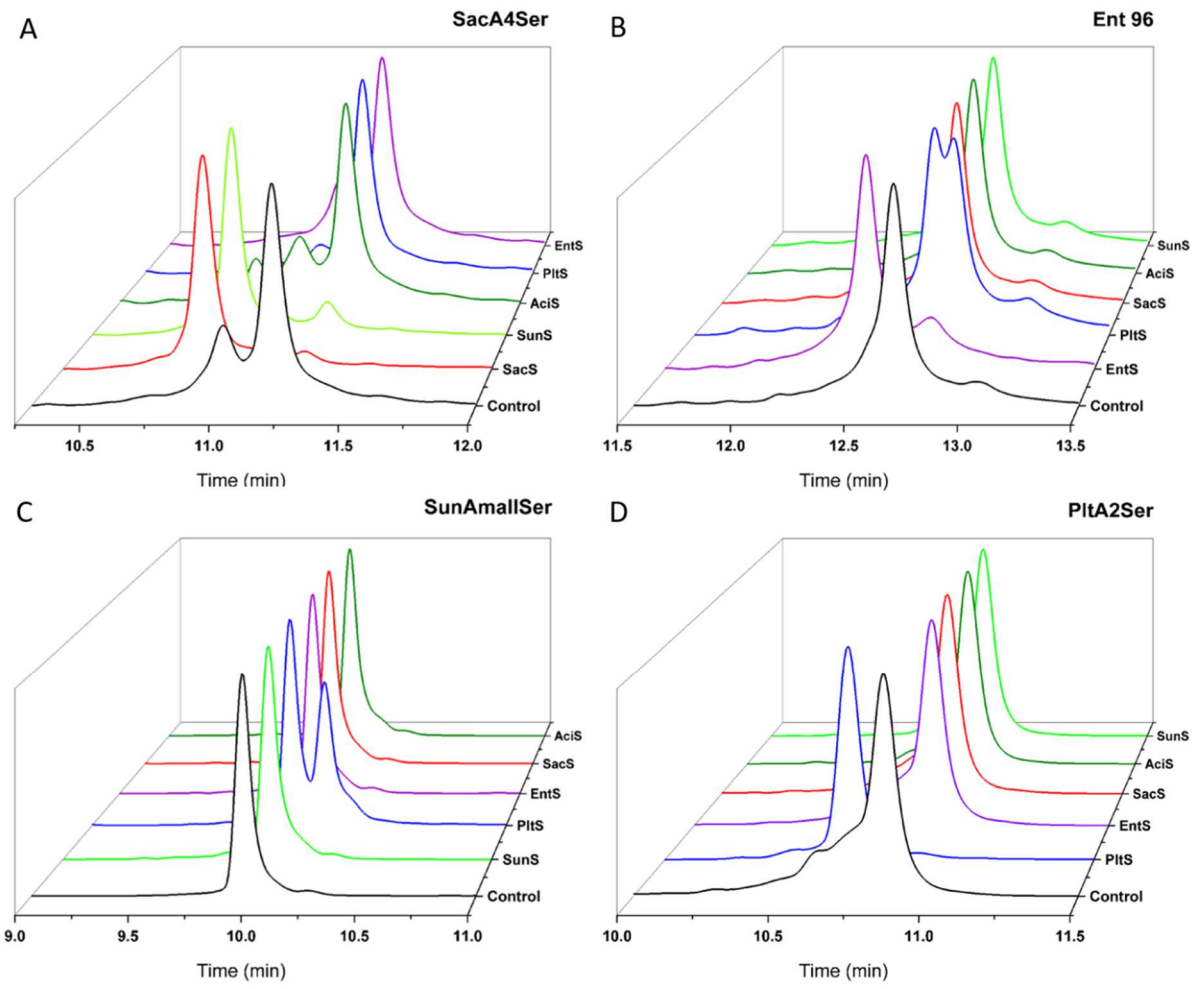


Figure 51 *Chromatograms of the peptides SacA4Ser (A), Ent 96 (B), SunAmallSer (C) and PltA2Ser (D) with different GTs. 1 mM UDP-Glc or UDP-GlcNAc in the case of PltS was used as sugar-donor. The reaction mixture lacking any GT was used as control.*

To investigate this phenomenon further, the SunAm wild-type peptide was incubated with all SunS-like GTs investigated in this study. While, as described earlier, SunS was capable of glycosylating this peptide, AciS and SacS did not cause a shift indicative of a hexose addition (Figure 52B). Surprisingly, SacS caused a different modification with an increase in hydrophobicity. It is so far unclear what this apparent shift represents.

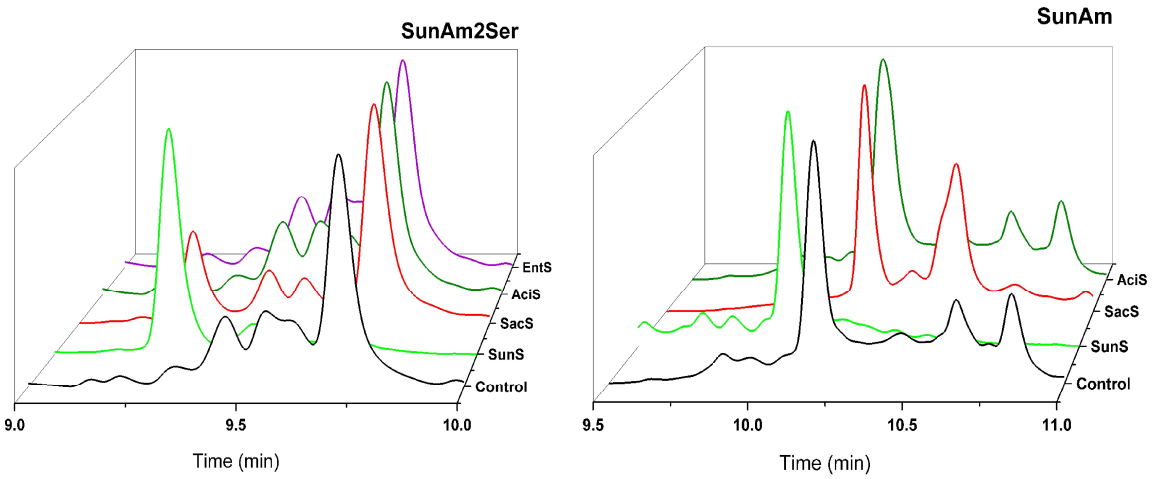


Figure 52 *Chromatogram of the SunAm peptide incubated with UDP-Glc and either SunS, SacS or AciS. Glycosylation could only be observed for SunS. A different, enzyme-dependent reaction occurred with SacS.*

3.7. Reaction kinetics for SacS

Initial attempts to determine the reaction kinetics of SacS were performed using endpoint analysis, detecting the amount of released UDP indirectly using a Malachite green phosphate assay (6.15.2). However, the detection limit of that assay was too high to obtain initial rates, when peptide in the low micromolar concentration range was used.

Thus, an LC-UV assay to determine the ratio of glycosylated peptide to peptide aglycone was developed. However, the detection of peaks at low conversion rates from aliquots of the reaction mixture, necessary to determine accurate initial rate constants, proved to be unreliable for an accurate determination of the enzymatic parameters.
To circumvent this, the reactions was quenched, concentrated *in vacuo*, and redissolved in dilute hydrochloric acid. The use of the full reaction mixture allowed the detection of the desired substrate at a good signal to noise ratio. Unfortunately, the wild type SacA tended to precipitate at basic pH. Thus, the much more soluble 4Ser-mutant was chosen for the assay. Still, I could observe a large drop in reproducibility of the reaction speed at peptide concentrations equal or greater than 100 µM, followed by a further drop in speed at even higher concentrations (Figure 53 A). This is most likely caused by aggregation or even precipitation at concentrations above 100 µM. Thus, for the determination of the kinetic constants only the concentration-range from 12.5 to 100 µM was included (Figure 53 B). For the 4Ser variant a K_m value of 27.5 µM and a k_{cat} of 6.04 s^{-1} was obtained after non-linear fitting in Origin. When I attempted to investigate the reaction kinetics of SacS for *O*-glycosylation, using the SacAallSer peptide under the same conditions as for the *S*-glycosylation mentioned above, I failed to observe any glycosylation. Adjusting the enzyme concentration allowed the determination of the kinetic constants in a in a similar manner to the *S*-glycosylation described above, I obtained a K_m value of 130 µM and a k_{cat} of 0.03 s^{-1}. Due to the higher solubility of the AllSer mutant compared to the peptide harbouring a cysteine, data at higher concentration ranges could be obtained. Nevertheless, even at the highest concentrations the characteristic plateau could not be obtained (Figure 53 C). Unfortunately, even this peptide substrate was insufficiently soluble to obtain values for higher concentrations under these conditions.

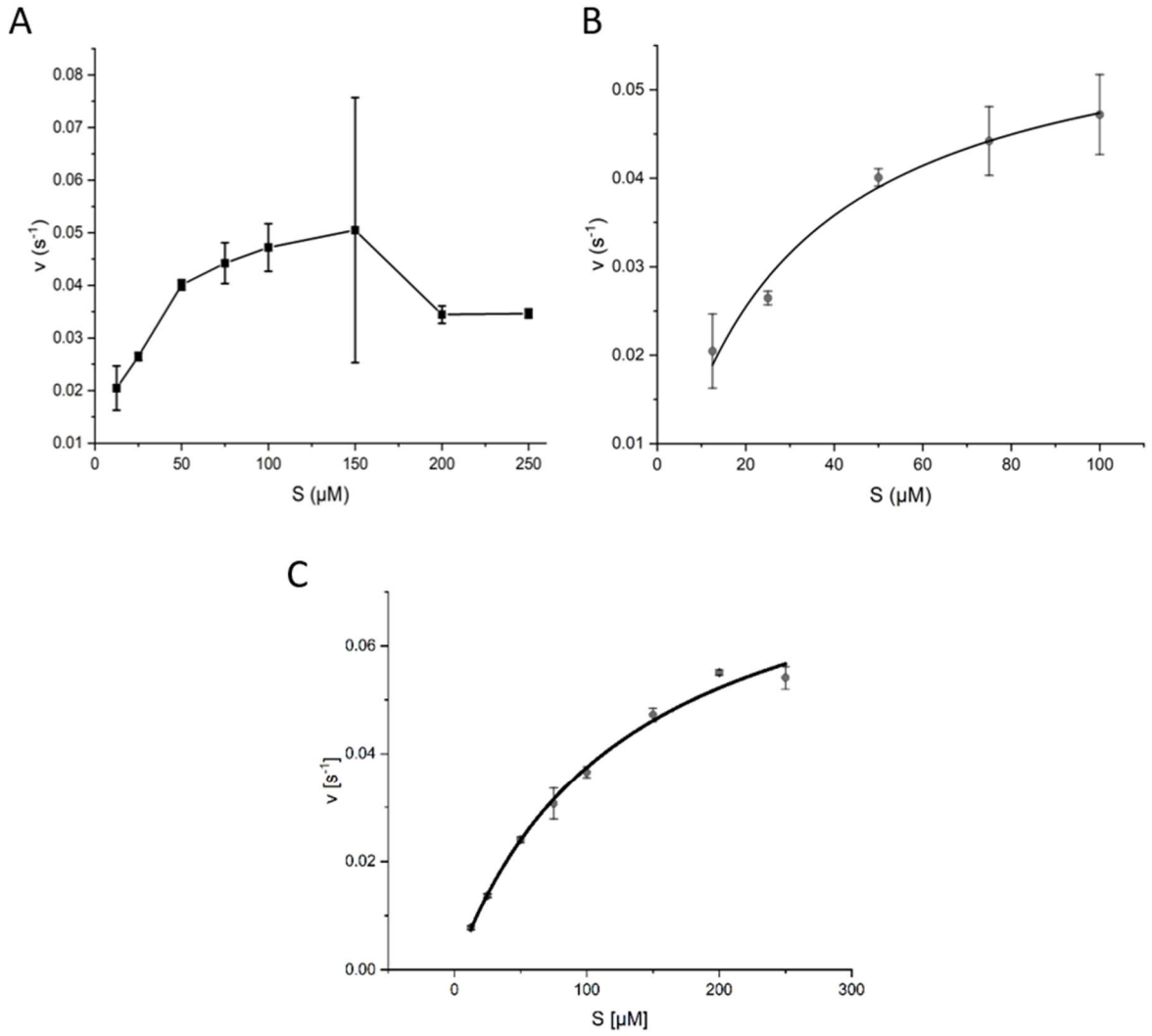


Figure 53 *Reaction rate of SacS for a given concentration of the SacA4Ser peptide (A) and the corresponding non-linear fit (B). Non-linear fit for O-glycosylation of the SacAallSer peptide (C). Error bars represent the standard deviation.*

3.8. An in vivo production method for glycocins

The chemoenzymatic peptide synthesis for full length glycocin usually resulted only in small amounts of peptide, due to the aforementioned challenges in the peptide synthesis and purification process (chapter 3.3). I envisioned, that a broadly applicable, fully biological way to produce said mature glycocins may be of considerable use to analyse their biological function and the *in vivo* occurring glycosylation specificity. Indeed a similar approach has been reported already for GccF[167], Sublancin[175], and Enterocin 96[75] though these either require the use of non-standard expression systems, or suffer from low yields or purification conditions that are not applicable to other glycocins, respectively

The main problem of the synthetic peptides was their tendency to aggregate and form insoluble precipitates. Multiple solubility enhancing tags are known and used in recombinant protein production. I chose the MBP-tag, as it is a known, strong facilitator of solubility[176] that would allow for easy, orthogonal purification as well. A similar approach has already been reported for Ent 96, which could be recombinantly produced as a chitin-binding-domain (CBD) -tagged protein[75]. Contrary to the approach to generate Ent 96 which used an intein-based strategy for tag-removal, which could disrupt the disulphide folding of the glycocins, I envisioned a protease cleavage site remove the tag. The TEV protease was chosen, due to the similarity of the cleavage sequence ENLYFQ|G to the natural leader sequence of most glycocins with its double-glycine motif. It has been shown that the GGT is the only essential enzyme to generate functional glycocins. Therefore, a Duet-vector with two open reading frames was chosen, modified with the MBP-Tag in orf1 for the glycocin and the second orf reserved for the GGT (Figure 54).

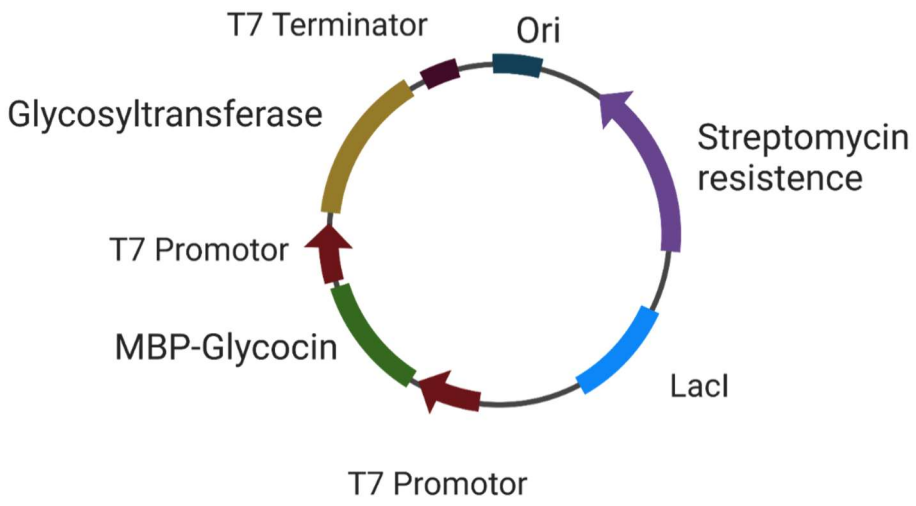


Figure 54 *Schematic representation of the pGlycocin expression vector.*

Initially expression was attempted in TB medium, similar to the expression of the Gtases alone (chapter 3.2). While the SacA and PltA expression could be detected under those conditions, attempts with AciA and SunA resulted in low yields after MBP-trap affinity purification, as well as a protein that was approximately 4 kDa too small, a size that corresponds to the mass of the glycocin. Additionally, the purified product proved resistant to treatment with TEV-protease, leaving the conclusion that either translation terminated before the glycocin was expressed, or the peptide was cleaved off from its MBP-tag prematurely *in cellulo*.

Since glycocins are usually expressed during the late-log to stationary phase of growth in their native hosts[54], it was speculated if expression during a similar phase within *E. coli* may be beneficial to the production of these peptides. Thus, I adapted an autoinduction medium from Studier 2014[177]. Indeed, expression levels showed a marked increase, while an additional band could be detected after MBP-trap, with a size corresponding well to the desired MBP- glycocin-construct (Figure 55).

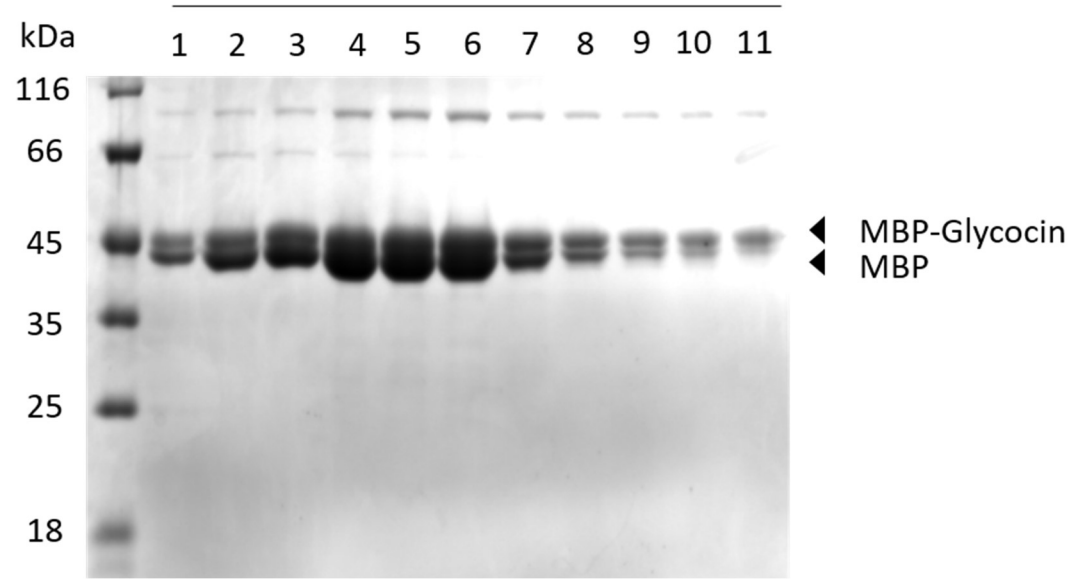


Figure 55 *The eluting fraction of the MBPtrap contains both MBP-Glycocin as well as MBP alone. Exemplified using SDS-Page of the MBP-AciA construct.*

Direct capture of the thus released glycosylated glycocin proved difficult, however, as no peptide of the corresponding size eluted from a size exclusion column and the glycocin could also not be detected in the flow-through of an MBP-trap column designed to retain the cleaved tag.

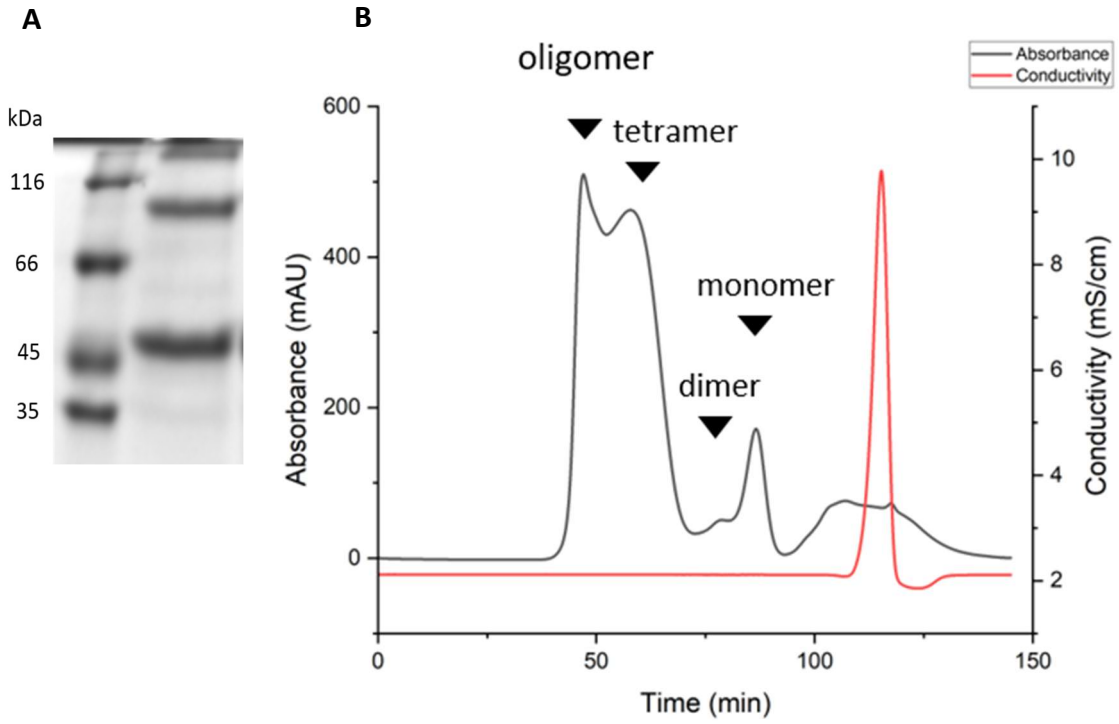


Figure 56 *Size exclusion chromatogram of MBP-SacA. The fusion protein elutes in two bimodal peaks, owing to the resolving power of the column. The first peak constitutes of a mixture of polymer and tetramer, the second of mono- and dimer.*

Given, that SacA expressed in the absence of its cognate GT tended to form soluble oligomers (Figure 56 A), I speculated that oligomerisation may be responsible for the marked absence of eluting SacA-Glc in the previous purification attempts. Indeed, when injecting the MBP-SacA onto a size exclusion column, several peaks, corresponding roughly to the mono-, di-, tetra-, and higher oligomers could be detected (Figure 56 B).

Concentrating the monomeric fraction, followed by treatment with TEV protease allowed for the separation of the glycocin from the tag and the protease via SEC (Figure 57), whereas treating the multimeric fractions in a similar manner did not.

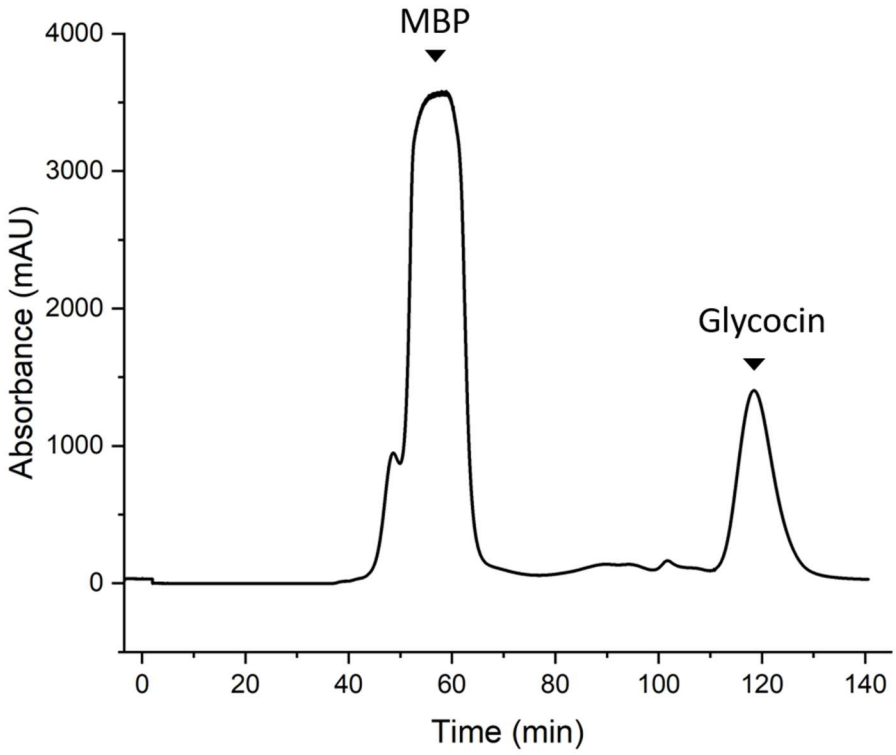


Figure 57 *Representative chromatogram of glycocin isolation after TEV digest. Absorbance was measured at 220 nm.*

Hypothesizing that the oligomeric state may be caused primarily by incorrect disulphide connectivity, due to the lack of the oxidoreductases usually found in glycocin synthesis operons, the multimeric fractions were exposed to a denaturing, highly reducing buffer in an attempt to perform an oxidative refolding. Indeed, after dialyzing the denatured MBP-SacA-Glc back against a volatile buffer allowed for the isolation of further monomeric MBP-SacA-Glc. This protein could likewise be digested by treatment with TEV protease to obtain additional mature glycocin. Using this methodology, it was possible to obtain SunA, as well as PltA. In contrast, the eluted fraction of MBP-AciA was severely contaminated with the MBP-protein alone, without the desired glycocin, making it challenging to be further purified. I noticed that AciA, akin to most Sublancin-like glycocins, contains an amino acid stretch highly enriched in aromatic amino acids, which should increase the hydrophobicity of the construct considerably and thereby allow for purification using hydrophobic interaction chromatography (HIC). Indeed, using this method it was possible to separate the MBP-AciA from the MBP lacking the glycocin and a pure solution of the MBP-TEV-AciA fusion construct could be produced (Figure 58).

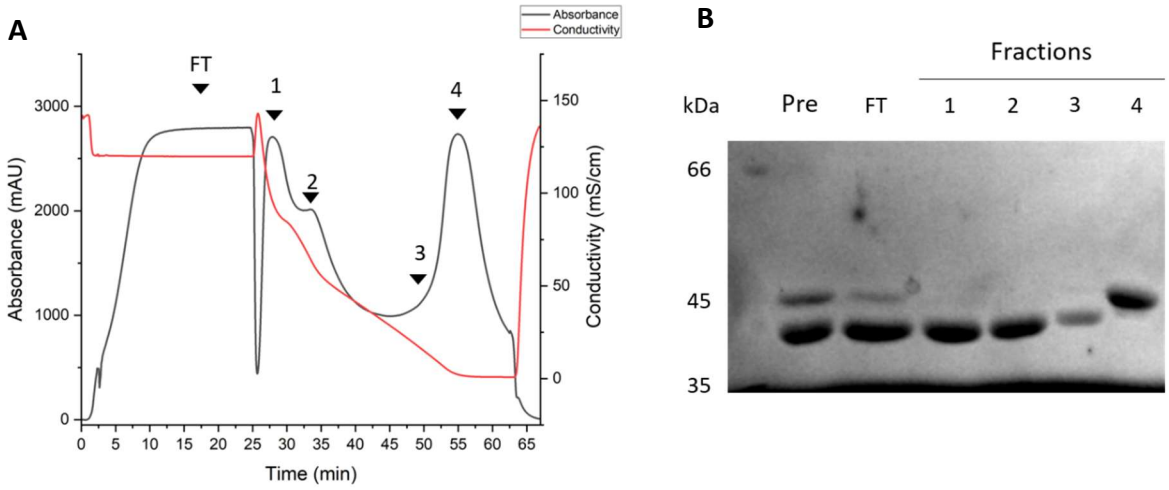


Figure 58 *IEX could be used to separate the MBP from the MBP-AciA construct (A). Indicated fractions and the mixture prior to IEX separation (Pre) were analysed by SDS-PAGE (B).*

However, while this protein could be cleaved by the TEV protease, no AciA appeared to elute from the SEC column. Wondering if this phenomenon may be explained due to the "stickiness" of AciA, interacting with to the column material, organic solvent was added to the elution buffer. Indeed, addition of 30% MeCN caused elution of AciA from the column.

Analysis of the eluting fractions from the SEC by RP-HPLC revealed several undesired side products for all of the purified glycocins, albeit at varying degrees (Figure 59). For SunA the major contaminant could be identified to be the GlcNAcylated glycovariant of the Glycocin.

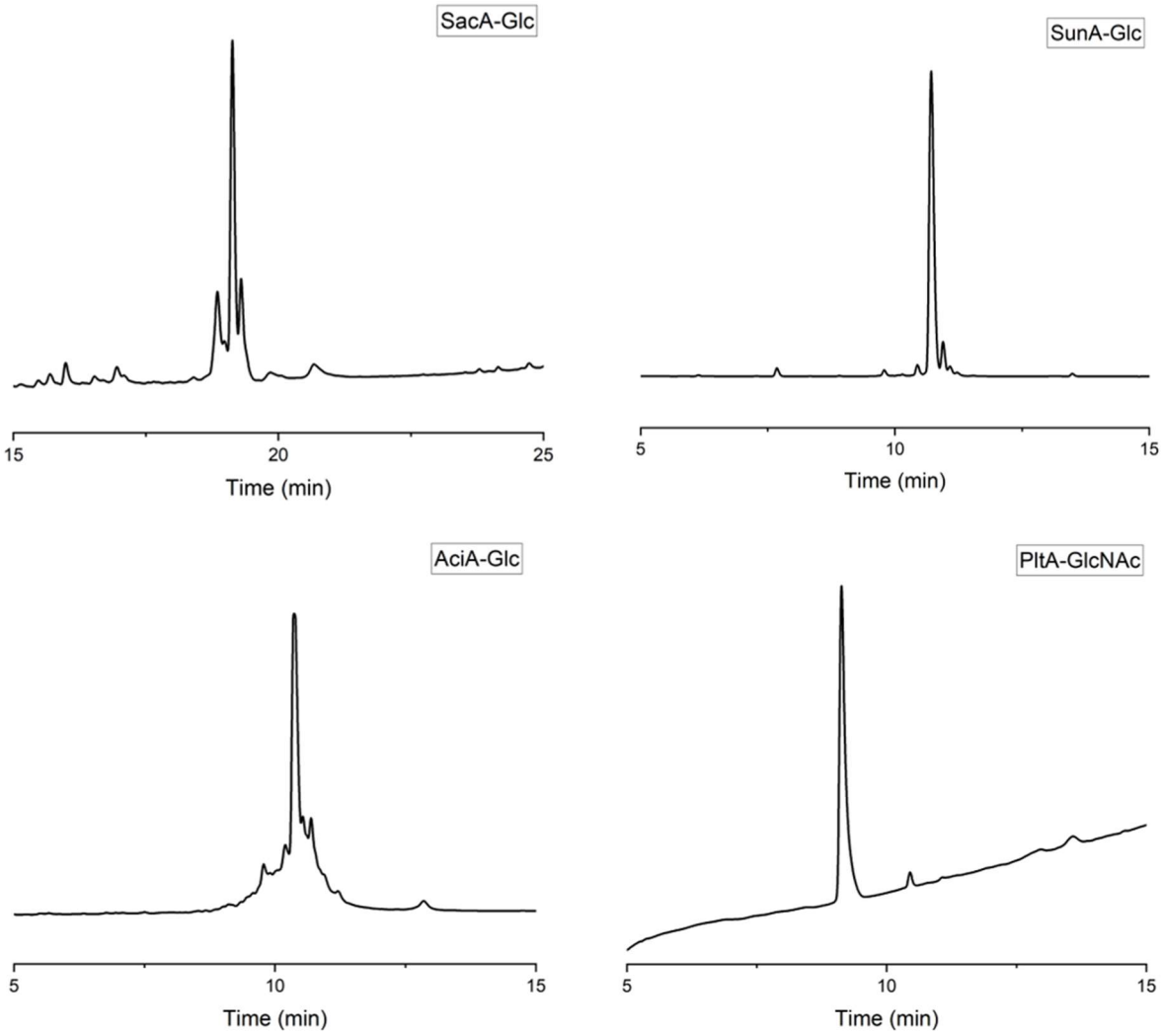


Figure 59 *RP-HPLC chromatograms of glycocins eluting from the size exclusion. Absorbance was measured at 220 nm.*

Thus, to ensure highly pure glycocins a preparative HPLC using a Water-MeCN gradient, containing 0.1% formic acid was chosen for the subsequent, final purification, akin to the purification of the glycosylated peptides that were produced *in vitro* (chapter 3.5). Mass spectrometry was used to verify the purified glycocin (Figure 60). In the case of Sublancin the reported mass corresponding to the glycosylated peptide containing two disulphides[42] could be detected. In the case of AciA and SacA the mass difference to the theoretical mass was +158 amu corresponding to a single hexose and the formation of two disulphide bonds. For PltA, a difference of +201, indicating a HexNAc modification as well as a single disulphide. The final amount of the isolated, pure glycocins are listed in Table 6.

Table 6 *Yields obtained for each glycocin.*

Glycocin	*Average yield per litre of culture*	*Glycosylation*
SunA	3 mg	S-Glc
SacA	5 mg	S-Glc
PltA	0.6 mg	S-GlcNAc
AciA	0.08 mg	S-Glc

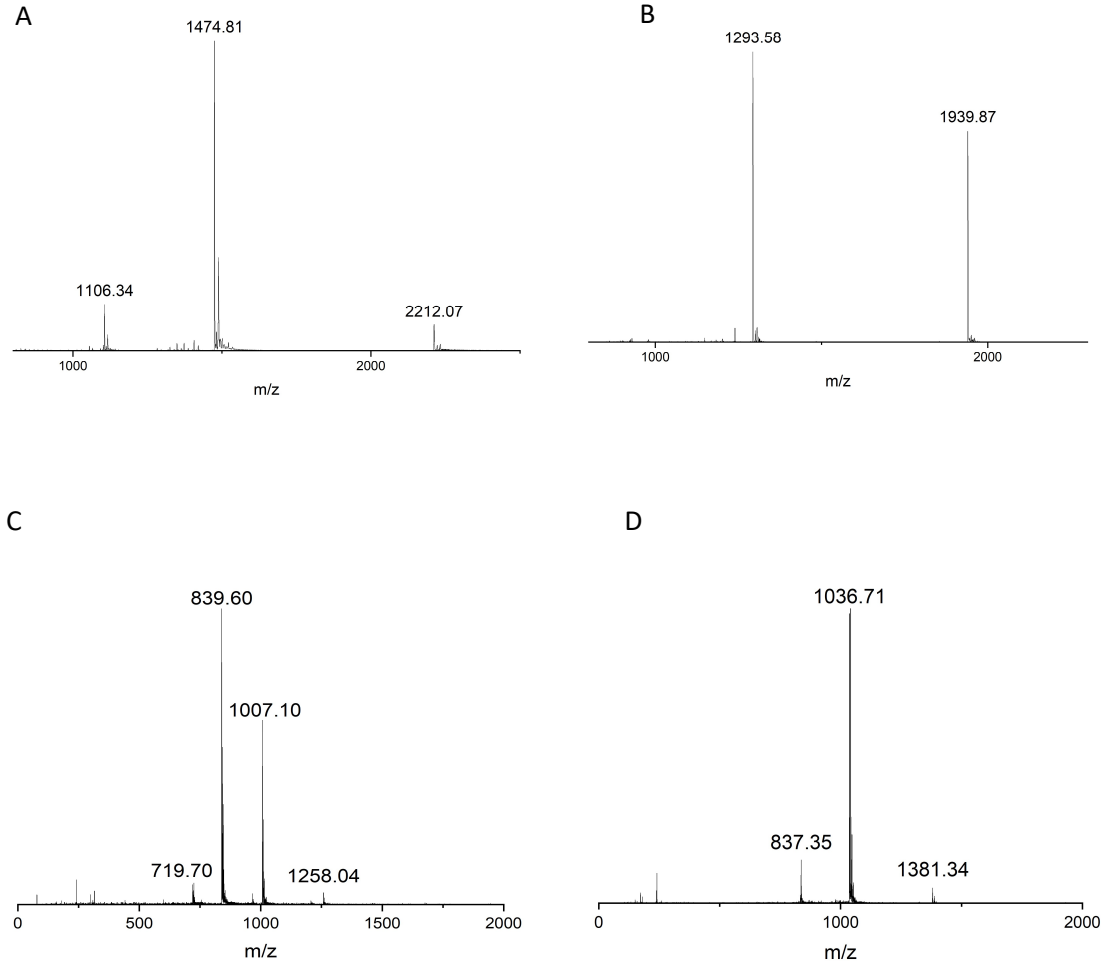


Figure 60 *Mass spectrum of SacA (A), SunA (B), PltA (C) and AciA (D) obtained after HPLC purification.*

3.9. Antimicrobial activity

To investigate the potential antimicrobial activity of the chemoenzymatically synthesized glycocin-peptides, I first performed an agar-well diffusion assay.

First attempts were performed using the full-length Enterocin 96 peptide against *Listeria monocytogenes*, known to be susceptible for this peptide[45], as a proof of concept.

Given that all other glycocins so far could be inhibited by addition of the hexose, with which they are decorated, an additional well was filled with a mixture containing 25 µM Enterocin 96 and 500 mM glucose. Surprisingly, the activity of Enterocin 96 was not inhibited in the presence of glucose (Figure 61). Taken together with the ability of this peptide to inhibit in its reduced form, contrary to all other glycocins described thus far, indicated that the mechanism of action of Enterocin 96 is substantially different from other glycocins.

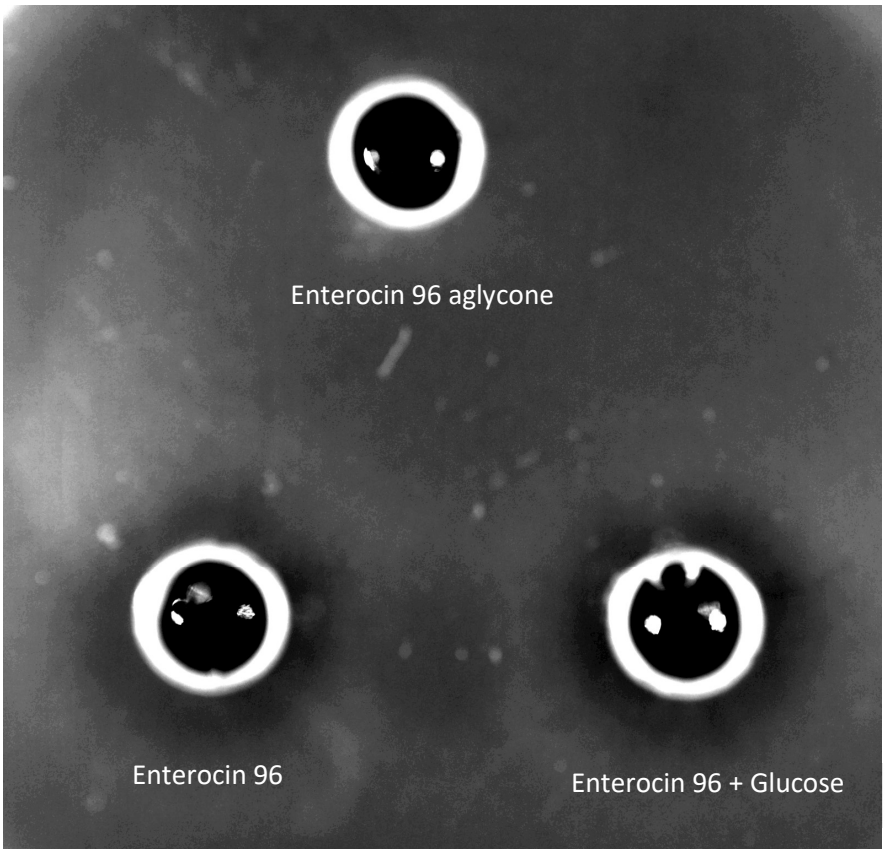


Figure 61 *Agar well diffusion assay of Enterocin 96 in a plate of L. monocytogenes. Zones of inhibition can be seen surrounding wells of glycosylated enterocin 96, regardless of glucose addition.*

A similar Agar-well diffusion assay was performed for the SacA and PltA peptides with the following bacteria:

- *Bacillus subtilis*
- *Bacillus amyloliquefaciens*
- *Alkalihalobacillus plakortidis*
- *Streptomyces Platensis*
- *Laceyella putidus*
- *Listeria monocytogenes*
- *Staphylococcus aureus* subsp. *Rosenbach*
- *Streptococcus suis*
- *Acinetobacter baumanii*
- *Pseudomonas aeruginosa*
- *Salmonella enterica* serovar Typhimurium

PltA showed no discernible activity against any of these strains in this assay, whereas SacA did produce a zone of inhibition for *L. putidus*. This result, however, could not be replicated, possibly due to peptide aging or variations in the media, as *L. putidus* grew substantially different depending on the supplier, the media components were sourced from. This result could also not be replicated using a broth-dilution assay. The amount of added peptide had no discernible impact on the growth of *L. putidus*. Two explanations for the apparent inactivity of SacA, given its high sequence similarity to the reportedly highly active SunA, seem plausible.

Firstly, looking at the gene cluster responsible for SacA production, there are no discernible candidates for an immunity protein that usually is present in such clusters. While a multidrug exporter is located directly downstream of the SacA production cluster, this protein is highly conserved between all domains of life. Without the presence of a protection mechanism against the own peptide, evolution would eventually lead to a modification of said peptide to become inactive.

A different explanation may be the high sugar content found in the medium used for *Thermoactinomycetales*. Given that glycocin activity is inhibited by the hexoses, they are decorated with, growing the target strains in media with moderate to high carbohydrate concentrations such as GYM or Oat flake agar, may explain the lack of activity.

3.10. Crystallization of glycosyltransferases

In order to explain the various differences in activities of the GGTs and characterise their reaction mechanism I crystallised the recombinant GGTs.

Remarkably, while the unstable, His-tagged variant of SacS readily crystalized in a multitude of different conditions (Figure 62).

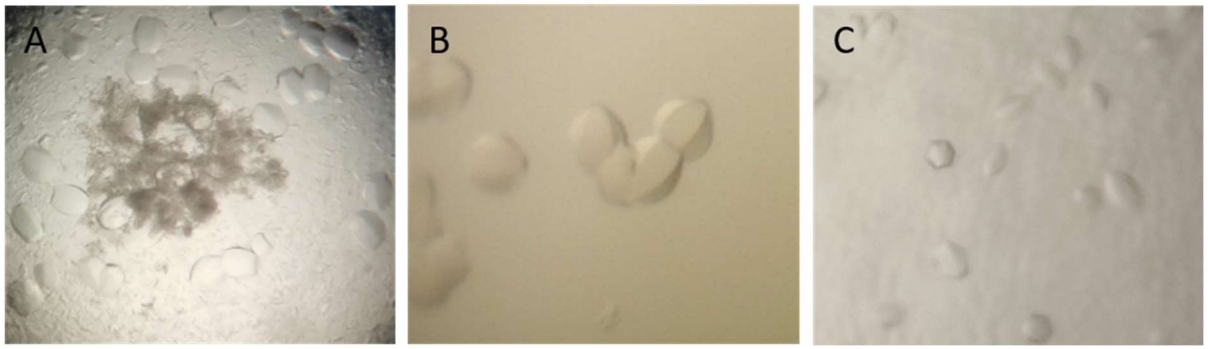


Figure 62 *Crystals of His-SacS grown in 600 mM NaOAc, 100 mM HEPES pH 7.5 (A), 270 mM sodium formiate, 100 mM Tris pH 8.5 (B), 20% PEG 3350, 20% Glycerol (C).*

Unfortunately, crystals obtained from His-SacS did not diffract X-rays to a noticeable amount, even after extensive attempts to optimize the crystals, for example by using different crystallization temperatures, different nucleotides, nucleotide sugars or SacA variants, different protein concentrations, crystal dehydration or change of the crystallization method to microbatch-under-oil. While changing the protein concentration to below 5 mg/ml and lowering the temperature from 293 to 278 K resulted in diffraction, the resolution limit of these crystals was at approx. 15 Å, making them unsuitable for structure determination. Given the destabilization of the protein by the His-tag in the absence of imidazole, His-SacS was desalted into an imidazole-based buffer prior to crystallization under similar conditions. This attempt afforded crystals that diffracted to 7-8 Å resolution. While noticeably better than previous attempts, the obtained resolution is still insufficient for structure determination.

As the His-tag was clearly involved in an interaction detrimental to crystal quality, further attempts were made to crystallize the SacS without the N-terminal His-tag. The much more stable SacSΔHis did not crystallize in the initial screens. Seeding with crystals obtained from His-SacS also did not yield in any crystals.

At 313 K it was possible to obtain crystals of SacSΔHis in the presence of UDP and SunA4Ser-Glc. However, these crystals proved hardly reproducible and also did not diffract beyond 6 Å.

With the advances in protein structure prediction with alphafold2[169], and the published structure of SunS[168], it was possible to predict the domains of SacS. Furthermore, it was shown the crystals of the catalytic domain of SunS diffracts to noticeable higher resolution than the full-length protein. Thus, SacS348, lacking the C-terminal dimerization domain was designed. While the SacS348 mutant did not afford any crystals during initial screens, the SacS348noSt mutant with an elongated C-terminus readily afforded reproducible, diffraction-quality crystals under a condition containing 140 mM TMAO, 60 mM Li_2SO_4, Tris pH 7.5, 14% PEG monomethyl ether (MME) 5000 and 7.5% PEG 3350 (Figure 63 A).

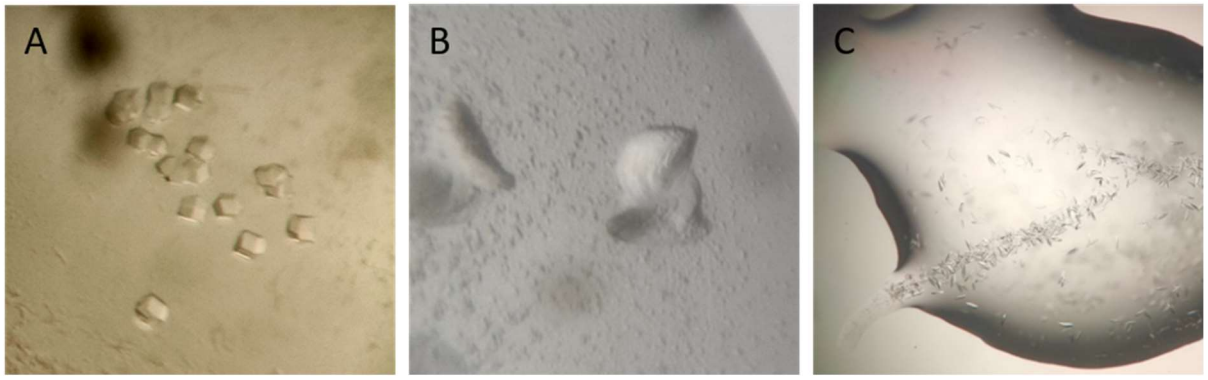


Figure 63 *SacS348noSt crystalized in the presence of 1 mM ADP-Glc in 140 mM TMAO, 60 mM Li_2SO_4, Tris pH 7.5, 14% PEG-MME 5000 and 7.5% PEG 3350 (A), SacS348 crystalized in the presence of 1 mM ADP-Glc in 100 mM Bis-Tris pH 6.5 2 M ammonium sulphate (B), SacS348 streak seeded (C) under the same conditions.*

While it was possible to solve the structure using a dataset with a maximum resolution of 3.5 Å (performed by Christian Roth, MPIKG), a higher quality dataset was required to gain insight into substrate binding and build a better model. Therefore, crystals of SacS348noSt were crushed and used for crystal seeding in a microseed-matrix screen. While seeding with these crystals gave 105 hits for crystallization in new conditions for SacS348noSt, few of these proved reproducible and none improved the diffraction quality. Seeding also resulted in growth of crystals of SacS348 in 4 different conditions, which were reproducible. Crystals of SacS348 obtained this way in a solution of Bis-Tris and ammonium sulphate, in the presence of ADP had the tendency to fuse into each other to form helical crystal bundles (Figure 63 B). Using streak seeding instead of transfer of seed crystals in a suspension of mother liquor, it was possible to obtain single crystals, though the tendency of growing into each other remained (Figure 63 C). With these crystals, one dataset up to a resolution of 2.1 Å could be obtained. Unfortunately, due to the presence of a second crystal lattice, data beyond 2.45 Å had to be excluded.

As the ability to bind to the natural acceptor peptide is largely predicated on the correct formation of the dimer for ThuS[168] and EntS (see 3.11.5), the lack of the dimerization domain of SacS348 did constitute a problem. While the residual glycosylation activity indicates the general possibility of peptide binding, limitations of the acceptor peptide solubility were likely to render a stable ternary complex of the GT impossible at the thusly reduced affinity. Encouraged by the observation that SacS can glycosylate SacA, even with large, N-terminal tags, and the position of the solved EntS-Ecm crystal structure (see chapter 3.11.5), a C-to-N fusion of SacS with SacA4Ser was designed with a simple GSSG-linker in between the protein and its cognate peptide (see chapter 3.2.2). Similar to the SacS experiment previously, this protein crystallized in a multitude of conditions while His-tagged, whereas crystallization without His-tag did not occur under standard crystallization conditions, even when seeded with crystals obtained from SacS348. Given that the SacSA-fusion protein appeared to elute from SEC as a monomer and that a sequential binding mode has been reported for many Leloir-type GTs (chapter 1.3), I hypothesized that the C-terminal acceptor peptide may be unable to enter the catalytic site and interfere with the dimerization instead. Indeed, incubating the SacSA-fusion protein at 50°C for 30 min in the presence of ADP or UDP resulted in a protein solution that readily underwent crystallization in conditions (Figure 64). Whether these crystals are of diffraction quality remains to be determined.

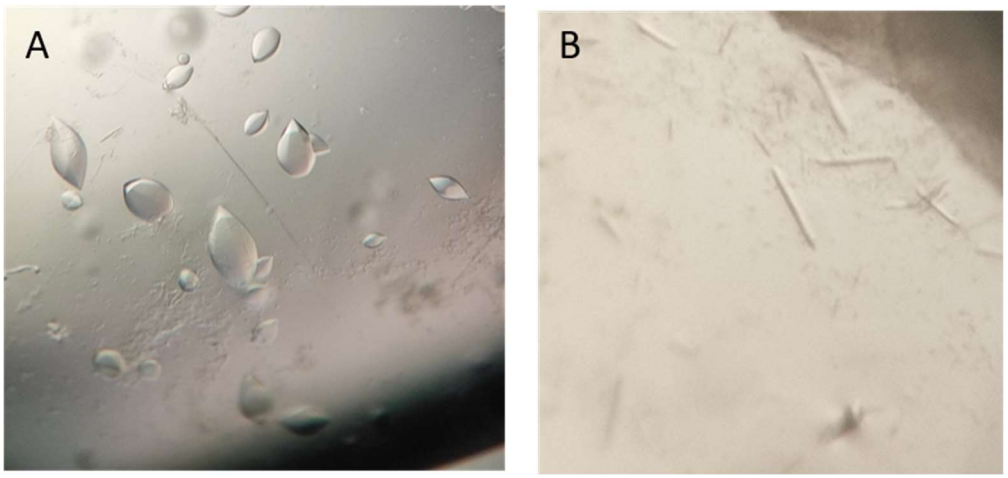


Figure 64 *SacSA-crystals grown in 100 mM Bis-Tris pH 6.5, 200 mM NaCl, 25% PEG 3350 (A) or 100 mM HEPES pH 7.5, 22% sodium polyacrylate 5.1K (B) after incubation at 50°C*

Initial attempts to crystalize EntS in the presence of UDP and $MnCl_2$ were successful in generating crystals in various conditions containing poly(ethylene glycol) (PEG) and salts of organic acids (Figure 65).

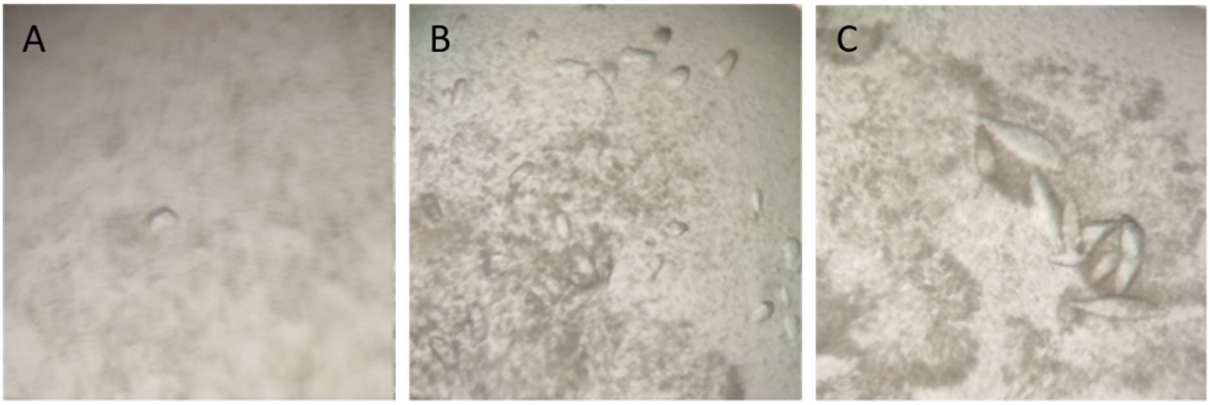


Figure 65 *EntS crystals grown in the presence of 1 mM UDP and 1 mM $MnCl_2$ in 20 % PEG 3350 supplemented with either 200 mM sodium potassium tartrate (A), 200 mM sodium malate (B) or 200 mM sodium acetate (C).*

However, these crystals were usually accompanied by large amounts of non-crystalline precipitate, were difficult to reproduce reliably and similar to SacS, did not diffract x-rays to a noticeable extent. Methods to improve the crystals, such as changing pH, temperature or protein concentration failed to improve the crystal quality. However, it was possible to use these crystals as seedstock for matrix seeding. This led to a number of new conditions, usually still involving PEG and carbonic acid salts. Nevertheless, despite extensive attempts to improve the crystal quality, these crystals failed to diffract to an observable extent as well. Addition of the glycosylated acceptor peptide to the crystallization mixture, noticeably improved the crystal quality, diffracting up to a resolution of 15 Å. Suspecting the main cause for this low diffraction to be insufficient stability of the crystals, cross-linked protein crystals (CLPC) were generated from these according to the protocol of Yan *et al* (2015)[178]. While this procedure did not immediately show any improvement in crystal quality, these cross-linked crystals proved to be excellent seed crystals for another round of matrix seeding. In this round another 49 different conditions have been identified. Most notably crystals were growing in a NaF, PEG condition (Figure 66 A), which grew in an easily reproducible manner and diffracted up to 6 Å.

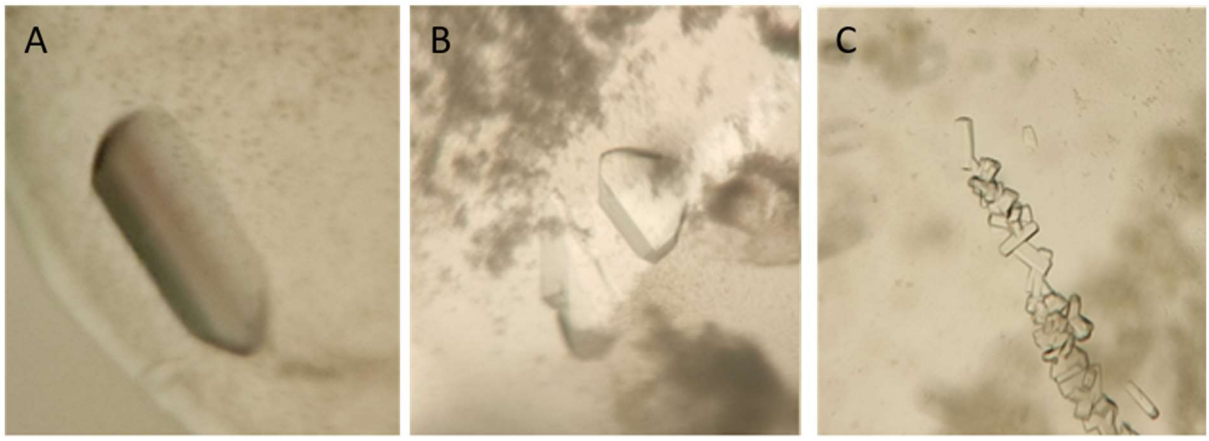


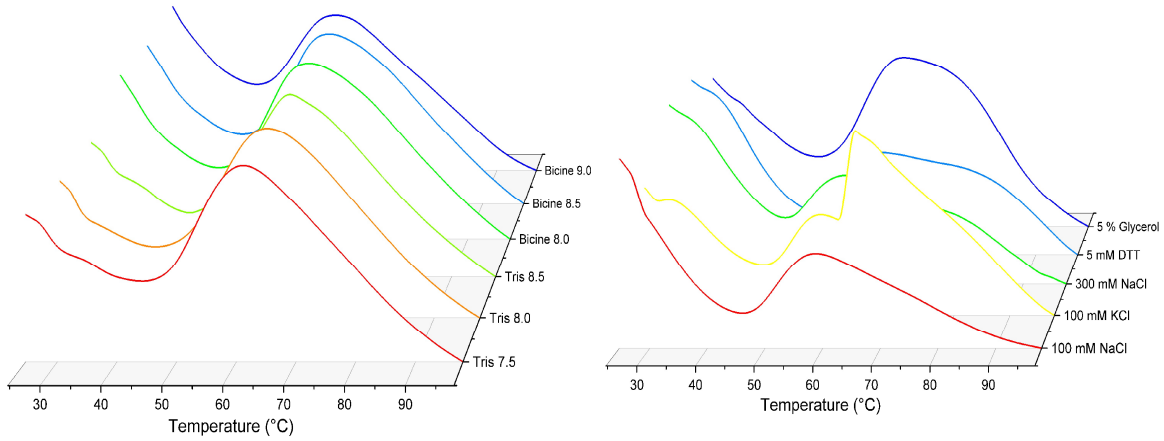
Figure 66 *EntS-Ecm-GlcGlc cocrystal seeded from CLPC in NaF, PEG 3350 (A), Cocrystal seeded from NaF to KF (B) and EntS-EC(Ala)m cocrystal seeded from NaF crystal (C).*

Using streak seeding from these crystals and changing the crystal setup temperature to 4°C prior to incubation at 293 K for crystal growth, allowed for the reliable generation of crystals in conditions containing PEG 3350 and either NaF, KF, or NH_4F (Figure 66 B). Here I was able to observe that crystals grown in the presence of KF regularly diffracted beyond 4 Å, whereas it was much rarer for crystals grown in the presence of other fluoride salts to diffract that well. Thus, I assumed sodium- and ammonium-ions to have a detrimental effect on crystal quality. Indeed, replacing the NaCl in the SEC buffer for KCl, it was possible to generate crystals diffracting beyond 3 Å resolution. Interestingly, EntS in complex with the EC(Ala)m peptide instead of the diglycosylated variant did not grow diffraction quality crystals under the same conditions. As all previous, successful attempts to generate EntS crystals had involved a PEG and organic acid salt, another screen optimizing such conditions with Ec(Ala)m was set up. Indeed, more than 20% of the investigated conditions afforded crystals, most of them at pH 7.5-8.0 with the best reliability when using MMT buffer. Additionally, while crystals did appear under condition including high molecular weight PEGs, crystals grew much more frequent in conditions involving low molecular weight PEGs. Unfortunately, all attempts at cryo-protection, such as the addition of glycerol, ethylene glycol, low Mw PEGs, or cryo-protecting salts, caused the obtained crystals to disintegrate. A method recommended by Hampton research[179], involving the addition of glucose and cryoprotection in a stepwise manner was attempted. While crystals still readily disintegrated when transferred into a solution containing 30% glucose, they remained intact with 15 % glucose as cryoprotectant. In order to reduce the amount of crystal handling, 15 % glucose was added to the crystallization mixture. EntS at 8 mg/ml in a buffer of 20 mM MMT pH 8.0, 150 mM KCl, 1mM EC(Ala)m, mixed in a 1:1 ratio with the optimized mother liquor consisting of 100 mM MMT pH 8.0, 200 mM KSCN, 24% PEG 1500, 30% Glucose, afforded crystals (Figure 66 C) which diffracted up to 2.6 Å.

In order to obtain a complex with unhydrolyzed donor, the EntS(H214A) mutant that should not be hydrolytically active was seeded with the previously diffracting crystals in an extensive crystallization screen. However, this mutant did not crystalize under any of the attempted conditions.

All attempts to generate SunS crystals from SunS dissolved in a Tris-based buffer through use of commercial screens failed. While the crystal structure of SunS and the required conditions for these have been published[168], they appeared irreproducible until the size exclusion buffer was changed to a HEPES-based buffer. With the addition of 1 mM SunAm-Glc and incubation at 277 K it was possible to generate diffraction-quality crystals, diffracting up to 2.6 Å. Unfortunately, while the obtained electron density shows the presence of a ligand at the position the acceptor peptide would be expected, the quality of this density was insufficient to determine its identity as the acceptor peptide.

PltS crystals grown initially from a solution of 20 mM Tris pH 7.5, 150 mM NaCl failed to produce single crystal in any condition. Instead, a frequent occurrence was the presence of spherical crystallites, which afforded a powder-diffraction pattern when introduced into the X-ray beam. Seeding with smashed crystallizes did not improve the diffraction behaviour in any noticeable way. Encouraged by previous results of both SunS and EntS emphasizing the importance of the initial buffer during crystallization, I used differential scanning fluorimetry (DSF) to obtain a more favourable initial buffer for PltS crystallization. While these experiments suffered from a low signal:noise ratio, a few key details could be observed (Figure 67): PltS was much more stable at pH 8.0 than 7.5, Bicine was superior to Tris, Glycerol and the addition of DTT greatly improved the stability, agreeing with previous observations of the oxygen-sensitivity of PltS.



	Melting temperature Tm [°C]
Buffers	
Tris pH 7.5	52.46
Tris pH 8.0	53.19
Tris pH 8.5	52.16
Bicine pH 8.0	54.07
Bicine pH 8.5	53.05
Bicine pH 9.0	49.37
Additives	
100 mM NaCl	51.87
100 mM KCl	56.32
300 mM NaCl	50.40
5 mM DTT	57.01
5% Glycerol	53.78

Figure 67 *Normalized absorbance spectrum of PltS stained with SYPRO Orange at different temperatures (A) and the corresponding melting temperatures based on maxima of the first derivative.*

Changing the PltS buffer to 20 mM Bicine pH 8.0, 100 mM KCl, 2 mM DTT, 5% Glycerol afforded large single crystals of PltS in complex with $MnCl_2$ and UDP-GlcNAc within two days in a Tacsimate-based crystallization condition as well as a succinate, malonate and tartrate-PEG condition (Figure 68).

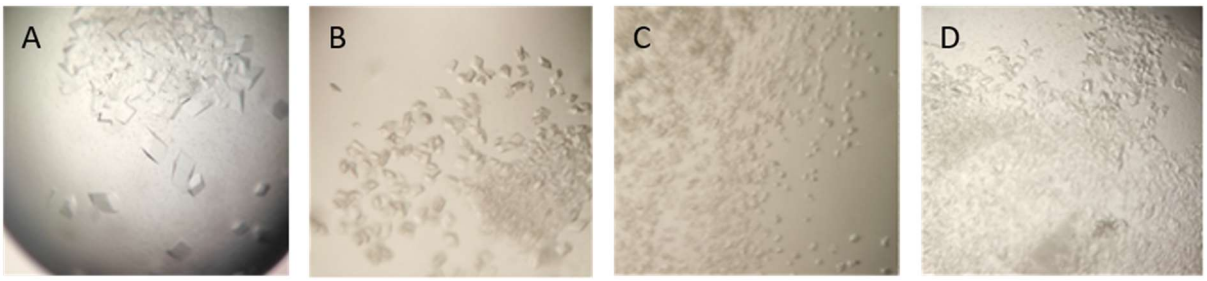


Figure 68 *PltS crystals grown in 35% Tacsimate pH 7.0 (A), 100 mM HEPES pH 7.0, 1.1 M sodium malonate, 0.5 % Jeffamine® ED-2001 (B), 800 mM sodium succinate pH 7.0 (C) or 100 mM Tris pH 8.5, 700 mM sodium potassium tartrate, 0.5% PEG MME 5000 (D).*

Similar to EntS, these crystals did not tolerate standard cryo-protecting approaches. While most crystals obtained from the Tacsimate-condition rapidly disintegrated after transferring them to a 3 M sodium malate solution, some appeared to retain their structural integrity. Fortuitously, these did diffract up to a resolution of 3.0 Å and the structure of PltS could be solved.

During the crystal harvesting it became apparent that the crystallization drops of PltS formed a "skin" on the air-water interface that greatly hampered the isolation of the crystals. I speculated that this skin might either consist of oxidatively damaged protein or be formed by denatured protein due to the hydrophobic air-contact. Therefore, I attempted to crystalize this protein either in the presence of large amounts of reducing agent by supplementing the reservoir solution with 1% mercaptoethanol or by removing the air-water interface by growing the crystals in a microbatch-under-oil. While growing PltS crystals under oil did not help, the addition of mercaptoethanol stopped the skin-formation, even after a week. Growth at lowered temperature further improved this behaviour. Simultaneously, attempts were made to improve the crystal quality by increasing or decreasing the precipitant and protein concentrations. It remains to be seen if these crystals show better order and diffract to higher resolution.

PltS crystals in complex with UDP and PltA2Ser-GlcNAc interestingly did not grow under the same condition. It was possible to observe crystal growth under a tartrate-based condition, which underwent similar optimization processes as PltS + UDP-GlcNAc. Whether these crystals diffract x-rays, needs to be determined.

3.11. Crystal structures of Glycocin Glycosyltransferases

After extensive attempts I could obtain datasets for all of the investigated GTs with the exception of AciS.

Table 7 *X-ray data collection and model refinement statistics*

	SunS	*PltS*	*EntS+ EcmGlc*	*EntS+ EC(Ala)m*	*SacS348*
Wavelength	0.9184	0.9184	0.9184	0.9184	0.9184
Reflections	561099	214022	197656	273765	127063
Unique	29334	38934	29922	40390	29748
Space group	$P3_121$	C121	$P12_11$	$P12_11$	I121
Cell dimensions					
a, b, c	118.673, 119.673, 108.854	174.726, 150.529, 80	75.2772, 94.5595, 92.8093	74.8696, 94.5989, 92.9751	98.568, 52.516, 166.811
α, β, γ	90, 90, 120	90, 107.60, 90	90, 100.605, 90	90, 101.181, 90	90, 106.933, 90
Resolution	40.11-2.55 (2.59-2.55)	57.76-3.00 (3.06-3.00)	34.45-2.85 (2.90-2.85)	52.75-2.57 (2.61-2.57)	47.19-2.45 (2.55-2.45)
Completeness	100 (100)	99.02 (96.13)	100 (99.27)	99.21 (97.29)	97.9 (98.4)
I/σ(I)	7.91 (0.22)	5.09 (0.26)	10.34 (0.41)	10.65 (0.28)	6.1 (2.3)
$CC_{1/2}$	0.991 (0.404)	0.990 (0.287)	0.994 (0.473)	0.999 (0.322)	0.989 (0.890)
R_{merge}	0.212 (4.703)	0.227 (3.125)	0.146 (2.145)	0.121 (3.309)	0.113 (0.414)
Multiplicity	19.13 (19.04)	5.50 (5.67)	6.61 (6.53)	6.78 (7.01)	4.3 (4.4)
Refinement					
Resolution	40.14-2.55	57.76-3.00	34.45-2.85	52.74-2.57	47.19-2.45
R_{work}/R_{free}	0.21/0.27	0.23/0.27	0.22/0.27	0.22/0.28	0.20/0.28
Number of GT monomers	1	2	2	2	2
Average B factors					
GT	79.19	43.51	101.05	96.05	41.78
Peptide	-	-	142.65	147.68	-
Ligands	63.84	35.61	84.14	30.0	66.52
Ions	73.71	101.15	69.1	40.6	59.59
Solvent	60.73	94.29	66.46	74.76	33.23
RMS deviations					
Bonds	0.0122	0.0125	0.0050	0.0055	0.0114
Angles	1.902	1.527	1.314	1.267	1.933
Ramachandran					
Favoured	378 (90%)	762 (88.4%)	838 (90.99%)	830 (90.51%)	627 (95%)
Allowed	29 (6.9%)	68 (7.89%)	52 (5.65%)	59 (6.43%)	24 (3.64%)
Disfavoured	13 (3.1%)	32 (3.71%)	31 (3.37%)	28 (3.05%)	9 (1.36%)

3.11.1. Overall fold

All GGTs show a similar, conserved fold, in agreement with the observed sequence similarity ranging from 36-59% (Figure 69). This fold is composed of a variable N-terminal region, a GT-A type Rossmann-fold, a TPR-like domain and a C-terminal dimerization domain.

Figure 69 *Alignment of the GTs. Degree of conservation is colour-coded with no (white), low (green), medium (yellow) or high (red) degrees of conservation between the investigated GTs. Amino acids involved in nucleotide binding are framed in orange, amino acids predicted to be involved in catalysis are framed in blue.*

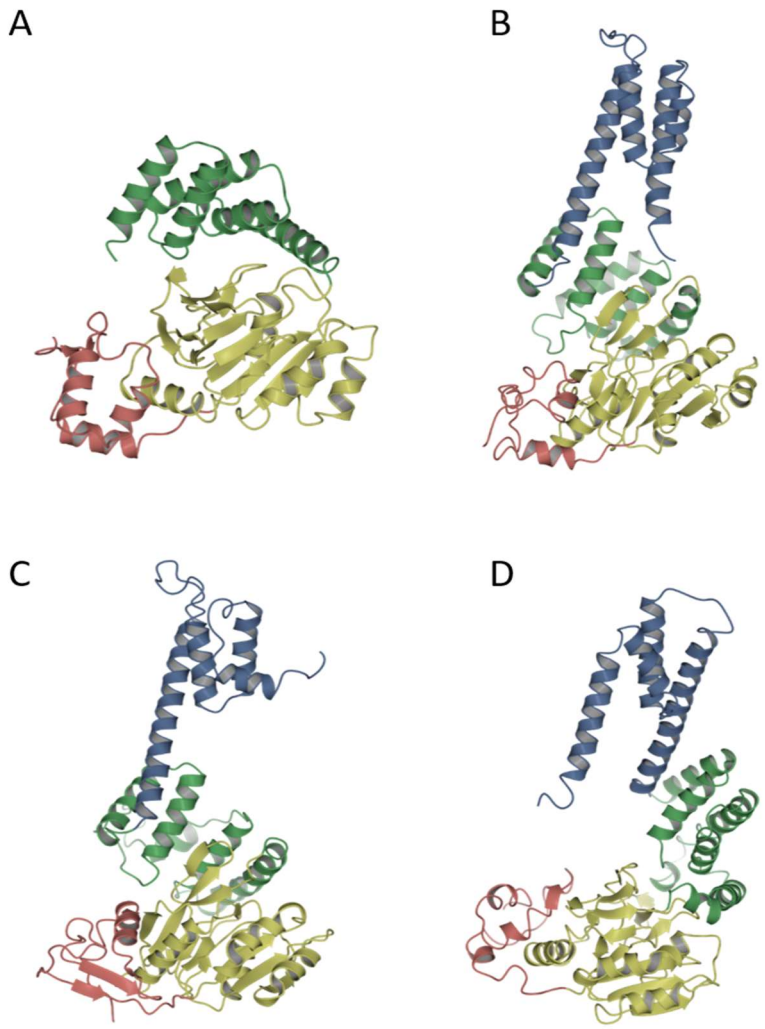


Figure 70 *Domain architecture of the investigated GTs. N-terminal domain in orange, catalytic Rossmann fold in yellow, TPR-like domain in green and helical dimerization domain in blue A) SacS348, B) SunS, C) EntS, D) PltS.*

The N-terminal domain ranges from amino acid 1-59 for SunS, 1-71 for EntS, 1-62 for PltS and 1-58 for SacS and typically consists of three α-Helices and up to three β-sheets. The Rossmann-like fold is characterized by an α/β/α sandwich surrounding a seven membered β-sheet, which ranges from residue 59-223 for SunS, 72-239 for EntS, 63-232 for PltS and 58-227 for SacS. This canonical GT fold is followed by a TPR like domain, consisting of 5 α-helices and ranging from 224-322 for SunS, 244-345 for EntS, 235-341 for PltS and 220-327 for SacS. The Sugar nucleotide binding site located at the interface of the Rossman-like fold and the TPR-like domain, with a small cleft, constituting the (putative) peptide binding site in between. This cleft extends towards Helix α1 of the dimerization domain, which consists of 4 α-helices and ranges from amino acid 321-422 for SunS, 346-446 for EntS, and 342-442 for PltS. While the electron density did not allow unambiguous placement of amino acids 443-456 of PltS, though it appears likely that these are part of the dimerization domain as well.

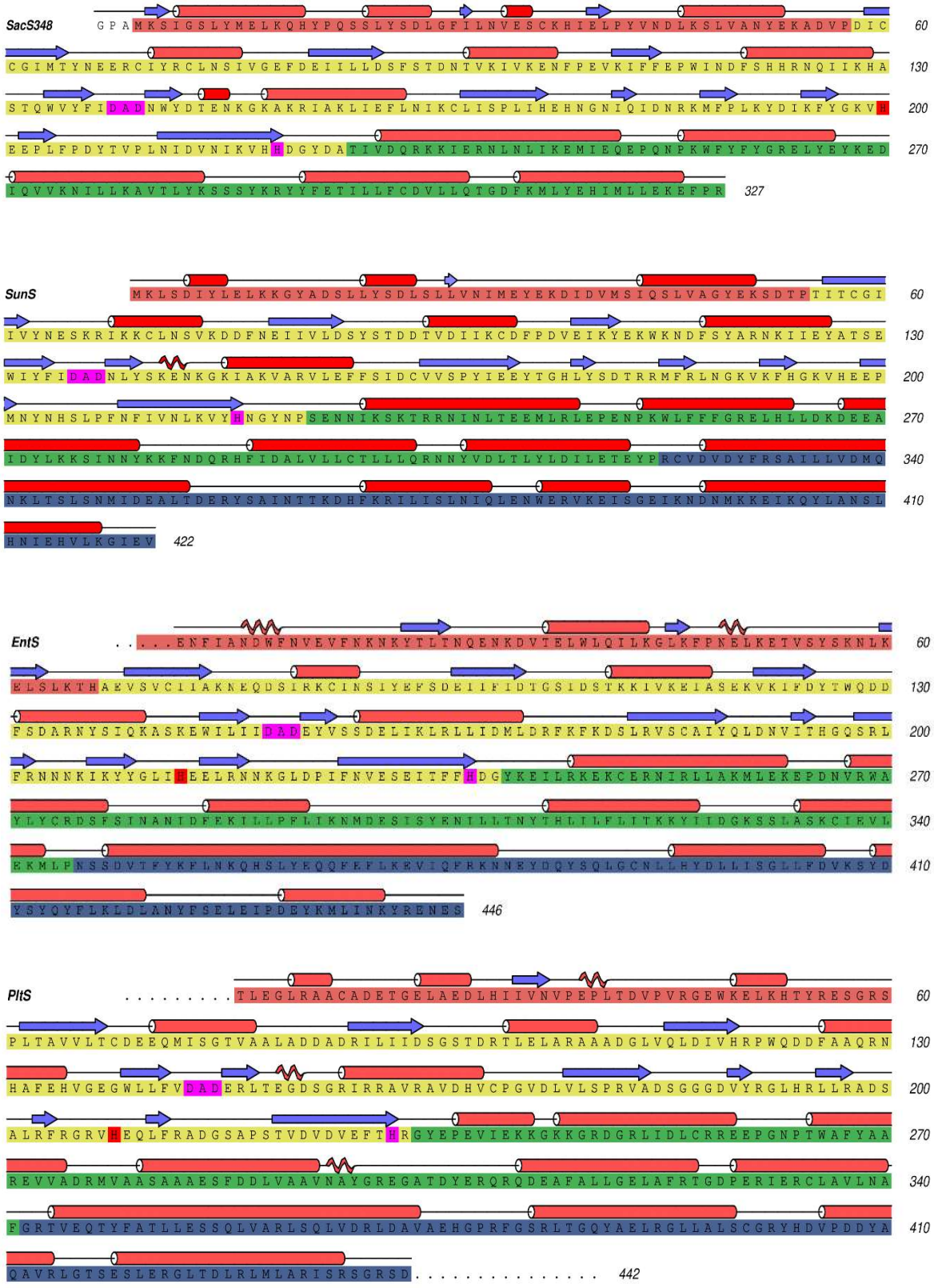


Figure 71 *Secondary structure of GGTs. N-terminal domain is marked in red, GT-domain in yellow, TPR-like domain in green and dimerization domain in blue. β-strands are marked with arrows, α-helices with cylinders and 3_{10}-helices with stylized helices. The DXD motif and metal-coordinating His are marked in purple. The putative catalytic base is marked in dark red.*

As stated previously, the overall fold of the GTs appears highly conserved, with the Rossmann-like domain and TPR-like domain superimposing with an R.m.s.d. of 0.32-0.35 Å (Figure 72 A). Similarly, the TPR and the dimerization domains can be superimposed (Figure 72 B) with an R.m.s.d. of 2.27-2.83 Å. Only EntS shows a distinctly different orientation of helices α4 and α5 in its dimerization domain.

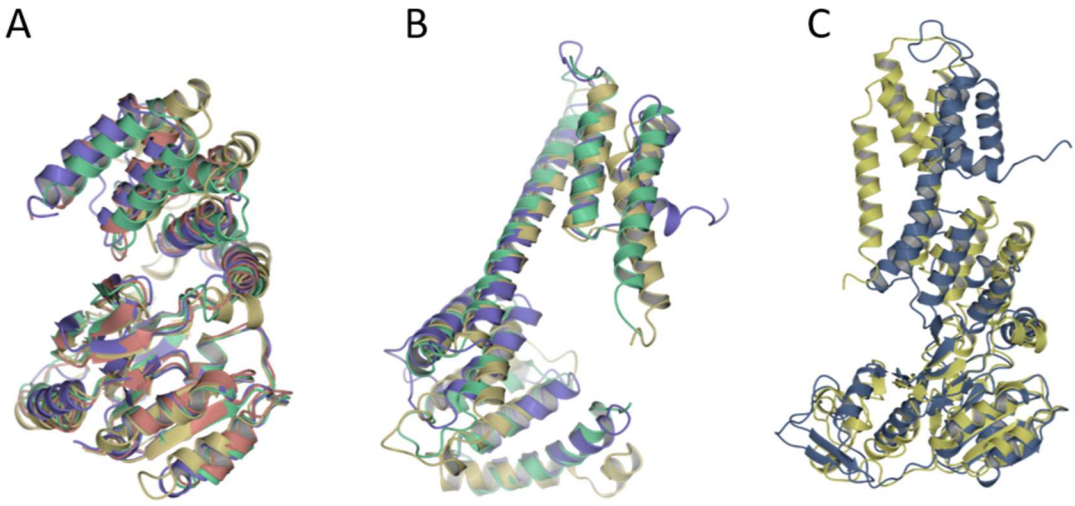


Figure 72 *A) superimposed image of the catalytic and TPR-like domains. Average RMSD is 0.32-0.35. B) Superimposed image of the TPR-like and dimerization domains. RMSD is 2.27 and 2.83 and 2.27 for EntS, PltS and SunS respectively. EntS in blue, SunS in green, PltS in yellow, SacS in Red. C) Overlay of the EntS (blue) and PltS (yellow) enzymes. While the GT-domain and TPR-like domain align, a pronounced difference in orientation of the dimerization domain can be seen.*

However, if the entirety of the protein is aligned against each other, a major difference of PltS is revealed: Due to its slightly longer TPR-like domain with an additional helix, the dimerization domain of PltS is twisted by several degrees with regard to the substrate binding domains, when compared to EntS, or SunS. Given the dimerization domain is essential for substrate binding in all known GGTs, this "twist", may have a profound effect on the activity of PltS.
With the crystal structures of the monomers in hand, a model of the functional dimer was be created from using PISA[180]. The resulting Van-der-Waals surface of the dimer revealed a large pocket, formed by two dimerization domains and one TPR-like domain, which leads into either a narrow groove in the case of PltS or a channel in the case of the other GGTs, both of which lead towards the GT-A type fold. In the cocrystal of EntS with the Ecm peptide this pocket appears to bind the N-terminal helical region of the peptide, with the disordered loop binding in the channel. A second pocket on the opposite side of the GT-A type fold is the sugar-nucleotide binding site. The L7 loop creates a lid over the diphosphate and the resulting groove connects the donor substrate binding site towards the inner

opening of the peptide channel (Figure 73). The modes of sugar-donor and –acceptor binding are described in the following sections (chapters 3.11.4 & 3.11.5).

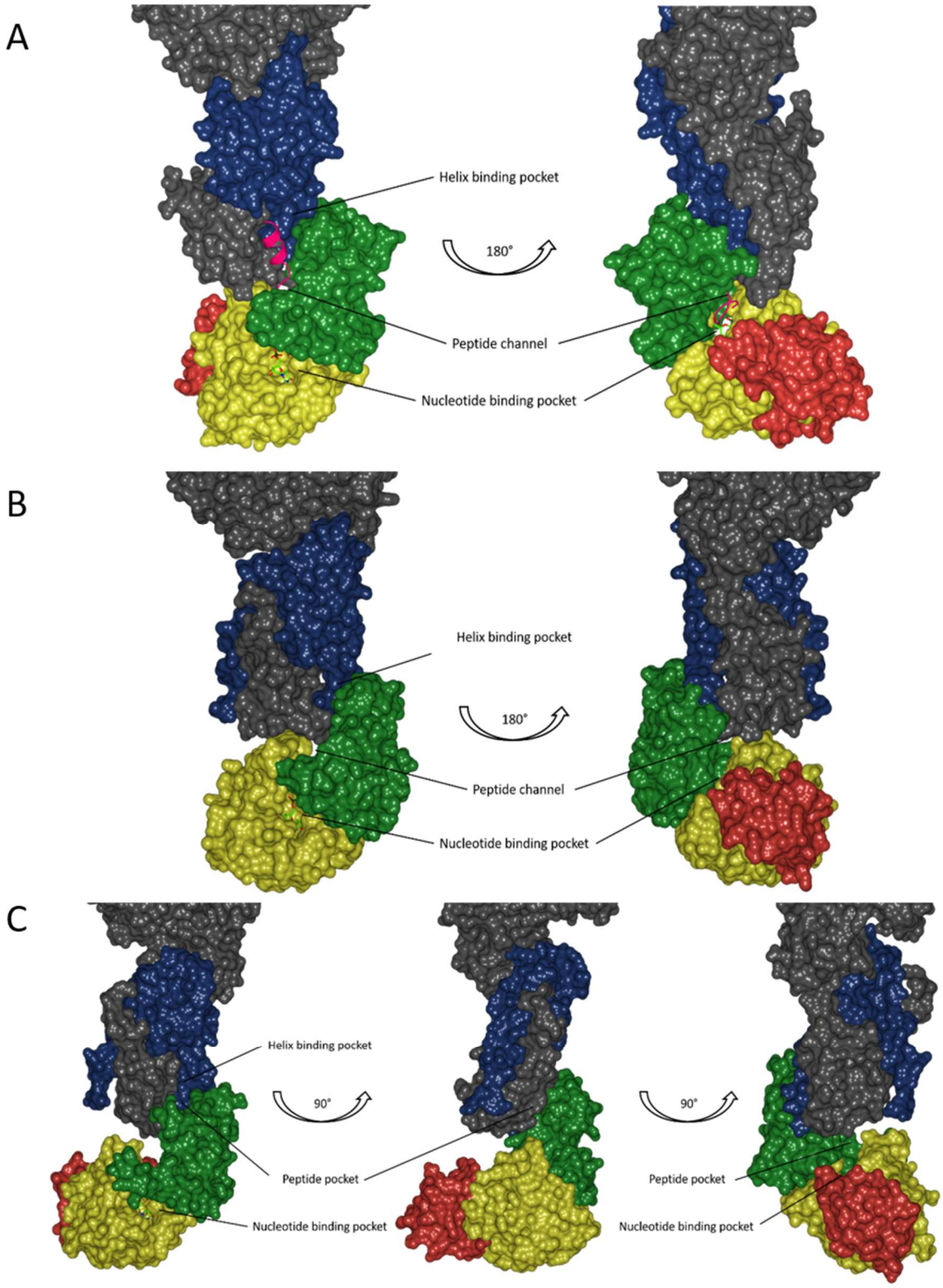


Figure 73 *Surface model of the dimers of EntS (A), SunS (B) or PltS (C) with the Nucleotide shown as cylinders and the acceptor peptide as cartoon (pink). Domains of the enzymes are coloured with the N-terminal domain in red, the GT-A fold domain in yellow, the TPR-like domain in green and the dimerization domain in blue. The dimerization partner of the respective enzymes is shown in grey.*

3.11.2. The GT-A fold domain

As mentioned above, the catalytic domain follows the canonical GT-A type fold, composed of a 7-stranded, twisted β-sheet of a 3214657 topology, where all strands except for β6 are parallel. Strand β4 and β4' enclose a loop, where the signature metal-binding DXD motif (here DAD) is located, with the β4' strand forming a second β-sheet with β7, β5 and a strand that is located within the variable region of the GT-A fold. The β7 strand appears to cross over the β5 strand to give this sheet the topology 5'574'. The last amino acid of β7 is usually a His, which helps coordinate the metal ion. The catalytic His residue is located in the loop of a β-turn region (Figure 74 A). While the strand β5 in EntS is contiguous, β7 displays a bulge that separates the strand in two (Figure 74 B). A similar bulge is found for strand β5 in PltS, SacS or SunS (Figure 74 C).

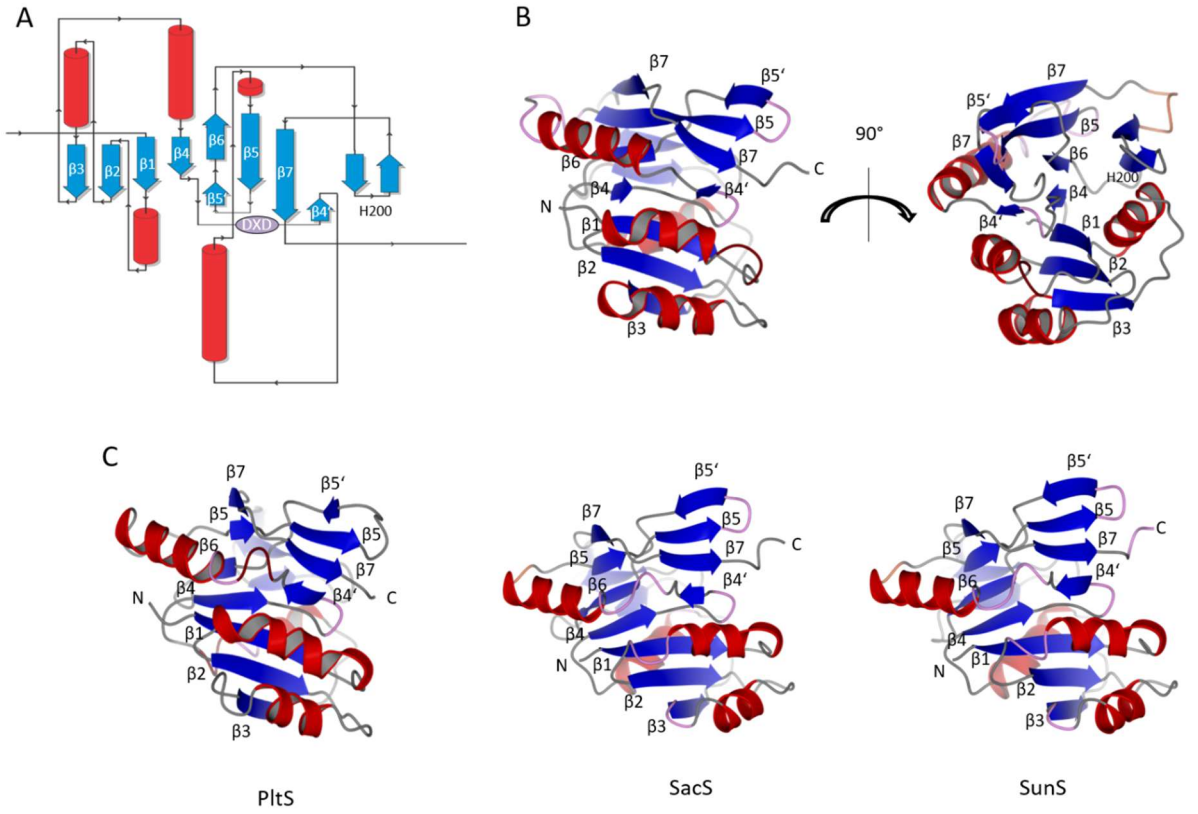


Figure 74 *(A) Common topology of GGT Rossmann-fold domains with EntS as example. (B) Organization of helices and strand in space of EntS. (C) In the structures of PltS, SacS and SunS, strand β5 and β7 show a distinct bulge, separating these strands into two.*

3.11.3. Dimerization domain

The dimerization domains consist of 5 α-helices. This domain of each monomer is rotated against the other dimerization domain by 180°, resulting in each helix being mirrored by its dimerization partner in an anti-parallel manner (Figure 75). These helices form a six helix-bundle with Helix 2 and 2* stacked almost on top of each other.

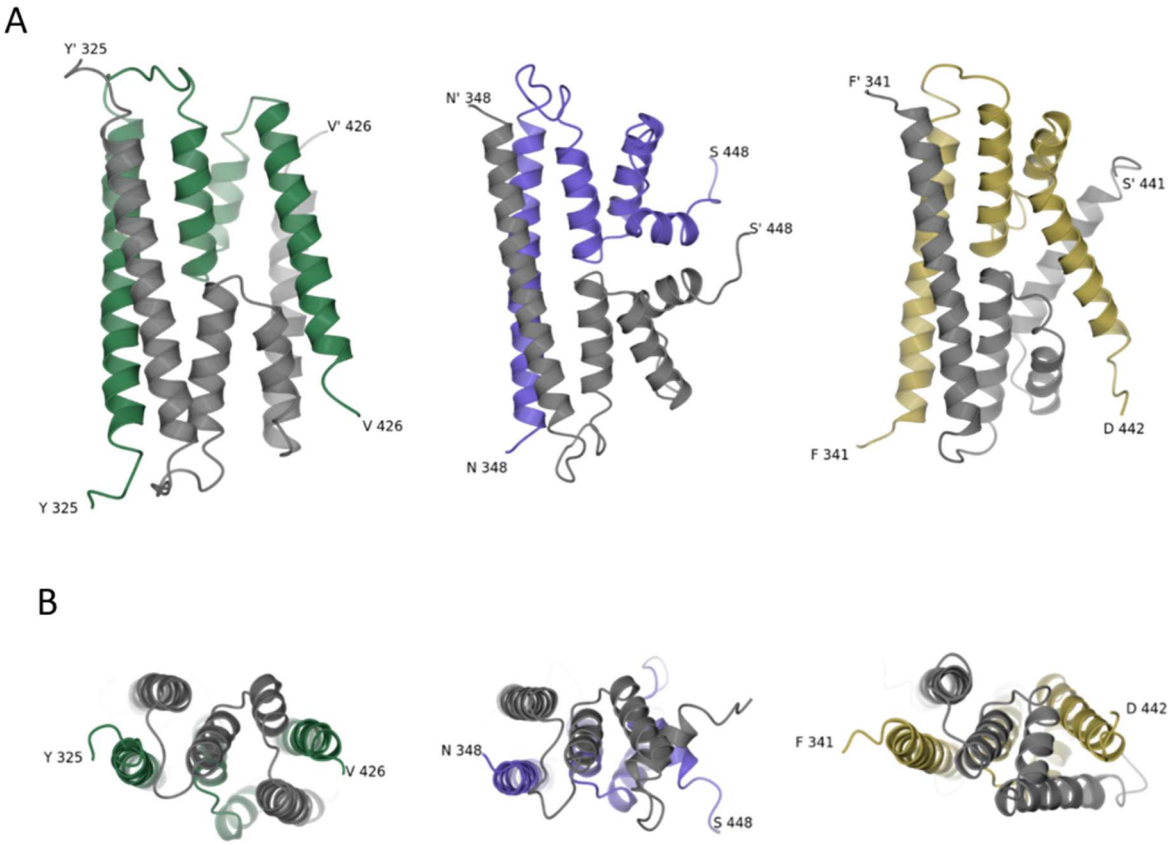


Figure 75 *Dimerization domain of SunS (green), EntS (blue) and PltS (yellow). Dimerization partners are shown in grey (A) and a top-down view of the helices (B).*

While for both SunS and PltS helix 5 is located pointing back at the GT-domain, for EntS there is a noticeable kink between helix 4 and 5, the latter pointing outwards at an almost 90° angle to interact with its dimerization partner. While a similar kink is also present in PltS, it is far less pronounced and the helix 5 interacts only in a very limited manner with helix 5*.

3.11.4. Donor binding

As mentioned previously, the affinity and specificity of the investigated GGTs for the nucleotides and nucleotide-sugars varies substantially (Table 4). A detailed view of the nucleotide binding sites reveals a few, key differences, described in detail below.

3.11.4.1. Nucleotide

Binding of the pyrophosphate and ribose moiety of the nucleotides occurs highly similar in all investigated GTs (Figure 76). The Mn^{2+} ion is coordinated by the C-terminal Asp of the DXD motif, and a His residue located in β7, with two further coordination sites serving as attachment points for the pyrophosphate. Further interactions of the phosphate moiety include a salt bridge to nearby Arg and Lys residues, with the latter predominantly interacting with the β-phosphate. A further interaction of this phosphate group is a hydrogen bond to a nearby Tyr residue. Interactions with the ribose appear equally conserved, with O3 forming a hydrogen bond with the backbone amide of the DXD-motif Ala as well as the backbone carbonyl of the Ile, Leu or Met residue for EntS/SunS, PltS or SacS, respectively. O2 is in turn bound by interactions with a nearby Glu residue, as well as a backbone amide that further interacts with the nucleobase. Here the most apparent differences in binding can be found. The pyrimidine or purine ring in SunS or SacS, respectively, is sandwiched between Y62 and W111, indicating π-π stacking interactions. Contrary, EntS harbours a lysine instead of Tyr with a hydrogen bridge forming between the positively charged Lys and the Uracil carbonyl. Interestingly, PltS has a substitution in the same position, with a cysteine, likely engaging in an S-π interaction (Figure 76 C). A conserved hydroxyl group of either a Ser (for SunS and PltS) or a Thr (for EntS) interacts with the carbonyl or amide of Uracil, which may help confer the necessary nucleotide specificity. Interestingly this Ser is also present in the same location for SacA, possibly forming a hydrogen bridge with an ADP amine, with a distance ranging from 3.3-4.2 Å, depending on the molecule in the asymmetric unit.

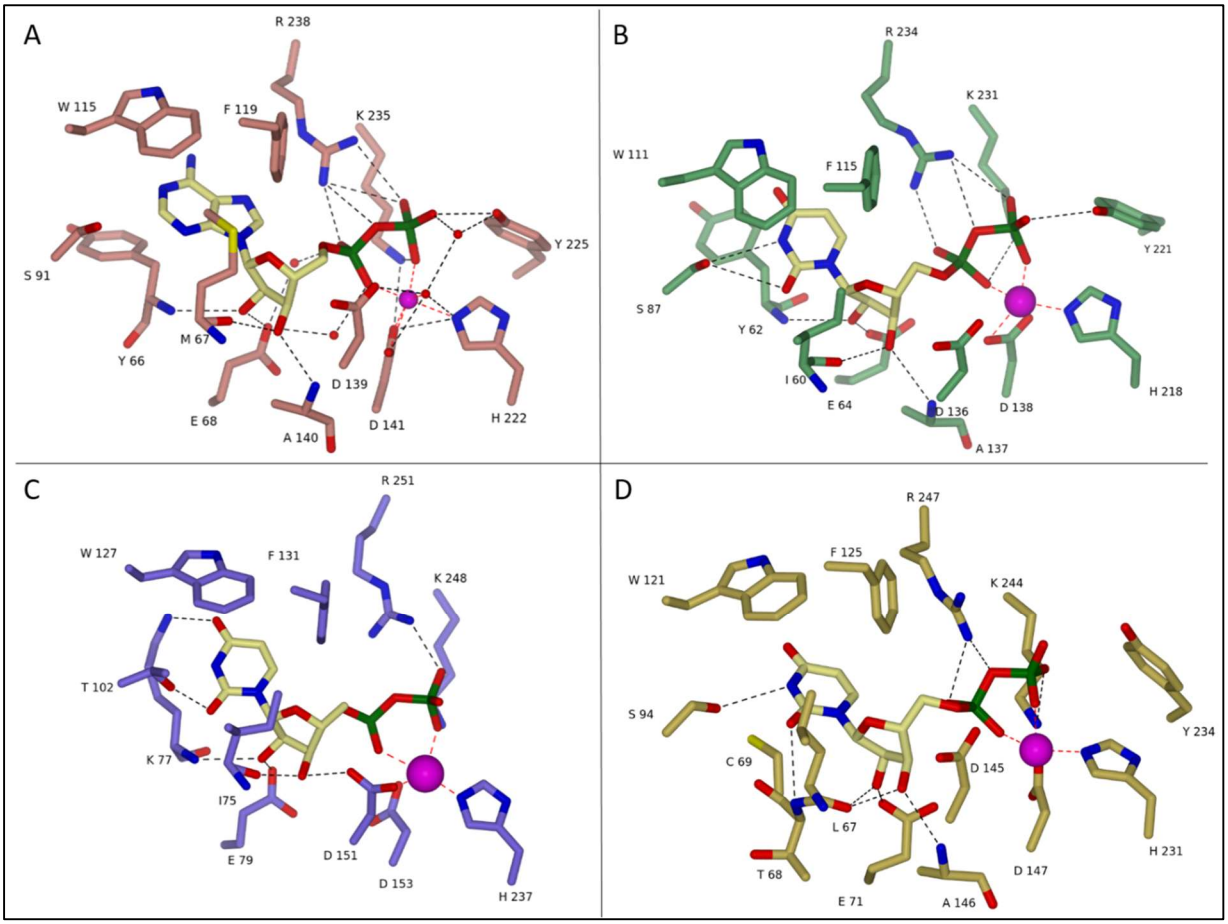


Figure 76 *Nucleotide binding of GGTs. A) SacS348 B) SunS, C) EntS and D) PltS. Nucleotides and amino acids within a 4 Å radius are shown as cylinders. Hydrogen bonds (black) and metal coordination is shown as dashed lines.*

Given the unusually high affinity of SacS (chapter 3.4) to ADP-Glucose, compared to the other investigated GTs, the nucleotide binding pocket was examined more closely, for differences to SunS, which shows a high sequence similarity to SacS. As mentioned above, the binding pocket of SacS is highly similar to that of SunS with most amino acids within 4 Å of the bound nucleotide being virtually identical (Figure 76). A notable difference is the sulphur-atom of M67. As mentioned above, the other GTs rely on hydrophobic interactions via leucine or isoleucine in this position instead. This methionine may undergo S-π interactions, further increasing the binding-energies of the nucleotide. As this Leu-to-Met substitution appears to be the only discernible difference between SacS and SunS, it is likely to be the cause of the former's ADP preference. Considering that the preference is much stronger for the sugar donor ADP-Glc, it is likely that the basis for the nucleotide preference is caused by subtle differences in the positioning and interaction of the sugar residue instead.

3.11.4.2. Carbohydrate-donor – Glucose

I was able to obtain a cocrystal structure of EntS in complex with UDP-Glc and the Ser-to-Ala mutant of its acceptor peptide. The electron density of the glucose moiety is poorly resolved. This is most likely caused by partial hydrolysis of the UDP-Glc molecule with water as the aberrant glycosyl-acceptor, which is commonly reported for Leloir-GTs (chapter 1.3). Nevertheless, while the exact conformation of the glucose is not entirely clear, the difference electron density allows a likely placement of the glucose with reasonable confidence. Given the diffuse electron density, two conformations of the glucose are possible (Figure 77).

In the first possible conformation the glucose is recognized primarily via hydrogen bonds with E215 forming a hydrogen bond to the vicinal diol of O3 and O4. O3 is further held in place by a hydrogen bridge with R135. O2 forms a hydrogen bridge with D151, while C6 may be involved in a hydrophobic interaction with I213. In this conformation F131 may promote the sugar-binding, being located at the hydrophobic face of the sugar.

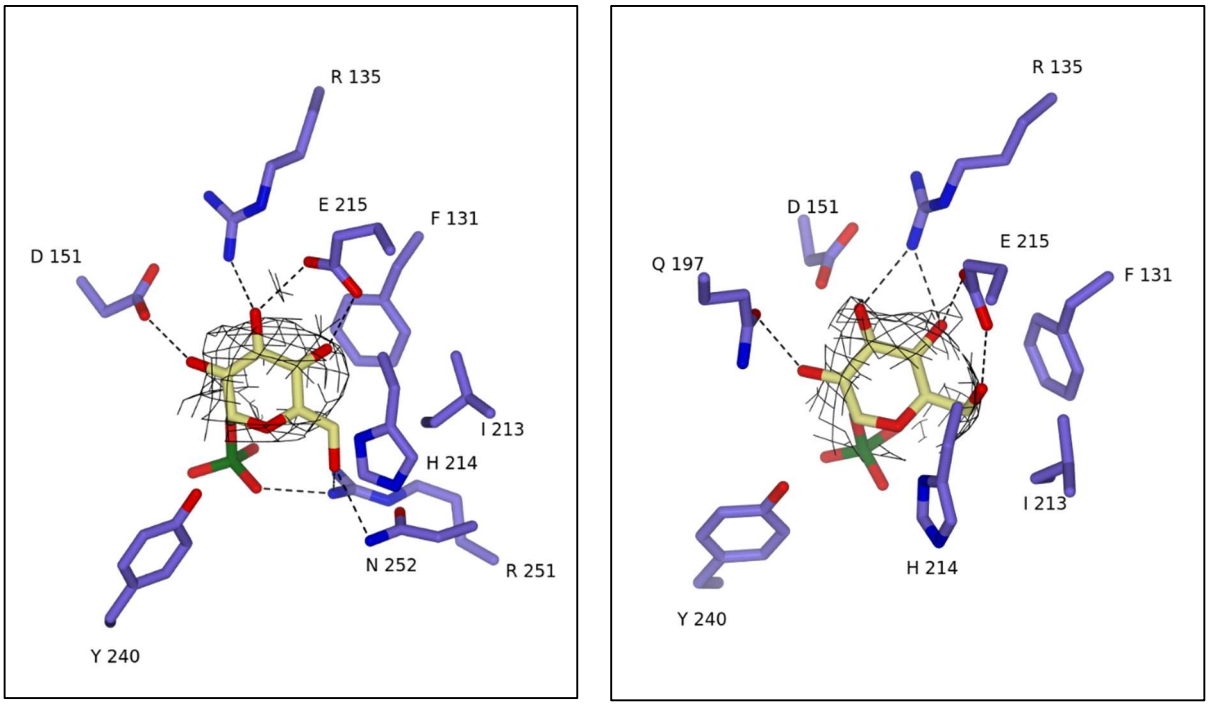


Figure 77 *Possible modes of glucose binding of EntS. The Glc and α-phosphate of the sugar-donor as well as amino acid side chains within 4 Å of the Glc moiety are shown as sticks. The electron density is shown at 0.05 e/Å*

While the second, possible conformation is very similar overall, a few key differences exist. Here E215 interacts with O4 and O6 instead, while R135 interacts with the vicinal diol of O3 and O4. O2 interacts with Q197 instead of a main chain carboxylate. Instead of being located at the hydrophobic face of the carbohydrate, F131 is located more towards the side of the sugar in this conformation and would interact primarily with the C5/C6 of the sugar.

In both of these conformations the O2 of the glucose is forming a hydrogen bond with a side chain. Q197 of these residues is poorly conserved between the investigated GTs and may explain the reported inability of EntS to use UDP-GlcNAc as sugar-donor. Other GTs have in its position a Thr or Leu with the loop further distant from the sugar moiety, allowing for a greater steric flexibility for O2 substitutions, especially PltS (see 3.11.4.3). D151 on the other hand is highly conserved between GTs, which would require the mode of sugar binding for EntS to be substantially different from the other GTs. Indeed, the glucose conformation in the bound state of SunS and ThuS has been reported to closely resemble the 2nd conformation seen here[168]. Furthermore, the lack of the C6-OH in UDP-Xyl severely reduces the affinity of EntS towards the sugar-nucleotide (chapter 3.4), indicating direct involvement of these atoms, making the 2nd binding mode more likely overall.

3.11.4.3. Carbohydrate donor – N-acetyl glucosamine

In vivo production of PltA lead to the identification of UDP-GlcNAc as the preferred sugar donor for PltS (chapter 3.8). Based on the crystal structure it is so far unclear as to why PltS appears to prefer UDP-GlcNAc over glucose *in vivo*, as there are limited potential interaction partners in the immediate surroundings of the N-acetyl group (Figure 78 A). A possible interaction may be R194 with the carbonyl, however the distance of 3.9 Å is outside the range of usual hydrogen bonds. A nearby Glu, E148, may facilitate this interaction by forming a hydrogen bond network with R194 and may help bind the 3.6 Å distant acetyl group through a water molecule. No electron density within the range indicative of an ordered water molecule could be found, however, though that may be due to insufficient resolution and data quality. A nearby Leu (L192) with a distance of 3.7 Å may constitute a hydrophobic interaction with the CH3-tail of the N-acetyl moiety. Interestingly, it is possible to see the nitrogen of the GlcNAc positioned directly above a phosphate group of the nucleotide-diphosphate, a phenomenon often described for Leloir-GTs and thought to contribute to the glycosylation activity by stabilizing the charge of the phosphate, similar to the nearby, basic sidechains of K244 and R247[106].

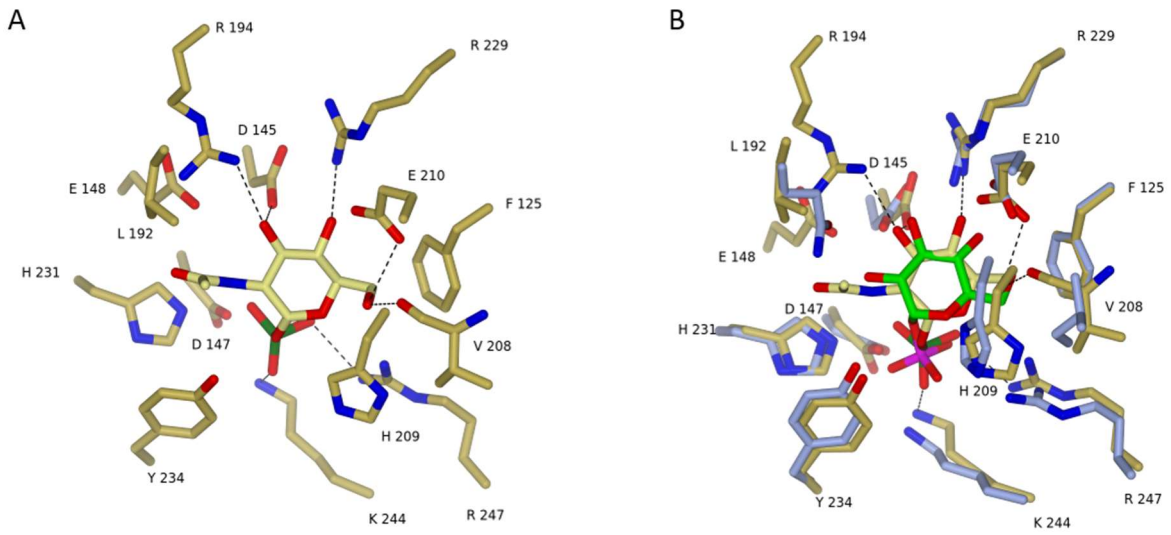


Figure 78 *A) GlcNAc-binding of PltS. The GlcNAc and the α-phosphate of the sugar donor as well as amino acid side chains within 4Å of the GlcNAc moiety are shown as sticks. B) The GlcNAc binding residues of PltS (gold) overlaid with the proposed Glc (green) binding residues of EntS (blue).*

Notably, while most interactions with the other hydroxyls, and the nearby phenylalanine appear to be similar in their position compared to the second proposed EntS-Glucose binding mode (see above), there are a few differences (Figure 78 B). The OH6-group is in position to form a hydrogen bond with the backbone carbonyl V208 (I 213 in EntS). Recognition of O3 appears to occur via D145 of the DXD motif of PltS and R194. Neither of such interactions are present in EntS. R229 forms a hydrogen bridge with O4. This interaction in addition to the 1.8 Å distant E210 is likely cause for the observed preference of Glc > Gal. A similar function is performed by R135 and E215 in EntS.

3.11.5. Acceptor peptide binding of EntS

Only for EntS a complex with the respective acceptor peptide could be obtained. The Ecm minimal peptide is held in place by interactions primarily located within the TPR-like and dimerization domain. Here helix α4 of the TPR interacts with the helical region of the acceptor peptide (amino acid 1-12) together with helix α1 of the dimerization domain. Interactions with the flexible part of the peptide (13-23) are formed primarily by α1 and α2 of the TPR-like domain as well as a loop between strands β5' and β6 of the Rossman-fold. The loop of the β-hairpin containing the H214 residue sticks out of the GT-A fold in the same plane as the glycosylated loop of the acceptor peptide.

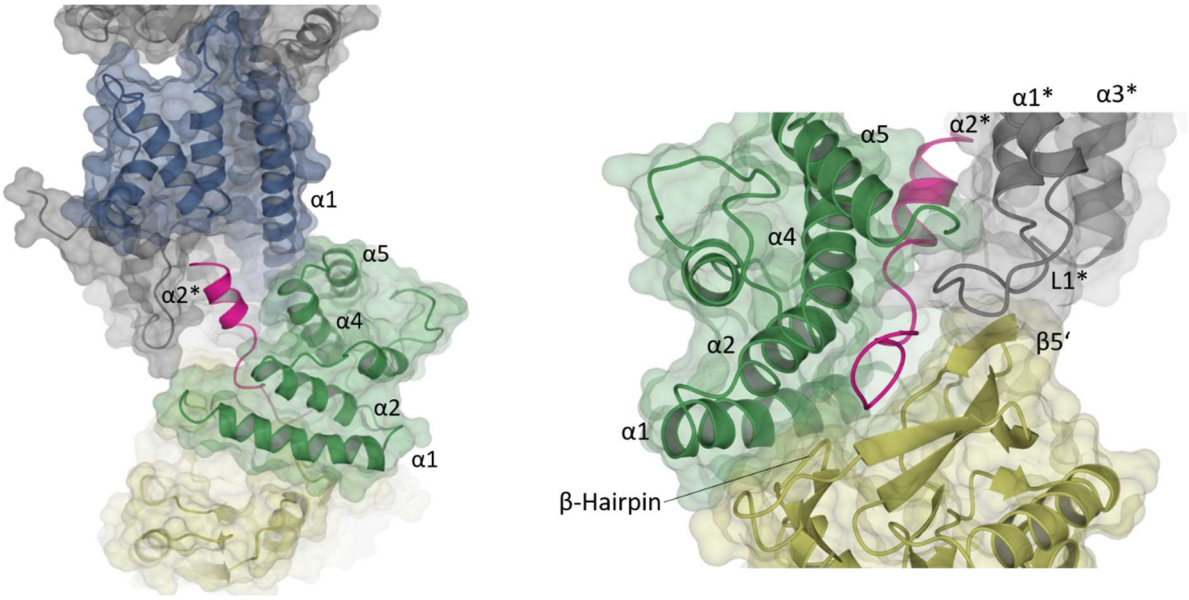


Figure 79 *Position of the acceptor peptide EC(Ala)m (pink) in the EntS dimer. GT-A fold is shown in yellow, TPR-like domain in green, dimerization domain in blue. The dimerization partner is shown in grey. A) the helical region of EC(Ala)m binds to a pocket composed of α1 and α2* of the dimerization domains, as well as α4 and α5 of the TPR-like region. B) The loop containing the glycosylation site enters a pore in the EntS protein enclosed by the loop L1* of the dimerization domain, α2 of the TPR-like region and β5' of the GT-A fold. The loop is sandwiched between TPR-α1 and the GT-A-β5' until it reaches the β-Hairpin of the GT-A region.*

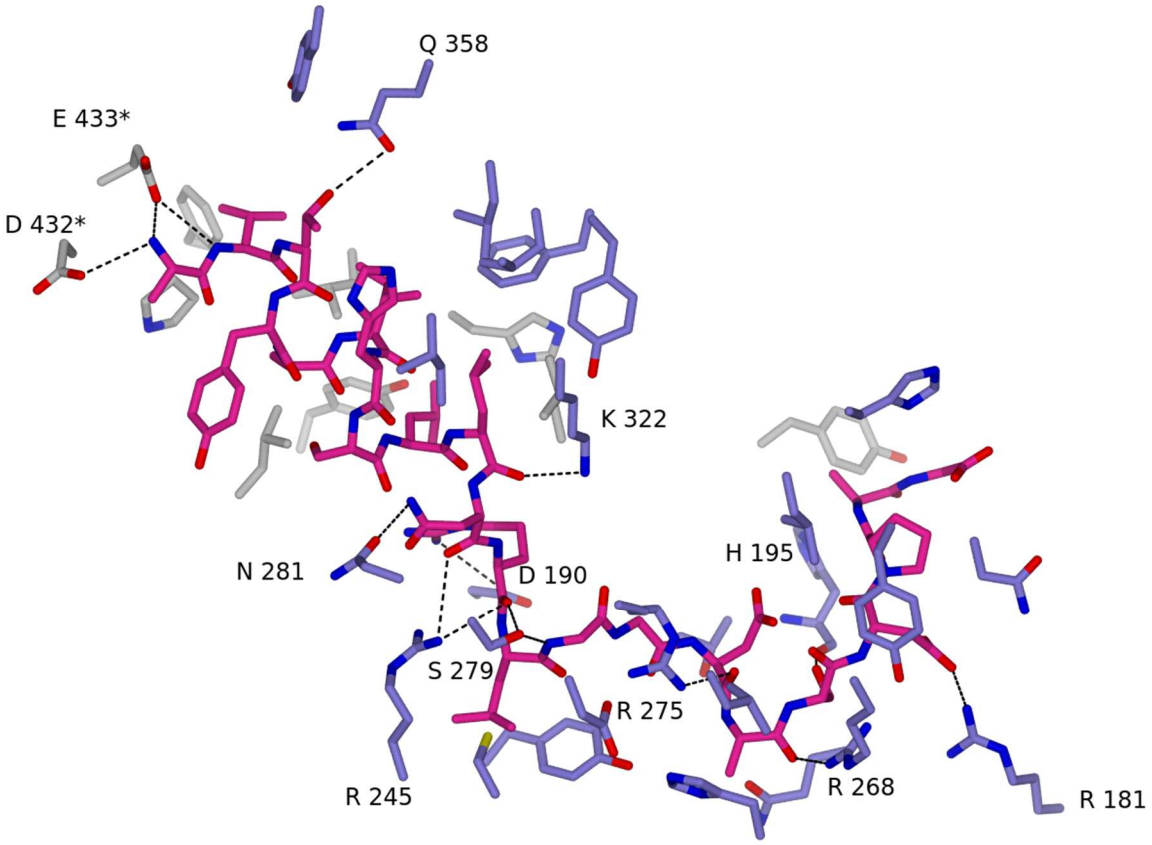


Figure 80 *Amino acids of EntS (blue) or its dimerization partner (grey) within 4 Å distance from the acceptor peptide (pink). Hydrogen bonds are shown in dashed lines and involved amino acids of the GT are labelled.*

A salt bridge between the N-terminus and a Q358-T3 hydrogen bond facilitate the interaction of the helical region of the acceptor peptide with its binding pocket. As the N-terminus-interaction is caused by the nature of the truncation construct, it is unlikely to contribute for the full-length peptide, however. Other interactions between the acceptor peptide helix and the GT are facilitated primarily by hydrophobic contacts, most of which are interactions with the dimerization-partner EntS*, as opposed to the catalytic EntS. In contrast, interaction with the loop contains more electrostatic interactions, many of which involve the peptide backbone. A list of peptide-enzyme interactions can be found in Table 8.

Table 8 *Residues involved in acceptor peptide binding of EntS.*

EC(Ala)m	***EntS***
Salt bridges	
N-term	D432*, E433*
R12	D190
D19	R181
Hydrogen bonds	
V2 amide	E433*
T3	Q358
N11	N281
S18	H195 carbonyl
L10 carbonyl	K322
N11 carbonyl	R245
R12 carbonyl	S279
L13 carbonyl	S279
G14 carbonyl	R275
G15 carbonyl	Y240
D16 carbonyl	R275
A17 carbonyl	R268
Hydrophobic interactions	
A1	I430*, P431*
V2	G401*, F416*, P431*
A5	L428*
I6	L356, L397*, L398*
L9	H394*, L428*, Y424*
L10	I325, I326, K322, L388*
L13	R245, Y240
A21	Y385*

Side chain interactions are limited to few hydrogen bridges between EntS-H314 and Ecm-D19, EntS-D190 with Ecm-R12. Further hydrogen bridges with main chain carbonyls can be found between the side chains of EntS-R245, EntS-279 and EntS-K322 with the carbonyls of Ecm-N11, Ecm-L13 and Ecm-L10, respectively.

3.11.5.1. Helicity of the Ecm peptide

Surprisingly, the Ecm peptide was bound to the EntS with a partially formed N-terminal helix. While this is consistent with secondary structure prediction, such as PSIPRED, Enterocin 96 is reported to adopt a disordered random coil conformation, prior to glycosylation[45]. Thus, I measured the CD spectrum of the Ecm peptide prior to glycosylation as well as the diglycosylated peptide. Neither of the two peptide variants showed the absorbance spectrum indicative of α-helicity (Figure 81). Addition of potassium fluoride to mimic the crystallization conditions did not change the absence of α-helix either. Thus, Ecm does not spontaneously adopt a secondary structure in aqueous solutions, and it is the binding to the enzyme that causes the peptide to show a helical conformation.

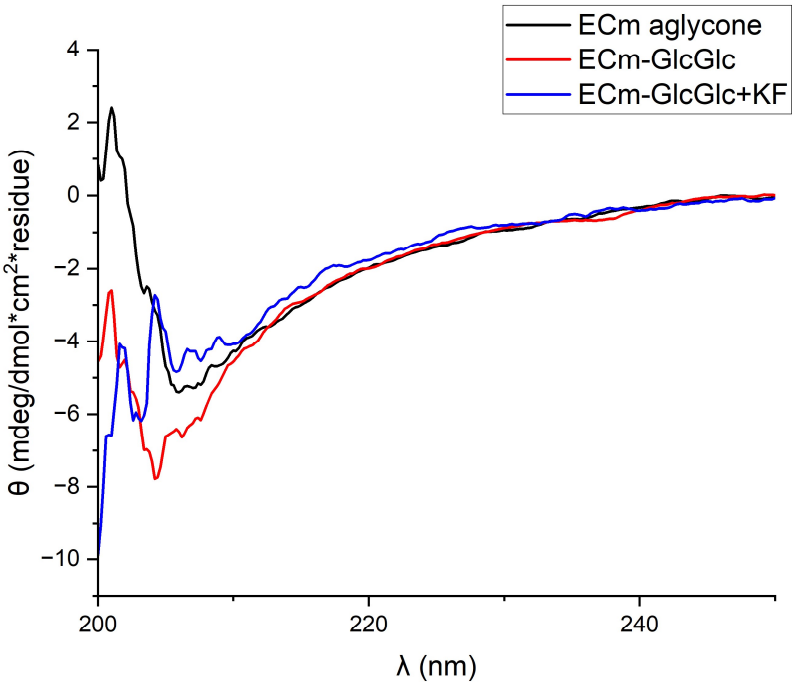


Figure 81 *Circular dichroism spectrum of the Enterocin 96 minimal peptide, Ecm.*

3.11.6. Snapshots of the reaction catalysed by EntS: Before the reaction and after the 1st reaction.

Interestingly, the main mode of binding of the Ecm peptide, facilitated by hydrophobic interactions appears largely the same for the unglycosylated peptide. The differences are located exclusively within the flexible loop (Figure 82).

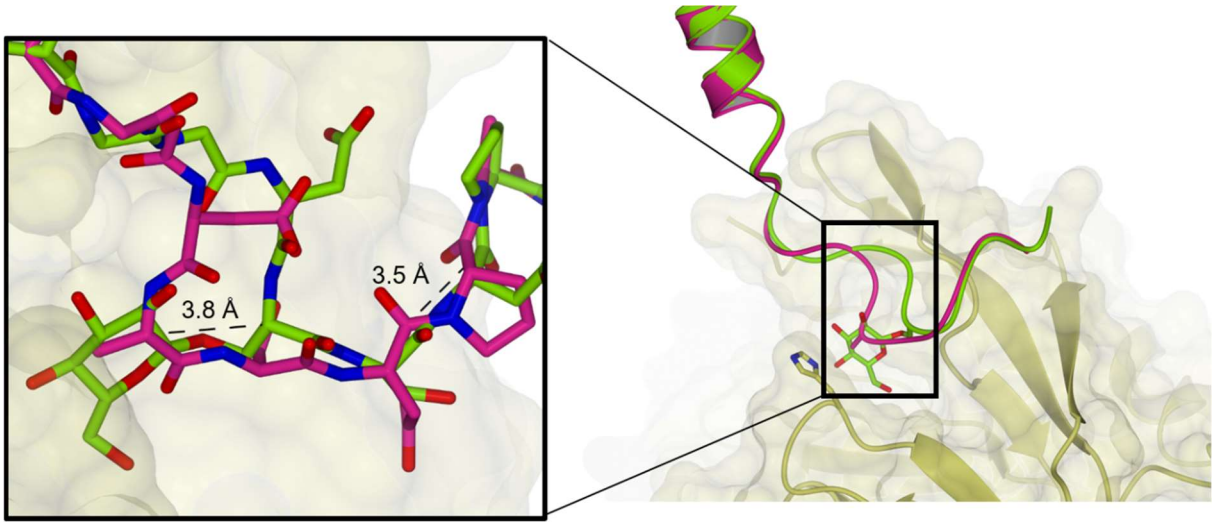


Figure 82 *Binding of the Ecm peptide before (pink) and after (green) glycosylation and zoomed in view (left). Sugar residue and putative catalytic base are shown as stick model, with the GT-A domain of EntS shown as cartoon with the surface overlaid. Other domains of the enzyme were removed for clarity. In the zoomed in view both peptides are shown as cylinder models with carbon atoms displayed in the colour of the corresponding peptide. Distances are measured from the C^{α} atoms of Ser 17 – Ala 17 and Asp 19 – Asp 19 of the respective peptide models.*

The Cα of Ser17 of the peptide moves back 3.8 Å, which allows the C4-hydroxyl group of the glucose to take the position previously held by the alanine (serine in wild-type sequence). This twist of the loop furthermore causes a shift of approx. 3.5 Å in the direction of the C-terminus of the peptide. This shift in turn causes a pronounced change in the hydrogen-bonding network of the glycosylated loop. D16 is now in position to form a hydrogen bridge with H195, while its carbonyl is no longer adjacent to R275. Similarly, the D19-R181 interaction is replaced by a D19-H314 salt bridge and a D19-T313 hydrogen bond. At the same time the S17 amide can now interact with H195 as opposed to the S18 hydroxyl group.

The glucose moiety interacts via O2 and O3 with R275 and O6 with R268. An additional hydrogen bridge of O4-H214 is formed in the position of the acceptor nucleophile. Together with the β-configuration of the first glucose as Ser17 and the relative position of the attached glucosyl moiety to

the bound UDP-Glc strongly suggests that the second glucose will also be attached in β-configuration, in line with the inverting mechanism of EntS[45].

3.11.7. Proposed reaction mechanism

The catalytic mechanism was of particular interest to see if differences between the *S*- and *O*-preferential GGTs exist. Two principal mechanistic proposals would be consistent with the experimental data.

Mechanism 1 starts with the destabilization of the oxygen-hydrogen bond of the attacking hydroxyl-group by H214, whose nucleophilicity is enhanced via a hydrogen bonding network involving D276 and the adjacent R275. The resulting δ^- charged oxygen then undergoes an S_N2-type reaction on the C1 of the glucose. Positively charged residues such as the Mn^{2+} ion and K248 stabilize the negative charge of the pyrophosphate leaving group.

Mechanism 2 occurs similar to Mechanism 1, but here the δ^+-charged H214 is stabilized by the hydrogen bonding network of N252 and R251. A hydrogen bond between R251 and the phosphodiester facilitates the reaction by protonating the charged phosphate. A similar reaction mechanism has been reported for the cleavage of a phosphodiester bond in a phospholipase[181]. While a His-Asn containing catalytic dyad/triad seem unusual compared to His-Asp, a similar motif has been reported for cysteine proteases such as papain[182].

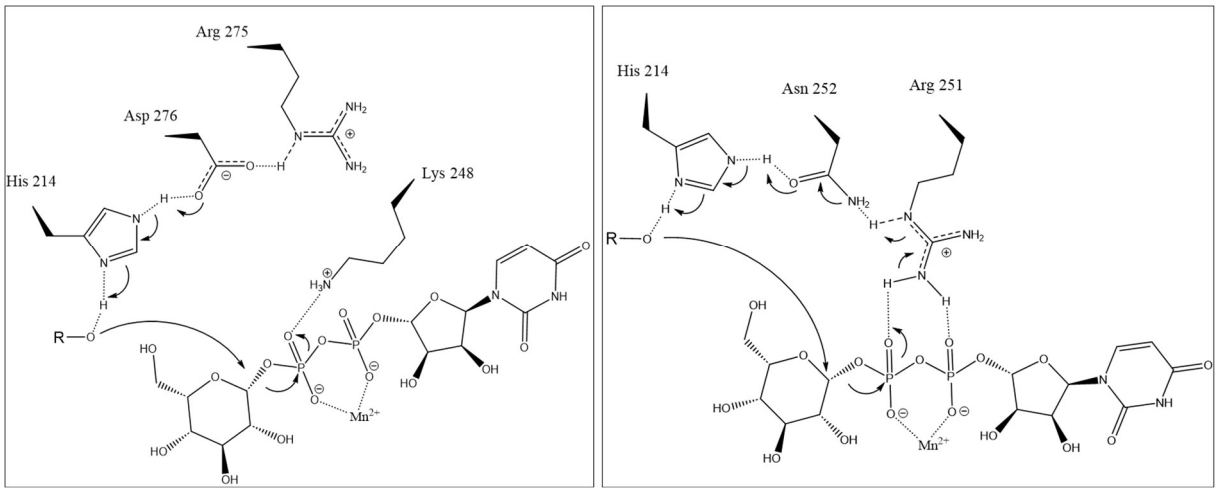


Figure 83 *Proposed reaction mechanism of EntS. A) The Serine or O4 hydroxy group is deprotonated by the catalytic triad of H214, D276, R275 and subsequently performs a nucleophilic attack on the C1 carbon of the glucose in an S_N2 type reaction. The resulting charge is stabilized by the Mn^{2+} ion and the nearby K248. B) The Ser or O4 hydroxy group is deprotonated by a H214, N252, R251 catalytic triad. The resulting charge is stabilized by the Mn^{2+} ion and the R251 side chain that is part of the catalytic triad.*

To verify the critical role of H214 as catalytic base, the His-to-Ala mutant, EntS(H214A) was generated. This mutant proved to be inactive (Figure 84), corroborating the importance of H214 for the catalytic mechanism, regardless which of these mechanisms, or combination thereof is followed.

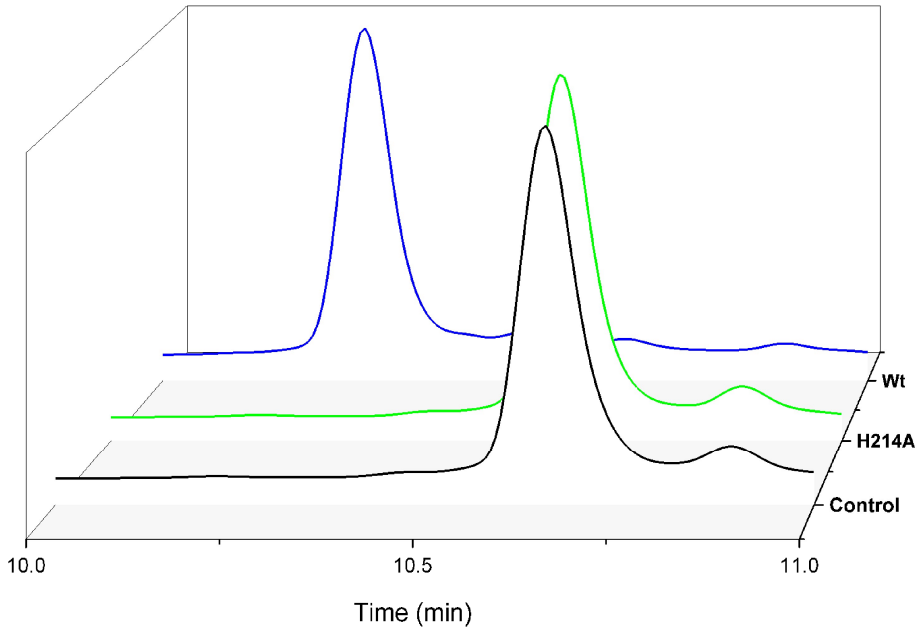


Figure 84 *Chromatogram of Ecm glycosylation after 24 hours catalyzed either by Wt EntS (blue), or a H214A mutant (green). A reaction mixture lacking any GT was used as control (black).*

It is of note that all the amino acids proposed here to be important for the catalysis of the reaction are highly conserved between the GTs, although a D276E (EntS notation) substitution is present in PltS, SacS and SunS, whereas N252D can be found in PltS only (Figure 69).

3.11.8. EntS hydrolyses diglycosylated Ecm peptides.

As mentioned before, the crystal structure of EntS cocrystalized with diglycosylated Ecm did not show any electron density corresponding to the second sugar moiety on the peptide. In an attempt to determine if this lack is caused by the flexibility of the carbohydrate or by a lack of this hexose, freshly prepared Ecm-GlcGlc was incubated with EntS and UDP in similar conditions compared to the crystallization conditions. The same solution, lacking the EntS enzyme and an Ecm-GlcGlc aliquot stored at -80°C were used as a control to determine if that effect may be due to peptide aging or spontaneous hydrolysis of the glycosidic bond at pH 8.0.

While no difference could be detected between Ecm-GlcGlc stored at -80°C and peptide stored at RT in solution at pH 8.0, the peptide stored in the presence of EntS showed a “shoulder” (Figure 85 A). Mass spectrometry revealed that the emerging peptide species has a shift in mass of -162 with respect to the diglycosylated peptide, indicating the hydrolysis of the glycosidic bond and loss of a glucose moiety (Figure 85B).

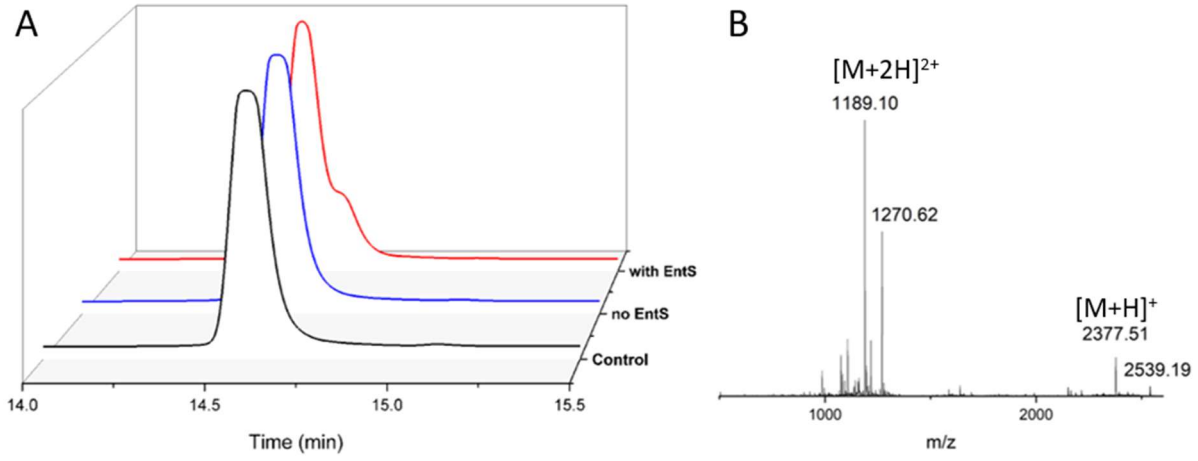


Figure 85 *Chromatogram (A) of Ecm-GlcGlc stored either at -80°C (black) at RT (blue) or at RT in the presence of EntS and UDP (Red). The mass spectrum of the emerging peptide species (B) reveals a loss of 162 amu.*

In an attempt to elucidate the mechanism of this surprising phenomenon, the Ecm-GlcGlc peptide was incubated in the presence or absence of UDP, as well as with the H214A mutant of EntS and samples were taken after 2 days. After 7 days a large amount of precipitate appeared and no Ecm peptide or glycovariant thereof could be detected in the supernatant. However after 2 days, evidence for the hydrolysis of the glycosidic bond could be detected. The hydrolysis was dependent on both, the nucleotide, as well as the residue H214 (Figure 86). While it is yet unclear if the nucleotide is only required for peptide binding, or if it acts as a nucleophile, the dependence of the histidine indicates a base catalysed S_N2-type reaction, congruent to the original transfer of the sugar to the peptide. While the large enthalpies of the nucleotide release indicate the former, such a phenomenon is known to occur for other GT proteins[125]. Throughout the experiment only the glycosidic bond between both glucoses seemed to be able to be hydrolysed by the enzyme. A complete deglycosylation was not observed.

Interestingly, full-length Enterocin 96 is likely to fold and form disulphides after glycosylation, after which it is unable to enter the catalytic site of EntS, which would effectively protect the peptide from unwanted hydrolysis.

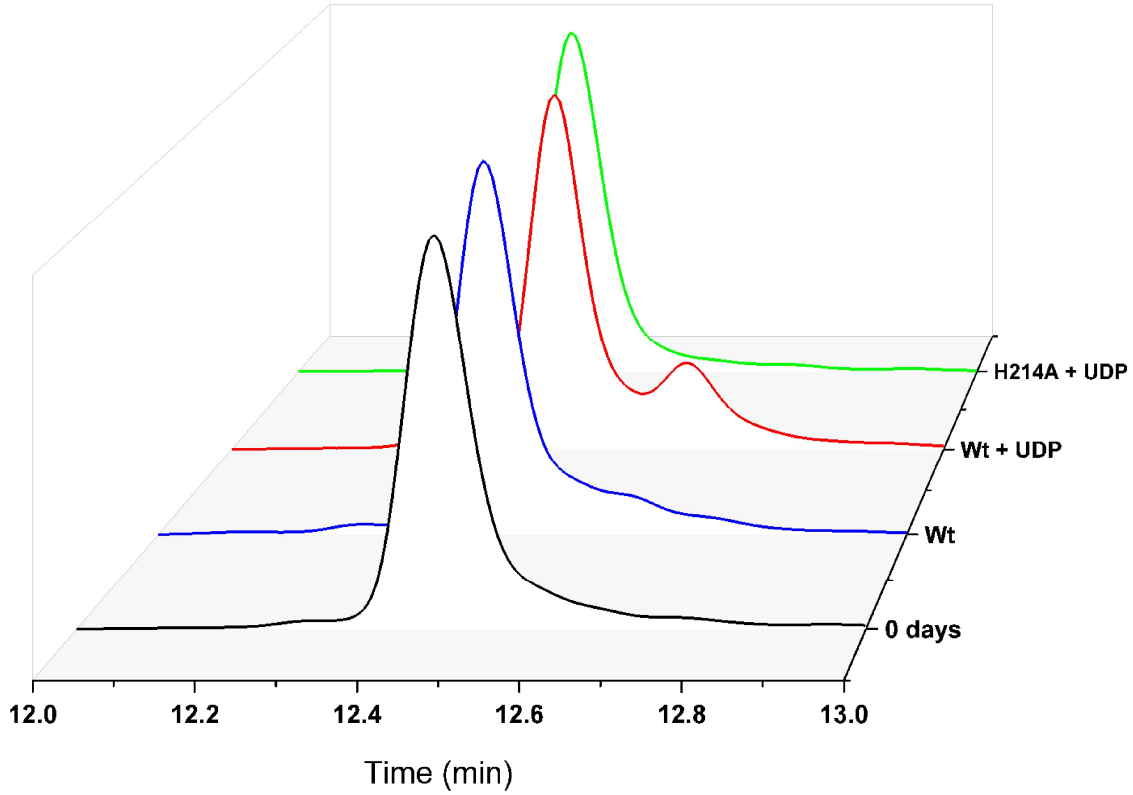


Figure 86 *Chromatogram of Ecm-GlcGlc incubated in the presence of EntS for two days. Absorbance intensities were normalized. The second peak corresponding to monoglycosylated Ecm only occurs for the Wt peptide and is dependent on the addition of UDP.*

3.11.9. Comparison between the S-preferential SunS/SacS and O-preferential EntS

As mentioned previously, the overall fold as well as the catalytic site appear to be highly conserved between GGTs. When comparing the surroundings of the catalytic His, only the D276E (EntS notation), L271F and a R268K substitutions stand out as noticeably different. The Phe is located in a parallel plane to the imidazole ring of the catalytic His, which may be caused by the different angle and distance of the catalytic base with regards to the sugar donor. In contrast, the angle and location of the Arg or respective Lys appear unsuitable to influence the reaction. In contrast, the carboxylate of the Asp/Glu residue is likely to be directly involved in the reaction mechanism. Indeed, for SunS its importance has already been verified[168]. While these residues remain well within distance to stabilize the catalytic His upon change from Asp to Glu, their conformation creates a slightly larger distance towards the nearby R275 (EntS notation), with 3.1 Å for EntS, 4.2 Å for SunS, and 3.6 Å for SacS. Similarly, the His-Asn-Arg interaction of the *S*-preferential GTs appears to

follow the same trend, with the His-Asn distance increasing from 3.3 Å to 4.2 or 4.3 Å for SunS and SacS, respectively. The Asn-Arg distance remains largely the same, as does its position. Interestingly for the highly similar GT ThuS, involved in the biosynthesis of the glycocin Thurandacin, it was shown that the Arg residue of this cascade is not vital to the reaction[168].

Thus far it is very much unclear however how, or even if, these minor changes in distance may cause such a pronounced difference in substrate preference. The reason for the apparent difference may also lie in a subtle shift in position and/or bond angles of the sulphur-nucleophile with regards to the His. The conserved Phe in lieu of the Leu of EntS may contribute towards the proper positioning of the catalytic His in this manner.

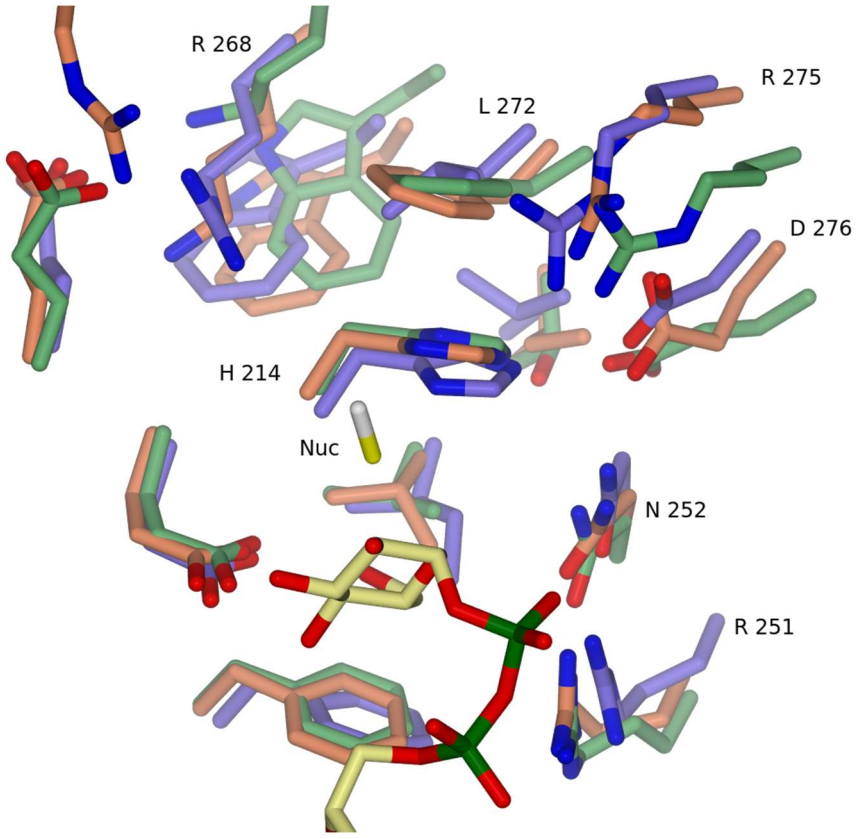


Figure 87 *Overlay of the catalytic site of SunS (green), SacS (red) and EntS (blue) by homology. Residues within 6 Å distance from the catalytic His are shown as sticks. The position of the nucleotide and sugar donor are inferred by homology to the EntS – Ecm-Glc or EntS-UDP-Glc cocrystal structures. Numbering of residues corresponds to the EntS sequence.*

The Histidine base is not necessary for Glycosylation of cysteines.

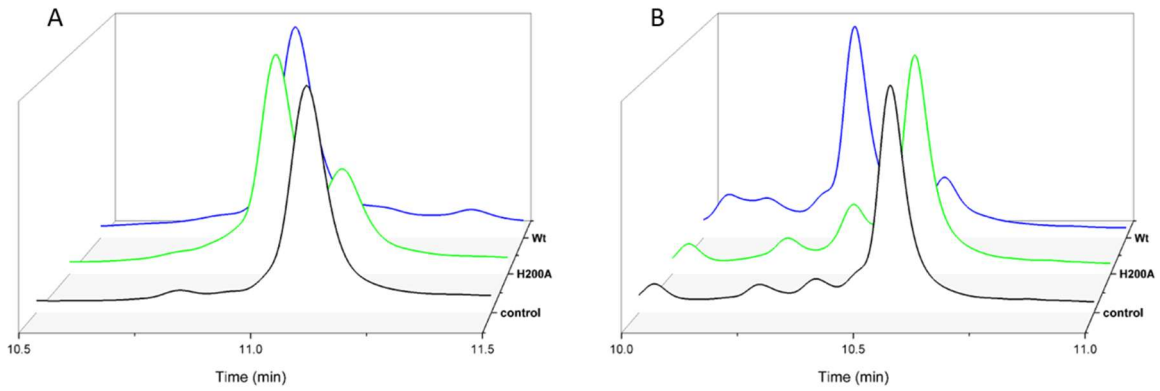


Figure 88 *Chromatogram of SacA4Ser (A) or SacAallSer (B) glycosylated with either Wt SacS (blue) or an H200A mutant (green) thereof. A reaction mixture lacking any kind of GT was used as control (black).*

As the H214 residue is highly conserved between glycocin glycosyltransferases (Figure 69), I also created a H200A (equivalent of H214 in EntS) mutant of SacS. Similar as was already reported for the Thurandacin-GT ThuS[168], this mutation only having a minor effect on its capacity to glycosylate sulfhydryl groups in SacA4Ser (Figure 88 A). Indeed, even some residual activity with regards to the hydroxyl-group of SacAallSer could be detected (Figure 88 B), indicating that the reaction mechanism of *S*-glycosylation is substantially different from the mechanism proposed for EntS. Without the His-residue deprotonating the Ser-nucleophile this activity, a different catalytic base is required for the reaction. While a Glu (E201) is located 5.2 Å distant from the putative position of the nucleotide (Figure 87**Figure 87**), this residue is conserved between all investigated GGTs (Figure 69) and thus is unlikely to contribute to this marked difference between EntS and SacS.

3.12. Insights into the folding of glycocins

CD spectroscopy revealed that the unglycosylated peptide SacA does not adopt an α-helical architecture. This is contrary to predictions using folding servers, such as PSIPRED[183], yet consistent to observations of the Rao group for Enterocin 96, which only adopts a secondary structure after glycosylation. Similarly the minimal peptide of SunA, SunAm did not show α-helicity in its non-glycosylated form, despite this N-terminal helix being reported as necessary as a recognition site for the cognate GGT[74].

To investigate the folding state, CD-spectroscopy was used to determine the secondary structure of the prepared peptides. With the exception of PltA, no aglycone showed any signal indicative of a secondary structure. In contrast, the mature, glycosylated peptide of SacA and PltA did show the expected signal for α-helices, similar to what has already been reported for GccF, SunS and Ent96. Surprisingly the glycosylated variant for SacA lacking disulphide bridges, SacA4Ser-Glc, does not show a secondary structure (Figure 89 A). This reinforces the idea that the disulphide bonds are essential for the stabilization of the 3D-structure and therefore, likely function. Similarly, the fold of the PltA peptide was influenced greatly by the presence of the singular disulphide bond. While unglycosylated PltA already shows some amount of secondary structure, it gets more pronounced and more helical upon glycosylation, a phenomenon that has also been reported for GccF[47]. However, only in the presence of the disulphide bond does the CD-spectrum indicate the helicity predicted with *in silico* methods.

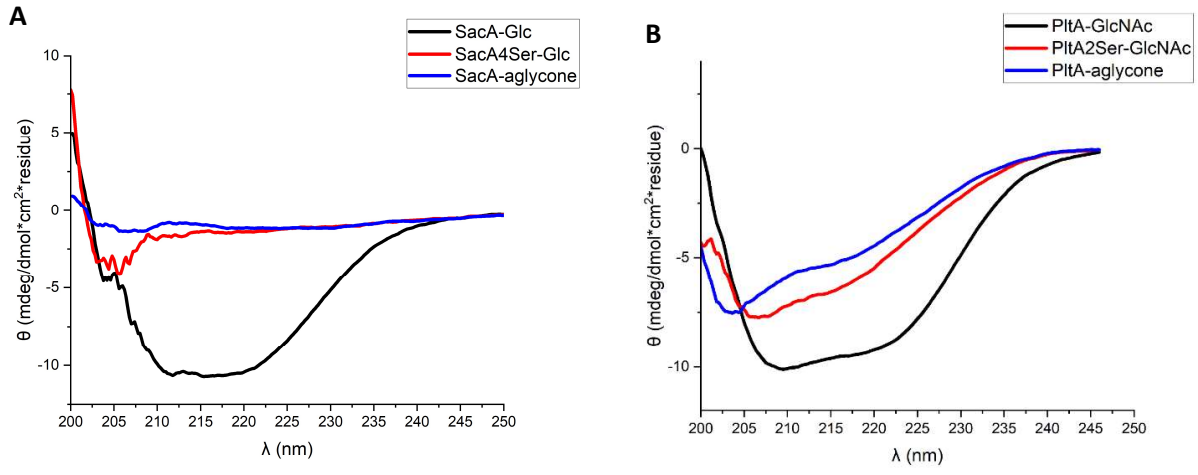


Figure 89 *Circular dichroism spectrum of SacA (A) and PltA (B). The aglycone is shown in blue, the glycosylated wild-type sequence in black and a glycosylated mutant unable to form disulphide bonds in red. Ellipticity indicative of α-helical structures is only present for the wild-type peptides.*

Taken together, with the helicity of the Ecm peptide only observed when in complex with the GT (chapter 3.11.5.1), it seems likely that the association with the helix-bundle of the GT introduces the initial helix formation within the glycocin peptide. This secondary structure is then stabilized by the hexose and subsequently the rapid formation of the disulphide bonds, which keeps the secondary structure stable. Indeed, with the correct disulphide-bonds the secondary structure seem to be retained even under denaturing conditions, used for HPLC analysis after synthesis.

4. Discussion

4.1. Discovery of glycocins

Several putative glycocin peptides could be discovered by sequence similarity to known glycocin peptides. As mentioned previously, the discovered peptides match well to the results of Singh&Rao[16]. Using a similar, yet more expansive and automated methodology, they were able to identify all of the here described putative glycocins as well, with the notable exception of SacA. Given the high sequence similarity of SacA to SunA, the identity as a glycocin or at least glycocin-like peptide appears likely, even without the inclusion of a putative immunity protein within the gene cluster. However, purely automatic identification and annotation methods might miss interesting putative glycocins as shown for the case of SacA.

Due to the manual nature of the genome mining performed in this work, the volume of putative glycocin genes and parent strains was comparatively low. This in turn may be reason for the difficulty clustering the observed sequences into distinct glycocin families, as seen by the generally low bootstrap values of the maximum-likelihood phylogenetic tree.

While the phylogenetic tree proposed by Singh&Rao[16] is made up from a much larger set of sequences, it shows far less distinctive and a greater amount of families of glycocins (Figure 90). Overall, both trees appear in agreement with each other, but some notable exceptions have been observed. Interestingly, in their assessment, *B. thuringiensis* does not cluster together with *B. subtilis*, but clusters in its own family of glycocins. The putative glycocin of *C. botulinum* clustered together with the *B. subtilis* group, whereas in my tree it was closer to *B. thuringiensis* (Figure 20). In a similar manner, *Streptomyces*-glycocins clustered into their own family, which did not include Enterocin F4-9 in the tree of Singh&Rao. The Glycocin F-family shows great overlap between my tree and the one of Singh&Rao. In both trees *S. plakortidis*, *S. hominis* and *T. thermosaccharolyticum* clustered together with GccF. Interestingly, *O. bavariensis* and *C. perfringens* NM49 B9-7 are absent from the tree of Singh&Rao, whereas they were part of the GccF cluster in my tree. A different putative *C. perfringens* glycocin was found by Singh&Rao which, however, clustered in a diverse family, containing, among others, *Yersinia*, *Ruminococcus* and *Bacillus* species. A possible reason for the discrepancies, aside from their inclusion of a vastly larger number of sequences, is the different method with which the data were analysed. While the phylogenetic tree of this work was inferred using a maximum-likelihood method, Singh&Rao used a neighbour-joining algorithm.

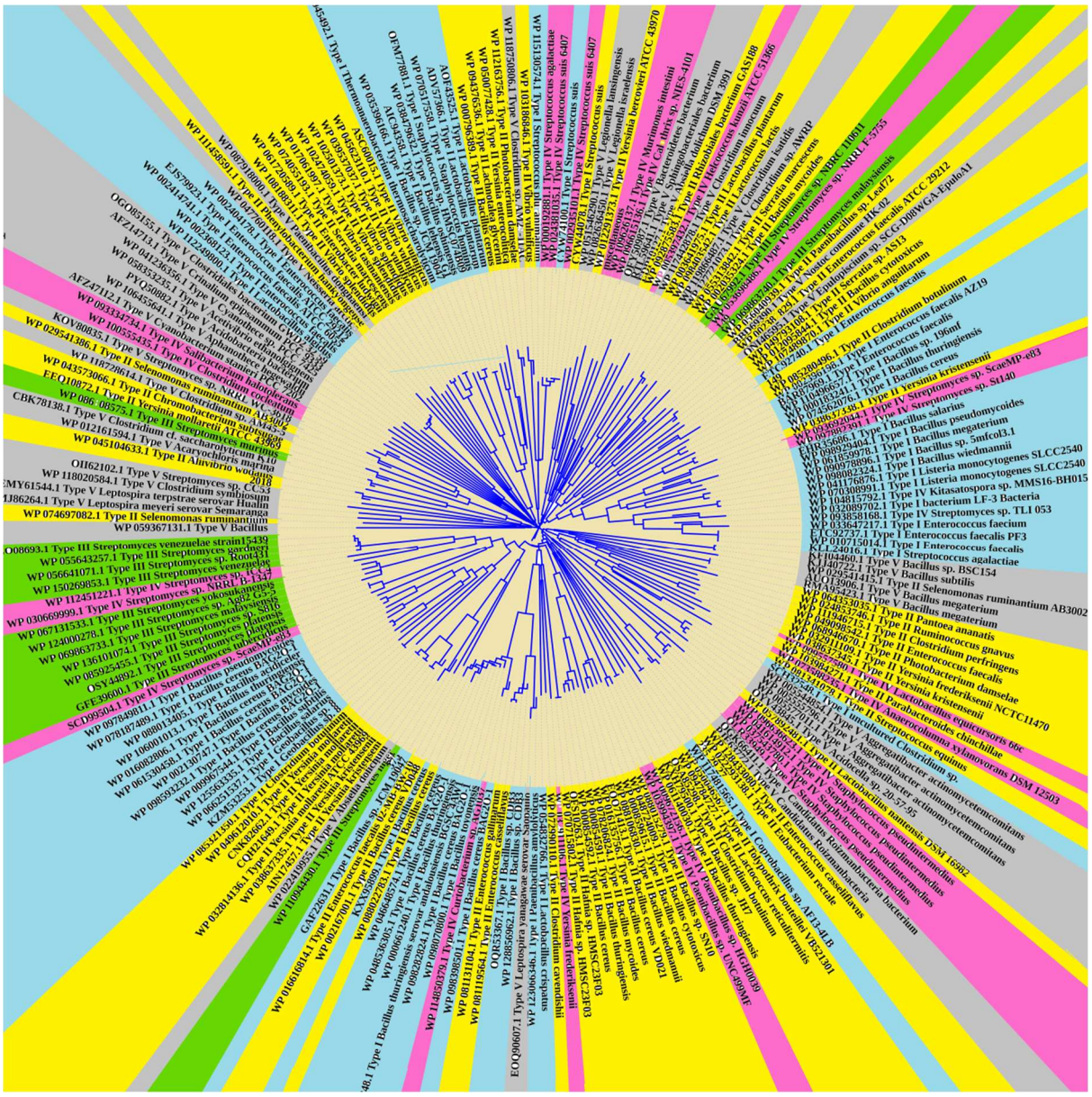


Figure 90 *Phylogenetic tree of Glycocins, published by Singh&Rao[16].*

Nevertheless, all putative peptides that were investigated in more detail, were indeed glycosylated by their cognate glycosyltransferase, corroborating their identity as glycocin-peptides. Given their sequence similarity to published glycocins, AciA, SacA and PltA are likely *S*-glycosylated, though there is as of yet a lack of experimental evidence for AciA and PltA. Likewise, no antimicrobial activity could be reliably assigned to any of the putative glycocins identified in this work. As glycocins are often reported to have a very narrow spectrum of activity and can be highly specific towards different strains of the same species (see chapter 1.1), it is thus far unclear whether a susceptible strain has not been found yet, or whether these peptides do not exhibit any antimicrobial activity. This matter is complicated further as species most likely to be inhibited by PltA and SacA are grown in carbohydrate-

rich media under laboratory conditions, which might inhibit the activity of the glycocins. Nevertheless, these peptides may also constitute a new class of glycocin-like peptides without antimicrobial activity, akin to non-lanthibiotic lanthipeptides, and have an unknown function.

4.2. Synthesis of glycocins

A major challenge in the investigation of glycocins has been the challenging preparation of pure glycocin-peptide. Chemical synthesis procedures for Sublancin[162, 184] and GccF[62, 163] by solid phase peptide synthesis (SPPS) and native chemical ligation (NCL) have been reported, though they suffer from poor yields of 0.4-6 % and challenging synthetic protocols. A different option is the isolation of these glycocins from the supernatants of the producer strains. Often, the exact conditions for the optimal production are not known, or challenging to reproduce in a laboratory setting, making this option not generally applicable. An interesting approach has been shown for Sublancin, where engineering of the *B. subtilis* 1A747 strain with the introduction of three strong promoters into the synthetic gene cluster allowed for the isolation of over 600 mg of pure peptide from one litre of media supernatant[185]. As this method requires intricate knowledge of the regulation of the involved genes, the most common approach used so far to produce glycocins is via pathway refactoring into *E. coli* as a heterologous host, which has proven successful for Listeriocytosin, Bacillicin, and Pallidocin[49, 50]. Unfortunately, *gccA*, the gene encoding for the GT of the GccF-peptide does not express well in *E. coli* and pathway refactoring of the GccF gene cluster needed to be performed within the gram-positive *L. plantarum* NC8 expression strain to be successful[167]. An improved approach was developed for Sublancin[175] and Ent 96[75], as it was discovered that the presence of the oxidoreductases was not necessary for the folding of these glycocins in *E. coli*, despite being required for efficient production in *B.subtilis*[186]. Co-expression of the glycocin tagged with either a His-tag or Chitin binding domain (CBD) together with the cognate GT proved sufficient for the isolation of mature peptide, though yields were generally poor for the His-tag-based approaches (up to 3 mg per litre of culture). Ent 96, produced with a C-terminal CBD tag was isolated in yields of up to 30 mg/ml, though it is unclear if that is the fusion construct or the peptide alone.

In this work two methods to obtain mature glycocin were explored: The chemoenzymatic route and a general biosynthetic route. While the exact nature of the occurring side reactions during the aglycone peptide synthesis part of the chemoenzymatic route was not investigated in detail, it was possible to reduce their occurrence to a level that SacA, PltA and Ent 96 could be obtained in acceptable yields of

3-8% for the aglycone, comparable with yields obtained for the SunA-aglycone using Microwave-assisted SPPS and NCL[162].

The synthesis procedure could likely be further optimized, especially with regards to the coupling of the last 10 amino acids of both PltA and Ent 96, during which most of the side reactions appeared (Data not shown). Additionally, the composition of the deprotection cocktail may be improved upon. Thioanisole has been implicated in several side reactions and replacement for Me_2S/trimethylsilyl chloride/triphenylphosphine, has been shown to have superior properties for protecting Cys and Met containing peptides[187]. Furthermore, strict monitoring of the reaction time and temperature may further reduce the amount of side reactions encountered.

Consistent with previous reports about the synthesis of Glycocin-peptides, aggregation posed a serious problem[163, 184]. Contrary to these reports, however, aggregation of SunAm, SacA and PltA was less prominent during the coupling of amino acids but occurred primarily during the purification and downstream processes. Surprisingly, this effect was largely dependent on the presence of trifluoroacetate, a common supplement to during HPLC purification of peptides for its peak-sharpening effects[188], as well as its contribution to efficient separation[189]. In stark contrast to reports for other peptides, such as Amyloid β, where the addition of fluorinated solvents efficiently disrupts aggregates and allows for solubilization[190], the addition of HFIP exacerbated the aggregation propensity for SacA- and SunA-derived peptides. Fluorinated solvents inducing the formation of nanostructures has already been reported for some peptides[191], and would be consistent with the observed drop in solubility. Fortuitously, it was possible to replace the trifluoroacetate with chloride to afford soluble peptides. While these peptides were generally still prone to aggregation and precipitation in a pH dependent manner, it was possible to decorate these peptides with carbohydrates in an enzyme-catalysed reaction.

Nevertheless, especially the glycosylation conditions for SacA may be optimized further, as pH dependent side-reactions were observed. Additionally, the wild-type SacA peptide precipitated at basic pH after a short amount of time, indicating that the concentration may need to be optimized further to allow for efficient glycosylation. Further precipitation may be due to the vast excess of reducing agent in the queried glycosylation condition, as the helical fold is not stabilized by the critical disulphide bridges at this point. This is further supported by the mass spectra of SacA-Glc showing the peptide to be present almost exclusively in its reduced form. At the same time, especially for the SacA peptide, a surprisingly large amount of unglycosylated peptide could be detected, even after extended periods of glycosylation. Considering the previously mentioned precipitation, this peptide fraction may represent the precipitate that was, due to its aggregated state, unavailable for glycosylation, but

dissolved during the quenching step at acidic pH. It may, however, also point to product inhibition, either by UDP or by the glycosylated, yet not folded peptide. Considering the vast excess of UDP-Glucose over the emergent UDP, product inhibition of this extent seems unlikely, even though UDP has higher affinity to the enzyme than its sugar-donor counterpart. Some of the problems regarding the concentration dependent aggregation and potential product inhibition might be alleviated by the use of methods from flow chemistry combined with immobilized enzymes. The product could be constantly removed and slowly brought into a slightly oxidizing environment to favour the formation of the stabilizing disulfide bonds. Furthermore, lower concentrations of peptide could be used, while the total amount of peptide could be kept the same. In this manner both aggregation and erroneous disulphide connectivity problems may be reduced. Nevertheless, with the currently developed method all investigated peptides could be glycosylated *in vitro* with acceptable yields (40-70 % in a one-step reaction) which could be further improved by recovering the starting material.

As the chemoenzymatic methods to obtain mature glycocin scaled poorly, a separate approach to generate them by heterologous production in *E. coli* was investigated. As mentioned previously, a similar approach was already published for Sublancin and Enterocin 96, making use of a His-tag or CBD-tag, respectively[74, 75, 175]. Given the major difficulty of glycocin synthesis being their limited solubility in aqueous media, the His-tag was considered unsuitable as a general method, as this tag has been reported to have a low, or even negative impact on protein solubility[192]. In line with the expression system of Enterocin 96, I developed and optimized an *in vivo* production method, with the MBP-tag conferring the required solubility. In contrast to the CBD-tag approach I chose a TEV cleavage site *in lieu* of an intein based strategy, as removal of the latter is performed under strongly reducing conditions and thus might be prone to disrupt the disulphide bridges vital for Glycocin-activity. Indeed, under reducing conditions the cysteines of SunA could be reacted with iodoacetamide[42], while glycosylation reactions *in vitro* under reducing conditions afforded the reduced peptide as primary product (see chapter 3.5), once more exemplifying the general susceptibility of the glycocin disulphides towards reducing environments. In contrast, all glycocins produced using the TEV-cleavage method afforded peptides with mass spectra corresponding to the predicted number of disulphides, with -4 amu for SacA, SunA and AciA, and -2 amu for PltA. Gratifyingly, even the AciA peptide that was completely insoluble after solid-phase peptide synthesis, could be obtained. Notably, the cleavage of the peptides from their MBP-tag required the isolation of the monomeric form of the construct. It appears likely that erroneous formation of disulphides creating a dense network of glycocin-peptide polymers, which remained in solution due to the solubilizing power of the MBP-tag. Such a problem has not been described to occur for SunA[175] or Ent96[75]. Indeed, earlier reports even showed the possibility to obtain reduced, unglycosylated SunA or ThuA peptides from

heterologous expression in *E. coli* [43, 74]. Expression in *B. subtilis* lacking the oxidoreductases did however lead to the erroneous disulphides[186] that may be the cause for the apparent oligomerization. Glycocins investigated herein also displayed a "stickiness" due to the pronounced hydrophobicity that has been described for other bacteriocins[193], though to varying degrees. That could be also a reason for the observed oligomerization. Additionally, these "stickiness" of glycocins, hampers the purification by interaction with the separation material used in the column purification. AciA exemplified how such a problem may be overcome by the addition of suitable amounts of organic solvent to the mobile phase during chromatography steps, though yields remained poor even with this method.

Currently the glycocin production method still suffers from certain limitations. Firstly, despite the production of Enterocin 96 being published using a similar method[75], the approach developed in this work has, so far, not afforded the HPLC-grade pure peptide of Enterocin 96. This might be related to the N-terminal location of the MBP-Tag as opposed to the C-terminal CBD-Tag used in the published method[75]. Secondly, the reason for the apparent premature termination of the translation after the MBP, resulting in an apparent expression of the MBP-tag alone is currently not understood. While this is not an issue for the SacA and PltA wild-type peptides, it occurs strongly for SunA, AciA and mutations of SacA that increase its aggregation propensity when introduced into synthetic peptides. Without knowledge about said mechanism it is difficult to foresee which peptides may prove problematic until attempts at their *in vivo* production have been made. An interesting observation is the dependence on the growth medium of this phenomenon. It is much more pronounced in the standard TB-medium compared to the autoinduction medium. It is yet unclear whether the causative factor lies within the composition of the medium or with the time of induction, as expression in TB is performed in the mid-log phase of growth, whereas the autoinduction media rather lead to expression during the stationary phase. Nevertheless, both point towards an influence of the regulatory genes of *E. coli.* Whether the lack of the glycocin peptide is caused by proteolytic cleavage or ribosomal stalling is, as of yet, undetermined.

Lasty, this method requires the ability of *E. coli* to efficiently translate and fold the cognate GT of the glycocin of interest. While this proved to be possible for all glycocins investigated in this study, glycocins of the GccF type have so far not been attempted with this method. All reported attempts to produce GccA recombinantly in *E. coli* failed[167]. Thus, a method to produce GccF recombinantly is not available. It remains to be seen if that is a general feature of this subclass of glycocins and their cognate glycosyltransferases or rather an unfortunate unique feature of the GccA/GccF pair.

4.3. Crystallization and structure determination

In this work the structure of three different GGTs, SunS, PltS and EntS, each with notable differences in their activity could be solved at a moderate to low resolution, as well as the structure of the catalytic domain of SacS. While the structure of the GGTs SunS and ThuS have already been published at a resolution of 3.0 and 2.1 Å respectively[168], the here presented structure of the SunS protein provides a slightly better resolution with the L7 loop showing less disorder.

As mentioned previously, the overall structures of the GTs investigated in this work are highly similar in their architecture and resemble the already reported structures of the GGTs ThuS and SunS[168]. An N-terminal Rossmann-fold followed by a helical substrate binding domain, termed TPR domain, is a common architecture amongst structures of the GT-2 family. The presence of the hook-shaped dimerization domain appears to be so far unique for GGT's within the GT-2 family. Surprisingly, the overall architecture is reminiscent of OGT, a GT-B fold GT containing a C-terminal, α-helical TPR domain region. This domain is, like the its equivalent in the structures investigated here, assisting the acceptor peptide recognition[194] and dimerization[195] of the enzyme, though the orientation and type of dimerization between the monomers is noticeably different to that of GGTs.

The mode of nucleotide binding appears to resemble the interactions described for other GTs of the metal-dependent GT-A type fold (see chapter 1.3). The metal ion is coordinated by the C-terminal Asp of the DXD-motif as well as a His, located at the C-term of the β7 strand. The loop L7 forms a lid over the sugar donor molecule upon binding. Similarly, the interactions of positively charged side chains with the phosphate backbone have also been reported in other GTs, though primarily in metal-independent proteins[106]. The same binding mode is described for ThuS and EntS proteins[45, 168]. Both of these required the His and Asp at the metal binding site for their function. Interestingly, the Arg and Lys side chains interacting with the phosphate improved the rate of activity, but neither were vital for the function of these proteins [45, 168].

Differences in the interactions with the aromatic base can be seen amongst the solved structures, with a π-π-π sandwich of SunS and SacS. Such rather unspecific binding mode corresponds well with the observed thermodynamic data and the displayed promiscuity towards the different nucleotides. Interestingly, SacS showed a much greater promiscuity towards several nucleotides with a 5-6-fold higher affinity to the purine bases ADP and GDP, while only showing a single amino acid difference within the nucleotide binding pocket, with a Met instead of an Ile near the nucleobase. While an S-π interaction is possible, a similar interaction should also take place for UDP, resulting in a generally

improved affinity. This was not observed, however. A further possibility is, that the choice of nucleotide shifts the sugar unit by several degrees, which may optimize the binding of the whole ligand. Indeed, such a mechanism would explain the increased affinity of SacS with regards to ADP-Glc, whereas the affinity to ADP alone is more similar to that of SunS. This would also explain why the affinity of SacS is higher to UDP alone than to UDP-Glc; a unique feature amongst the investigated GTs. Due to the complete hydrolysis of the sugar-nucleotide within the crystal, the positioning of the glucose within the active site could so far not be determined. Non hydrolysable analogs of the donor may help to gain insight how the carbohydrate may influence the positioning of the nucleotide handle in the binding site. Interestingly, for many of the amino acids surrounding the nucleobase, the aforementioned experiments of ThuS revealed that many of these are not vital to the functionality of the GT. For instance, either of the aromatic side chains involved in the uracil π-stack could be replaced by an Ala, without compromising the activity of the enzyme.

In contrast, the Trp forming the singular π-π stacking interaction of EntS was vital to its function[46], agreeing with observations that the nucleotide binding of this enzyme is weaker in general. This weaker, and more specific binding can easily be explained by its structure, showing two hydrogen bonds between the uracil and a nearby Lys or Thr, in lieu of the π-π sandwich of SunS-like GTs. PltS on the other hand shows a different mode of binding for the purine base, with a Cys located beneath the Uracil, contributing in an S-π binding. While it is unclear how this residue influences the specificity towards the donor molecule, it seems likely that the overall binding towards the nucleotide-part of the donor is decreased, in turn increasing the importance of the sugar-residue facilitated interactions. Comparing the binding pockets of the more promiscuous EntS with that of the donor restricted PltS, the binding pocket of the latter appears narrower and displays additional interaction partners, D145 and R194. The corresponding residues in EntS are not in position to interact with the carbohydrate. Interestingly, in the donor binding sites of SunS, SacS and ThuS[168] show greater similarity with PltS in this regard than EntS. This is surprising, considering the greater donor tolerance of all three enzymes, in particular SacS.

Using a mixture of ITC and functional assays it was possible to expand the knowledge of accepted and preferred sugar-donors of GGTs. For both SunS and EntS the obtained affinities correspond well to the observed activities. For SunS the clear preference for sugar nucleotides is Glc with Glc>Gal/Man/GlcNAc>Xyl, which corresponds with the described activity *in vitro*[42]. Similarly, EntS showed a clear preference for Glc, whereas other sugar-nucleotides such as UDP-GlcNAc, UDP-Gal and GDP-Man bound with much lower affinities. While EntS has been reported to use UDP-Gal *in vitro*[45], it is unable to utilize UDP-GlcNAc and GDP-Man, despite binding these sugar-nucleotides with

micromolar affinities as well. For GDP-Man, this may be explained by the O2 of the Man pointing towards the acceptor peptide, causing an electrostatic repulsion and thereby preventing the reaction to take place. No such effect should be present for UDP-GlcNAc, however. Steric repulsion of the Gln or Asp in contact with O2 for EntS (chapter 3.11.4.2) may cause the sugar to bind in a twisted conformation, increasing the distance of the nucleophile to the anomeric carbon of the sugar, thereby limiting the reaction, though no data as of yet was obtained that could corroborate this hypothesis.

While the binding affinities for SacS showed a clear preference of UDP-Glc>GDP-Man>UDP-Gal≈UDP-GlcNAc>Xyl, there are differences to the glycosylation reaction *in vitro*. UDP-GlcNAc is used with greater efficiency than UDP-Gal, despite similar affinities, while GDP-Man with superior affinity is not used as a substrate, in stark contrast to the activity reported for the highly similar SunS[42]. The inability of SacS to use GDP-Man can also not be explained by the structural data. Similar to the structure of SunS and ThuS[168], no amino acid in the active site would lead to a steric or electrostatic repulsion of the O2 of a mannose residue. Given the C2-O2 bond of this mannose is predicted to point towards the catalytic His, repulsion of the mannose may originate from the acceptor peptide instead. The inability of SacS to use ADP-Rib as donor despite low micromolar affinity is likely caused to the unusual architecture of this molecule with the nucleotide being attached to the nonreducing end of the sugar. As such the connecting carbon is comparatively more electron-rich and may be insufficiently electrophilic for the S_N2-type reaction of the GT. Interestingly, ADP-Glc could be shown to be a good sugar donor for SacS, binding with a superior dissociation constant than even UDP-Glc. To the best of my knowledge this makes SacS the first ADP-Glc preferential peptide glycosyltransferase described, with this particular sugar-nucleotide usually being used for energy storage via the formation of glycogen. Whether the observed affinities towards ADP-Glc as opposed to the more commonly described UDP-Glc translates towards an effect *in vivo* is as of yet uncertain, as no data about the abundance of ADP-Glc in *Actinomycetales* could be found. Given that in *E. coli* UDP-Glc and UDP-GlcNAc are two to three orders of magnitude more abundant than ADP-Glc[196], these affinities likely play only a minor role towards sugar specificity during the developed *in vivo* production method. Given the good affinity of SacS to ADP-Glc, ADP-Rib and GDP-Man, the purine base appears to be the determining factor for binding of a potential donor molecule. However, the carbohydrate moiety seems to determine if a productive complex with the acceptor peptide is formed, and a transfer takes place. A further class of ADP-linked sugars are ADP-Heptoses such as ADP-D-*glycero*-β-D-*manno*-heptose. So far these sugars have only been found in gram-negative bacteria and some *Streptomyces* strains[197]. Given that no such side-product was observed, when SacA was expressed in *E. coli,* as well as the structure of the SacS donor binding site

heavily indicating a preference for α-linked sugar-nucleotides, ADP-heptoses are unlikely to be used as donor substrates in SacS catalysed glycosylations.

Unfortunately, no thermodynamic data could be obtained *via* ITC for PltS, due to its inherent instability under experimental conditions. In functional assays it proved to be the least promiscuous of all investigated GGTs, only using UDP-Glc, ADP-Glc and UDP-GlcNAc as sugar donor. As the only observed modification of PltA during a glycosylation reaction *in vivo*, was a HexNAc, it seems likely that the natural substrate of PltS is UDP-GlcNAc. Though no data exists to rule out the possibility of GalNAc or ManNac modifications, the clear preference Glc over Man or Gal implies the importance of the position of O4 and N2 of the sugar. While the dense hydrogen bonding network seen in the PltS structure helps explain the general hexose preference, no definitive conclusion for the observed preference of GlcNAc over Glc can be drawn. The only potential hydrogen-bond donor for the acetyl-group is R194, which is too far away for such an interaction. Interestingly, in higher resolution structures of homologous proteins, such as SacS348 or ThuS, a water molecule in the Mn-coordination sphere can be seen located in between the Arg side chain and where a N-acetyl group would be located. Such a water molecule would be able to confer a hydrogen bond and improve the GlcNAc binding. A nearby Leu is within distance of the acetyl group to further assist by hydrophobic interaction. Given this residue is replaced by a Gln, Asn or Thr for EntS, SacS and SunS, respectively, who prefer Glc over GlcNAc in a decreasing order, whereas the R194 is highly conserved between all GGTs, the L192 seems much more likely to be the cause for the preferential activity towards UDP-GlcNAc.

With the exception of the EntS-Ecm cocrystal structures, it was, unfortunately not possible to obtain further complexes of a GGT with its acceptor peptide for comparative analysis. Such a problem has already been reported for SunS and ThuS and is likely caused by the low affinity and solubility of these peptides[168]. Furthermore, given that the position of the nucleophile within the sequence of the glycosylated loop can be shifted by several amino acids while not interfering with its glycosylation in SunS-like Gtases[74], it is likely that the peptide is not well ordered, even when bound to the active site. Correspondingly, the position of the peptide in the EntS protein is not amenable to shifts in the position of the nucleophilic amino acid within the peptide chain[45] and therefore forms a more stable complex. Considering the structural similarity, it was possible to predict the homologous regions, important for peptide binding in the other investigated GTs. My structural data of the helix binding pocket, formed by helices from both monomers, corresponds well to the importance of dimer formation for SacS, with the SacS348 truncated enzyme showing greatly reduced efficiency (see chapter 3.5.1). A similar behaviour has been reported for ThuS, which displays a marked drop in efficiency upon deletion of its dimerization domain. Interestingly, the truncated ThuS showed activity

comparable to that of the wild-type towards an acceptor peptide lacking the N-terminal helix, further corroborating the identity of the proposed N-terminal helix-binding pocket[168]. No data are available regarding the binding of the C-terminal part of the peptide. The Ecm minimal peptide lacks the C-terminal putative helix and it remains unclear whether this region of the peptide is directed "upwards" towards the backside of α1 and α1* of the dimerization domain, whether it would bind in an "trans" configuration, extending towards the N-terminal domain of the GT, or whether it does not interact with the enzyme in a meaningful way. Given the necessity to form disulphide bond and reducing the chance of intermolecular bond formation cis binding would appear more likely. Mutagenesis studies, targeting the putative binding site at α1 and α1* may shed light on that question. Similarly, the conformation of the remaining part of the N-terminal region of the Ent96 peptide is so far unknown. It may leave the enzyme either towards the C-terminus of the enzyme at dimerization helix α4 and α4* or may be located at the boundary of the dimerization and TPR-like domain. A final possibility is its binding along the dimerization domain in a helix-bundle-like fashion along the entire length of helix α1.

Several striking differences in the preference towards the acceptor peptide could be observed. The inability of EntS to glycosylate the Sublancin-like peptides SunAm and SacA, or the *Streptomyces* peptide PltA is not unexpected given the vast differences in the acceptor peptide sequence. EntS requires its acceptor peptide to contain an acidic amino acid in the -1 position[45], in stark contrast to the GCGG motif found in Sublancin-like peptides or the HGCVP putative glycosylation site of PltA. Surprisingly, while the loop-binding region of EntS is enriched in positively charged residues such as Arg and His, these appear to interact with the Asp in the +2 position more so than with the Asp in the -1 position. Indeed, the Asp16 side chain in the -1 position is within distance to form a hydrogen bridge with a nearby His only after the first glycosylation has occurred. The interaction may also be facilitated by a structural water, but no electron density that would correspond to such could be detected in the measured datasets. Interestingly, the C-terminal region was reported to have only a comparatively minor influence on EntS activity, which agrees with the structural data, considering the shift in peptide position upon the first glycosylation only influences the interactions of the C-terminal part of the peptide.

Despite their reported promiscuity, SunS-like GTs failed to glycosylate the Enterocin 96 peptide. In contrast, within the SunS-like GT-family, the peptides themselves appear to be largely interchangeable with both SunS and AciS showing activity towards SacA. This is well in agreement with reports showing the promiscuity of SunS in regard to changes in the peptide sequence[42, 74], which proposed that the only requirement for the glycosylation were an N-terminal helix and a Cys within a

disordered loop[74]. Interestingly, while no activity of SacS towards the SunA-minimal peptide SunAm could be observed, SacS proved capable of glycosylating the Cys-to-Ser mutant of SunAm2Ser, albeit at low efficiency. Given the slow speed of the observed glycosylation, aberrant disulphide formation and aggregation most likely outpaces the glycosylation reaction for the SunAm peptide containing the cysteines involved in disulphide formation in Sublancin. Surprisingly, however, neither SacS, nor SunS were capable of *O*-glycosylating SunAmallSer. While the inability of SunS to glycosylate such a mutant of SunA has been described[42], SacS is evidently capable of both *O*- and *S*-glycosylation of SacA and *S*-glycosylation of SunAm2Ser. Why this activity appears to be absent for the SunAmallSer-peptide is currently unknown, but it may be caused by an incorrect positioning of the nucleophile within the active site pocket.

Despite the previously mentioned reports towards the promiscuity of the SunS-like GTs, neither SacS, nor SunS showed any discernible activity towards PltA, despite its predicted capability to form an N-terminal α-helix and a disordered loop containing a GCV sequence. While such a sequence motif was not tested for SunS-activity, the enzyme was amenable to FCG and GCF sequences[74], suggesting sufficient steric flexibility to allow for a Val substitution. It seems more likely that either the His or Pro residue within the loop interfere with the positioning of the nucleotide within the active pocket. Alternatively, there may be undiscovered recognition motifs required for the binding of the N-terminal helix.

So far, the most tolerant GGT in respect of acceptor peptide specificity is PltS. Indeed, PltS appears to adopt a far more open conformation compared to SunS or EntS. The peptide channel appears to be laterally opened, forming an elongated, narrow groove instead. This openness of the PltS active site may be responsible for the astonishing promiscuity towards acceptor peptides, showing activity towards all peptides investigated. Such promiscuity has likewise been reported towards the SvGGT, with a minimal five amino acid long sequon Y(G/A/K/Q/E)(C/S/T)(G/A)G[51]. Whether the minimal peptide of PltS is equally short remains to be determined. Similarly, the ThuS structure appears more open than the SunS structure and likewise displays a greatly increased promiscuity and a shorter minimal sequence[168].

Based on the crystal structure, a reaction mechanism for EntS *O*-glycosylation involving a catalytic His-Asp or His-Asn dyad could be proposed. A His-Asp catalytic dyad has been well described for other inverting GTs, though it appears more frequent in the GT-B type fold than the GT-A fold[106]. A His-Asn dyad on the other hand has, to the best of my knowledge, not been described for GTs, though there are reports of such a dyad activating a nucleophile in proteases[181] and ribonucleases[182]. In line with other GT-A type folds, the leaving phosphate is stabilized by a divalent metal ion, coordinated to an Asp of the DXD motif and a His located at the C-terminus of the β7 strand. Interestingly, the phosphate is also stabilized by a Lys, Arg and Tyr residue, whose importance has, like the His-Asp dyad, been more commonly described for GT-B type folds rather than GT-A[106].

The observed complex with one glucose bound to the peptide allows to draw some conclusion of the likely configuration of the glycosidic bond of the second glycosylation, which resisted identification[45]. The angle of the first attached glucose within the active site, together with the location of the UDP-Glc in the nucleotide binding pocket, leaves only one plausible configuration of the second glycosidic linkage in the disaccharide of Ent 96 as β1-4 linkage.

The observed hydrolysis of the glycosidic bond in the presence of a simultaneously bound nucleotide is not unprecedented[125], though it is so far unclear, whether this reaction constitutes a full reverse reaction of carbohydrate transfer back on to the nucleotide handle, or whether the nucleotide is merely required for the binding of the peptide. Given the importance of the catalytic His residue for the hydrolysis reaction of the EntS-disaccharide, it appears likely that it follows a similar reaction mechanism as the glycosylation. This observed phenomenon leads to the question whether the observed monoglycosylated species of the Ecm and full-length peptides are caused by a dissociative mechanism of the EntS enzyme as described by Nagar&Rao[45], or whether it is rather caused by the observed hydrolysis of the glycosidic bond. This phenomenon may also point towards a biological function of the disulphides in the Ent 96 peptide, as their formation prevents the peptide from entering the GT and thereby protects the glycosidic bond.

The currently proposed model for GGT activity, however, still does not explain the activity of SacS and *S*-glycosylation in general. This preference towards *S*-glycosylation over *O*-glycosylation could be shown for SacS, with a much greater catalytic efficiency observed for the SacA4Ser peptide than the SacAallSer variant. The observed catalytic parameters for the *S*-glycosylation are well in line with that of other published Leloir GTs[118], though these are active primarily on small molecules as opposed to polypeptides. Comparing the kinetic constants with GTs active on peptides, such as OGT[198] or SunS[168]

reveals pronounced differences. Both k_{cat} and K_m of the SacS *S*-glycosylation were observed to be one order of magnitude higher than the parameters described for the *S*-glycosylation activity of SunS (k_{cat} of 0.88, K_m of 1.5 µM) or the *O*-glycosylation of OGT (k_{cat} of 0.07-0.9, K_m 7-8 µM). A possible confounding factor in this discrepancy is that the measurement of SacS kinetics was performed at 50°C, whereas SunS and OGT measurements were performed at room temperature or 37°C, respectively.

Interestingly, it could be observed that SacS remains active without the catalytic His-base, albeit with a much lower activity, with a similar phenomenon being reported for ThuS[168]. This may be caused by the higher acidity of cysteine (pKa 8.14)[199] compared to serine (pKa 13.6)[200], which, under the used reaction conditions, would cause a notable portion of cysteine residues to exist in a deprotonated state and would render the Cys-His interaction obsolete. Additionally, the Arg-phosphate hydrogen bond-based weakening of the sugar nucleotide bond may be sufficient to drive the reaction forward, with the catalytic His adopting only a supportive role as proton acceptor. A similar mechanism has been proposed for a transacylating enzyme with a catalytic Cys-His dyad, which retained its activity upon His-to-Ala substitution[201].

However, this does not explain the difference in the observed K_m values for *O*- and *S*-glycosylation. At the same time, the laid-out principles should lead to better *S*- than *O*-glycosylation for EntS, which is evidently not the case. Likewise, this model does not fully explain the residual *O*-glycosylation activity of SacS(H200A) either. As of yet, it is still unclear to which extent the respective glycosylation reactions are slowed down. Considering that *O*-glycosylation was already much slower than the *S*-glycosylation reaction, it could as well be envisioned that the His-mutant disrupts both reactions equally. In the case of SacS(H200A) it is possible that the nucleophile is positioned is such a way that the nearby Glu can perform the deprotonation instead, yet this amino acid is conserved between all investigated GGTs and should therefore also show this activity in the EntS(H214A) mutant, which it does not. Alternatively, a water chain, or the α-phosphate may serve as catalytic base as proposed for OGT[120], a GT-B type glycosyltransferase, lacking an obvious catalytic base. Further investigations are required to determine the exact molecular origin of the residual O-glycosylation activity of SacS(H200A).

5. Conclusion and Outlook

In this work glycocins, glycosylated antimicrobial peptides, a class of ribosomally synthesised and postranslationally modified peptides, have been investigated in regard to their synthesis, as well as the structure and function of the essential glycosyltransferase. To that end, genome mining was used to identify several potential glycocin producer strains. These peptides were successfully synthesized using Fmoc-based solid phase peptide synthesis, an approach that could be optimized to obtain good yields for most peptides investigated. A major issue with these peptides is their limited solubility in aqueous media. This issue could be alleviated by preparing the HCl-salts. Additionally, I have developed a broadly applicable method for the recombinant production of glycocins in *E. coli*. An N-terminal MBP-tag was used to confer the necessary solubility to the putative mature glycocin, while the His-tagged cognate glycosyltransferase on a Duet vector for glycosylated the MBP-tagged aglycone. With this method all peptides could be produced in their mature, glycosylated form. The newly discovered glycosyltransferases AciS, SacS and PltS as well as the previously known glycosyltransferases SunS and EntS were investigated with regards to their activity. It could be determined that AciS, SacS, EntS and SunS preferentially transfer a glucose unit, whereas PltS transfers an N-acetyl glucosamine instead. To investigate their structure and function, diffraction quality crystals could be obtained for PltS, SunS, EntS and a truncated form of SacS. The obtained crystal structures afforded insights into the rationale behind the observed donor specificities. For the first time I could obtain a structure of a GGT, EntS, with the acceptor peptide, providing several snapshots along the reaction pathway of the glycosylation reaction, providing unprecedented insight into the catalytic mechanism for the EntS *S*/*O*-glycosylation. The identity of the nearby His as catalytic residue could be confirmed. In particular its critical role in *O*-glycosylation was shown. In contrast, the *S*-glycosyltransferase SacS(H200A) still showed *S*-glycosylation activity without this vital His, showing that the reaction mechanism of SacS and EntS is noticeably different.

In order to help determine a detailed structure-function relationship of *S*-glycosyltransferases, higher quality crystals, especially of cocrystals in complex with their acceptor peptide would be of considerable use, helping to determine the positioning of the S^- nucleophile. To that end PltS appears to be a promising candidate, crystallizing in a reproducible manner in the presence of the PltA2Ser peptide. Use of non-cleavable sugar-donor analogues could help stabilise the pre-reaction donor-acceptor complex sufficiently to probe these in structural studies. Another approach is cryo electron microscopy, which may help investigate states of the protein that are difficult to crystallize. Furthermore, both the already obtained structural data, as well as the data yet to be obtained from

co-crystals of *S*-glycosyltransferases with their cognate acceptor-peptides will have to be experimentally validated using corresponding mutants. Obtaining the required peptides to this end may benefit from further optimization of the peptide synthesis. Similarly, the *in vivo* method may benefit from further optimization such as determining the rationale behind the observed MBP-glycocin truncation. Expansion of this method towards Enterocin-like and GccF-like peptides would render this method even more useful as a general and scalable method for the production of glycocins. Finally, given the obtained structural and functional data, it should be possible to produce glycocin-peptides containing a reactive handle, either by genetic code expansion, by inclusion of an unnatural amino acid during the solid phase synthesis or by using an unnatural UDP-sugar analogue during *in vitro* or *in vivo* glycosylation, which would be of considerable use to help deduce the function, localization and targets of the respective glycocins.

6. Materials and Methods

6.1. Reagents

All reagents were obtained in the highest quality possible unless otherwise specified. Salts, buffers and polymers were obtained from Carl Roth GmbH, Sigma-Aldrich (Merck Millipore) or ThermoScientific. Likewise, components of media were obtained from Carl Roth GmbH or Merck Millipore unless otherwise specified. DMF and Acetonitrile were obtained from Thermo fisher scientific. Ultrapure water was generated using a Milli-Q-8 Direct system (Merck Millipore).

6.2. Growth media

Agar plates were made by mixing the respective media with 1.5 % (w/v) of Agar prior to autoclaving. For GYM Streptomyces agar an additional 0.2 % (w/v) of $CaCO_3$ was added.

Table 9 *Media used in this study.*

Name	***Composition / Source***
Lysogenic Broth (LB Miller)	Carl Roth
Terrific Broth (TB)	1.2 % Tryptone from casein 2.4 % Yeast extract 0.4 % Glycerol 10 % TB phosphate buffer
Super optimal broth (SOB)	2 % Tryptone from casein 0.5 % Yeast extract 8.5 mM NaCl 2.5 mM KCl 1 mM $MgCl_2$
SOC	SOB supplemented with 20 mM Glucose
Autoinduction media	1 % Tryptone from casein 0.5 % Yeast extract 10 % M9 salts 2 % Autoinduction sugars 0.15 % Trace elements 1 mM $MgSO_4$
M9 minimal media	10 % M9 Salts 0.4 % Glucose 0.01 % Biotin 0.01 % Thiamine 2 mM $MgSO_4$ 100 µM $CaCl_2$
GYM Streptomyces media	0.4 % Glucose 0.4 % Yeast extract (BD Biosciences) 1 % Malt extract (BD Biosciences) Adjusted to pH 7.2 with NaOH
BHI Media	BD Biosciences

6.3. Buffers

Table 10 *Composition of buffers used in this study.*

Name	***Composition***
SDS loading dye	0.25 M Tris pH 6.8 5 % (v/v) beta-Mercaptoethanol 50 % (v/v) Glycerol 10 % (w/v) SDS
SDS running buffer	25 mM Tris 192 mM Glycine 0.1 % SDS
SDS-PAGE stain	70 µM Coomassie Brilliant Blue G250 0.13% HCl
TAE buffer	100 mM Tris base 18 mM acetic acid 1 mM EDTA
TB phosphate buffer	0.17 M KH_2PO_4, 0.72 M K_2HPO_4
M9 Salts	477 mM Na_2HPO_4, 220 mM KH_2PO_4, 200 mM NH_4Cl, 100 mM NaCl
Trace elements (1000x)	134 mM EDTA, 31 mM $FeCl_3$, 6.2 mM $ZnCl_2$, 1.6 mM H_3BO_3, 760 µM $CuSO_4$, 810 µM $MnCl_2$, 420 µM $CoCl_2$, 100 µM $NiCl_2$, 100 µM Na_2MoO_4
Autoinduction sugars	5% (w/v) Galactose 5% (w/v) Lactose 2.5% (w/v) Glucose 25% (v/v) Glycerol
Inoue transformation buffer	10 mM PIPES pH 6.7 55 mM $MnCl_2$ 15 mM $CaCl_2$ 250 mM KCl
HisTrap Wash buffer	50 mM Tris pH 8.0 500 mM NaCl 10 mM Imidazole 0.5 mM DTT
HisTrap Wash buffer for AciS	50 mM Tris pH 7.5 500 mM NaCl 10 mM Imidazole 10% (v/v) Glycerol
HisTrap Elution buffer	50 mM Tris pH 8.0 500 mM NaCl 500 mM Imidazole 0.5 mM DTT
HisTrap Elution buffer for AciS	50 mM Tris pH 7.5 500 mM NaCl 500 mM Imidazole 10% (v/v) Glycerol
StrepTrap Wash buffer	50 mM Tris pH 7.5 150 mM NaCl 0.5 mM DTT
StrepTrap Elution buffer	50 mM Tris pH 7.5

	150 mM NaCl 2.5 mM Desthiobiotin 0.5 mM DTT
MBPtrap wash buffer	50 mM Tris pH 8 250 mM NaCl
MBPtrap elution buffer	50 mM Tris pH 8 250 mM NaCl 10 mM Maltose
TBS	50 mM Tris pH 7.5 150 mM NaCl
TBST	TBS 0.1 % Tween 20
IEX wash buffer	20 mM Tris pH 8
IEX elution buffer	20 mM Tris pH 8 1 M NaCl
Volatile buffer	20 mM NH_4HCO_3 pH 8.0 (adjusted with NH_4OH)
Glycocin lysis buffer	50 mM Tris pH 8 250 mM NaCl 5 mM DTT 0.1% (v/v) Triton X-100
Refolding buffer	20 mM Tris pH 8.5 6 M Gdn*HCl 10 mM DTT
SacS buffer	20 mM Tris pH 8 150 mM NaCl
PltS buffer	20 mM Bicine pH 8 150 mM KCl 2 mM DTT
AciS buffer	20 mM Tris pH 7.5 150 mM NaCl 5% Glycerol
EntS buffer	20 mM Tris pH 7.5 150 mM NaCl
SunS buffer	20 mM HEPES pH 7.5 100 mM NaCl 10% Glycerol
Protease buffer	25 mM Tris pH 7.5 150 mM NaCl
MMT buffer (1M)	200 mM DL-malic acid 400 mM MES 400 mM Tris pH adjusted with NaOH to 8.0

6.4. Antibiotics

Table 11 Antibiotics used in this study.

Antibiotic	*Concentration in media*	*Source*
Kanamycin sulphate	50 µg/ml	Serva
Chloramphenicol	34 µg/ml	Carl Roth
Streptomycin sulphate	50 µg/ml	Sigma Aldrich

6.5. Primers

Primers were obtained from Integrated DNA technologies. A list of primers used can be found in Table 12. The nomenclature of extension primers follows Gene_Overhang_direction, whereas vector-linearization primers are named Veclin_ORFname_direction.

Table 12 Primers used in this study.

Primer Name	*Sequence (5' – 3')*
	Vector linearization
Veclin_Duet1_fwd	TAATGCTTAAGTCGAACAGAAAG
Veclin_Duet1_rev	CATGGTATATCTCCTTATTAAAG
Veclin_Duet2_fwd	AGAAACCGCTGCTGC
Veclin_Duet2_rev	CATATGTATATCTCCTTCTTATACTTAAC
Veclin_3C_fwd	CGCGCCTTCTCCTCACATATG
Veclin_3C_rev	TGCTGGTCCCTGGAACAGAAC
Veclin_HisCterm_fwd	CTCGAGCACCACCAC
Veclin_HisCterm_rev	CATGGTATATCTCCTTCTTAAAG
Veclin_MBP_fwd	GGTACCGGATCCGAATTC
Veclin_MBP_rev	GGCGCCCTGAAAATAAAG
Veclin_Sumo_fwd	TAGGTATTTATTCGGCG
Veclin_Sumo_rev	TCCGCTACCACCAATCTGTTCTC
	Extension primers
MBP_duet1_fwd	TTAATAAGGAGATATACCATGAAAATCGAAGAAGGTAAAC
MBP(TEV)_rev	GCCCTGAAAATAAAGATTCTC
SacA_TEV_fwd	GAATCTTTATTTTCAGGGCTTTACTGCAGCTC
SacA_duet1_rev	CTGTTCGACTTAAGCA TTATTTGCAATATTGATCATAGAG
Ent96_TEV_fwd	GAATCTTTATTTTCAGAGCAAACGTGATTGTAAC
Ent96_duet1_rev	CTGTTCGACTTAAGCA TTATTTACAATATTTTCTTAC
AciA_TEV_fwd	GAATCTTTATTTTCAGGGAAATGGAGCAC
AciA_duet1_rev	CTGTTCGACTTAAGCATTAATAGCAGTAACTTC
SunA_TEV_fwd	GAATCTTTATTTTCAGGGATTAGGAAAAGCTC
SunA_duet1_rev	CTGTTCGACTTAAGCATTATCTGCAGAATTGAC
PltA_TEV_fwd	GAATCTTTATTTTCAGGGCATGAGCAAGGC
PltA_duet1_rev	CTGTTCGACTTAAGCATTAGAGCTTCGGCCCC
Gtase_duet2_fwd	GTATAAGAAGGAGATATACAT ATGGGCAGCAGCC
SacS_duet2_rev	GCAGCAGCGGTTTCTGGCGCGTTATAAATCAATC

AciS_duet2_rev	GCAGCAGCGGTTTCTTTATAAAACCTCGTTTAG
SunS_duet2_rev	GCAGCAGCGGTTTCTTCATACTTCAATTCCTTTC
EntS_duet2_rev	GCAGCAGCGGTTTCTTCATGATTCATTCTCCC
PltS_duet2_rev	GCAGCAGCGGTTTCTTCACGGAGCTGGGG
PltS_duet2_fwd	GTATAAGAAGGAGATATACAT ATGGGCAGCAGCTG
SacS_3CLIC_fwd	CCAGGGACCAGCAATGAAAAGCATAGGTAGTC
SacS_3CLIC_rev	GAGGAGAAGGCGCGTTATAAATCAATCTTTTTTATTTC
PltS_3CLIC_fwd	CCAGGGACCAGCAATGAAAGCCGCCACGAG
PltS_3CLIC_rev	GAGGAGAAGGCGCGTCACGGAGCTGGGGAC
AciS_3CLIC_fwd	CCAGGGACCAGCAATGGGGGACTTAAAGAATC
AciS_3CLIC_rev	GAGGAGAAGGCGCGTTATAAAACCTCGTTTAGTTTTTC
AciS_HisCterm_fwd	AGAAGGAGATATACCATGGGGGACTTAAAGAATC
AciS_HisCterm_rev	GTGGTGGTGCTCGAGTAAAACCTCGTTTAGTTTTTC
EntS_3CLIC_fwd	CCAGGGACCAGCAATGTATTCTGAAAATTTTATTGCT
EntS_3CLIC_rev	GAGGAGAAGGCGCGTCATGATTCATTCTCCCG
GccA_3CLIC_fwd	CCAGGGACCAGCAATGAAAAACCGTCAGAACGAGATCGACAG
GccA_3CLIC_rev	GAGGAGAAGGCGCGTTAGTTGTTTTCGTTCTTGCTATCGCCTTGC
GccA_MBP_fwd	TATTTTCAGGGCGCCAAAAACCGTCAGAACG
GccA_MBP_rev	CGGATCCGGTACCTTAGTTGTTTTCGTTCTTG
SunS_SUMO_fwd	GATTGGTGGTAGCGGA ATGAAACTGAGTGATATTTATTTG
SunS_SUMO_rev	CGCCGAATAAATACCTATACTTCAATTCCTTTCAG
StrepPltS_Sumo_fwd	GATTGGTGGTAGCGGATGGGCAGCAGCTGG
StrepPltS_Sumo_rev	CCTAAGGTCTTTA TCACGGAGCTGGGGAC
	Mutagenesis and truncation
StrepTag_fwd	TCGAACTGCGGGTGGCTCCAGGCTGCTGCCCATGG
StrepTag_rev	AGCCACCCGCAGTTCGAAAAAAGCAGCGGCCTGG
3C-His_fwd	AGGTTCTGTTCCAGGGACCA CACCACCACCACCACCACTG
3C-AciS_rev	GTCCCTGGAACAGAAC CTCGAGTAAAACCTCG
EntS_H214A_fwd	TGGTCTAATAGCGGAAGAGTTGAGGAACAAC
EntS_H214A_rev	CTTCCGCTATTAGACCATAATACTTAATTTTATTATTG
SacS_H200A_fwd	CGGTAAGGTTGCGGAAGAACCTCTTTTTCC
SacS_H200A_rev	TTCCGCAACCTTACCGTAAAACTTAATATC
SacS_348noSt_rev	GAGGAGAAGGCGCGGGGGAACTCTTTTTCC
SacS_348_rev	GAGGAGAAGGCGCGTTAGGGGAACTCTTTTTCC

6.6. Bacterial strains

Table 13 *Bacterial strains used in this study.*

Name	Relevant Genotype / Strain	Source
E. coli DH5α	*fhuA2 Δ(argF-lacZ) U169 phoA glnV44 ϕ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
E. coli BL21 (DE3) Gold	F^- *ompT hsdS*($r_B^-m_B^-$) dcm^+ Tet^r *gal λ(DE3) endA Hte*	Fisher Scientific
E. coli BL21 (DE3) Rosetta	F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R)	Merck Millipore
E. coli BL21 (DE3) Arctic Xpress	F^- *ompT hsdS*($r_B^-m_B^-$) dcm^+ Tet^r *gal* λ(DE3) *endA* Hte (*cpn10 cpn60* $Gent^R$)	Agilent
Shouchella plakortidis	DSM 40041	DSMZ
Acinetobacter baumanii	K3-ATCC 17978	ATCC
Bacillus amyloliquefaciens	DSM 7	DSMZ
Bacillus subtilis	DSM 10	DSMZ
Laceyella putidus	DSM 44608	DSMZ
Listeria monocytogenes	DSM 20600	DSMZ
Salmonella enterica serovar Typhimurium	DSM 554	DSMZ
Staphylococcus aureus subsp. Rosenbach	ATCC25923	ATCC
Streptomyces platensis	DSM 40041	DSMZ
Streptococcus suis	DSM 9682	DSMZ

6.7. Genome mining

Amino acid sequences of all confirmed glycocins and their cognate GGTs were used as a query for a BLASTp search[202], using the NCBI webserver with the corresponding interface (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). Only results with an e-value below $1e^{-9}$ were used for further analysis. Results with a large genomic distance between the putative glycocin and glycosyltransferase were excluded. Only organisms in which an ABC-type transporter clustered with the putative glycocin and GT were used for further analysis. MEGA-X[203] was used to infer a maximum likelihood phylogenetic tree.

6.8. Molecular biology methods

6.8.1. Isolation of Genomic DNA

Genomic DNA of *G. acidiceleris, E. faecalis and L. sacchari* was obtained from the DSMZ, Germany.

Freeze dried bacteria of *S. platensis and B. subtilis* were obtained from the DSMZ. These bacteria were reconstituted and grown in GYM Streptomyces and LB media, respectively. Their genomic DNA was isolated using a Bacteria DNA Preparation kit (Jena Bioscience), according to the manufacturer's instructions.

6.8.2. Polymerase chain reaction

PCR was performed with a C1000 Thermal cycler (Bio-Rad) using Q5 Polymerase (NEB) to prepare DNA fragments for cloning. DreamTaq Polymerase (Fisher Scientific) was used instead for colony PCR and bacteria of a single colony were transferred into this reaction mixture as a template. An example for a PCR mixture composition can be found in Table 14. The GC-enhancer was added for reactions in which the desired product contained a GC-content >60%.

Table 14 *Example of a PCR mixture composition*

Master mix	2 µl
GC-enhancer	2 µl
8 mM dNTP mix (2 mM each)	1 µl
Template	0-3 µl (5-60 ng of DNA)
Primer (10 µM each)	1µl
Polymerase	0.1 µl
Water	To 10 µl

Temperatures and elongation times were chosen according to the manufacturer's recommendations. Typically, PCR was performed with 30 cycles. Vector linearization PCR for ligation was performed with 40 cycles instead.

6.8.3. Purification of DNA

After vector linearization PCR the remaining parent plasmid was first digested by addition of DpnI (NEB) to the crude PCR mixture, according to the manufacturer's instructions. The incubation time was extended to one hour to ensure full digestion of the parent DNA.

All DNA fragments have been additionally purified by agarose gel electrophoresis in a 2% Agarose gel for PCR products below 500 bp, 0.7% for products above 4,000 bp, and 1% for all other DNA lengths. DNA was visualised by addition of 0.01% (v/v) SYBR safe DNA stain to the gel (Fisher Scientific).

Bands in the gel, corresponding to the expected size, were excised and purified using either a GenElute™ Gel extraction kit (Sigma-Aldrich) or NucleoSpin™ Gel extraction kit (Macherey Nagel) according to the manufacturer's instructions.

6.8.4. Ligation

Ligation of amplified DNA was performed using the HiFi-Assembly kit (NEB) according to the manufacturer's instructions. A list of Plasmids prepared this way can be found in Table 15. Ligation products used for transformation of ultracompetent *E. coli* DH5α without prior purification (chapter 6.8.5).

Table 15 *Plasmids created in this study.*

Name	***Gene***	***Antibiotic resistance***	***Parent Plasmid***
pET-3CLIC-SacS	His-3C-SacS	Kanamycin	pET-3CLIC[a]
pET-3CLIC-EntS	His-3C-EntS	Kanamycin	pET-3CLIC
pET-3CLIC-SunS	His-3C-SunS	Kanamycin	pET-3CLIC
pET-3CLIC-PltS	His-3C-PltS	Kanamycin	pET-3CLIC
pET3CLIC-GccA	His-3C-GccA	Kanamycin	pET3CLIC
pET-YSBL-AciS	His-SacS	Kanamycin	pET-YSBL[b]
pET-MBP-GccA	MBP-TEV-GccA	Kanamycin	pET-MBP1b[c]
pET-Strep3CLIC-PltS	Strep-3C-PltS	Kanamycin	pET-3CLIC-PltS
pChampion-SUMO-SunS	His-SUMO-SunS	Kanamycin	Champion pET-SUMO[d]
pChampion-SUMO-PltS	His-SUMO-Strep-3C-PltS	Kanamycin	Champion pET-SUMO
pET-YSBL-AciS-His	AciS-His	Kanamycin	pET-YSBL
pET-YSBL-AciS-3CHis	AciS-3C-His	Kanamycin	pET-YSBL-AciS-His
pET-3CLIC-SacS348noSt	His-3C-SacS348-noStop	Kanamycin	pET-3CLIC-SacS
pET-3CLIC-SacS348	His-3C-SacS348	Kanamycin	pET-3CLIC-SacS
pET-3CLIC-SacSA-fused	His-3C-SacS-GSSG-SacA	Kanamycin	pET-3CLIC-SacS

pGlycocin-SacA-SacS	ORF1: MBP-TEV-SacA ORF2: His-3C-SacS	Streptomycin	pCDFduet-1[e]
pGlycocin-SunA-SunS	ORF1: MBP-TEV-SunA ORF2: His-SUMO-SunS	Streptomycin	pCDFduet-1
pGlycocin-AciA-AciS	ORF1: MBP-TEV-AciA ORF2: His-AciS	Streptomycin	pCDFduet-1
pGlycocin-PltA-PltS	ORF1: MBP-TEV-PltA ORF2: His-3C-PltS	Streptomycin	pCDFduet-1
pGlycocin-EntA-EntS	ORF1: MBP-TEV-Ent96 ORF2: His-3C-EntS	Streptomycin	pCDFduet-1

[a] a gift from YSBL[204] , [b] a gift from Markus Wahl (FU Berlin), [c]a gift from Nediljko Budisa (TU Berlin), [d]Thermofisher Scientific, [e] Merck Millipore

6.8.5. Transformation

Ultracompetent *E.coli* DH5α were prepared according to the Inoue method[205]. Briefly, cells were grown at 18°C in SOB media until an OD_{600} of 0.55 had been reached. The cell suspension was subsequently incubated on ice for 10 min. In the following steps the cells were kept on ice and centrifuged at 4°C. Cells were harvested by centrifugation at 2,500 x g and washed twice with ice-cold Inoue transformation buffer (see Table 10). The cells were suspended in the same buffer, and 0.75 % (v/v) of DMSO was added as cryoprotectant. 50 µl aliquots were prepared, snap-frozen in liquid nitrogen and stored at -80°C until further use.

Competent *E. coli* of expression strains were prepared by growth in LB media at 37°C until an OD_{600} of 0.45. The cells were harvested, washed once an ice-cold solution of each 100 mM $MgCl_2$ and 100 mM $CaCl_2$. Finally, they were suspended in the 100 mM $CaCl_2$ solution and 15% (v/v) of glycerol was added as a cryoprotectant. 50 µl aliquots were prepared, snap-frozen in liquid nitrogen and stored at -80°C until further use.

Transformation of ultracompetent *E. coli* DH5α with ligation products was performed according to the Inoue protocol[205]. Briefly the ligation mixture was added to the aliquot of competent cells in a 1:10 ratio, and the suspension mixed by gentle agitation. The mixture was incubated on ice for 30 min, followed by a heat shock at 42°C for 45 seconds. Cells were immediately plunged into an ice bath afterwards and incubated for 3 min. Subsequently 1 ml of SOC media was added, and cells incubated shaking at 37°C for 1h. Cells were transferred to LB agar plates containing the appropriate antibiotics. Colonies usually appeared after growth at 37°C for 12h. *E.coli* DH5α cells transformed with pGlycocin plasmids were usually grown for 30 hours until colonies were visible.

Single colonies of ligation products were picked and the presence of the gene of interest was confirmed using a colony (see chapter 6.8.2).

Heat shock competent *E. coli* expression strains were similarly transformed using 20-100 ng of purified plasmid with confirmed sequence.

6.8.6. Plasmid purification

Plasmids were isolated from *E. coli* DH5α competent cells using a HP GenElute plasmid miniprep kit (Sigma-Aldrich) according to the manufacturer's protocol. The sequence identity and fidelity was verified via Sanger-Sequencing (Microsynth, Germany).

6.8.7. Site Directed mutagenesis.

Site directed mutagenesis was performed using a standard QuikChange site directed mutagenesis, by using two overlapping primers harbouring the mutation in a PCR, to afford a circular DNA product. Following DpnI digest and purification, the thus created plasmids were introduced into *E. coli* DH5α as described previously.

6.9. Biochemical methods

6.9.1. Determination of protein concentration

Protein and peptide concentration was determined using a DS-11 FX+ spectrophotometer/fluorometer (DeNovix). Absorption was measured at 280 nm, and concentration calculated by applying Lambert-Beer's law:

$$A_\lambda = \log\left(\frac{I_0}{I}\right) = \varepsilon_\lambda * c * d$$

Where I_0 is the initial light intensity, I is the measured light intensity, ε_λ is the extinction coefficient at wavelength λ, *c* the concentration and *d* the pathlength of light.

The Expasy-ProtParam[206] webserver (https://web.expasy.org/protparam/) was used to calculate the extinction coefficient of the proteins or peptides. Cysteines were assumed to be reduced for this calculation, except for glycosylated wild-type peptides of Ent96, SacA, AciA and PltA.

6.9.2. SDS-PAGE

SDS-page was performed using standard methods for denaturing gel electrophoresis according to Laemmli (1970)[207] using a Tris-Glycine-SDS buffer. Gels were prepared as a discontiguous stack of a 4 % acrylamide stacking gel atop a 12% separating gel. The recipe for each can be found below:

Table 16 *Recipe for SDS-Polyacrylamide gels*

	Stacking gel (4%)	Separating gel (12%)
Water	7.25 ml	8.4 ml
1 M Tris pH 6.8	1.3 ml	-
1.5 M Tris pH 8.8	-	3.4 ml
30% Acrylamide/Bisacrylamide (37.5:1)	1.3 ml	8 ml
10 % (w/v) SDS	100 µl	200 µl
10 % (w/v) Ammoniumperoxodisulphate	100 µl	200 µl
TEMED	10 µl	20 µl

Gels were stained using the method of Lawrence *et al.* (2009)[208]: Gels were washed with hot water to remove SDS and subsequently submerged in SDS-PAGE staining solution, heated to a boil and incubated for 5-10 minutes at RT. Afterwards destaining was performed by submerging the gels in a 10% EtOH solution for 20 minutes.

6.9.3. Western blotting

Western blotting was performed using the Trans-Blot Turbo™ transfer system (Bio-Rad), transferring the proteins to a Trans-Blot Turbo Mini 0.2 µm PVDF membrane (Bio-Rad), according to the manufacturer's instructions. The membrane was subsequently blocked by incubation in 5% milk in TBS buffer for 1 hour. The membrane was washed thrice with TBST-buffer and submerged in a solution containing mouse α-His antibody (Invitrogen) in a 1:5,000 dilution in a 5 % bovine serum albumin (BSA)

solution of TBS and incubated overnight. The membrane was washed thrice with TBST and incubated for 1 hour with goat α-Mouse-HRP antibody (Dianova) in a 1:3,000 ratio in 5% BSA in TBS. After a final wash with TBST, the membrane was treated with ECL™ western blotting detection reagents (cytiva) according to the manufacturer's instructions. Luminescence was detected using an ImageQuant™ LAS-4000 mini luminescent image analyzer (Fujifilm).

6.9.4. Isothermal titration calorimetry

ITC experiments were performed on a MicroCal ITC (Malvern) or an AffinityITC (TA instruments) at 25°C unless specified otherwise. Stirring was performed at 550 or 125 rpm, respectively. Water, containing 8 mM sodium azide was used as reference.

Proteins were loaded into the cell in concentrations, ranging from 10 to 60 µM. Nucleotides and nucleotide sugars were injected in concentrations of 300 uM to 2 mM. Typically, a first injection of 0.4 µl was made, followed by 18 injections of 1.5-2 µl titrant. All experiments were performed in triplicates.

Data were analysed using NanoAnalyze (TA Instruments). The first injection was excluded from analysis. Stoichiometry of binding for ligands with a projected Kd > 10 uM were manually set to 1 before fitting the data.

6.9.5. Differential scanning fluorimetry

Protein melt curves were measured using a QuantStudio 5 real-time PCR machine (applied biosystems). The protein of interest at a concentration of 0.1 mg/ml unless otherwise specified with a 5x concentration of SYPRO Orange Protein gel stain (Sigma-Aldrich) and the buffers at a concentration of 100 mM. Additives were added at concentrations ranging from 5 mM to 300 mM. The temperature was increased over the course of the measurement from 25°C to 98°C at a rate of 0.016°C/s, measuring the fluorescence with λ_{ex} = 470 nm and λ_{em} = 520 nm. Data was analysed using the Protein Thermal shift software v1.3 (applied biosciences).

6.10. Lyophilization

Samples were first snap-frozen by plunging them into liquid nitrogen, before subjecting them to lyophilisation.

Resins were dried using an ALPHA 1-2 LD freeze dryer (Christ, Gefriertrocknungsanlagen GmbH), typically at 0.3-1 mbar. All other samples were dried using an ALPHA 2-4 LDplus (Christ, Gefriertrocknungsanlagen GmbH), typically at 0.04 mbar.

6.11. Production and purification of Glycosyltransferases

6.11.1. Protein production

LB, containing the appropriate antibiotics, was inoculated with a bacterial culture, harbouring the expression plasmid of interest (Table 15). After overnight growth at 37°C, expression media containing the appropriate antibiotics was inoculated with this preculture in a 1:100 ratio.
The Expression media of choice was generally TB-medium (Table 9) unless otherwise specified. Cells were allowed to grow at 37°C, 250 rpm to an OD_{600} of 2.0, after which gene expression was induced by addition of 0.1 mM Isopropyl-β -D-thiogalactopyranosid (IPTG). The culture was further incubated at 16°C, 250 rpm for another 18 h. Cells were subsequently harvested by centrifugation at 6,000 x *g* for 15 min at 4°C. The cells were either used directly or stored at -20°C until further use.

6.11.1.1. SeMet containing proteins.

For the production of SeMet containing proteins, expression took place in M9 minimal media, according to the protocol of Walden (2010)[209]. Similar to the normal expression described above, LB was inoculated with bacteria harbouring a plasmid encoding for the gene of interest. After growth over night, bacteria were collected by centrifugation at 2000 x g and the pellet washed once with M9 minimal media. M9 minimal media (Table 9) supplemented with 5 g/l glucose was inoculated with this pellet and grown until OD_{600} of 0.6 was reached. At this point 100 mg phenylalanine, 100 mg lysine, 100 mg threonine, 50 mg leucine, 50 mg isoleucine and 50 mg of valine were added for each litre of

culture to reduce the activity of the Methionine biosynthesis of E. coli. 100 mg/l of $_{DL}$-Selenomethionine were added and the culture incubated for 15 min. Induction of expression with IPTG, harvesting of cells and purification of the protein was performed as described for wild-type proteins.

6.11.2. Lysis of cells

The cell pellet was resuspended in an appropriate lysis buffer (Table 10) and lysed using an Emulsiflex C-3 (Avestin) homogenizer at 15000 psi in the presence of 3 mM $MgCl_2$ and catalytic amounts of DNase I. Cell debris was removed via centrifugation at 45,000 x *g*, 4°C for 30 min.

6.11.3. Affinity chromatography

6.11.3.1. HisTrap

The supernatant, obtained after cell lysis, was loaded onto a HisTrap™ FF column (GE Healthcare), washed with 5 column volumes (CV) of wash buffer and eluted with a gradient of Imidazole from 10 to 500 mM over 5 CV. The flowthrough of the loading step, the wash as well as fractions over the whole gradient were collected.

6.11.3.2. StrepTrap

The supernatant, obtained after cell lysis, was loaded onto a StrepTrap™ FF column (GE Healthcare), washed with 10 CV of wash buffer and eluted with a gradient of desthiobiotin from 0 to 2.5 mM over 5 CV. The flowthrough of the loading step, the wash as well as fractions over the whole gradient were collected.

6.11.3.3. MBPTrap

The supernatant, obtained after cell lysis, was loaded onto an MBPTrap™ HP column (GE Healthcare), washed with 5 CV of wash buffer and eluted with a gradient of maltose from 0 to 10 mM over 5 CV. Fractions with an absorbance above 200 mAU at 280 nm were collected. The flowthrough was re-injected into the column up to 13 times for a complete capture of MBP-tagged protein.

6.11.4. Size exclusion chromatography

Size exclusion chromatography was performed either using a HiLoad 16/600 Superdex 200pg or a HiLoad 16/600 Superdex 75pg column (Cytiva) for molecules above or below a 30 kDa size, respectively. Injection volumes were kept between 1-2% of the bed volume of the column. Elution was performed in an isocratic manner over 1.3 CV unless specified otherwise. Absorbance was measured at 260, 280 and 220 nm. Fractions were collected between 0.4-1.3 CV.

6.11.5. Ion exchange chromatography

A MonoQ 4.6/100 PE column (Cytiva) equilibrated with IEX wash buffer was charged with the protein solution to be purified via IEX, washed with 5 CV of wash buffer retained proteins were eluted with a linear NaCl gradient from 100 mM to 1 M over 12 CV. The flowthrough of the loading step, the wash as well as fractions over the whole gradient were collected.

6.11.6. Proteases

In this study 3 proteases were used for tag-removal of recombinantly produced polypeptides:

Tobacco etch virus protease (TEV), Human rhinovirus 3C protease (3C) and Ubiquitin-like-specific protease 1 (Ulp1).

TEV[210] and Ulp1[211] were produced as a His-tagged construct in *E. coli* BL21(DE3) Gold and from a pET28a(+) (a kind gift from Nedjiliko Budisa, TU Berlin) or pHYRS52A vector (a gift from the YSBL), respectively. They were purified using His-Trap followed by SEC using their respective buffers (Table 10).
His-MBP-tagged 3C protease was expressed and purified from a pMal-2X plasmid (a kind gift from the YSBL) in a similar manner by Christian Roth, MPIKG.

6.11.7. Purification of GccA

GccA was expressed as an MBP-TEV-GT construct in a pET-MBP1b plasmid in *E. coli* BL21 (DE3) Gold. The protein of interest was isolated from the cell lysate by MBPTrap as described above (6.11.3.3). Fractions containing the protein of interest were combined and dialyzed against GccA-cleavage buffer (Table 10). The resulting precipitate was collected and re-dissolved in 6 M guanidinium hydrochloride (Gdn). Refolding of GccA was performed by rapid dilution into GccA refolding buffer (Table 10) and residual Gdn was removed by injecting the supernatant onto a HiPrep™ 26/10 desalting column (Cytiva) using GccA refolding buffer as mobile phase. The resulting, diluted solution of GccA was concentrated using a centrifugal concentrator as described previously.

6.11.8. Purification of SacS

SacS was expressed as His-3C-GT constructs in a pET-3CLIC plasmid in *E. coli* BL21 (DE3) Gold. SacS was isolated from the cell lysate via HisTrap as described above (6.11.3.1). To obtain the His-SacS protein, fractions were concentrated using a centrifugal concentrator and purified further by SEC (6.11.4).

For SacSΔHis and mutants thereof, the fractions of the His-trap containing the protein of interest were combined, 3C protease was added in a ratio of 1:1,000 and the mixture was dialysed against 3C-protease cleavage buffer (Table 10) at 4°C for a simultaneous exchange of the buffer and proteolytic digest. Subsequently, arising precipitate was removed by centrifugation at 17,000 x g for 20 min and the cleared supernatant injected onto a His-trap™ FF column. The flowthrough containing the cleaved tagless target protein was collected and concentrated as previously described and purified further using SEC with a HiLoad 16/60 Superdex 200pg size exclusion column. SacS buffer (Table 10) was used as mobile phase. Fractions, containing the protein of interest were combined, concentrated, aliquoted and stored at -80°C.

6.11.9. Purification of SunS

SunS was expressed as His-Sumo-GT constructs in a pChampion-pET-Sumo plasmid in *E. coli* BL21 (DE3) Rosetta. Sumo-SunS was isolated from the cell lysate via HisTrap described above (6.11.3.1). The fractions of the His-trap containing the protein of interest were combined, Ulp1 protease was added in a ratio of 1:200 and the mixture dialysed against 3C-protease cleavage buffer (Table 10) at 4°C for a simultaneous exchange of the buffer and proteolytic digest. Protease and the Sumo tag were separated off by SEC (6.11.4) using a HiLoad 16/60 Superdex 200pg size exclusion column. SunS buffer (Table 10) was used as mobile phase. Fractions, containing the protein of interest were combined, concentrated, aliquoted and stored at -80°C.

6.11.10. Purification of AciS

AciS was expressed as GT-3C-His construct in a pET-3CLIC plasmid in *E. coli* BL21 (DE3) Rosetta cells in autoinduction medium. AciS was isolated from the cell lysate via HisTrap (6.11.3.1) using AciS-HisTrap buffers.

Fractions containing the protein of interest were pooled and the tag was cleaved off via treatment with 3C protease during an overnight dialysis against TBS (Table 10), supplemented with 0.5 mM DTT and 10% Glycerol. Nucleotide and proteinaceous impurities could be removed by IEX (6.11.5), collecting the flow-through. The resulting protein solution was concentrated as described previously and purified further using SEC (6.11.4) with AciS buffer as mobile phase. Fractions containing the protein of interest of sufficient purity were combined, concentrated, aliquoted and finally stored at -80°C.

6.11.11. Purification of EntS

EntS was expressed as His-3C-GT constructs in a pET-3CLIC plasmid in *E. coli* BL21 (DE3) Rosetta. EntS was isolated from the cell lysate via HisTrap described above (6.11.3.1). The fractions of the His-trap containing the protein of interest were combined, 3C protease was added in a ratio of 1:1,000 and the mixture was dialysed against 3C-protease cleavage buffer (Table 10) at 4°C for a simultaneous exchange of the buffer and proteolytic digest. Subsequently, arising precipitate was removed by

centrifugation at 17,000 x g for 20 min and the cleared supernatant injected onto a His-trap™ FF column. The flowthrough containing the cleaved tagless target protein was collected and concentrated as previously described. For the EntSH200A mutant nucleotide-based impurities were subsequently removed by IEX (6.11.5). Final purification was performed using SEC (6.11.4) with EntS buffer (Table 10) as mobile phase unless specified otherwise. Fractions, containing the protein of interest in the dimeric state (based on elution time) were combined, concentrated, and reapplied to the SEC column for complete isolation of the dimeric state from higher oligomers. Dimeric EntS fractions were pooled, concentrated, aliquoted and stored at -80°C.

6.11.12. Purification of PltS

PltS was expressed as a His-Sumo-StrepTag-3C-GT construct in a pChamion-pET-Sumo plasmid in *E. coli* Bl21 (DE3) Rosetta cells growing in autoinduction medium. Tagged PltS was isolated from the cell lysate via HisTrap (6.11.3.1), with the column charged with Cobalt instead of Nickel. The tags were removed by treatment with 3C protease while dialysing against TBS (Table 10), containing 1 mM DTT for simultaneous buffer exchange and digest. The GT was separated from the His-SUMO tag by injecting the dialysate onto a HisTrap FF column, collecting the flow-through.

The resulting protein solution was concentrated using an Amicon-Ultra centrifugal concentrator with a 10 kDa cutoff (Merck). Final purification was performed using SEC (6.11.4) with PltS buffer (Table 10) as mobile phase. Fractions, containing the protein of interest were combined, concentrated, aliquoted and stored at -80°C.

6.11.13. Storage of proteins

Expressed proteins were concentrated using a centrifugal concentrator to a concentration of at least 20 mg/ml, snap-frozen in liquid nitrogen and stored at -80°C until further use.

6.12. Peptide synthesis

Microwave assisted solid phase peptide synthesis was performed in an Fmoc based strategy, using a LibertyBlue™ peptide synthesizer (CEM Corporation). Peptides were coupled to either a ProTide® Cl-MPA resin (CEM) or a Wang resin preloaded with Fmoc-Lysine (CEM). A list of used amino acids can be found in (Table 17). Standard amino acids were obtained from Carbolution (Carbolution chemicals GmbH), non-standard amino acids from CEM. Details for the methods during coupling, deprotection, final deprotection and purification are described in the following sections.

Table 17 *List of amino acids used for solid-phase peptide synthesis in this study.*

Standard amino acids	***Non-standard amino acids***
Fmoc-Ala-OH	Fmoc-His(Boc)-OH
Fmoc-Arg(Pbf)-OH	Fmoc-Asp(OMpe)-OH
Fmoc-Asn(Trt)-OH	
Fmoc-Asp(OtBu)-OH	
Fmoc-Cys(Trt)-OH	
Fmoc-Phe-OH	
Fmoc-Gln(Trt)-OH	
Fmoc-Glu(OtBu)-OH	
Fmoc-Gly-OH	
Fmoc-His(Trt)-OH	
Fmoc-Ile-OH	
Fmoc-Leu-OH	
Fmoc-Lys(Boc)-OH	
Fmoc-Met-OH	
Fmoc-Pro-OH	
Fmoc-Ser(tBu)-OH	
Fmoc-Thr(tBu)-OH	
Fmoc-Trp(Boc)-OH	
Fmoc-Tyr(tBu)-OH	
Fmoc-Val-OH	

6.12.1. Microwave methods and cycles.

Despite the differences in their synthesis methods, as described further in section 6.14, all synthesis steps followed standard microwave cycles, which are described in Table 18 in conjunction with Table 19.

Table 18 *Composition of Microwave cycles used in this study.*

Cycle name	Method
chloride-loading	DCA (Trityl loading) 4 x Wash
Single coupling	Deprotection 4 x Wash Coupling Wash
double coupling	Deprotection 4 x Wash 2 x Coupling Wash
double deprotection, single coupling	2 x Deprotection 4 x Wash 1 x Coupling Wash
double coupling extra washes	Deprotection 4 x Wash 2 x Coupling 2 x Wash
single coupling, capping	Deprotection 4 x Wash Coupling 2 x Wash Acetylation 3 x Wash

Wash steps were performed using DMF. Potassium iodide and Di-isopropyl ethyl amine (DIPEA) were used in a Finkelstein-like reaction to facilitate the coupling of the first amino acid to the Cl-MPA resin, at a concentration of 0.125 M and 1 M, respectively. Standard couplings were performed using the Carbomax™ approach[212], using 0.5 M DIC as activator and a mixture of 0.5 M OxymaPure® supplemented with 0.05 M DIPEA as base. Amino acids were used in 5-fold excess concentration (0.2 M) in DMF. Fmoc-deprotection was performed as described in Collins *et al* (2014)[138], using piperazine 10% (w/v) in a 9:1 mixture of N-methyl pyrrolidone (NMP):Ethanol (EtOH), supplemented with 0.1 M

HOBt or OxymaPure® to suppress racemisation reactions. Acetylation was performed using 10 % (v/v) acetic anhydride in DMF, with 10 % (v/v) DIPEA added as base.

Table 19 *Heating scheme for microwave methods of Table 18*

Method	*temperature (°C)*	*power (W)*	*hold time (s)*	*delta T (°C)*
capping	65	45	35	5
	65	0	50	5
	65	50	35	5
	65	0	45	5
DCA loading	80	75	60	2
	90	20	540	1
coupling 10 min at 50°C	25	0	120	2
	50	35	480	1
coupling 4 min	75	217	15	2
	90	43	225	2
deprotection	75	155	15	2
	90	35	120	1

6.12.2. Reaction control

To determine the success of the peptide synthesis, the resin was first washed with dichloromethane (DCM), subsequently dried under air and then *in vacuo*.
A small amount of the dried resin was transferred to a syringe reactor and 200 µl of a suitable cleavage cocktail, usually consisting of 94% TFA, 2.5% H_2O, 2.5% Ethylene di-thiol (EDT), 1% Tri-isopropyl silane (TIS), were added. For peptides containing methionine the cleavage cocktail was supplemented with 1% Thioanisole.

The cleavage was allowed to occur for 2-3 hours for peptides above 20 amino acids in length, and 90 minutes for smaller peptides. Resin-material was removed by filtration off, and the solution concentrated under a nitrogen stream. Peptides were precipitated by the addition of 1 ml ice-cold diethyl ether and collected by centrifugation at 4,000 x g. Peptides deprotected using a cocktail containing thioanisole were washed an additional time with diethyl ether.

The resulting precipitate was dissolved in a 1:1 mixture of water:acetonitrile, containing 0.1% TFA. This solution was subsequently analysed by analytical RP-HPLC as described in 6.13.1.

6.12.3. Full cleavage and deprotection of peptides

Peptides were cleaved from the resin and amino acid side chains were deprotected in a one-pot reaction using the same deprotection cocktail described above (chapter 6.12.2). Typically, 3 ml of reaction cocktail for 200 mg of dried resin was added. Deprotection and cleavage was performed under gentle agitation at RT for 2.5 hours. Afterwards the dissolved peptide was filtered through the frit of the syringe reactor, and the resin washed thrice with TFA. Volatiles were evaporated under reduced pressure at room temperature for wild-type peptides, and 40°C for truncated peptides and Cys-to-Ser mutants.

30 ml of ice-cold diethyl ether were added, precipitating the crude peptide. The crude was washed thrice with ice cold diethyl ether, suspended in dilute hydrochloric acid and dried via lyophilization. The powder was dissolved in a 60:40 mixture of acetonitrile:water, containing 0.1% formic acid. Peptides containing tryptophanes were incubated in this mixture over night to allow for complete deprotection of the Trp-carbamates.

6.13. Peptide purification and analysis

6.13.1. Analytical HPLC

Analytical HPLC was performed either on a VWR-Hitachi Chromaster HPLC system (VWR International GmbH) or on a Primaide HPLC system (Hitachi-VWR International GmbH), both equipped with a Kinetex® C18 column (5 µM, 100 Å, 250 x 4.6 mm, Phenomenex®), containing a C18 cartridge (4 x 3.0 mm, Phenomenex®) for column protection.

Data were collected and evaluated using either the EZChrome *Elite* 3.3.2.a software (Agilent Technologies) or OpenLab EZChrome A.04.10 (Agilent Technologies).
For sample separation, a linear gradient of water-acetonitrile, both containing 0.1% TFA was used at a flow rate of 1 ml/min. A list of used gradients can be seen in Table 20:

Table 20 *List of HPLC gradients used for peptide analysis.*

Name	***Time***	***%Acetonitrile***
10-80_5minIso	0	10
	5	10
	23	80
	24	100
	27	100
	28	10
	32	10
10-80	0	10
	18	80
	19	100
	22	100
	23	10
	25	10
10-60	0	10
	18	60
	19	100
	22	100
	23	10
	25	10

Absorbance was detected at 220 and 280 nm. Eluting peptides were collected and analysed further using mass spectrometry.

6.13.2. Mass spectrometry

High-resolution mass spectra were obtained on an Agilent 6230 ESI-ToF LC-MS device (Agilent Technologies Inc.). Data collection was done on Mass Hunter Workstation version B.08.00 and analysed using MestreNova version 14.1 (Mestrelab Research).

Peptides obtained using the in vivo production method were analysed using an amazon ETD mass spectrometer (Bruker) in a direct infusion mode. Fragmenting mass spectra were typically obtained by isolating the $[M+4H]^{4+}$ mass and subsequent collision-induced dissociation (CID).

Monoisotopic masses and mass to charge ratios were calculated using the peptide mass calculator web-service version 3.2 (http://rna.rega.kuleuven.be/masspec/pepcalc.htm).

6.13.3. Preparative HPLC

Crude peptides were purified using a LaPrep∑ HPLC system (VWR International GmbH), equipped with a Kinetex® RP-C18 endcapped (5 µM, 100 Å, 250 x 21.2 mm, Phenomenex®, USA) column with a SecuritaGuard™ cartridge kit as a precolumn.

Samples were separated with a linear gradient (Table 21) of water-acetonitrile, containing 0.1% TFA for crude peptides. For the purification of glycosylated peptides, the TFA was substituted with 0.1% formic acid. Eluting peptides were detected via absorbance at 220 nm. Data was collected and analysed with the EZChrome *Elite* 3.3.3 SP2 software (Agilent Technologies).

Table 21 *Gradients used for the purification of peptides in preparative HPLC.*

Name	***Time***	***%Acetonitrile***	***Flow speed [ml/min]***
10-80_5minIso	0	10	10
	5	10	15
	23	80	15
	24	100	15
	27	100	15
	28	10	15
	32	10	15
10-80	0	10	15
	18	80	15
	19	100	15
	22	100	15
	23	10	15
	25	10	15
10-60	0	10	15
	18	60	15
	19	100	15
	22	100	15
	23	10	15
	25	10	15

Purity and identity of collected fractions were determined by HPLC-UV and MS respectively. Fractions, containing the peptide in sufficient purity, were pooled, snap-frozen in liquid nitrogen and volatiles were evaporated *in vacuo* by lyophilization. TFA was removed using the protocol of Andrushchenko *et al*. [213]Briefly, the peptide was dissolved in 10 mM hydrochloric acid, incubated at RT for 10 min, snap frozen in liquid nitrogen and lyophilized. This procedure was repeated thrice. For WT SacA peptide the first addition of hydrochloric acid did not cause the peptide to dissolve. Therefore, the first incubation step was performed as a slurry under strong agitation. Subsequent incubations were performed as described above with the peptide in solution.

6.14. Prepared peptides

Microwave cycles, and purification protocols of the prepared peptides can be found in the following sections. For clarity, amino acid positions are counted in the direction of synthesis, that is, the C-terminal amino acid is counted as position 1.

6.14.1. SacA derived peptides

Table 22 *Schematic representation of the synthesis methods for the SacA peptide*

Name	**Sequence**
SacA	H_2N-GFTAAQCAAF FVQCASGGTI GCGGMWHGRP AACDLYDQYC K-COOH
Resin	Lys-Wang 0.3 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	3-20 – 4 min single coupling 21-41 – 4 min double coupling, extra washes
Special cycles	Cys(2 & 9) – 10 min single coupling, capping Fmoc-His(Trt) – 10 min single coupling 50°C Fmoc-Arg(Pbf) – 4 min double coupling Fmoc-Cys(Trt) – double deprotection, double coupling, capping
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3-4 h
HPLC method	10-80_5minIso
Yield	8 mg

Table 23 *Schematic representation of the synthesis methods for the SacA4Ser peptide*

Name	**Sequence**
SacA4Ser	H_2N-GFTAAQSAAF FVQSASGGTI GCGGMWHGRP AASDLYDQYS K-COOH
Resin	Lys-Wang 0.3 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	3-20 – 4 min single coupling 21-41 – 4 min double coupling, extra washes
Special cycles	Ser(2) – 10 min single coupling, capping Ser(9) – 4 min double coupling Fmoc-His(Trt) – 10 min single coupling 50°C Fmoc-Arg(Pbf) – 4 min double coupling Fmoc-Cys(Trt) – double deprotection, double coupling, capping
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3-4 h
HPLC method	10-80_5minIso
Yield	12 mg

Table 24 *Schematic representation of the synthesis methods for the SacAallSer peptide*

Name	**Sequence**
SacAallSer	H_2N-GFTAAQSAAF FVQSASGGTI GSGGMWHGRP AASDLYDQYS K-COOH
Resin	Lys-Wang 0.3 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	3-20 – 4 min single coupling 21-41 – 4 min double coupling, extra washes
Special cycles	Ser(2) – 10 min single coupling, capping Ser(9) – 4 min double coupling Fmoc-His(Trt) – 10 min single coupling 50°C Fmoc-Arg(Pbf) – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3-4 h
HPLC method	10-80_5minIso
Yield	21 mg

Table 25 *Schematic representation of the synthesis methods for the Leader-TEV-SacA peptide*

Name	**Sequence**
Leader-TEV-SacA	H_2N-MDQLFKELKL EELENLYFQG FTAAQCAAFF VQCASGGTIG CGGMWHGRPA ACDLYDQYCK-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-60 – 4 min double coupling
Special cycles	Lys (1) – Chloride loading Fmoc-His (Trt) (14) – 10 min single coupling 50°C Fmoc-Arg(Pbf) (12) – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3-4 h
HPLC method	10-80_5minIso
Yield	1 mg

6.14.2. SunA derived peptides

Table 26 *Schematic representation of the synthesis methods for the SunAm peptide*

Name	**Sequence**
SunAm	H_2N-GKAQCAALWL QCASGGTIGC GGGAV-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-25 – 4 min double coupling
Special cycles	Val (1) – Chloride loading
Cleavage	TFA/TIS/H_2O/EDT 3 h
HPLC method	10-80_5minIso
Yield	1 mg

Table 27 *Schematic representation of the synthesis methods for the SunAm2Ser peptide*

Name	**Sequence**
SunAm2Ser	H_2N-GKAQSAALWL QSASGGTIGC GGGAV-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-25 – 4 min double coupling
Special cycles	Val (1) – Chloride loading
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3 h
HPLC method	10-80_5minIso
Yield	9 mg

Table 28 *Schematic representation of the synthesis methods for the SunAmallSer peptide*

Name	**Sequence**
SunAmAllSer	H_2N-GKAQSAALWL QSASGGTIGS GGGAV-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-25 – 4 min double coupling
Special cycles	Val (1) – Chloride loading
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3 h
HPLC method	10-80_5minIso
Yield	17 mg

6.14.3. Enterocin 96 derived peptides

Table 29 *Schematic representation of the synthesis methods for the full-length Enterocin 96 peptide*

Name	**Sequence**
Enterocin 96	H2N-MSKRDCNLMK ACCAGQAVTY AIHSLLNRLG GDSSDPAGCN DIVRKYCK-COOH
Resin	Lys-Wang 0.3 mmol/g
Scale	0.05 mmol
Non-standard amino acids	Fmoc-His(Boc) - 22 Fmoc-Asp(OMpe) - 31
Microwave Method	1-21 – 4 min single coupling 22-48 – 4 min double coupling
Special cycles	Fmoc-Met (9)– Double deprotection 4 min double coupling Fmoc-His(Boc) – 10 min single coupling 50°C Fmoc-Arg(Pbf) – 4 min double coupling Fmoc-Cys(Trt) – Double deprotection, 10 min coupling, capping
Cleavage	TFA/TIS/H_2O/EDT/Thioanisol 3-4 h
HPLC method	10-80_5minIso
Yield	8 mg

Table 30 *Schematic representation of the synthesis methods for the Ecm peptide*

Name	**Sequence**
Ecm	H_2N-AVTYAIHSLL NRLGGDSSDP AG-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	Fmoc-His(Boc) - 22 Fmoc-Asp(OMpe) - 31
Microwave Method	2-22 – 4 min single coupling
Special cycles	Gly (1) – Chloride loading Fmoc-His(Boc) – 10 min coupling Fmoc-Arg(Pbf) – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT 2 h
HPLC method	10-80
Yield	30 mg

Table 31 *Schematic representation of the synthesis methods for the EC(Cys)m peptide*

Name	**Sequence**
EC(Cys)m	H_2N-AVTYAIHSLL NRLGGDCSDP AG-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	Fmoc-His(Boc) - 22
	Fmoc-Asp(OMpe) - 31
Microwave Method	2-22 – 4 min single coupling
Special cycles	Gly (1) – Chloride loading
	Fmoc-His(Boc) – 10 min coupling
	Fmoc-Arg(Pbf) – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT 2.5 h
HPLC method	10-80
Yield	16 mg

Table 32 *Schematic representation of the synthesis methods for the EC(Ala)m peptide*

Name	**Sequence**
EC(Ala)m	H_2N-AVTYAIHSLL NRLGGDASDP AG-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	Fmoc-His(Boc) - 22
	Fmoc-Asp(OMpe) - 31
Microwave Method	2-22 – 4 min single coupling
Special cycles	Gly (1) – Chloride loading
	Fmoc-His(Boc) – 10 min coupling
	Fmoc-Arg(Pbf) – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT 2.5 h
HPLC method	10-80_5minIso
Yield	32 mg

6.14.4. PltA derived peptides

Table 33 *Schematic representation of the synthesis methods for the PltA peptide*

Name	**Sequence**
PltA	H2N-GMSKAECTYL YNLITTGATS SHGCVPSSNY LDLYRSNCKG KGPKL-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-30 – 4 min double coupling Synthesized in 5 blocks 1-10, 11-20, 21-25, 26-32, 33-45
Special cycles	Leu (1) – Chloride loading Fmoc-His(Trt) – 10 min single coupling 50°C Fmoc-Arg(Pbf) – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT 3 h
HPLC method	10-80_5minIso
Yield	7 mg

Table 34 *Schematic representation of the synthesis methods for the PltA2Ser peptide*

Name	**Sequence**
PltA2Ser	H2N-GMSKAESTYL YNLITTGATS SHGCVPSSNY LDLYRSNSKG KGPKL-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-45 – 4 min double coupling
Special cycles	Fmoc-Tyr(tBu)-OH (16) – double deprotection 4 min single coupling Fmoc-Cys(Trt)-OH – double deprotection, 10 min coupling, capping Fmoc-His(Trt)-OH – 10 min single coupling 50°C Fmoc-Arg(Pbf)-OH – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT/Thioanisole 3 h
HPLC method	10-60
Yield	11 mg

Table 35 *Schematic representation of the synthesis methods for the PltAallSer peptide*

Name	**Sequence**
PltAallSer	H2N-GMSKAESTYL YNLITTGATS SHGSVPSSNY LDLYRSNSKG KGPKL-COOH
Resin	Cl-MPA Protide® 0.16 mmol/g
Scale	0.05 mmol
Non-standard amino acids	-
Microwave Method	2-20 – 4 min single coupling 21-30 – 4 min double coupling 31-45 – 4 min double coupling, extra wash
Special cycles	Fmoc-Tyr(tBu)-OH (16) – double deprotection 4 min single coupling Fmoc-His(Trt)-OH – 10 min single coupling 50°C Fmoc-Arg(Pbf)-OH – 4 min double coupling
Cleavage	TFA/TIS/H_2O/EDT 3 h
HPLC method	10-60 HPLC had to be performed twice.
Yield	3mg

6.15. Activity tests of GT

6.15.1. Qualitative activity test

HPLC-purified and dried peptide was dissolved in water or 10 mM TCEP*HCl for cysteine containing peptides. This peptide solution was diluted to a concentration of 100 µM in the appropriate enzyme buffer, containing 1 mM $MnCl_2$. GT was added to a final concentration of 5 µM, and sugar-nucleotide to a concentration of 500 µM. For peptides containing cysteines, an additional 1 mM TCEP was added. The reaction was allowed to take place over night, before it was quenched by acidification with dilute hydrochloric acid to a pH < 3. Insoluble material was removed via centrifugation at 17,000 x g for 10 min and the supernatant was analysed via HPLC.

6.15.2. Malachite green assay

The malachite green assay to quantify released phosphates was performed as described by Baykov & Avaeva[214]. Briefly, a malachite green stock containing 1.2 g of malachite green chloride in 1l of a 16% sulphuric acid solution was prepared. 10 ml of this solution was combined with 2.5 ml of 380 mM sodium molybdate and 0.2 ml of 11% Tween 20 to afford a 4x working solution.

The GT at a concentration of 10-1000 nM of interest was combined with 10 U apyrase (NEB) in the presence of 1 mM sugar-nucleotide and 0-250 µM of the peptide acceptor and incubated at RT for 5-60 minutes. The reaction was quenched by the addition of the malachite green working solution to afford green malachite green – phosphomolybdate. The concentration of which could be measured at a wavelength of 405 and 630 nm using a NanoQuant infinite M200 microplate reader (Tecan).

6.15.3. Enzyme kinetics

Enzyme kinetics for SacS were obtained using HPLC purified peptides, harbouring Cys-to-Ser mutations, therefore lacking the ability to form disulphide bonds.

These peptides were prepared for glycosylation as described above with a peptide concentration of 12.5-250 µM, an enzyme concentration of either 10 nM or 3 µM, and without the addition of the sugar-nucleotide. The reaction mixture was set to equilibrate against a temperature of 50°C before f.c. 1 mM UDP-Glucose was added. The reaction occurred at 50°C under vigorous shaking for 3-13 minutes and quenched by acidification with hydrochloric acid. Samples were snap-frozen in liquid nitrogen and lyophilized. Dried samples were subsequently dissolved in 10 mM HCl and analysed via HPLC, measuring absorbance at 280 nm. Glycosylated and non-glycosylated species were identified based on retention time compared to pure peptide species. Measurements were performed in triplicate.

Data was analysed using OriginPro 2021b version 9.85 (OriginLab Corporation) by graphical integration of peaks and subsequent fitting of data against the Michaelis-Menten equation:

$$v = \frac{v_{max} * S}{K_m + S}$$

with $v_{max} = k_{cat} * [Enzyme]$.

6.15.4. Hydrolysis assay

Diglycosylated Ecm peptide was dissolved in water and diluted in 50 mM HEPES pH 7.5 to a concentration of 100 µM. 1 mM $MnCl_2$, and either 1 mM UDP, 5 µM EntS or both were added, and the reaction incubated at RT for 7 days. On day zero, one, two, and seven a sample of 100 µl was taken, acidified with hydrochloric acid and analysed via HPLC and mass spectroscopy.

6.16. Preparation of glycosylated glycocins

The glycocin of interest was dissolved in water or 1 mM TCEP and diluted in the appropriate buffer to a concentration of 100-200 µM. 2 mol eq. of nucleotide sugar donor and 5 µM of the cognate GT were added and the reaction mixture was incubated at RT for 16 h.

The reaction was quenched by acidification to pH 3 with hydrochloric acid and filtered through a 0.45 µm Whatman filter (Cytiva). GT, nucleotides and unreacted peptide were separated from the product by RP-HPLC. The glycosylated peptide was desalted, as described in 6.13.3 and dried by lyophilisation. The resulting powder was stored at -20°C until further use.

6.17. In vivo production of glycocins

6.17.1. Coexpression of glycocin with its cognate GT

For the coexpression of *E. coli* BL21(DE3) Rosetta were transformed with the pCDFduet_MBP-TEV-Glycocin_GTase plasmid via heat-shock and selected for using LB agar, containing 34 µg/ml Chloramphenicol and 50 µg/ml Streptomycin.

A single colony was picked from the agar plate and use to inoculate LB medium, and the culture was incubated for 12 to 14 h. For protein production, autoinduction medium (adapted from Studier, 2013), was inoculated with the preculture in a ratio of 1:100 and grown at 30°C for 30 hours.

6.17.2. Glycocin purification

After protein production, cells were harvested by centrifugation at 6,000 g for 15 min at 4°C. The cell pellet was resuspended in lysis buffer (Table 10) and the cells lysed and cell debris removed as described previously (chapter 6.11.2). MBP-tagged proteins were isolated using an MBPtrap HP column as described in 6.11.3.3.

Co-eluting MBP lacking any glycocin could be removed by hydrophobic interaction chromatography using a HiTrap Phenyl FF (HS) (Cytiva) column. For this, 1 M ammonium sulphate was added to the pooled fractions and the solution was loaded on the column. The proteins were eluted with a volatile buffer without ammonium sulphate using a linear gradient from 70 %-100% over 10 CV followed by elution with 100% volatile buffer for another 10 CV. This step proved to be necessary only for AciA.

The resulting MBP-Glycocin solution was concentrated using a centrifugal concentrator with 10 kDa cutoff (Merck) and subjected to size exclusion chromatography (see 6.11.4) using the volatile glycocin size exclusion buffer (Table 10). Fractions containing monomeric MBP-Glycocins (as determined by retention time) were collected.

Disulphide reshuffling of the multimeric fractions was performed by dialyzing them against a solution of refolding buffer (Table 10), incubating them for 4 hours, followed by a dialysis against volatile buffer for 16 hours at 4°C. Afterwards a second size exclusion run was performed. Fractions, containing the monomer were collected and combined with the previous fractions.

The pooled monomer was concentrated *in vacuo*. The MBP-Glycocin was resuspended in a cleavage buffer, supporting TEV activity and the solubility of glycocin. The MBP tag was cleaved off by incubation with a 1:10,000 ratio of TEV protease at room temperature for 12 hours.

Insoluble material, occurring at this stage, was separated via centrifugation and the supernatant purified via size exclusion (see 6.11.4) with the volatile size exclusion buffer (Table 10) as mobile phase. Fractions containing the glycocin were pooled and concentrated *in vacuo*, dissolved in dilute hydrochloric acid and the separated from undesired glycoforms and further impurities via preparative HPLC, desalted and dried, as mentioned in 6.13.3,. The resulting pure glycocin was stored at -80°C until further use.

6.18. Test for antimicrobial activity.

Agar well diffusion assays were performed using a standard protocol. Briefly, the appropriate medium for the tested bacterial strain was inoculated with a single colony and grown overnight at 37C. These cultures were diluted to an OD_{600} of 1.5 and 15 µl of this dilution added to 50 ml of an appropriate nutrient agar. 20 ml of the mixture was poured in a plate and left to solidify. Holes were punched into the agar using the backside of a sterile 1 ml pipette tip. The resulting holes were backfilled with 50 µl of fresh nutrient agar to create an agar well.

As filamentous bacteria did not grow in a manner amenable for homogeneous casting in an agar suspension, multiple colonies were transferred to the new nutrient agar plate via streak-inoculation before creating agar wells in the abovementioned manner instead.

25 µl of peptide at a concentration of 25-100 µM in a 50 mM HEPES buffer either containing or not containing glucose at a concentration of 500 mM were added to the wells and the cells left to grow for 24h at 37°C.

6.19. Crystallization

Initial hits were obtained using the Peg/Salt, JCSG++ and Basic kits (Jena Bioscience) and the Hampton Index kit. 50 µl of the kit's reservoir solution were added to a SwissCl2Drop plate (Jena Bioscience). Drops with a total volume of 0.3 µl, consisting of a 1:1 mixture of enzyme and reservoir solutions were placed using an Oryx4 pipetting robot (Douglas Instruments). The concentration of the enzyme solution was typically 10-20 mg/ml. The plate was subsequently sealed and incubated at 293 K unless otherwise specified. Crystals typically appeared within two days.

6.19.1. Crystallization of SacS348

Initial screens of both SacS348 and SacS348noSt were performed as described above. While SacS348 resulted in no crystallization, SacS348noSt readily crystalized. Further optimization of crystallization conditions resulted in a crystallization mixture composed of SacS348noSt at 10 mg/ml combined in a 1:1 ratio with the mother liquor of 140 mM TMAO, 60 mM Li_2SO_4, 100 mM Tris pH 7.5, 14% PEG MME 5000 and 7.5% PEG 3350.

Crystals from this condition were crushed, diluted in a 1:100 ratio with mother liquor and the resulting suspension used as seedstock for the generation of further crystals in a screen using the Hampton Index, Peg/Salt, JCSG++ and Basic kits mentioned above. Amongst the conditions in these screens SacS348 crystalized in a condition composed of 100 mM Bis-Tris pH 5.5, 2M $(NH_4)_2SO_4$. Crystals of that screen were collected without further optimization. Cryoprotection was performed by overlaying the crystal with silicone oil (Jena Biosciences).

6.19.2. Crystallization of SunS

Crystallization of SunS was performed following the protocol of Fujinami *et al* (2021)[168] with some modifications. Briefly, SunS lacking the His-tag in a 6 mg/ml solution of 20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM HEPES, 10% Glycerol was combined with f.c. 1 mM SunAm-Glc peptide, 1 mM $MnCl_2$ and 1 mM UDP. The resulting mixture was combined in a 1:1 mixture of the reservoir solution composed of 100 mM HEPES, 20% PEG 3350 and incubated at 278 K for two days. Afterwards the crystallization

mixture was incubated at 293 K. Crystals typically appeared within two weeks. Mother liquor supplemented with 20 % (v/v) glycerol was used as a cryoprotectant. Cryo-protected crystals were collected, plunged into liquid nitrogen and stored therein until data collection.

6.19.3. Crystallization of EntS

EntS crystallization was performed with a protein concentration of 12-15 mg/ml. Initial screening resulted in the sporadic occurrence of non-reproducible crystals that did not diffract. Optimization of the EntS storage buffer (Table 10) and addition of 500 µM Ecm-GlcGlc resulted in a crystal in a hanging drop vapour diffusion experiment with 180 mM tri-potassium citrate pH 8.3, 19 % PEG. These crystals were crosslinked by vapor diffusion with 25% glutaraldehyde, washed with mother liquor and subsequently crushed, following the protocol of Yan *et al* (2015)[178]. Microseed matrix screening using the crushed and crosslinked crystals afforded replicable crystals in the condition 200 mM NaF, 20% PEG 3350 in a mixture of 3:2:1 of protein solution (12 mg/ml) containing 1 mM Ecm-GlcGlc : Precipitant solution : Seedstock.

Streak seeding of crystals obtained in this manner into a 1:1 mixture of 200 mM KF, 18% PEG 3350 with the aforementioned enzyme solution afforded diffraction quality crystals. These could be optimized, by using a further optimized buffer for EntS (Table 10), which reproducibly afforded crystals with a good diffraction quality under the following condition: 100 mM MMT pH 7.5, 200 mM KF, 23% PEG 1500.

Cocrystals with the EC(Ala)m peptide were generated in a similar manner, but the final crystallization condition was changed to 100 mM MMT pH 8.0, 200 mM KSCN, 18% PEG 3350 and UDP-Glc was used in lieu of UDP.

Cryoprotection of the EntS crystals was performed by the addition of 15% Glucose (w/v) to the crystallization drop. Cryoprotected crystals were plunged into liquid nitrogen and stored therein until further use.

6.19.4. Crystallization of PltS

Initial screens of PltS were described as mentioned above. Optimization of the crystallization conditions gave the condition PltS at 8-12 mg/ml in a 1:1 ratio with 25-30% tacsimate. Crystals typically appeared within 2 days at 293 K. Cryoprotection was performed by submerging the crystals into a solution of 3 M malic acid 1 mM UDP-GlcNAc, followed by incubation for 5 min. Crystals were collected, plunged into liquid nitrogen and stored therein. PltS-PltA cocrystals were obtained from a mixture of 10 mg/ml PltS, 1 mM PltA2Ser-Glc, 1 mM $MnCl_2$, 1 mM UDP in a 1:1 ratio with a crystallization buffer composed of 100 mM Bicine pH 8.0, 600 mM sodium potassium tartrate 10 mM β-mercaptoethanol, 7.5% PEG 3350 and 15% Glucose. Crystals typically appeared within 2 days at 278 K.

6.20. X-ray diffraction

6.20.1. Data collection

Diffraction data were collected at Berlin BESSY II, beamline 14-1 or beamline 14-2 at 100 K. Measurements were performed with a wavelength of 0.9184 Å.
The measured diffraction pattern was visualized using Adxv (Scripps research institute). For crystals showing diffractions up to sufficient resolution, space group estimates and optimal data-collection strategies were predicted using iMosflm[215]. Exposure times per angle were typically 0.1 s/degree and a full 360° rotation measurement was recorded.

6.20.2. Structure solution.

Data were processed via the CCP4i2 software suite[216]. Xia2/DIALS[217] was used for scaling and merging the data. MR-Phaser[218] was used to solve the structures. For SunS the published structure with PDB ID 7MSN[168] was used as a search model. For SacS, PltS and EntS a predicted structure generated with alphafold2[169] colab[219] was used as initial search model. Structures of mutants or from crystals with higher quality were solved using the refined models of the initial, lower quality dataset as a search model.

BUCCANEER[220] was used for initial placement of peptide chains, followed by several iterations of manual model building in Coot[221] and refinement via Refmac5[222]. Covalent linkages were created via AceDRG[223]. Restraints for sugars, if applicable, were generated using Privateer[224]. The PDB-redo webserver[225] with paired refinement was used to determine whether higher resolution data should be included as well. Structures were visualized using CCP4mg[226]. PISA[180] was used to determine the most likely dimerization interface. PDBsum[227] was used to generate topological maps of the protein structures.

6.21. CD Spectroscopy

CD spectra were obtained using a JASCO-810 spectropolarimeter (JASCO Deutschland GmbH). For temperature control a JASCO PTC-423S Peltier element and a HAAKE WKL water recirculator (Thermo Electron GmbH) was connected to the instrument. Measurements were performed at a constant nitrogen stream at 3.0 l/min. Samples were a concentration range of 50-100 µM in Tris buffer at pH 8.0.

Spectra were measured in triplicate for a wavelength range between 190-250 nm. The baseline was corrected using the corresponding buffer as a blank. Data was collected using the software JWS-510-J-800 Spectra Manager 2 (JASCO Deutschland GmbH). Analysis of data was performed using Origin 2021b. Due to the presence of chloride ions in the buffer, values obtained between 190 and 200 nm were discarded. Residual data were normalized to mean residue molar ellipticity [θ] according to the equation:

$$[\theta] = \frac{\theta_{obs}}{c * n * l * 10}\ mdeg\ dmol^{-1}\ cm^2\ residue^{-1}$$

With θ_{obs} corresponding to the measured ellipticity, c the concentration in mmol/l, n the number of amide bonds and l the pathlength of the cuvette in cm.

7. Bibliography

(1) Romero-García, J.; Francisco, C.; Biarnés, X.; Planas, A. Structure-Function Features of a Mycoplasma Glycolipid Synthase Derived from Structural Data Integration, Molecular Simulations, and Mutational Analysis. *PLOS ONE* **2013**, *8* (12), e81990. DOI: 10.1371/journal.pone.0081990.

(2) Lewis, K. The Science of Antibiotic Discovery. *Cell* **2020**, *181* (1), 29-45. DOI: 10.1016/j.cell.2020.02.056.

(3) Li, J. W.; Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **2009**, *325* (5937), 161-165. DOI: 10.1126/science.1168243.

(4) Rogers, L. A. The Inhibiting Effect of Streptococcus Lactis on Lactobacillus Bulgaricus. *J Bacteriol* **1928**, *16* (5), 321-325. DOI: 10.1128/jb.16.5.321-325.1928.

(5) Cavera, V. L.; Arthur, T. D.; Kashtanov, D.; Chikindas, M. L. Bacteriocins and their position in the next wave of conventional antibiotics. *Int J Antimicrob Agents* **2015**, *46* (5), 494-501. DOI: 10.1016/j.ijantimicag.2015.07.011.

(6) Gradisteanu Pircalabioru, G.; Popa, L. I.; Marutescu, L.; Gheorghe, I.; Popa, M.; Czobor Barbu, I.; Cristescu, R.; Chifiriuc, M. C. Bacteriocins in the Era of Antibiotic Resistance: Rising to the Challenge. *Pharmaceutics* **2021**, *13* (2). DOI: 10.3390/pharmaceutics13020196.

(7) Russell, A. H.; Truman, A. W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput Struct Biotechnol J* **2020**, *18*, 1838-1851. DOI: 10.1016/j.csbj.2020.06.032.

(8) Ayikpoe, R. S.; Shi, C.; Battiste, A. J.; Eslami, S. M.; Ramesh, S.; Simon, M. A.; Bothwell, I. R.; Lee, H.; Rice, A. J.; Ren, H.; et al. A scalable platform to discover antimicrobials of ribosomal origin. *Nat Commun* **2022**, *13* (1), 6135. DOI: 10.1038/s41467-022-33890-w.

(9) Leeds, J. A.; Sachdeva, M.; Mullin, S.; Dzink-Fox, J.; Lamarche, M. J. Mechanism of action of and mechanism of reduced susceptibility to the novel anti-Clostridium difficile compound LFF571. *Antimicrob Agents Chemother* **2012**, *56* (8), 4463-4465. DOI: 10.1128/AAC.06354-11. Mullane, K.; Lee, C.; Bressler, A.; Buitrago, M.; Weiss, K.; Dabovic, K.; Praestgaard, J.; Leeds, J. A.; Blais, J.; Pertel, P. Multicenter, randomized clinical trial to compare the safety and efficacy of LFF571 and vancomycin for Clostridium difficile infections. *Antimicrob Agents Chemother* **2015**, *59* (3), 1435-1440. DOI: 10.1128/AAC.04251-14.

(10) Klaenhammer, T. R. Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbiol Rev* **1993**, *12* (1-3), 39-85. DOI: 10.1111/j.1574-6976.1993.tb00012.x.

(11) Cotter, P. D.; Hill, C.; Ross, R. P. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol* **2005**, *3* (10), 777-788. DOI: 10.1038/nrmicro1273.

(12) Heng, N. C. K.; Wescombe, P. A.; Burton, J. P.; Jack, R. W.; Tagg, J. R. The Diversity of Bacteriocins in Gram-Positive Bacteria. In *Bacteriocins: Ecology and Evolution*, Riley, M. A., Chavan, M. A. Eds.; Springer Berlin Heidelberg, 2007; pp 45-92.

(13) Cotter, P. D.; Ross, R. P.; Hill, C. Bacteriocins — a viable alternative to antibiotics? *Nature Reviews Microbiology* **2013**, *11* (2), 95-105. DOI: 10.1038/nrmicro2937.

(14) Alvarez-Sieiro, P.; Montalban-Lopez, M.; Mu, D.; Kuipers, O. P. Bacteriocins of lactic acid bacteria: extending the family. *Appl Microbiol Biotechnol* **2016**, *100* (7), 2939-2951. DOI: 10.1007/s00253-016-7343-9.

(15) Simons, A.; Alhanout, K.; Duval, R. E. Bacteriocins, Antimicrobial Peptides from Bacterial Origin: Overview of Their Biology and Their Impact against Multidrug-Resistant Bacteria. *Microorganisms* **2020**, *8* (5). DOI: 10.3390/microorganisms8050639.

(16) Singh, V.; Rao, A. Distribution and diversity of glycocin biosynthesis gene clusters beyond Firmicutes. *Glycobiology* **2021**, *31* (2), 89-102. DOI: 10.1093/glycob/cwaa061.

(17) Montalban-Lopez, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; et al. New developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* **2021**, *38* (1), 130-239. DOI: 10.1039/d0np00027b.

(18) Nes, I. F.; Brede, D. A.; Diep, D. B. Chapter 16 - Class II Non-Lantibiotic Bacteriocins. In *Handbook of Biologically Active Peptides (Second Edition)*, Kastin, A. J. Ed.; Academic Press, 2013; pp 85-92.

(19) van Belkum, M. J.; Worobo, R. W.; Stiles, M. E. Double-glycine-type leader peptides direct secretion of bacteriocins by ABC transporters: colicin V secretion in Lactococcus lactis. *Mol Microbiol* **1997**, *23* (6), 1293-1301. DOI: 10.1046/j.1365-2958.1997.3111677.x.

(20) McAuliffe, O.; Ross, R. P.; Hill, C. McAuliffe, O, Ross, RP & Hill, C. Lantibiotics: structure, biosynthesis and mode of action. FEMS Microbiol. Rev. 25, 285-308. *FEMS microbiology reviews* **2001**, *25*, 285-308. DOI: 10.1111/j.1574-6976.2001.tb00579.x.

(21) Tan, S.; Ludwig, K. C.; Muller, A.; Schneider, T.; Nodwell, J. R. The Lasso Peptide Siamycin-I Targets Lipid II at the Gram-Positive Cell Surface. *ACS Chem Biol* **2019**, *14* (5), 966-974. DOI: 10.1021/acschembio.9b00157. Munch, D.; Muller, A.; Schneider, T.; Kohl, B.; Wenzel, M.; Bandow, J. E.; Maffioli, S.; Sosio, M.; Donadio, S.; Wimmer, R.; et al. The lantibiotic NAI-107 binds to bactoprenol-bound cell wall precursors and impairs membrane functions. *J Biol Chem* **2014**, *289* (17), 12063-12076. DOI: 10.1074/jbc.M113.537449.
(22) van Heusden, H. E.; de Kruijff, B.; Breukink, E. Lipid II induces a transmembrane orientation of the pore-forming peptide lantibiotic nisin. *Biochemistry* **2002**, *41* (40), 12171-12178. DOI: 10.1021/bi026090x.
(23) Marki, F.; Hanni, E.; Fredenhagen, A.; van Oostrum, J. Mode of action of the lanthionine-containing peptide antibiotics duramycin, duramycin B and C, and cinnamycin as indirect inhibitors of phospholipase A2. *Biochem Pharmacol* **1991**, *42* (10), 2027-2035. DOI: 10.1016/0006-2952(91)90604-4. Hasim, S.; Allison, D. P.; Mendez, B.; Farmer, A. T.; Pelletier, D. A.; Retterer, S. T.; Campagna, S. R.; Reynolds, T. B.; Doktycz, M. J. Elucidating Duramycin's Bacterial Selectivity and Mode of Action on the Bacterial Cell Envelope. *Front Microbiol* **2018**, *9*, 219. DOI: 10.3389/fmicb.2018.00219.
(24) Destoumieux-Garzon, D.; Peduzzi, J.; Thomas, X.; Djediat, C.; Rebuffat, S. Parasitism of iron-siderophore receptors of Escherichia coli by the siderophore-peptide microcin E492m and its unmodified counterpart. *Biometals* **2006**, *19* (2), 181-191. DOI: 10.1007/s10534-005-4452-9.
(25) Bhunia, A. K.; Johnson, M. C.; Ray, B. Purification, characterization and antimicrobial spectrum of a bacteriocin produced by Pediococcus acidilactici. *J Appl Bacteriol* **1988**, *65* (4), 261-268. DOI: 10.1111/j.1365-2672.1988.tb01893.x.
(26) Johnsen, L.; Fimland, G.; Nissen-Meyer, J. The C-terminal domain of pediocin-like antimicrobial peptides (class IIa bacteriocins) is involved in specific recognition of the C-terminal part of cognate immunity proteins and in determining the antimicrobial spectrum. *J Biol Chem* **2005**, *280* (10), 9243-9250. DOI: 10.1074/jbc.M412712200.
(27) Zhu, L.; Zeng, J.; Wang, C.; Wang, J. Structural Basis of Pore Formation in the Mannose Phosphotransferase System by Pediocin PA-1. *Appl Environ Microbiol* **2022**, *88* (3), e0199221. DOI: 10.1128/aem.01992-21.
(28) Hamada, T.; Matsunaga, S.; Fujiwara, M.; Fujita, K.; Hirota, H.; Schmucki, R.; Guntert, P.; Fusetani, N. Solution structure of polytheonamide B, a highly cytotoxic nonribosomal polypeptide from marine sponge. *J Am Chem Soc* **2010**, *132* (37), 12941-12945. DOI: 10.1021/ja104616z.
(29) Sanchez-Barrena, M. J.; Martinez-Ripoll, M.; Galvez, A.; Valdivia, E.; Maqueda, M.; Cruz, V.; Albert, A. Structure of bacteriocin AS-48: from soluble state to membrane bound state. *J Mol Biol* **2003**, *334* (3), 541-549. DOI: 10.1016/j.jmb.2003.09.060.
(30) Thennarasu, S.; Lee, D. K.; Poon, A.; Kawulka, K. E.; Vederas, J. C.; Ramamoorthy, A. Membrane permeabilization, orientation, and antimicrobial mechanism of subtilosin A. *Chem Phys Lipids* **2005**, *137* (1-2), 38-51. DOI: 10.1016/j.chemphyslip.2005.06.003 From.
(31) Kaur, H.; Jakob, R. P.; Marzinek, J. K.; Green, R.; Imai, Y.; Bolla, J. R.; Agustoni, E.; Robinson, C. V.; Bond, P. J.; Lewis, K.; et al. The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase. *Nature* **2021**, *593* (7857), 125-129. DOI: 10.1038/s41586-021-03455-w.
(32) Cao, L.; Do, T.; Link, A. J. Mechanisms of action of ribosomally synthesized and posttranslationally modified peptides (RiPPs). *J Ind Microbiol Biotechnol* **2021**, *48* (3-4). DOI: 10.1093/jimb/kuab005.
(33) Parks, W. M.; Bottrill, A. R.; Pierrat, O. A.; Durrant, M. C.; Maxwell, A. The action of the bacterial toxin, microcin B17, on DNA gyrase. *Biochimie* **2007**, *89* (4), 500-507. DOI: 10.1016/j.biochi.2006.12.005.
(34) Metelev, M.; Osterman, I. A.; Ghilarov, D.; Khabibullina, N. F.; Yakimov, A.; Shabalin, K.; Utkina, I.; Travin, D. Y.; Komarova, E. S.; Serebryakova, M.; et al. Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel. *Nat Chem Biol* **2017**, *13* (10), 1129-1136. DOI: 10.1038/nchembio.2462.
(35) Harms, J. M.; Wilson, D. N.; Schluenzen, F.; Connell, S. R.; Stachelhaus, T.; Zaborowska, Z.; Spahn, C. M.; Fucini, P. Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol Cell* **2008**, *30* (1), 26-38. DOI: 10.1016/j.molcel.2008.01.009.
(36) Kobayashi, Y.; Ichioka, M.; Hirose, T.; Nagai, K.; Matsumoto, A.; Matsui, H.; Hanaki, H.; Masuma, R.; Takahashi, Y.; Omura, S.; et al. Bottromycin derivatives: efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. *Bioorg Med Chem Lett* **2010**, *20* (20), 6116-6120. DOI: 10.1016/j.bmcl.2010.08.037.
(37) Heffron, S. E.; Jurnak, F. Structure of an EF-Tu complex with a thiazolyl peptide antibiotic determined at 2.35 A resolution: atomic basis for GE2270A inhibition of EF-Tu. *Biochemistry* **2000**, *39* (1), 37-45. DOI: 10.1021/bi9913597.
(38) Metlitskaya, A.; Kazakov, T.; Kommer, A.; Pavlova, O.; Praetorius-Ibba, M.; Ibba, M.; Krasheninnikov, I.; Kolb, V.; Khmel, I.; Severinov, K. Aspartyl-tRNA synthetase is the target of peptide nucleotide antibiotic Microcin C. *J Biol Chem* **2006**, *281* (26), 18033-18042. DOI: 10.1074/jbc.M513174200.

(39) Van de Vijver, P.; Vondenhoff, G. H.; Kazakov, T. S.; Semenova, E.; Kuznedelov, K.; Metlitskaya, A.; Van Aerschot, A.; Severinov, K. Synthetic microcin C analogs targeting different aminoacyl-tRNA synthetases. *J Bacteriol* **2009**, *191* (20), 6273-6280. DOI: 10.1128/jb.00829-09.
(40) Gavrish, E.; Sit, C. S.; Cao, S.; Kandror, O.; Spoering, A.; Peoples, A.; Ling, L.; Fetterman, A.; Hughes, D.; Bissell, A.; et al. Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chem Biol* **2014**, *21* (4), 509-518. DOI: 10.1016/j.chembiol.2014.01.014.
(41) Norris, G. E.; Patchett, M. L. The glycocins: in a class of their own. *Current Opinion in Structural Biology* **2016**, *40*, 112-119. DOI: https://doi.org/10.1016/j.sbi.2016.09.003.
(42) Oman, T. J.; Boettcher, J. M.; Wang, H.; Okalibe, X. N.; van der Donk, W. A. Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nature Chemical Biology* **2011**, *7*, 78. DOI: 10.1038/nchembio.509
(43) Wang, H.; Oman, T. J.; Zhang, R.; Garcia De Gonzalo, C. V.; Zhang, Q.; van der Donk, W. A. The Glycosyltransferase Involved in Thurandacin Biosynthesis Catalyzes Both O- and S-Glycosylation. *Journal of the American Chemical Society* **2014**, *136* (1), 84-87. DOI: 10.1021/ja411159k.
(44) Maky, M. A.; Ishibashi, N.; Zendo, T.; Perez, R. H.; Doud, J. R.; Karmi, M.; Sonomoto, K. Enterocin F4-9, a Novel O-Linked Glycosylated Bacteriocin. *Applied and Environmental Microbiology* **2015**, *81* (14), 4819-4826. DOI: 10.1128/aem.00940-15.
(45) Nagar, R.; Rao, A. An iterative glycosyltransferase EntS catalyzes transfer and extension of O- and S-linked monosaccharide in enterocin 96. *Glycobiology* **2017**. DOI: 10.1093/glycob/cwx042.
(46) Hanchi, H.; Hammami, R.; Fernandez, B.; Kourda, R.; Ben Hamida, J.; Fliss, I. Simultaneous Production of Formylated and Nonformylated Enterocins L50A and L50B as well as 61A, a New Glycosylated Durancin, by Enterococcus durans 61A, a Strain Isolated from Artisanal Fermented Milk in Tunisia. *J Agric Food Chem* **2016**, *64* (18), 3584-3590. DOI: 10.1021/acs.jafc.6b00700.
(47) Stepper, J.; Shastri, S.; Loo, T. S.; Preston, J. C.; Novak, P.; Man, P.; Moore, C. H.; Havlicek, V.; Patchett, M. L.; Norris, G. E. Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett* **2011**, *585* (4), 645-650. DOI: 10.1016/j.febslet.2011.01.023.
(48) Main, P.; Hata, T.; Loo, T. S.; Man, P.; Novak, P.; Havlíček, V.; Norris, G. E.; Patchett, M. L. Bacteriocin ASM1 is an O/S-diglycosylated, plasmid-encoded homologue of glycocin F. *FEBS Letters* **2019**, *n/a* (n/a). DOI: 10.1002/1873-3468.13708 (acccessed 2020/01/24).
(49) Kaunietis, A.; Buivydas, A.; Citavicius, D. J.; Kuipers, O. P. Heterologous biosynthesis and characterization of a glycocin from a thermophilic bacterium. *Nat Commun* **2019**, *10* (1), 1115. DOI: 10.1038/s41467-019-09065-5.
(50) Ren, H.; Biswas, S.; Ho, S.; van der Donk, W. A.; Zhao, H. Rapid Discovery of Glycocins through Pathway Refactoring in Escherichia coli. *ACS Chem Biol* **2018**, *13* (10), 2966-2972. DOI: 10.1021/acschembio.8b00599.
(51) Sharma, Y.; Ahlawat, S.; Rao, A. Biochemical characterization of an inverting S/O-HexNAc-transferase and evidence of S-linked glycosylation in Actinobacteria. *Glycobiology* **2022**, *32* (2), 148-161. DOI: 10.1093/glycob/cwab089.
(52) Ortiz-Lopez, F. J.; Carretero-Molina, D.; Sanchez-Hidalgo, M.; Martin, J.; Gonzalez, I.; Roman-Hurtado, F.; de la Cruz, M.; Garcia-Fernandez, S.; Reyes, F.; Deisinger, J. P.; et al. Cacaoidin, first member of the new lanthidin RiPP family. *Angew Chem Int Ed Engl* **2020**. DOI: 10.1002/anie.202005187.
(53) Iorio, M.; Sasso, O.; Maffioli, S. I.; Bertorelli, R.; Monciardini, P.; Sosio, M.; Bonezzi, F.; Summa, M.; Brunati, C.; Bordoni, R.; et al. A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. *ACS Chem Biol* **2014**, *9* (2), 398-404. DOI: 10.1021/cb400692w.
(54) Ahn, S.; Stepper, J.; Loo, T. S.; Bisset, S. W.; Patchett, M. L.; Norris, G. E. Expression of Lactobacillus plantarum KW30 gcc genes correlates with the production of glycocin F in late log phase. *FEMS Microbiol Lett* **2018**, *365* (23). DOI: 10.1093/femsle/fny261.
(55) Drower, K. R. The bacteriostatic diglycocylated bacteriocin glycocin F targets a sugar-specific transporter : a thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Biochemistry at Massey University, Manawatu, New Zealand. Masters, Massey University, 2014. http://hdl.handle.net/10179/6021.

(56) Dubois, J. Y.; Kouwen, T. R.; Schurich, A. K.; Reis, C. R.; Ensing, H. T.; Trip, E. N.; Zweers, J. C.; van Dijl, J. M. Immunity to the bacteriocin sublancin 168 Is determined by the SunI (YolF) protein of Bacillus subtilis. *Antimicrob Agents Chemother* **2009**, *53* (2), 651-661. DOI: 10.1128/AAC.01189-08.
(57) Maky, M. A.; Ishibashi, N.; Nakayama, J.; Zendo, T. Characterization of the Biosynthetic Gene Cluster of Enterocin F4-9, a Glycosylated Bacteriocin. *Microorganisms* **2021**, *9* (11). DOI: 10.3390/microorganisms9112276.
(58) Kelly, W. J.; Asmundson, R. V.; Huang, C. M. Characterization of plantaricin KW30, a bacteriocin produced by Lactobacillus plantarum. *Journal of Applied Bacteriology* **1996**, *81* (6), 657-662. DOI: https://doi.org/10.1111/j.1365-2672.1996.tb03561.x.

(59) Hata, T.; Tanaka, R.; Ohmomo, S. Isolation and characterization of plantaricin ASM1: A new bacteriocin produced by Lactobacillus plantarum A-1. *International Journal of Food Microbiology* **2010**, *137* (1), 94-99. DOI: https://doi.org/10.1016/j.ijfoodmicro.2009.10.021.
(60) Venugopal, H.; Edwards, P. J.; Schwalbe, M.; Claridge, J. K.; Libich, D. S.; Stepper, J.; Loo, T.; Patchett, M. L.; Norris, G. E.; Pascal, S. M. Structural, dynamic, and chemical characterization of a novel S-glycosylated bacteriocin. *Biochemistry* **2011**, *50* (14), 2748-2755. DOI: 10.1021/bi200217u.
(61) Bisset, S. W.; Yang, S. H.; Amso, Z.; Harris, P. W. R.; Patchett, M. L.; Brimble, M. A.; Norris, G. E. Using Chemical Synthesis to Probe Structure-Activity Relationships of the Glycoactive Bacteriocin Glycocin F. *ACS Chem Biol* **2018**, *13* (5), 1270-1278. DOI: 10.1021/acschembio.8b00055.
(62) Amso, Z.; Bisset, S. W.; Yang, S. H.; Harris, P. W. R.; Wright, T. H.; Navo, C. D.; Patchett, M. L.; Norris, G. E.; Brimble, M. A. Total chemical synthesis of glycocin F and analogues: S-glycosylation confers improved antimicrobial activity. *Chem Sci* **2018**, *9* (6), 1686-1691. DOI: 10.1039/c7sc04383j.
(63) Hemphill, H. E.; Gage, I.; Zahler, S. A.; Korman, R. Z. Prophage-mediated production of a bacteriocinlike substance by SP beta lysogens of Bacillus subtilis. *Can J Microbiol* **1980**, *26* (11), 1328-1333. DOI: 10.1139/m80-220.
(64) Garcia De Gonzalo, C. V.; Zhu, L.; Oman, T. J.; van der Donk, W. A. NMR structure of the S-linked glycopeptide sublancin 168. *ACS Chem Biol* **2014**, *9* (3), 796-801. DOI: 10.1021/cb4008106.
(65) Lazarevic, V.; Dusterhoft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D. Nucleotide sequence of the Bacillus subtilis temperate bacteriophage SPbetac2. *Microbiology (Reading)* **1999**, *145 (Pt 5)*, 1055-1067. DOI: 10.1099/13500872-145-5-1055.
(66) Paik, S. H.; Chakicherla, A.; Hansen, J. N. Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by Bacillus subtilis 168. *J Biol Chem* **1998**, *273* (36), 23134-23142. DOI: 10.1074/jbc.273.36.23134.
(67) Wu, C.; Biswas, S.; Garcia De Gonzalo, C. V.; van der Donk, W. A. Investigations into the Mechanism of Action of Sublancin. *ACS Infect Dis* **2019**, *5* (3), 454-459. DOI: 10.1021/acsinfecdis.8b00320.
(68) Biswas, S.; Wu, C.; van der Donk, W. A. The Antimicrobial Activity of the Glycocin Sublancin Is Dependent on an Active Phosphoenolpyruvate-Sugar Phosphotransferase System. *ACS Infect Dis* **2021**. DOI: 10.1021/acsinfecdis.1c00157.
(69) Kouwen, T. R. H. M.; Trip, E. N.; Denham, E. L.; Sibbald, M. J. J. B.; Dubois, J.-Y. F.; van Dijl, J. M. The Large Mechanosensitive Channel MscL Determines Bacterial Susceptibility to the Bacteriocin Sublancin 168. *Antimicrobial Agents and Chemotherapy* **2009**, *53* (11), 4702-4711. DOI: 10.1128/aac.00439-09.
(70) Garcia De Gonzalo, C. V.; Denham, E. L.; Mars, R. A.; Stulke, J.; van der Donk, W. A.; van Dijl, J. M. The phosphoenolpyruvate:sugar phosphotransferase system is involved in sensitivity to the glucosylated bacteriocin sublancin. *Antimicrob Agents Chemother* **2015**, *59* (11), 6844-6854. DOI: 10.1128/aac.01519-15.
(71) Wang, S.; Wang, Q.; Zeng, X.; Ye, Q.; Huang, S.; Yu, H.; Yang, T.; Qiao, S. Use of the Antimicrobial Peptide Sublancin with Combined Antibacterial and Immunomodulatory Activities To Protect against Methicillin-Resistant Staphylococcus aureus Infection in Mice. *J Agric Food Chem* **2017**, *65* (39), 8595-8605. DOI: 10.1021/acs.jafc.7b02592. Wang, S.; Ye, Q.; Wang, K.; Zeng, X.; Huang, S.; Yu, H.; Ge, Q.; Qi, D.; Qiao, S. Enhancement of Macrophage Function by the Antimicrobial Peptide Sublancin Protects Mice from Methicillin-Resistant Staphylococcus aureus. *J Immunol Res* **2019**, *2019*, 3979352. DOI: 10.1155/2019/3979352.
(72) Izquierdo, E.; Wagner, C.; Marchioni, E.; Aoude-Werner, D.; Ennahar, S. Enterocin 96, a novel class II bacteriocin produced by Enterococcus faecalis WHE 96, isolated from Munster cheese. *Appl Environ Microbiol* **2009**, *75* (13), 4273-4276. DOI: 10.1128/AEM.02772-08.
(73) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **2015**, *12* (1), 7-8. DOI: 10.1038/nmeth.3213.
(74) Wang, H.; van der Donk, W. A. Substrate selectivity of the sublancin S-glycosyltransferase. *J Am Chem Soc* **2011**, *133* (41), 16394-16397. DOI: 10.1021/ja2075168.
(75) Choudhary, P.; Rao, A. SELECT-GLYCOCIN: a recombinant microbial system for expression and high-throughput screening of glycocins. *Glycoconj J* **2021**, *38* (2), 233-250. DOI: 10.1007/s10719-020-09960-w.
(76) Hanchi, H.; Hammami, R.; Gingras, H.; Kourda, R.; Bergeron, M. G.; Ben Hamida, J.; Ouellette, M.; Fliss, I. Inhibition of MRSA and of Clostridium difficile by durancin 61A: synergy with bacteriocins and antibiotics. *Future Microbiol* **2017**, *12*, 205-212. DOI: 10.2217/fmb-2016-0113.
(77) Lafite, P.; Daniellou, R. Rare and unusual glycosylation of peptides and proteins. *Nat Prod Rep* **2012**, *29* (7), 729-738. DOI: 10.1039/c2np20030a.
(78) *Essentials in glycobiology*; Cold Spring Harbor Laboratory Press, 2022. DOI: 10.1101/9781621824213.

(79) Shental-Bechor, D.; Levy, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences* **2008**, *105* (24), 8256-8261. DOI: doi:10.1073/pnas.0801340105.
(80) Knop, M.; Hauser, N.; Wolf, D. H. N-glycosylation affects endoplasmic reticulum degradation of a mutated derivative of carboxypeptidase yscY in yeast. *Yeast* **1996**, *12* (12), 1229-1238. DOI: https://doi.org/10.1002/(SICI)1097-0061(19960930)12:12<1229::AID-YEA15>3.0.CO;2-H.
(81) Ferreira, I. G.; Pucci, M.; Venturi, G.; Malagolini, N.; Chiricolo, M.; Dall'Olio, F. Glycosylation as a Main Regulator of Growth and Death Factor Receptors Signaling. *Int J Mol Sci* **2018**, *19* (2). DOI: 10.3390/ijms19020580.
(82) Varki, A. Biological roles of glycans. *Glycobiology* **2017**, *27* (1), 3-49. DOI: 10.1093/glycob/cww086.
(83) Guerry, P.; Szymanski, C. M. Campylobacter sugars sticking out. *Trends Microbiol* **2008**, *16* (9), 428-435. DOI: 10.1016/j.tim.2008.07.002.
(84) Schäffer, C.; Messner, P. Emerging facets of prokaryotic glycosylation. *FEMS Microbiology Reviews* **2017**, *41* (1), 49-91. DOI: 10.1093/femsre/fuw036 (acccessed 10/9/2023).
(85) Singh, D. G.; Lomako, J.; Lomako, W. M.; Whelan, W. J.; Meyer, H. E.; Serwe, M.; Metzger, J. W. beta-Glucosylarginine: a new glucose-protein bond in a self-glucosylating protein from sweet corn. *FEBS Lett* **1995**, *376* (1-2), 61-64. DOI: 10.1016/0014-5793(95)01247-6.
(86) Lassak, J.; Keilhauer, E. C.; Fürst, M.; Wuichet, K.; Gödeke, J.; Starosta, A. L.; Chen, J. M.; Søgaard-Andersen, L.; Rohr, J.; Wilson, D. N.; et al. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat Chem Biol* **2015**, *11* (4), 266-270. DOI: 10.1038/nchembio.1751.
(87) Li, S.; Zhang, L.; Yao, Q.; Li, L.; Dong, N.; Rong, J.; Gao, W.; Ding, X.; Sun, L.; Chen, X.; et al. Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. *Nature* **2013**, *501* (7466), 242-246. DOI: 10.1038/nature12436. Pearson, J. S.; Giogha, C.; Ong, S. Y.; Kennedy, C. L.; Kelly, M.; Robinson, K. S.; Lung, T. W. F.; Mansell, A.; Riedmaier, P.; Oates, C. V. L.; et al. A type III effector antagonizes death receptor signalling during bacterial gut infection. *Nature* **2013**, *501* (7466), 247-251. DOI: 10.1038/nature12524.
(88) Bewley, C. A.; Faulkner, D. J. Theonegramide, an Antifungal Glycopeptide from the Philippine Lithistid Sponge Theonella swinhoei. *The Journal of Organic Chemistry* **1994**, *59* (17), 4849-4852. DOI: 10.1021/jo00096a028.
(89) He, H.; Williamson, R. T.; Shen, B.; Graziani, E. I.; Yang, H. Y.; Sakya, S. M.; Petersen, P. J.; Carter, G. T. Mannopeptimycins, Novel Antibacterial Glycopeptides from Streptomyces hygroscopicus, LL-AC98. *Journal of the American Chemical Society* **2002**, *124* (33), 9729-9736. DOI: 10.1021/ja020257s.
(90) Li, J. S.; Cui, L.; Rock, D. L.; Li, J. Novel glycosidic linkage in Aedes aegypti chorion peroxidase: N-mannosyl tryptophan. *J Biol Chem* **2005**, *280* (46), 38513-38521. DOI: 10.1074/jbc.M508449200.
(91) Wandall, H. H.; Nielsen, M. A. I.; King-Smith, S.; de Haan, N.; Bagdonaite, I. Global functions of O-glycosylation: promises and challenges in O-glycobiology. *FEBS J* **2021**, *288* (24), 7183-7212. DOI: 10.1111/febs.16148.
(92) Rodriguez, I. R.; Whelan, W. J. A novel glycosyl-amino acid linkage: Rabbit-muscle glycogen is covalently linked to a protein via tyrosine. *Biochemical and Biophysical Research Communications* **1985**, *132* (2), 829-836. DOI: https://doi.org/10.1016/0006-291X(85)91206-9.
(93) Halim, A.; Brinkmalm, G.; Ruetschi, U.; Westman-Brinkmalm, A.; Portelius, E.; Zetterberg, H.; Blennow, K.; Larson, G.; Nilsson, J. Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. *Proc Natl Acad Sci U S A* **2011**, *108* (29), 11848-11853. DOI: 10.1073/pnas.1102664108.
(94) Trinidad, J. C.; Schoepfer, R.; Burlingame, A. L.; Medzihradszky, K. F. N- and O-glycosylation in the murine synaptosome. *Mol Cell Proteomics* **2013**, *12* (12), 3474-3488. DOI: 10.1074/mcp.M113.030007.
(95) Zarschler, K.; Janesch, B.; Pabst, M.; Altmann, F.; Messner, P.; Schäffer, C. Protein tyrosine O-glycosylation—A rather unexplored prokaryotic glycosylation system. *Glycobiology* **2010**, *20* (6), 787-798. DOI: 10.1093/glycob/cwq035 (acccessed 10/9/2023).
(96) Steiner, K.; Pohlentz, G.; Dreisewerd, K.; Berkenkamp, S.; Messner, P.; Peter-Katalinić, J.; Schäffer, C. New insights into the glycosylation of the surface layer protein SgsE from Geobacillus stearothermophilus NRS 2004/3a. *J Bacteriol* **2006**, *188* (22), 7914-7921. DOI: 10.1128/jb.00802-06.
(97) Harris, C. M.; Kopecka, H.; Harris, T. M. Vancomycin: structure and transformation to CDP-I. *Journal of the American Chemical Society* **1983**, *105* (23), 6915-6922. DOI: 10.1021/ja00361a029.
(98) Plaza, A.; Gustchina, E.; Baker, H. L.; Kelly, M.; Bewley, C. A. Mirabamides A–D, Depsipeptides from the Sponge Siliquariaspongia mirabilis That Inhibit HIV-1 Fusion. *Journal of Natural Products* **2007**, *70* (11), 1753-1760. DOI: 10.1021/np070306k.

(99) Minakata, S.; Manabe, S.; Inai, Y.; Ikezaki, M.; Nishitsuji, K.; Ito, Y.; Ihara, Y. Protein C-Mannosylation and C-Mannosyl Tryptophan in Chemical Biology and Medicine. *Molecules* **2021**, *26* (17). DOI: 10.3390/molecules26175258.
(100) Shcherbakova, A.; Tiemann, B.; Buettner, F. F.; Bakker, H. Distinct C-mannosylation of netrin receptor thrombospondin type 1 repeats by mammalian DPY19L1 and DPY19L3. *Proc Natl Acad Sci U S A* **2017**, *114* (10), 2574-2579. DOI: 10.1073/pnas.1613165114.
(101) Lote, C. J.; Weiss, J. B. Identification in urine of a low-molecular-weight highly polar glycopeptide containing cysteinyl-galactose. *Biochem J* **1971**, *123* (4), 25P. DOI: 10.1042/bj1230025pa.
(102) Weiss, J. B.; Lote, C. J.; Bobinski, H. New low molecular weight glycopeptide containing triglucosylcysteine in human erythrocyte membrane. *Nat New Biol* **1971**, *234* (44), 25-26. DOI: 10.1038/newbio234025a0.
(103) Maynard, J. C.; Burlingame, A. L.; Medzihradszky, K. F. Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), A New Post-translational Modification in Mammals. *Mol Cell Proteomics* **2016**, *15* (11), 3405-3411. DOI: 10.1074/mcp.M116.061549.
(104) Drula, E.; Garron, M. L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **2022**, *50* (D1), D571-d577. DOI: 10.1093/nar/gkab1045.
(105) Coutinho, P. M.; Deleury, E.; Davies, G. J.; Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **2003**, *328* (2), 307-317. DOI: 10.1016/s0022-2836(03)00307-3.
(106) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **2008**, *77*, 521-555. DOI: 10.1146/annurev.biochem.76.061005.092322.
(107) Mestrom, L.; Przypis, M.; Kowalczykiewicz, D.; Pollender, A.; Kumpf, A.; Marsden, S. R.; Bento, I.; Jarzebski, A. B.; Szymanska, K.; Chrusciel, A.; et al. Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. *Int J Mol Sci* **2019**, *20* (21). DOI: 10.3390/ijms20215263.
(108) Cardini, C. E.; Paladini, A. C.; Caputto, R.; Leloir, L. F. Uridine Diphosphate Glucose: the Coenzyme of the Galactose–Glucose Phosphate Isomerization. *Nature* **1950**, *165* (4188), 191-192. DOI: 10.1038/165191a0.
(109) Mikkola, S. Nucleotide Sugars in Chemistry and Biology. *Molecules* **2020**, *25* (23). DOI: 10.3390/molecules25235755.
(110) Taujale, R.; Zhou, Z.; Yeung, W.; Moremen, K. W.; Li, S.; Kannan, N. Mapping the glycosyltransferase fold landscape using interpretable deep learning. *Nat Commun* **2021**, *12* (1), 5656. DOI: 10.1038/s41467-021-25975-9.
(111) Chang, A.; Singh, S.; Phillips, G. N., Jr.; Thorson, J. S. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol* **2011**, *22* (6), 800-808. DOI: 10.1016/j.copbio.2011.04.013.
(112) Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol* **2014**, *28*, 131-141. DOI: 10.1016/j.sbi.2014.08.012.
(113) Breton, C.; Snajdrova, L.; Jeanneau, C.; Koca, J.; Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **2006**, *16* (2), 29R-37R. DOI: 10.1093/glycob/cwj016.
(114) Qasba, P. K.; Ramakrishnan, B.; Boeggeman, E. Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem Sci* **2005**, *30* (1), 53-62. DOI: 10.1016/j.tibs.2004.11.005.
(115) Osawa, T.; Sugiura, N.; Shimada, H.; Hirooka, R.; Tsuji, A.; Shirakawa, T.; Fukuyama, K.; Kimura, M.; Kimata, K.; Kakuta, Y. Crystal structure of chondroitin polymerase from Escherichia coli K4. *Biochemical and Biophysical Research Communications* **2009**, *378* (1), 10-14. DOI: https://doi.org/10.1016/j.bbrc.2008.08.121. Guerin, M. E.; Kordulakova, J.; Schaeffer, F.; Svetlikova, Z.; Buschiazzo, A.; Giganti, D.; Gicquel, B.; Mikusova, K.; Jackson, M.; Alzari, P. M. Molecular Recognition and Interfacial Catalysis by the Essential Phosphatidylinositol Mannosyltransferase PimA from Mycobacteria*. *Journal of Biological Chemistry* **2007**, *282* (28), 20705-20714. DOI: https://doi.org/10.1074/jbc.M702087200.
(116) Pesnot, T.; Jorgensen, R.; Palcic, M. M.; Wagner, G. K. Structural and mechanistic basis for a new mode of glycosyltransferase inhibition. *Nat Chem Biol* **2010**, *6* (5), 321-323. DOI: 10.1038/nchembio.343. Schuman, B.; Persson, M.; Landry, R. C.; Polakowski, R.; Weadge, J. T.; Seto, N. O.; Borisova, S. N.; Palcic, M. M.; Evans, S. V. Cysteine-to-serine mutants dramatically reorder the active site of human ABO(H) blood group B glycosyltransferase without affecting activity: structural insights into cooperative substrate binding. *J Mol Biol* **2010**, *402* (2), 399-411. DOI: 10.1016/j.jmb.2010.07.036.
(117) Qasba, P. K.; Ramakrishnan, B.; Boeggeman, E. Structure and function of beta -1,4-galactosyltransferase. *Curr Drug Targets* **2008**, *9* (4), 292-309. DOI: 10.2174/138945008783954943. Lü, W.; Du, J.; Stahl, M.; Tzivelekidis, T.; Belyi, Y.; Gerhardt, S.; Aktories, K.; Einsle, O. Structural Basis of the Action of Glucosyltransferase Lgt1 from Legionella pneumophila. *Journal of Molecular Biology* **2010**, *396* (2), 321-331. DOI: https://doi.org/10.1016/j.jmb.2009.11.044.

(118) Nidetzky, B.; Gutmann, A.; Zhong, C. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. *ACS Catalysis* **2018**, *8* (7), 6283-6300. DOI: 10.1021/acscatal.8b00710.
(119) Quirós, L. M.; Carbajo, R. J.; Braña, A. F.; Salas, J. A. Glycosylation of macrolide antibiotics. Purification and kinetic studies of a macrolide glycosyltransferase from Streptomyces antibioticus. *J Biol Chem* **2000**, *275* (16), 11713-11720. DOI: 10.1074/jbc.275.16.11713.
(120) Schimpl, M.; Zheng, X.; Borodkin, V. S.; Blair, D. E.; Ferenbach, A. T.; Schüttelkopf, A. W.; Navratilova, I.; Aristotelous, T.; Albarbarawi, O.; Robinson, D. A.; et al. O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nature Chemical Biology* **2012**, *8* (12), 969-974. DOI: 10.1038/nchembio.1108.
(121) Kozmon, S.; Tvaroska, I. Catalytic mechanism of glycosyltransferases: hybrid quantum mechanical/molecular mechanical study of the inverting N-acetylglucosaminyltransferase I. *J Am Chem Soc* **2006**, *128* (51), 16921-16927. DOI: 10.1021/ja065944o. Krupicka, M.; Tvaroska, I. Hybrid quantum mechanical/molecular mechanical investigation of the beta-1,4-galactosyltransferase-I mechanism. *J Phys Chem B* **2009**, *113* (32), 11314-11319. DOI: 10.1021/jp904716t.
(122) Tvaroška, I. Atomistic insight into the catalytic mechanism of glycosyltransferases by combined quantum mechanics/molecular mechanics (QM/MM) methods. *Carbohydrate Research* **2015**, *403*, 38-47. DOI: https://doi.org/10.1016/j.carres.2014.06.017.
(123) Nunez, H. A.; Barker, R. The metal ion catalyzed decomposition of nucleoside diphosphate sugars. *Biochemistry* **1976**, *15* (17), 3843-3847. DOI: 10.1021/bi00662a030.
(124) Moréra, S.; Imberty, A.; Aschke-Sonnenborn, U.; Rüger, W.; Freemont, P. S. T4 phage beta-glucosyltransferase: substrate binding and proposed catalytic mechanism. *J Mol Biol* **1999**, *292* (3), 717-730. DOI: 10.1006/jmbi.1999.3094.
(125) Zhang, C.; Griffith, B. R.; Fu, Q.; Albermann, C.; Fu, X.; Lee, I. K.; Li, L.; Thorson, J. S. Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **2006**, *313* (5791), 1291-1294. DOI: 10.1126/science.1130028.
(126) Fan, C.; Deng, Q.; Zhu, T. F. Bioorthogonal information storage in L-DNA with a high-fidelity mirror-image Pfu DNA polymerase. *Nat Biotechnol* **2021**, *39* (12), 1548-1555. DOI: 10.1038/s41587-021-00969-6.
(127) Apptech Technical Notes. Apptech, complete peptide product source. https://www.peptide.com/resources/solid-phase-peptide-synthesis/ (accessed August 2023).
(128) Yang, Y. Chapter 6 - Intramolecular Cyclization Side Reactions. In *Side Reactions in Peptide Synthesis*, Yang, Y. Ed.; Academic Press, 2016; pp 119-161.
(129) Sandmeier, E.; Hunziker, P.; Kunz, B.; Sack, R.; Christen, P. Spontaneous Deamidation and Isomerization of Asn108 in Prion Peptide 106–126 and in Full-Length Prion Protein. *Biochemical and Biophysical Research Communications* **1999**, *261* (3), 578-583. DOI: https://doi.org/10.1006/bbrc.1999.1056.
(130) Zhu, J.; Marchant, R. E. Solid-phase synthesis of tailed cyclic RGD peptides using glutamic acid: unexpected glutarimide formation. *Journal of Peptide Science* **2008**, *14* (6), 690-696. DOI: https://doi.org/10.1002/psc.975.
(131) Grzetic, J.; Oomens, J. Spectroscopic Identification of Cyclic Imide b2-Ions from Peptides Containing Gln and Asn Residues. *J Am Soc Mass Spectrom* **2013**, *24* (8), 1228-1241. DOI: 10.1007/s13361-013-0661-6.
(132) Bodanszky, M.; Sigler, G. F.; Bodanszky, A. Structure of the peptide antibiotic amphomycin. *Journal of the American Chemical Society* **1973**, *95* (7), 2352-2357. DOI: 10.1021/ja00788a040.
(133) Radkiewicz, J. L.; Zipse, H.; Clarke, S.; Houk, K. N. Accelerated Racemization of Aspartic Acid and Asparagine Residues via Succinimide Intermediates: An ab Initio Theoretical Exploration of Mechanism. *Journal of the American Chemical Society* **1996**, *118* (38), 9148-9155. DOI: 10.1021/ja953505b.
(134) Yang, Y. Doctoral Dissertation, Molecular recognition of integrin a3b1 and inorganic compounds by tailor-made peptides. Bielefeld University, 2008, 2008.
(135) Mergler, M.; Dick, F.; Sax, B.; Weiler, P.; Vorherr, T. The aspartimide problem in Fmoc-based SPPS. Part I. *J Pept Sci* **2003**, *9* (1), 36-46. DOI: 10.1002/psc.430.
(136) Yang, Y.; Sweeney, W. V.; Schneider, K.; Thörnqvist, S.; Chait, B. T.; Tam, J. P. Aspartimide formation in base-driven 9-fluorenylmethoxycarbonyl chemistry. *Tetrahedron Letters* **1994**, *35* (52), 9689-9692. DOI: https://doi.org/10.1016/0040-4039(94)88360-2.
(137) Palasek, S. A.; Cox, Z. J.; Collins, J. M. Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. *Journal of Peptide Science* **2007**, *13* (3), 143-148. DOI: https://doi.org/10.1002/psc.804.
(138) Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (HE-SPPS). *Organic Letters* **2014**, *16* (3), 940-943. DOI: 10.1021/ol4036825.

(139) Souza, M. P.; Tavares, M. F. M.; Miranda, M. T. M. Racemization in stepwise solid-phase peptide synthesis at elevated temperatures. *Tetrahedron* **2004**, *60* (21), 4671-4681. DOI: https://doi.org/10.1016/j.tet.2004.03.070.
(140) Jones, J. H.; Ramage, W. I.; Witty, M. J. Mechanism of racemisation of histidine derivatives in peptide synthesis. *Int J Pept Protein Res* **1980**, *15* (3), 301-303. DOI: 10.1111/j.1399-3011.1980.tb02581.x.
(141) Yang, Y. Chapter 11 - Peptide Racemization. In *Side Reactions in Peptide Synthesis*, Yang, Y. Ed.; Academic Press, 2016; pp 257-292.
(142) Kovacs, J.; Mayers, G. L.; Johnson, R. H.; Cover, R. E.; Ghatak, U. R. Racemization of amino acid derivatives. III. Rate of racemization and peptide bond formation of cysteine active esters. *The Journal of Organic Chemistry* **1970**, *35* (6), 1810-1815. DOI: 10.1021/jo00831a021.
(143) Han, Y.; Albericio, F.; Barany, G. Occurrence and Minimization of Cysteine Racemization during Stepwise Solid-Phase Peptide Synthesis1,2. *The Journal of Organic Chemistry* **1997**, *62* (13), 4307-4312. DOI: 10.1021/jo9622744.
(144) Boulègue, C.; Musiol, H. J.; Prasad, V.; Moroder, L. Synthesis of cystine-rich peptides. *Chimica Oggi* **2006**, *24* (4), 24-36, Article. Scopus.
(145) Katsoyannis, P. G.; Tometsko, A. M.; Zalut, C.; Fukuda, K. Insulin Peptides. XV. The Synthesis of the A Chain of Sheep Insulin and Its Combination with Synthetic or Natural B Chain to Produce Insulin1,2. *Journal of the American Chemical Society* **1966**, *88* (23), 5625-5635. DOI: 10.1021/ja00975a050. Hiskey, R. G.; Upham, R. A.; Beverly, G. M.; Jones, W. C., Jr. Sulfur-containing polypeptides. X. .beta.-Elimination of mercaptides from cysteine peptides. *The Journal of Organic Chemistry* **1970**, *35* (2), 513-515. DOI: 10.1021/jo00827a053.
(146) Wang, Z.; Rejtar, T.; Zhou, Z. S.; Karger, B. L. Desulfurization of cysteine-containing peptides resulting from sample preparation for protein characterization by mass spectrometry. *Rapid Communications in Mass Spectrometry* **2010**, *24* (3), 267-275. DOI: https://doi.org/10.1002/rcm.4383.
(147) Volkin, D. B.; Klibanov, A. M. Thermal destruction processes in proteins involving cystine residues. *Journal of Biological Chemistry* **1987**, *262* (7), 2945-2950. DOI: https://doi.org/10.1016/S0021-9258(18)61451-6.
(148) Norris, K.; Halstrøm, J.; Brunfeldt, K.; Pihlaja, K.; Schaumburg, K.; Ehrenberg, L. Stability of Methionyl Residues Towards Oxidation During Solid Phase Peptide Synthesis. *Acta Chemica Scandinavica - ACTA CHEM SCAND* **1971**, *25*, 945-954. DOI: 10.3891/acta.chem.scand.25-0945.
(149) Huang, H.; Rabenstein, D. L. A cleavage cocktail for methionine-containing peptides. *The Journal of Peptide Research* **1999**, *53* (5), 548-553. DOI: https://doi.org/10.1034/j.1399-3011.1999.00059.x.
(150) Nicolás, E.; Vilaseca, M.; Giralt, E. A study of the use of NH4I for the reduction of methionine sulfoxide in peptides containing cysteine and cystine. *Tetrahedron* **1995**, *51* (19), 5701-5710. DOI: https://doi.org/10.1016/0040-4020(95)00234-Y.
(151) King, D. S.; Fields, C. G.; Fields, G. B. A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int J Pept Protein Res* **1990**, *36* (3), 255-266. DOI: 10.1111/j.1399-3011.1990.tb00976.x.
(152) Albericio, F.; Kneib-Cordonier, N.; Biancalana, S.; Gera, L.; Masada, R. I.; Hudson, D.; Barany, G. Preparation and application of the 5-(4-(9-fluorenylmethyloxycarbonyl)aminomethyl-3,5-dimethoxyphenoxy)-valeric acid (PAL) handle for the solid-phase synthesis of C-terminal peptide amides under mild conditions. *The Journal of Organic Chemistry* **1990**, *55* (12), 3730-3743. DOI: 10.1021/jo00299a011.
(153) Sieber, P. Modification of tryptophan residues during acidolysis of 4-methoxy-2,3,6-trimethylbenzenesulfonyl groups. Effects of scavengers. *Tetrahedron Letters* **1987**, *28* (15), 1637-1640. DOI: https://doi.org/10.1016/S0040-4039(00)95379-6.
(154) Isidro-Llobet, A.; Álvarez, M.; Albericio, F. Amino Acid-Protecting Groups. *Chemical Reviews* **2009**, *109* (6), 2455-2504. DOI: 10.1021/cr800323s.
(155) Pearson, D. A.; Blanchette, M.; Baker, M. L.; Guindon, C. A. Trialkylsilanes as scavengers for the trifluoroacetic acid deblocking of protecting groups in peptide synthesis. *Tetrahedron Letters* **1989**, *30* (21), 2739-2742. DOI: https://doi.org/10.1016/S0040-4039(00)99113-5.
(156) Franzén, H.; Grehn, L.; Ragnarsson, U. Synthesis, properties, and use of Nin-Boc-tryptophan derivatives. *Journal of the Chemical Society, Chemical Communications* **1984**, (24), 1699-1700, 10.1039/C39840001699. DOI: 10.1039/C39840001699.
(157) White, P. Peptides, chemistry and biology. In *Twelfth American Peptide Symposium*, Leiden, The Netherlands, 1992; JA Smith, J. R., Ed.; pp 537–538.
(158) Giraud, M.; Cavelier, F.; Martinez, J. A side-reaction in the SPPS of Trp-containing peptides. *Journal of Peptide Science* **1999**, *5* (10), 457-461. DOI: https://doi.org/10.1002/(SICI)1099-1387(199910)5:10<457::AID-PSC215>3.0.CO;2-7.

(159) Yang, Y. Chapter 3 - Peptide Global Deprotection/Scavenger-Induced Side Reactions. In *Side Reactions in Peptide Synthesis*, Yang, Y. Ed.; Academic Press, 2016; pp 43-75.
(160) Stathopoulos, P.; Papas, S.; Pappas, C.; Mousis, V.; Sayyad, N.; Theodorou, V.; Tzakos, A. G.; Tsikaris, V. Side reactions in the SPPS of Cys-containing peptides. *Amino Acids* **2013**, *44* (5), 1357-1363. DOI: 10.1007/s00726-013-1471-7.
(161) Piszkiewicz, D.; Landon, M.; Smith, E. L. Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. *Biochemical and Biophysical Research Communications* **1970**, *40* (5), 1173-1178. DOI: https://doi.org/10.1016/0006-291X(70)90918-6.
(162) Hsieh, Y. S. Y.; Wilkinson, B. L.; O'Connell, M. R.; Mackay, J. P.; Matthews, J. M.; Payne, R. J. Synthesis of the Bacteriocin Glycopeptide Sublancin 168 and S-Glycosylated Variants. *Organic Letters* **2012**, *14* (7), 1910-1913. DOI: 10.1021/ol300557g.
(163) Brimble, M. A.; Edwards, P. J.; Harris, P. W.; Norris, G. E.; Patchett, M. L.; Wright, T. H.; Yang, S. H.; Carley, S. E. Synthesis of the antimicrobial S-linked glycopeptide, glycocin F. *Chemistry* **2015**, *21* (9), 3556-3561. DOI: 10.1002/chem.201405692.
(164) Yamashiro, D.; Blake, J.; Hao Li, C. The use of trifluoroethanol for improved coupling in solid-phase peptide synthesis. *Tetrahedron Letters* **1976**, *17* (18), 1469-1472. DOI: https://doi.org/10.1016/S0040-4039(00)71285-8.
(165) Bacsa, B.; Bősze, S.; Kappe, C. O. Direct Solid-Phase Synthesis of the β-Amyloid (1−42) Peptide Using Controlled Microwave Heating. *The Journal of Organic Chemistry* **2010**, *75* (6), 2103-2106. DOI: 10.1021/jo100136r.
(166) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **2013**, *30* (1), 108-160. DOI: 10.1039/c2np20085f.
(167) Drummond, B. J.; Loo, T. S.; Patchett, M. L.; Norris, G. E. Optimized Genetic Tools Allow the Biosynthesis of Glycocin F and Analogues Designed To Test the Roles of gcc Cluster Genes in Bacteriocin Production. *J Bacteriol* **2021**, *203* (7). DOI: 10.1128/jb.00529-20.
(168) Fujinami, D.; Garcia de Gonzalo, C. V.; Biswas, S.; Hao, Y.; Wang, H.; Garg, N.; Lukk, T.; Nair, S. K.; van der Donk, W. A. Structural and mechanistic investigations of protein S-glycosyltransferases. *Cell Chem Biol* **2021**. DOI: 10.1016/j.chembiol.2021.06.009.
(169) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2.
(170) *Superdex prep grade and prepacked HiLoad columns SIZE EXCLUSION CHROMATOGRAPHY. Technical notes*; Cytiva, 2020.
(171) Butt, T. R.; Edavettal, S. C.; Hall, J. P.; Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* **2005**, *43* (1), 1-9. DOI: 10.1016/j.pep.2005.03.016.
(172) Nagar, R.; Rao, A. In Vitro Synthesis of Bioactive Glycovariants of Enterocin 96, an Antimicrobial Peptide from Enterococcus faecalis. *Methods Mol Biol* **2019**, *1954*, 279-296. DOI: 10.1007/978-1-4939-9154-9_22.
(173) Yang, Y. Chapter 1 - Peptide Fragmentation/Deletion Side Reactions. In *Side Reactions in Peptide Synthesis*, Yang, Y. Ed.; Academic Press, 2016; pp 1-31.
(174) Jacobson, J.; Melander, W.; Vaisnys, G.; Horvath, C. Kinetic study on cis-trans proline isomerization by high-performance liquid chromatography. *The Journal of Physical Chemistry* **1984**, *88* (20), 4536-4542. DOI: 10.1021/j150664a018. Sui, Q.; Rabenstein, D. L. Cis/trans isomerization of proline peptide bonds in the backbone of cyclic disulfide-bridged peptides. *Peptide Science* **2018**, *110* (6), e24088. DOI: https://doi.org/10.1002/pep2.24088.
(175) Biswas, S.; Garcia De Gonzalo, C. V.; Repka, L. M.; van der Donk, W. A. Structure-Activity Relationships of the S-Linked Glycocin Sublancin. *ACS Chem Biol* **2017**, *12* (12), 2965-2969. DOI: 10.1021/acschembio.7b00819.
(176) Kapust, R. B.; Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **1999**, *8* (8), 1668-1674. DOI: https://doi.org/10.1110/ps.8.8.1668.
(177) Studier, F. W. Stable Expression Clones and Auto-Induction for Protein Production in E. coli. In *Structural Genomics: General Applications*, Chen, Y. W. Ed.; Humana Press, 2014; pp 17-32.
(178) Yan, E.-K.; Cao, H.-L.; Zhang, C.-Y.; Lu, Q.-Q.; Ye, Y.-J.; He, J.; Huang, L.-J.; Yin, D.-C. Cross-linked protein crystals by glutaraldehyde and their applications. *RSC Advances* **2015**, *5* (33), 26163-26174, 10.1039/C5RA01722J. DOI: 10.1039/C5RA01722J.
(179) Stoddard, B. Fred Hutchinson Cancer Research Center Personal communication.

(180) Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007**, *372* (3), 774-797. DOI: 10.1016/j.jmb.2007.05.022.
(181) Kubiak, R. J.; Yue, X.; Hondal, R. J.; Mihai, C.; Tsai, M.-D.; Bruzik, K. S. Involvement of the Arg−Asp−His Catalytic Triad in Enzymatic Cleavage of the Phosphodiester Bond. *Biochemistry* **2001**, *40* (18), 5422-5432. DOI: 10.1021/bi002371y.
(182) Vernet, T.; Tessier, D. C.; Chatellier, J.; Plouffe, C.; Lee, T. S.; Thomas, D. Y.; Storer, A. C.; Ménard, R. Structural and Functional Roles of Asparagine 175 in the Cysteine Protease Papain (∗). *Journal of Biological Chemistry* **1995**, *270* (28), 16645-16652. DOI: https://doi.org/10.1074/jbc.270.28.16645.
(183) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research* **2019**, *47* (W1), W402-W407. DOI: 10.1093/nar/gkz297 (acccessed 10/10/2023).
(184) Katayama, H.; Asahina, Y.; Hojo, H. Chemical synthesis of the S-linked glycopeptide, sublancin. *Journal of Peptide Science* **2011**, *17* (12), 818-821. DOI: https://doi.org/10.1002/psc.1406.
(185) Ji, S.; Li, W.; Baloch, A. R.; Wang, M.; Cao, B. Improved production of sublancin via introduction of three characteristic promoters into operon clusters responsible for this novel distinct glycopeptide biosynthesis. *Microbial Cell Factories* **2015**, *14* (1), 17. DOI: 10.1186/s12934-015-0201-0.
(186) Dorenbos, R.; Stein, T.; Kabel, J.; Bruand, C.; Bolhuis, A.; Bron, S.; Quax, W. J.; van Dijl, J. M. Thiol-Disulfide Oxidoreductases Are Essential for the Production of the Lantibiotic Sublancin 168*. *Journal of Biological Chemistry* **2002**, *277* (19), 16682-16688. DOI: https://doi.org/10.1074/jbc.M201158200.
(187) Nandhini, K. P.; Alhassan, M.; Veale, C. G. L.; Albericio, F.; de la Torre, B. G. Methionine-Containing Peptides: Avoiding Secondary Reactions in the Final Global Deprotection. *ACS Omega* **2023**, *8* (17), 15631-15637. DOI: 10.1021/acsomega.3c01058.
(188) McCalley, D. V. Effect of buffer on peak shape of peptides in reversed-phase high performance liquid chromatography. *Journal of Chromatography A* **2004**, *1038* (1), 77-84. DOI: https://doi.org/10.1016/j.chroma.2004.03.038.
(189) Guo, D. C.; Mant, C. T.; Hodges, R. S. Effects of ion-pairing reagents on the prediction of peptide retention in reversed-phase high-performance liquid chromatography. *J Chromatogr* **1987**, *386*, 205-222. DOI: 10.1016/s0021-9673(01)94598-4.
(190) Stine, W. B.; Dahlgren, K. N.; Krafft, G. A.; LaDu, M. J. In Vitro Characterization of Conditions for Amyloid-β Peptide Oligomerization and Fibrillogenesis*. *Journal of Biological Chemistry* **2003**, *278* (13), 11612-11622. DOI: https://doi.org/10.1074/jbc.M210207200.
(191) Pachahara, S. K.; Chaudhary, N.; Subbalakshmi, C.; Nagaraj, R. Hexafluoroisopropanol induces self-assembly of β-amyloid peptides into highly ordered nanostructures. *Journal of Peptide Science* **2012**, *18* (4), 233-241. DOI: https://doi.org/10.1002/psc.2391.
(192) Waugh, D. S. Making the most of affinity tags. *Trends in Biotechnology* **2005**, *23* (6), 316-320. DOI: https://doi.org/10.1016/j.tibtech.2005.03.012. Woestenenk, E. A.; Hammarström, M.; van den Berg, S.; Härd, T.; Berglund, H. His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors. *J Struct Funct Genomics* **2004**, *5* (3), 217-229. DOI: 10.1023/b:jsfg.0000031965.37625.0e
(193) Ali, A.; Happel, D.; Habermann, J.; Schoenfeld, K.; Macarrón Palacios, A.; Bitsch, S.; Englert, S.; Schneider, H.; Avrutina, O.; Fabritz, S.; et al. Sactipeptide Engineering by Probing the Substrate Tolerance of a Thioether-Bond-Forming Sactisynthase. *Angewandte Chemie International Edition* **2022**, *61* (45), e202210883. DOI: https://doi.org/10.1002/anie.202210883.
(194) Lazarus, M. B.; Nam, Y.; Jiang, J.; Sliz, P.; Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **2011**, *469* (7331), 564-567. DOI: 10.1038/nature09638.
(195) Meek, R. W.; Blaza, J. N.; Busmann, J. A.; Alteen, M. G.; Vocadlo, D. J.; Davies, G. J. Cryo-EM structure provides insights into the dimer arrangement of the O-linked β-N-acetylglucosamine transferase OGT. *Nature Communications* **2021**, *12* (1), 6508. DOI: 10.1038/s41467-021-26796-6.
(196) Sajed, T.; Marcu, A.; Ramirez, M.; Pon, A.; Guo, A. C.; Knox, C.; Wilson, M.; Grant, J. R.; Djoumbou, Y.; Wishart, D. S. ECMDB 2.0: A richer resource for understanding the biochemistry of E. coli. *Nucleic Acids Research* **2015**, *44* (D1), D495-D501. DOI: 10.1093/nar/gkv1060 (acccessed 10/12/2023).
(197) Sidor, K.; Skirecki, T. A Bittersweet Kiss of Gram-Negative Bacteria: The Role of ADP-Heptose in the Pathogenesis of Infection. *Microorganisms* **2023**, *11* (5), 1316.
(198) Shen, D. L.; Gloster, T. M.; Yuzwa, S. A.; Vocadlo, D. J. Insights into O-linked N-acetylglucosamine ([0-9]O-GlcNAc) processing and dynamics through kinetic analysis of O-GlcNAc transferase and O-GlcNAcase activity on protein substrates. *J Biol Chem* **2012**, *287* (19), 15395-15408. DOI: 10.1074/jbc.M111.310664.
(199) CRC Handbook of Chemistry and Physics: A Ready-Reference of Chemical and Physical Data, 85th ed Edited by David R. Lide (National Institute of Standards and Technology). CRC Press LLC: Boca Raton, FL. 2004. 2712 pp.

$139.99. ISBN 0-8493-0485-7. *Journal of the American Chemical Society* **2005**, *127* (12), 4542-4542. DOI: 10.1021/ja041017a.
(200) Bruice, T. C.; Fife, T. H.; Bruno, J. J.; Brandon, N. E. Hydroxyl Group Catalysis. II. The Reactivity of the Hydroxyl Group of Serine. The Nucleophilicity of Alcohols and the Ease of Hydrolysis of Their Acetyl Esters as Related to Their pKa'*. *Biochemistry* **1962**, *1* (1), 7-12. DOI: 10.1021/bi00907a002.
(201) Davies, C.; Heath, R. J.; White, S. W.; Rock, C. O. The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from Escherichia coli. *Structure* **2000**, *8* (2), 185-195. DOI: https://doi.org/10.1016/S0969-2126(00)00094-0.
(202) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421. DOI: 10.1186/1471-2105-10-421.
(203) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **2018**, *35* (6), 1547-1549. DOI: 10.1093/molbev/msy096 (acccessed 10/13/2023).
(204) Fogg, M. J.; Wilkinson, A. J. Higher-throughput approaches to crystallization and crystal structure determination. *Biochem Soc Trans* **2008**, *36* (Pt 4), 771-775. DOI: 10.1042/BST0360771.
(205) Inoue, H.; Nojima, H.; Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **1990**, *96* (1), 23-28. DOI: https://doi.org/10.1016/0378-1119(90)90336-P.
(206) Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the Expasy Server. In *The Proteomics Protocols Handbook*, Walker, J. M. Ed.; Humana Press, 2005; pp 571-607.
(207) Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227* (5259), 680-685. DOI: 10.1038/227680a0.
(208) Lawrence, A. M.; Besir, H. U. Staining of proteins in gels with Coomassie G-250 without organic solvent and acetic acid. *J Vis Exp* **2009**, (30). DOI: 10.3791/1350.
(209) Walden, H. Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D* **2010**, *66* (4), 352-357. DOI: doi:10.1107/S0907444909038207.
(210) Wei, L.; Cai, X.; Qi, Z.; Rong, L.; Cheng, B.; Fan, J. In vivo and in vitro characterization of TEV protease mutants. *Protein Expr Purif* **2012**, *83* (2), 157-163. DOI: 10.1016/j.pep.2012.03.011.
(211) Muona, M.; Aranko, A. S.; Iwai, H. Segmental Isotopic Labelling of a Multidomain Protein by Protein Ligation by Protein Trans-Splicing. *ChemBioChem* **2008**, *9* (18), 2958-2961. DOI: https://doi.org/10.1002/cbic.200800604.
(212) Collins, J. M.; Singh, S. K. Use of excess carbodiimide for peptide synthesis at elevated temperatures. EP3290430A1, 2017.
(213) Andrushchenko, V. V.; Vogel, H. J.; Prenner, E. J. Optimization of the hydrochloric acid concentration used for trifluoroacetate removal from synthetic peptides. *J Pept Sci* **2007**, *13* (1), 37-43. DOI: 10.1002/psc.793
(214) Baykov, A. A.; Evtushenko, O. A.; Avaeva, S. M. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal Biochem* **1988**, *171* (2), 266-270. DOI: 10.1016/0003-2697(88)90484-8.
(215) Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (Pt 4), 271-281. DOI: 10.1107/s0907444910048675.
(216) Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; et al. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D* **2018**, *74* (2), 68-84. DOI: doi:10.1107/S2059798317016035.
(217) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **2011**, *67* (4), 235-242. DOI: doi:10.1107/S0907444910045749. Beilsten-Edmands, J.; Winter, G.; Gildea, R.; Parkhurst, J.; Waterman, D.; Evans, G. Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallographica Section D* **2020**, *76* (4), 385-399. DOI: doi:10.1107/S2059798320003198. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **2010**, *43* (1), 186-190. DOI: doi:10.1107/S0021889809045701. Winter, G.; Waterman, D. G.; Parkhurst, J. M.; Brewster, A. S.; Gildea, R. J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I. D.; et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D* **2018**, *74* (2), 85-97. DOI: doi:10.1107/S2059798317017235.

(218) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *Journal of Applied Crystallography* **2007**, *40* (4), 658-674. DOI: doi:10.1107/S0021889807021206.
(219) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nature Methods* **2022**, *19* (6), 679-682. DOI: 10.1038/s41592-022-01488-1.
(220) Cowtan, K. Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallogr D Biol Crystallogr* **2012**, *68* (Pt 4), 328-335. DOI: 10.1107/s0907444911039655. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **2006**, *62* (Pt 9), 1002-1011. DOI: 10.1107/s0907444906022116.
(221) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 4), 486-501. DOI: 10.1107/s0907444910007493.
(222) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* **2011**, *67* (4), 355-367. DOI: doi:10.1107/S0907444911001314.
(223) Long, F.; Nicholls, R. A.; Emsley, P.; Graæulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol* **2017**, *73* (Pt 2), 112-122. DOI: 10.1107/s2059798317000067.
(224) Agirre, J.; Iglesias-Fernández, J.; Rovira, C.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D. Privateer: software for the conformational validation of carbohydrate structures. *Nature Structural & Molecular Biology* **2015**, *22* (11), 833-834. DOI: 10.1038/nsmb.3115.
(225) Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **2014**, *1* (4), 213-220. DOI: doi:10.1107/S2052252514009324.
(226) McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D* **2011**, *67* (4), 386-394. DOI: doi:10.1107/S0907444911007281.
(227) Laskowski, R. A.; Jabłońska, J.; Pravda, L.; Vařeková, R. S.; Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Science* **2018**, *27* (1), 129-134. DOI: https://doi.org/10.1002/pro.3289.

8. Annex I

8.1. Peptide Data

8.1.1. SacA derived peptides

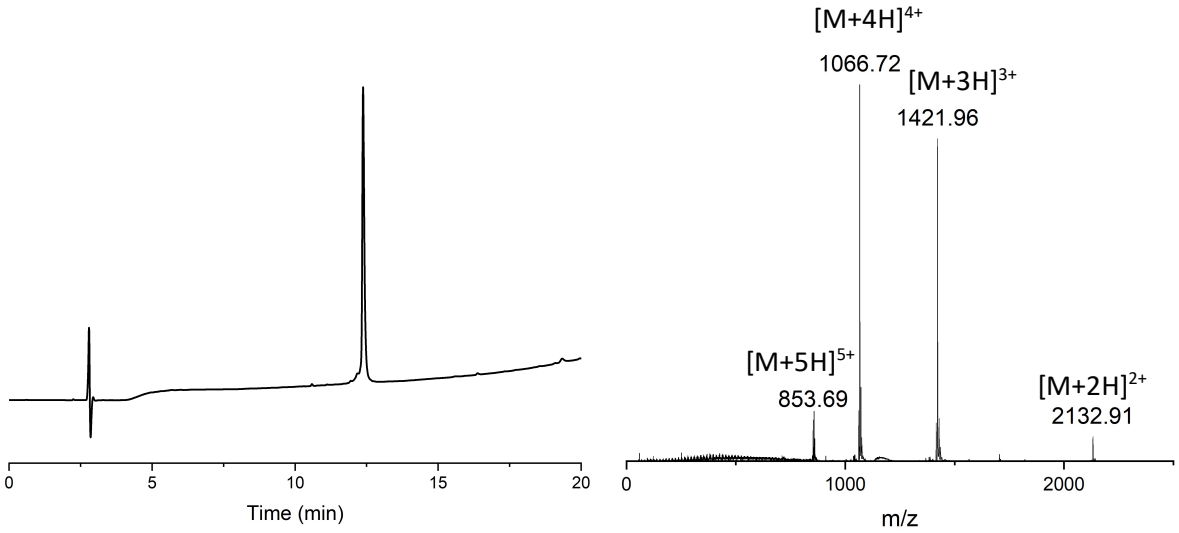


Figure 91 *Analytical data of the SacA peptide Retention time: 12.4 min (18 min linear gradient 10-80% MeCN). Deconvoluted mass: 4263.863, calculated mass: 4263.878*

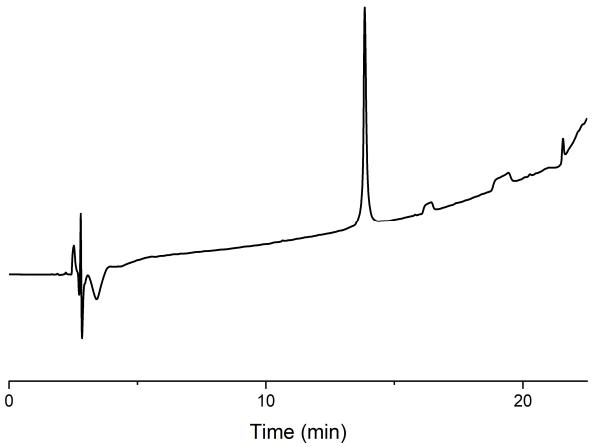


Figure 92 *Analytical data of the Leader-TEV-SacA peptide. Retention time: 13.85 min (linear gradient 10-80% MeCN)*

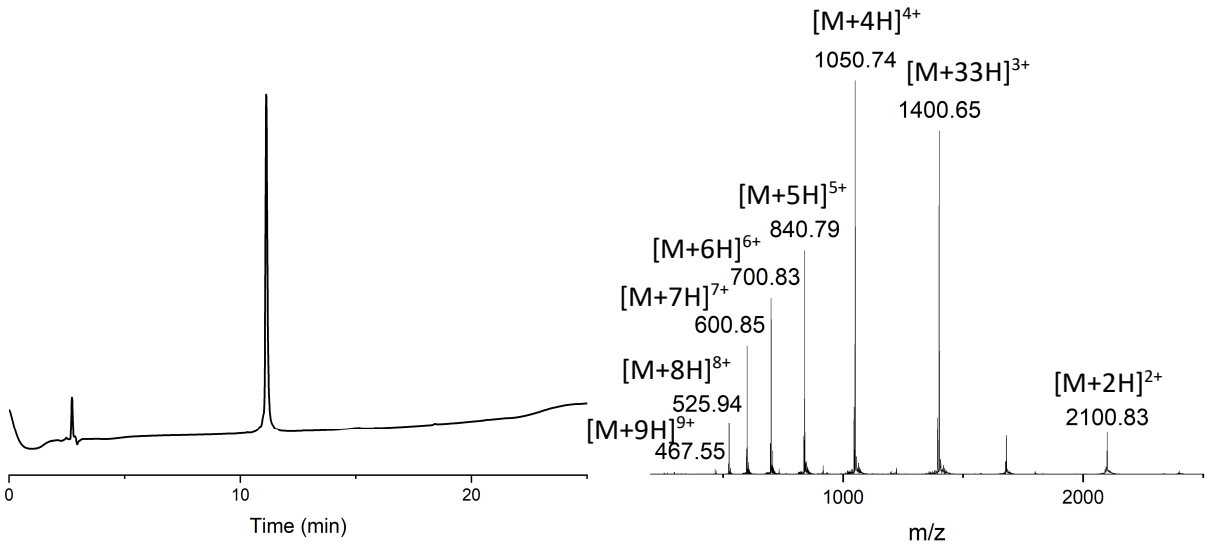


Figure 93 *Analytical data of the SAcA4Ser peptide. Retention time: 11.12 min (10-80 linear gradient MeCN in H_2O). Deconvoluted mass: 4198.919, calculated mass: 4199.6114*

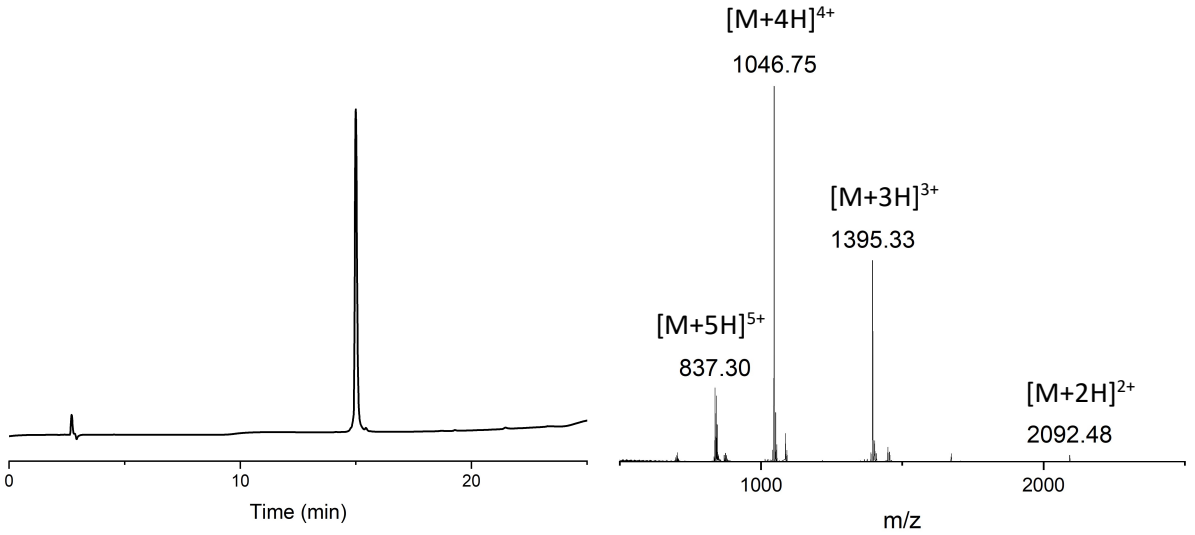


Figure 94 *Analytical data of the SacAallSer peptide. Retention Time: 15.0 min (5 min 10% MeCN isocratic, 18 min linear gradient 10-80% MeCN). Deconvoluted Mass: 4182.966, Calculated Mass: 4183.545*

8.1.2. SunA derived peptides

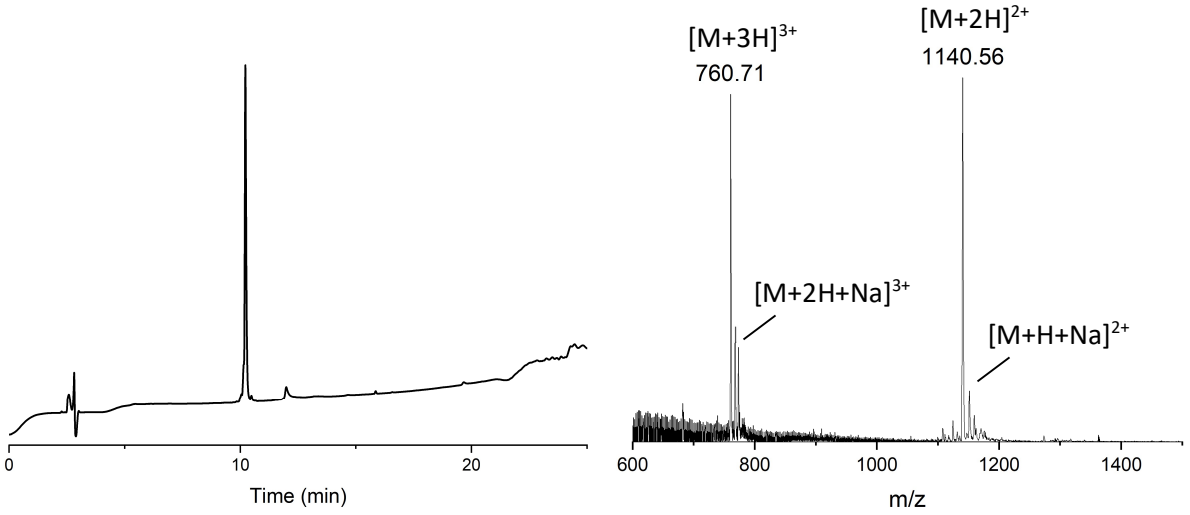


Figure 95 *Analytical data for the SunAm peptide. Retention time: 10.2 min (linear gradient 10-80% MeCN). Deconvoluted mass: 2279.107, calculated mass: 2279.651. Additional mass peaks correspond to sodium-adducts.*

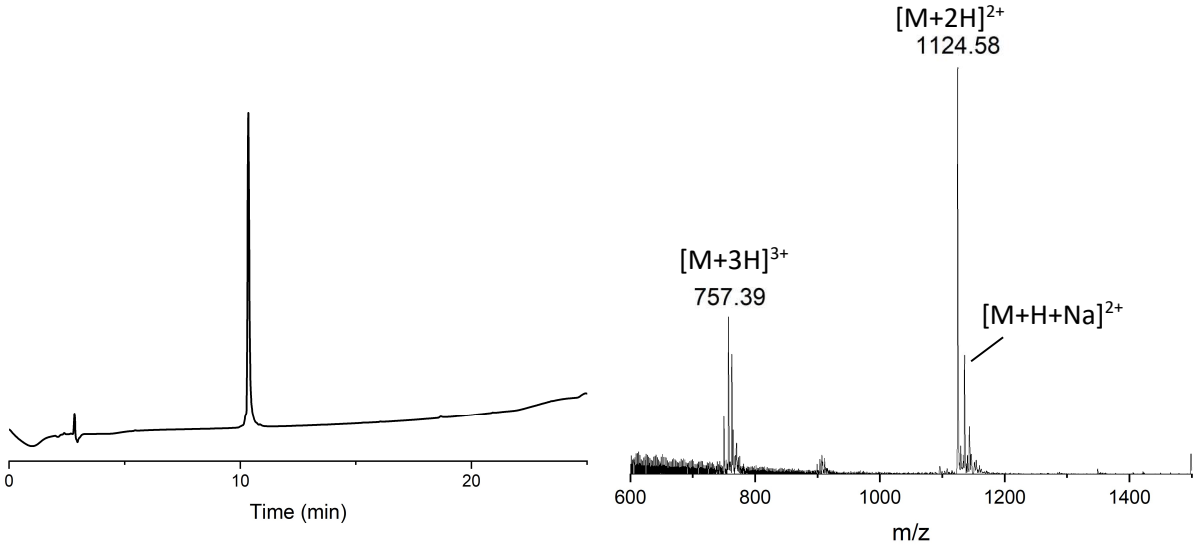


Figure 96 *Analytical data of the SunAm2Ser peptide. Retention time: 10.35 min (linear gradient 10-80% MeCN). Deconvoluted mass: 2247.156, calculated mass: 2247.518. Additional peaks correspond to sodium adducts.*

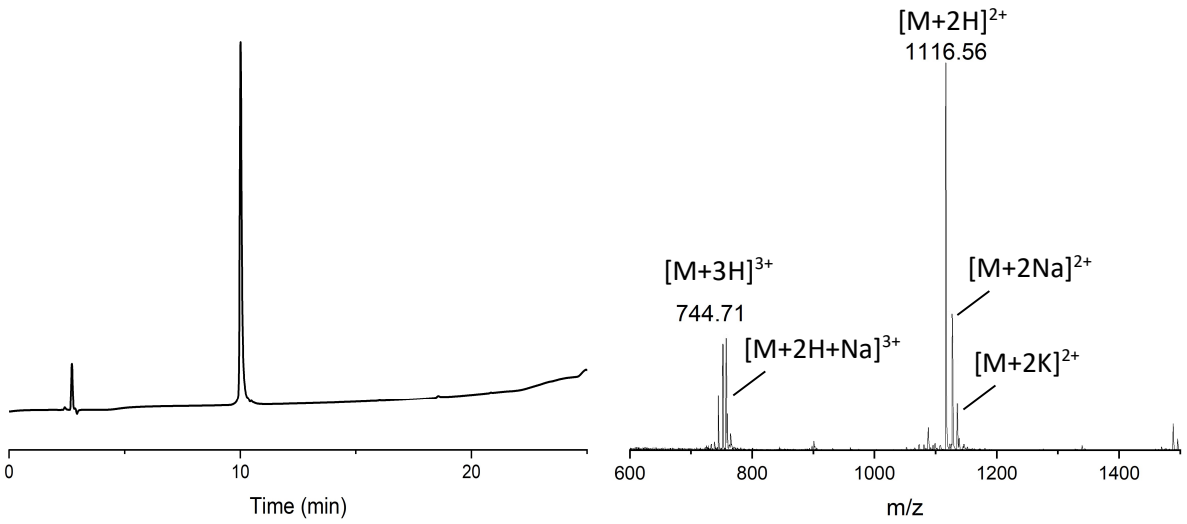


Figure 97 *Analytical data of the SunAmallSer peptide. Retention time: 10.0 min (linear gradient 10-80% MeCN). Deconvoluted mass: 2231.215, calculated mass: 2231.451. Additional peaks correspond to +Na and +K adducts respectively.*

8.1.3. Enterocin derived peptides

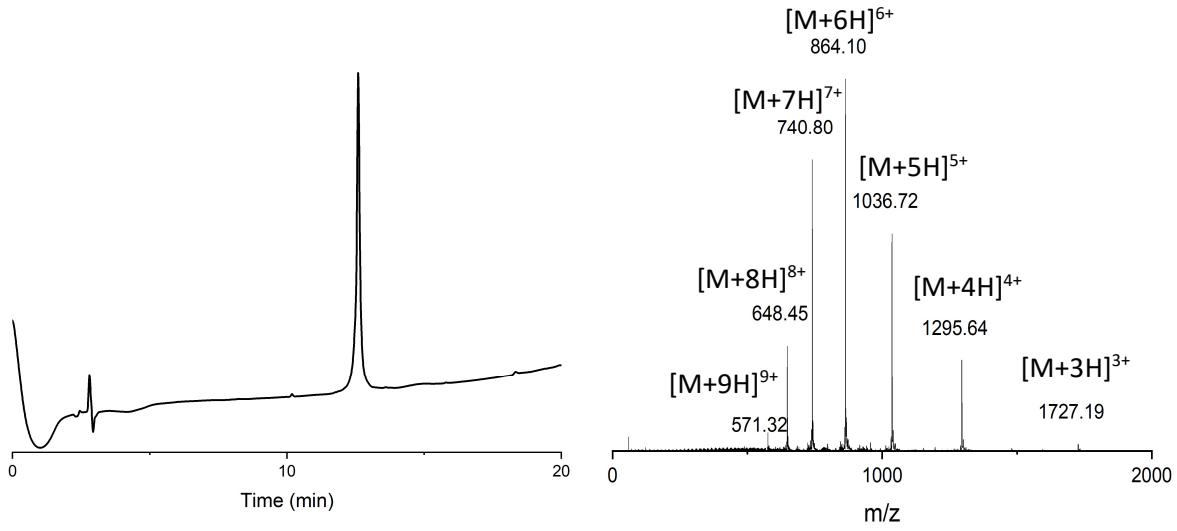


Figure 98 *Analytical data for the full length Enterocin 96 aglycone. Retention time 12.6 min (linear gradient 10-80% MeCN). Deconvoluted mass: 5178.600; calculated mass: 5179.0449*

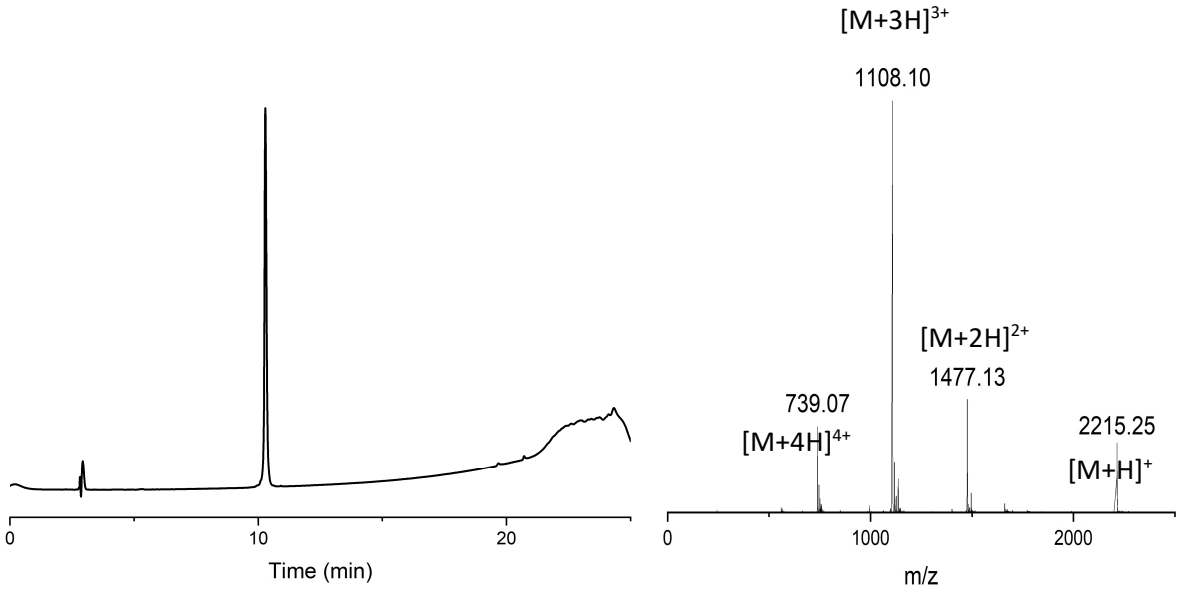


Figure 99 *Analytical Data for the Ecm peptide. Retention time: 10.28 min (linear gradient 10-80% MeCN). Deconvoluted mass: 2214.183, Calculated mass: 2214.420*

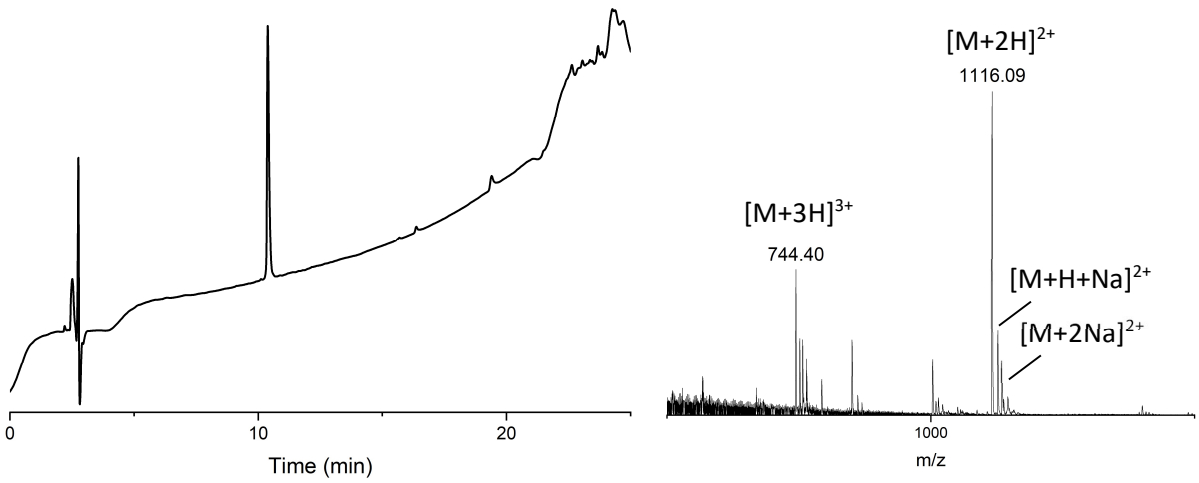


Figure 100 *Analytical Data of the EC(Cys)m peptide. Retention Time: 10.40 min (linear gradient 10-80% MeCN). Deconvoluted mass: 2230.168, Calculated mass: 2230.487. Additional peaks correspond to sodium adducts.*

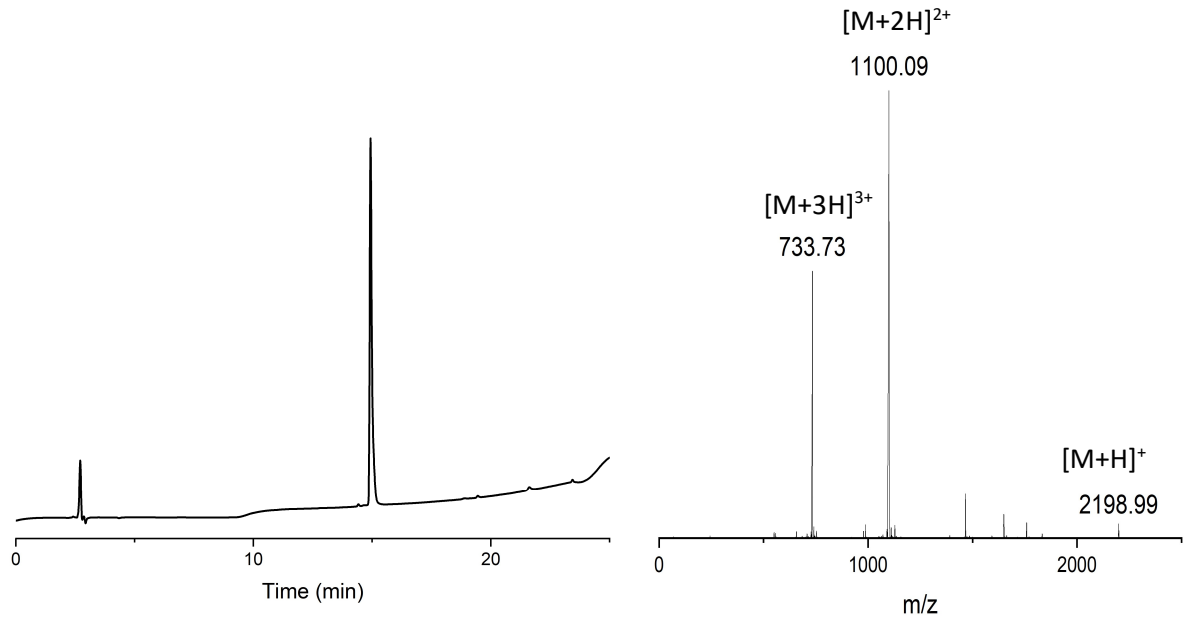


Figure 101 *Analytical Data of the EC(Ala)m peptide. Retention Time: 14.9 min (5 min 10% MeCN isocratic, 18 min linear gradient 10-80% MeCN). Deconvoluted Mass: 2198.164, Calculated Mass: 2198.421*

8.1.4. PltA derived peptides

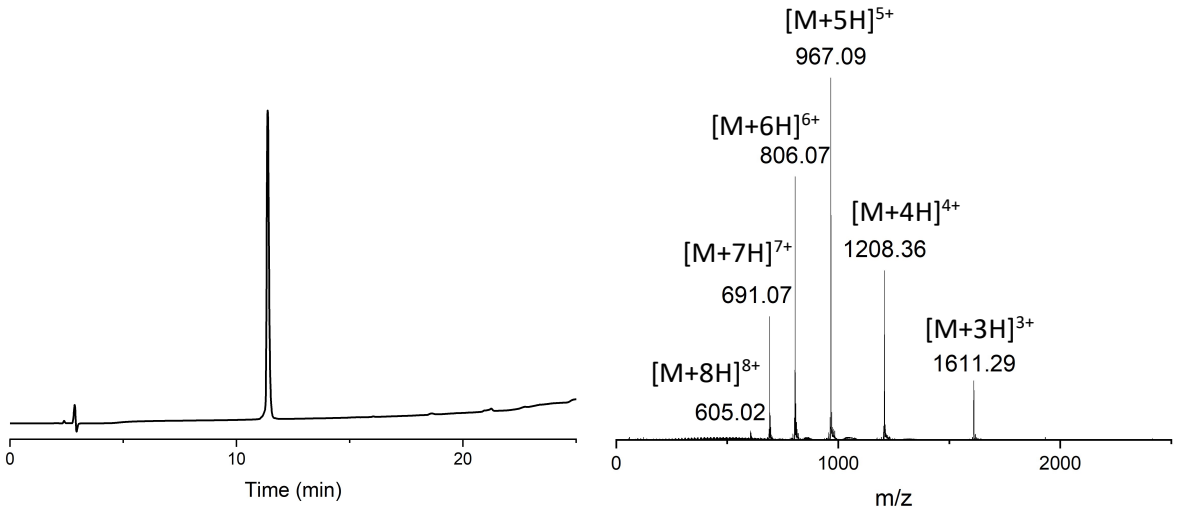


Figure 102 *Analytical data of the PltA peptide. Retention time: 11.38 min (linear gradient 10-60% MeCN). Deconvoluted mass: 4830.401, calculated mass: 4830.524*

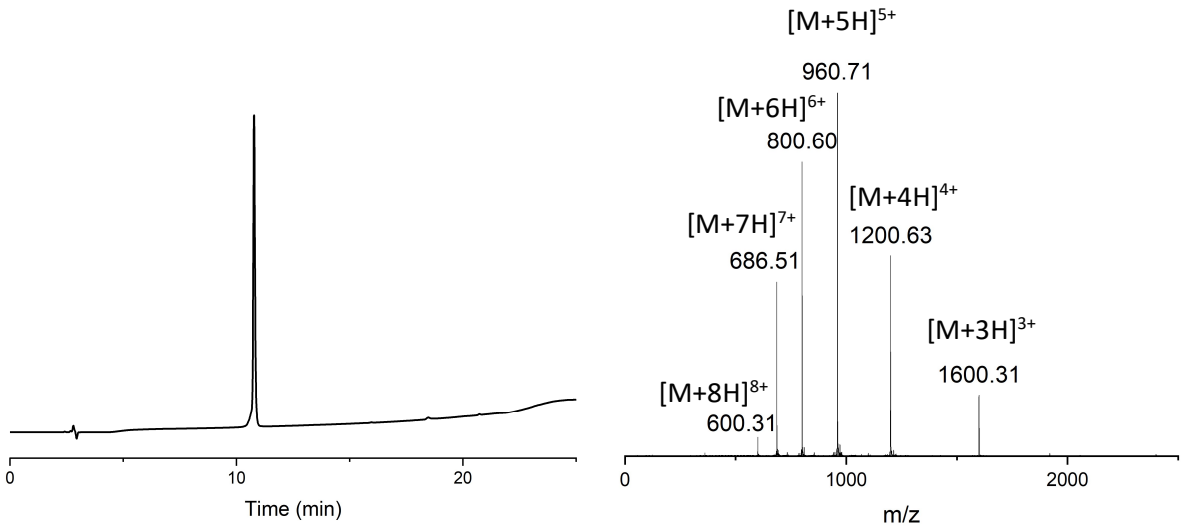


Figure 103 *Analytical data of the PltA2Ser peptide. Retention time: 10.8 min (linear gradient 10-60% MeCN). Deconvoluted mass: 4798.514, calculated mass: 4798.391*

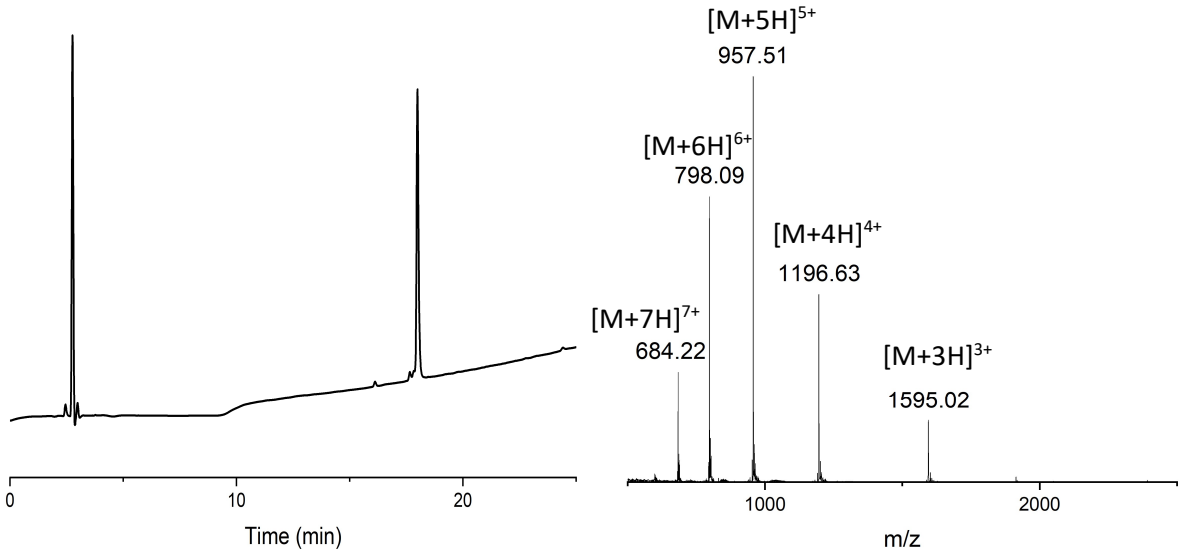


Figure 104 *Analytical data of the PltAallSer peptide. Retention time: 18.0 min (5 min isocratic 10% MeCN, 18 min 10-60% MeCN). Deconvoluted mass: 4782.493; calculated mass: 4782.324*

8.2. Glycosylated Peptides

8.2.1. SacA derived peptides

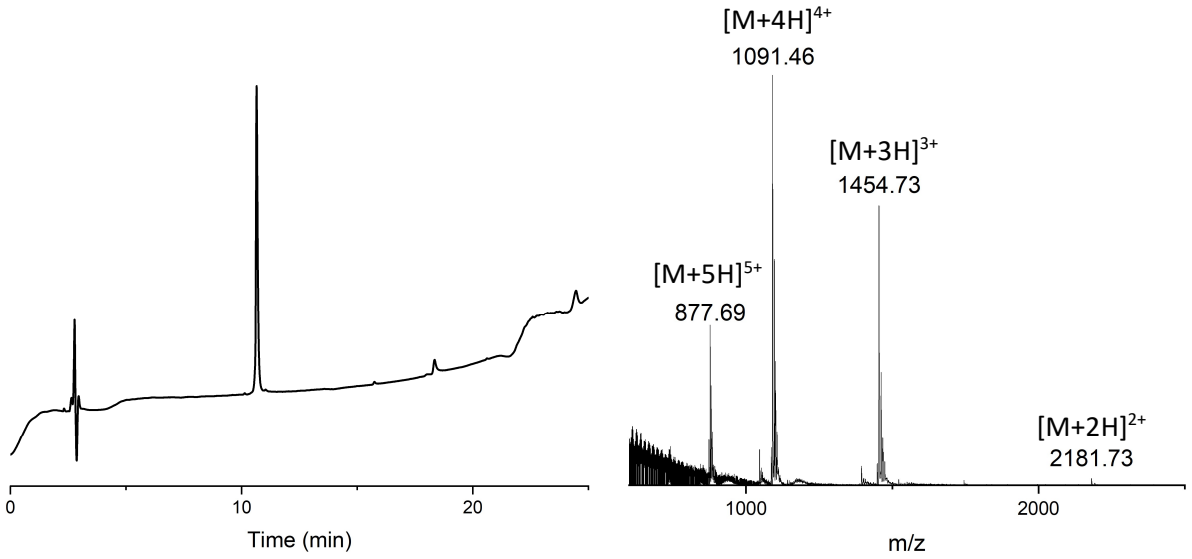


Figure 105 *Analytical data for the glycosylated SacA4Ser peptide. Retention time: 10.68 min (linear gradient 10-80% MeCN). Deconvoluted mass: 4361.007, calculated mass: 4361.752*

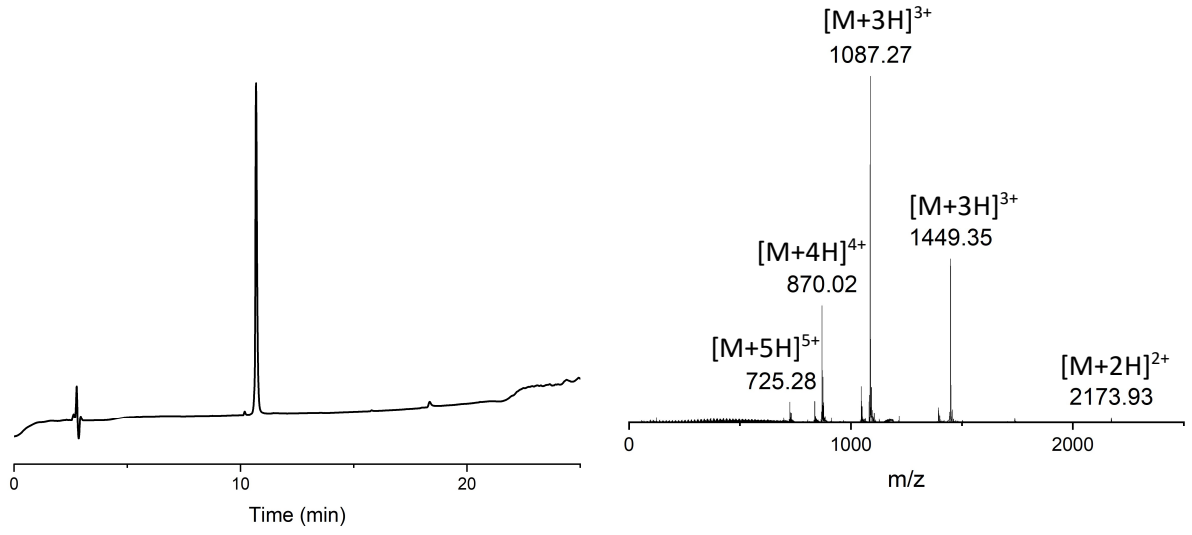


Figure 106 *Analytical data of the glycosylated SacAallSer peptide. Retention time: 10.65 min (linear gradient 10-80% MeCN). Deconvoluted mass: 4345.040, calculated mass: 4345.686*

8.2.2. SunA derived peptides

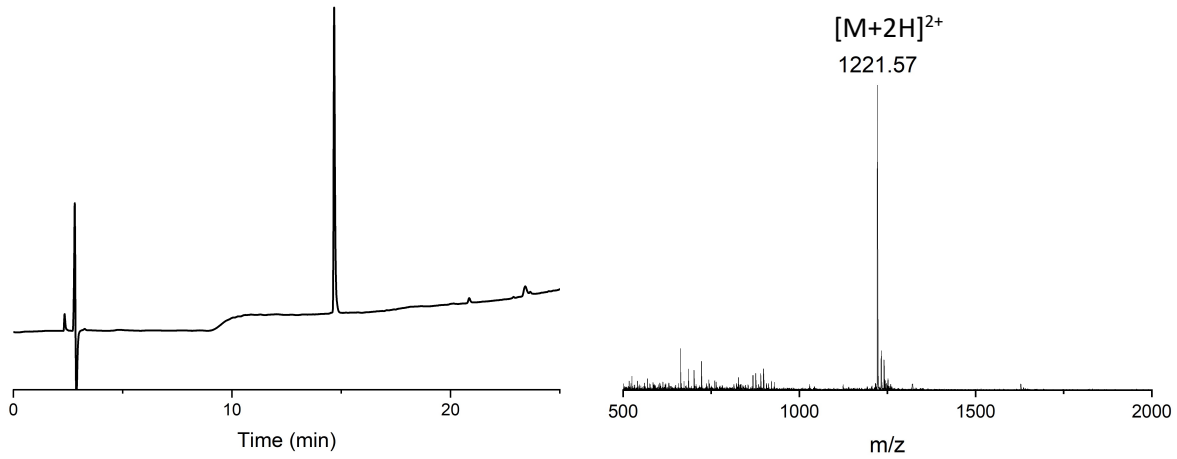


Figure 107 *Analytical data of the glycosylated SunAm peptide. Retention time: 14.7 min (5 min 10% MeCN isocratic, 18 min linear gradient 10-80% MeCN). Deconvoluted mass: 2441.140; calculated mass: 2441.651*

8.2.3. Enterocin 96 derived peptides

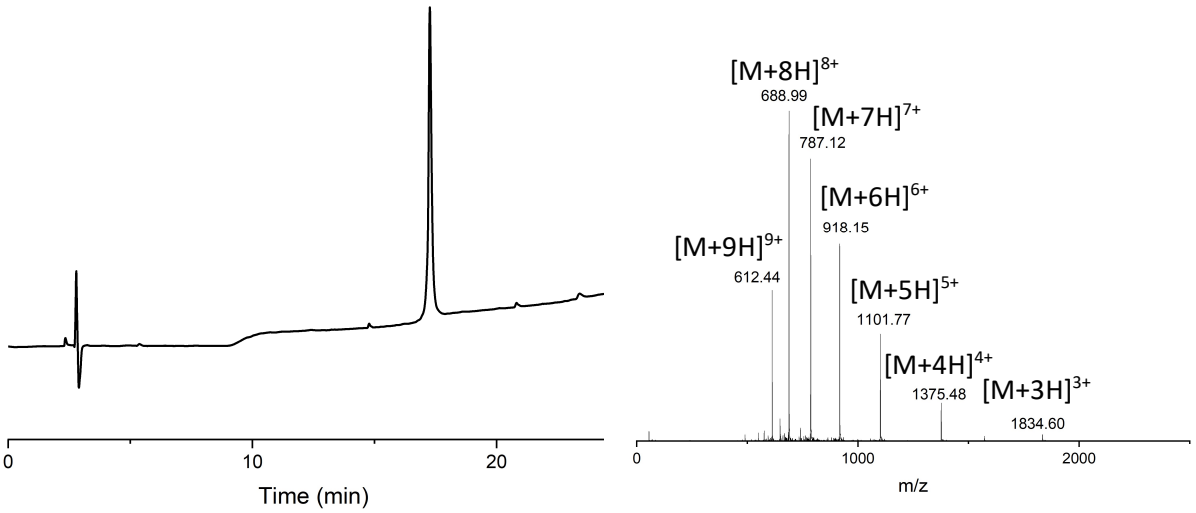


Figure 108 *Analytical data of glycosylated Enterocin 96. Retention time: 17.26 min (5 min 10% MeCN isocratic, 18 min linear gradient 10-80% MeCN). Deconvoluted Mass: 5503.900; calculated mass: 5499.326 (oxidized) 5503.327 (reduced).*

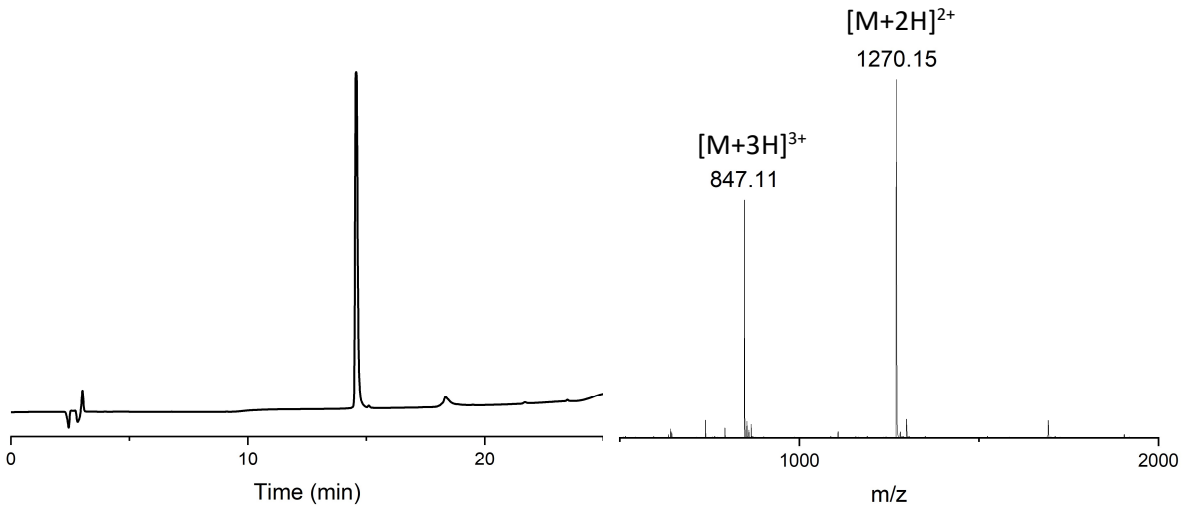


Figure 109 *Analytical data of the diglycosylated Ecm minimal peptide. Retention time: 14.58 min (5 min isocratic 10% MeCN, linear gradient 10-80% MeCN). Deconvoluted mass:2538.295, calculated mass: 2538.703*

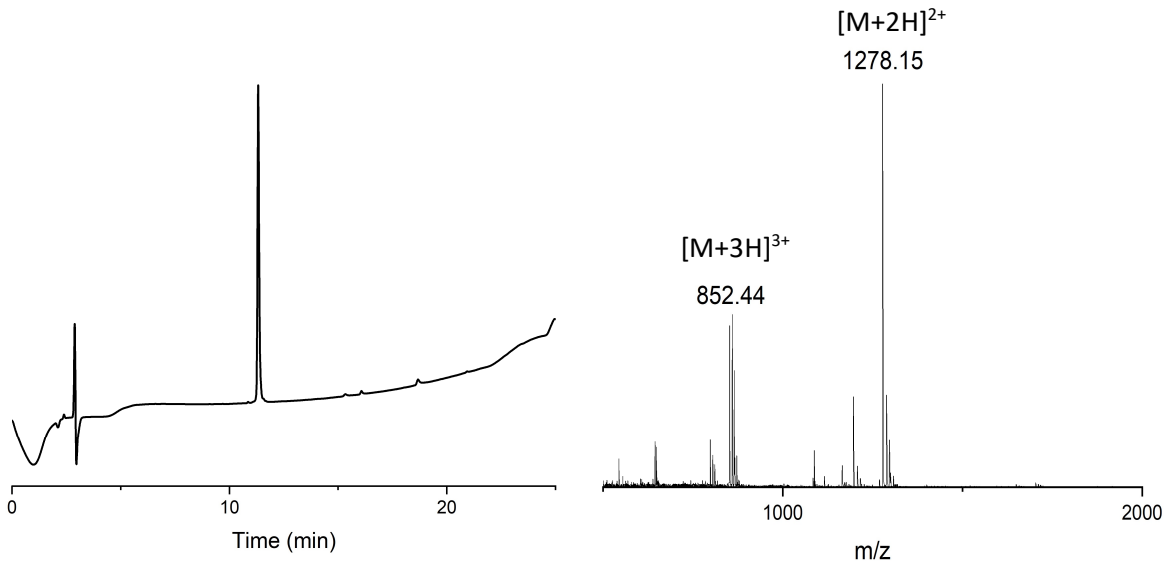


Figure 110 *Analytical data of the diglycosylated EC(Cys)m peptide. Retention time: 11.3 min (18 min Linear Gradient 10-80% MeCN). Deconvoluted mass: 2554.294; calculated mass: 2554.769*

8.2.4. PltA derived peptides

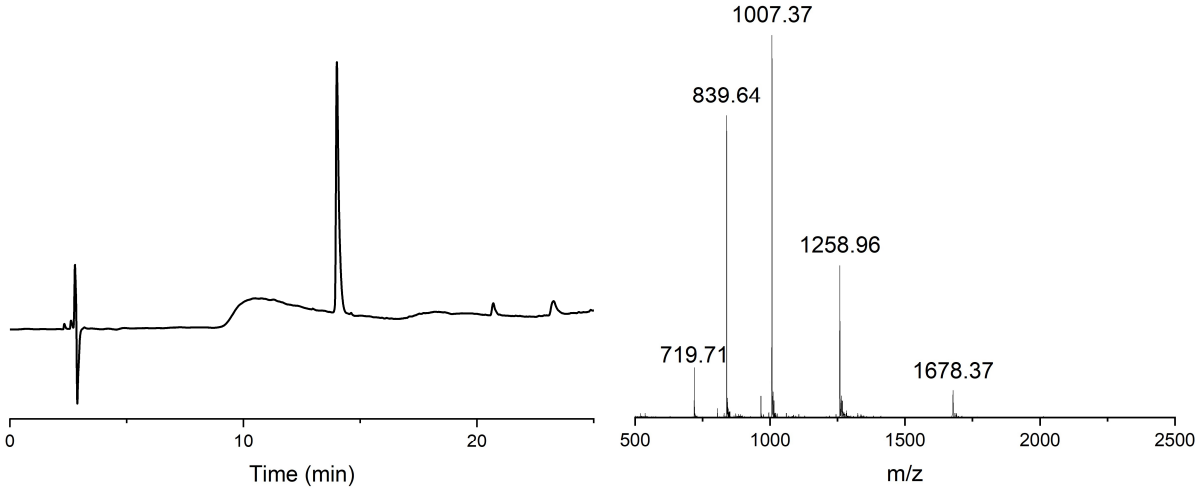


Figure 111 *Analytical data for GlcNAc-ylated PltA peptide. Retention time: 14.0 min (5 min isocratic, 18 min 10-80% MeCN). Deconvoluted mass: 5030.868, calculated mass: 5031.719.*

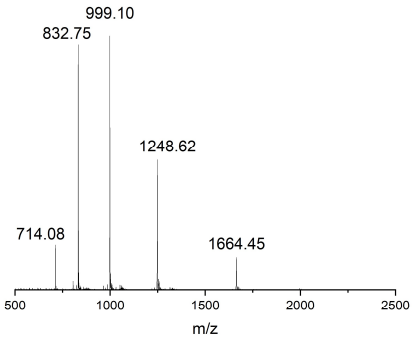


Figure 112 *Mass spectrum for PltA-Glc. Deconvoluted mass: 4990.478, calculated mass: 4990.524*

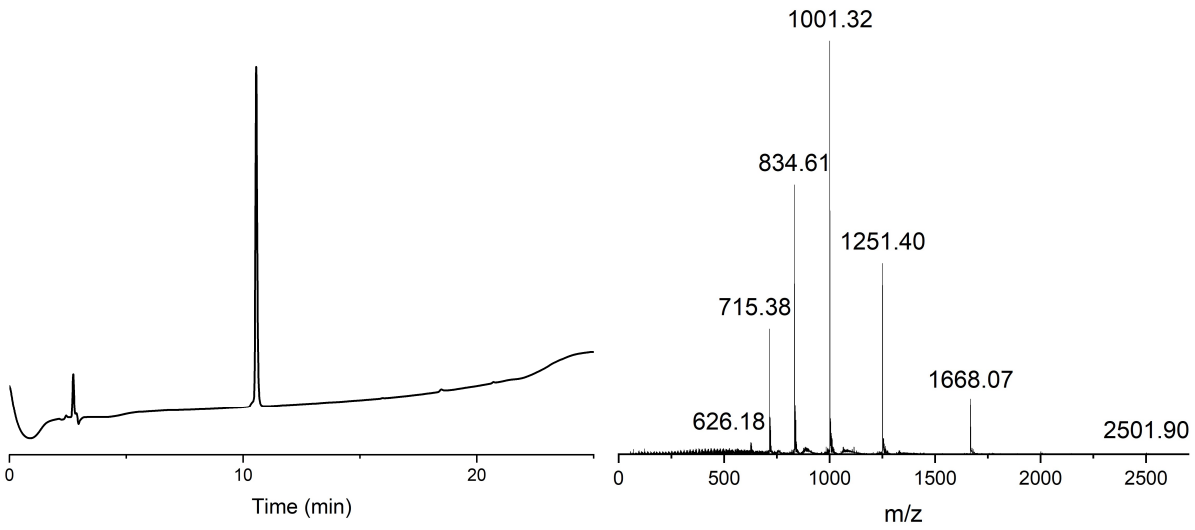


Figure 114 *Chromatogram and mass spectrum of GlcNAc-ylated PltA2Ser. Retention time: 10.56 (18 min linear gradient 10-60% MeCN). Deconvoluted mass: 5001.583, calculated mass: 5001.586*

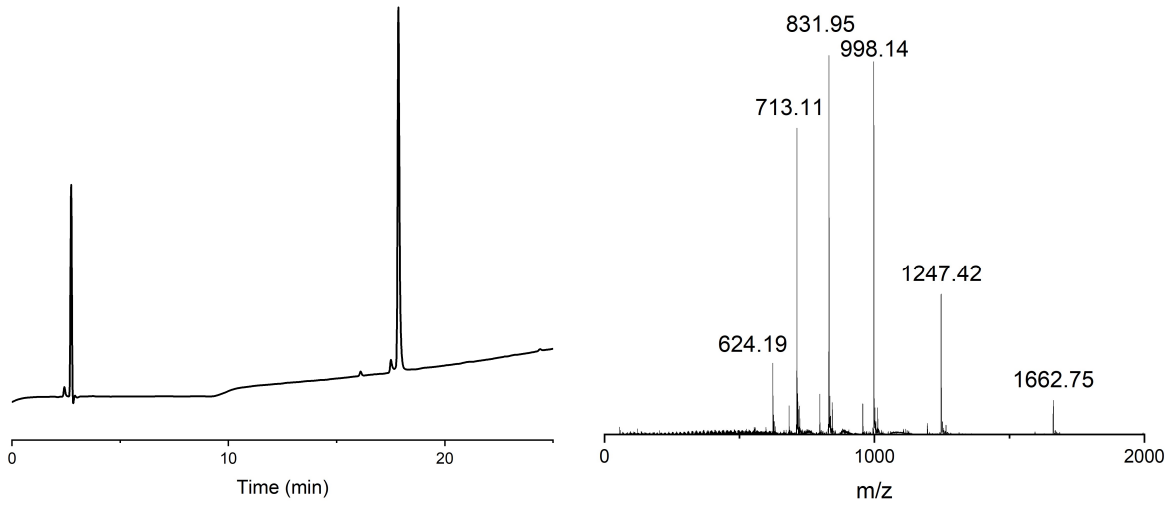


Figure 113 *Analytical Data of GlcNAc-ylated PltAallSer. Retention time: 17.85 min (5 min isocratic, 18 min 10-60% MeCN). Deconvoluted mass: 4985.651, calculated mass: 4985.519*