Fundamental Biochemistry of the Early N-Glycosylation Pathway



Alice Eddershaw

Keble College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary 2024

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. The work has not been submitted for any degree or professional qualification except as specified.

Impact of COVID-19 statement

This DPhil was undertaken during the COVID-19 pandemic which severely disrupted my progress at the start of the project. The total lost time away from the laboratory due to lockdowns or due to social distancing requirements was 20.25 weeks. The reduced operational capacity of both Departments, lack of physical supervision on new equipment, delayed access to other university Departments for equipment access, and a laboratory move, meant disruption to the DPhil exceeded the 6 months of extension funding I was fortunate to receive.

Alice Eddershaw, Hilary term 2024

Word count: ~50,000 excluding references, appendices, diagrams and tables

Abstract

Oligosaccharides destined for use as N-linked glycans are synthesised on phosphorylated dolichols in the endoplasmic reticulum (ER) by a series of glycosyltransferase enzymes known as ALG proteins. Mutations in any of these synthesis enzymes can lead to severe, multisystem syndromes known as congenital disorders of glycosylation (CDG). To date, a number of these glycosyltransferases remain uncharacterised, with most work focussing on yeast homologs. This has been due to the difficulty of accessing either the membrane-associated enzymes or their lipid-based substrates. By combining techniques in biochemistry, molecular biology and synthetic chemistry we aimed to elucidate the mechanisms and kinetics of key glycosyltransferase proteins in the pathway, utilising a library of polyisoprenol probes.

The enzymes DPAGT1 (ALG7) and ALG13/ALG14 perform the first phase of early N-glycosylation to yield dolichol pyrophosphate chitobiose on the cytoplasmic leaflet of the ER. Using a baculovirus expression system in insect cells, we present evidence that the human proteins form a multi-enzyme complex that together catalyse this initial step of lipid-linked oligosaccharide synthesis. The kinetics of these enzymes were assessed both when alone and in complex using two experimental formats which has allowed insight into the control and dynamics at the start of the pathway.

Iterative attempts were made at structural elucidation by cryoEM of the complex in its unbound and lipid acceptor bound state. Furthermore, we showed that DPAGT1 is highly selective for native dolichol lipids over those typically used in bacterial membranes, whereas ALG13/ALG14 can utilise lipid acceptor substrates of all saturation states. Finally, study of CDG patient mutants helped to inform function of these key enzymes and their pathogenic mechanism. Together, this work aids our understanding on an essential and highly conserved process.

Acknowledgments

I feel very fortunate to have been surrounded by many talented people during my DPhil studies and am indebted to them all for their part in helping me present this thesis. I would like to give a heartfelt thank you to:

- My supervisor Dr Yin Dong for the opportunity to become immersed in the fascinating world of N-glycosylation and CDGs, and for going above and beyond to support and guide me throughout the DPhil process- I am honoured to be your first student!
- To Prof Ben Davis for allowing me the opportunity to grow as a scientist and person, and for the freedom to learn such a breadth of techniques. For teaching me that there is no such thing as a pure Biologist or Chemist – instead a scientist who can apply both.
- Dr Andrew Giltrap for so patiently teaching me the skills of the chemistry lab, for encouragement and support at every stage of my DPhil, for expertly proof-reading my entire thesis, and for friendship away from the lab.
- Jing Wang for her expert assistance with cell culture, baculovirus cloning and purification, but especially for her kindness, listening ear and wise advice.
- Dr Peter Harrison for tireless support and assistance not only practically with membrane protein purification, grid preparation and cryoEM, but also thesis writing and mentorship.
- Members of the Yin Dong group for insightful discussions, support in the lab and making Monday mornings so enjoyable. Thank you to Judy, Richard, Susan, Imre, Liang, Setareh, Anna, Lenny, Lara, Lesley, Angela, David.
- Members of the Membrane Protein Lab at Diamond in particular Dr Andrew Quigley for sponsorship and expert input across my project, and to Monika and Harish for support with detergent screens, as well as Charlie, James, Tereza and Rebecca.

- Members of the ‘Sugar Team’ meetings at RFI at which I learnt so much - in particular to Prof Jim Naismith, Dr Audrey Le Bas and Dr Phil Ward.
- Members of the Davis group for your incredible chemistry knowledge and many fun and cherished memories from the days in ICL, our move to RFI and various summer and Christmas socials. I will remember my time at Oxford so fondly because of you all. In particular thanks to Andrew, Adeline, Brian, Mateusz, Tim, Xuanxiao, Carla, Effie, Dimitrios, Jack, Yana, Mellisa, Chao, George and Ajay.
- To the support staff in all institutes I was fortunate to be a part of, in particular: Dr Zoe Christodoulou and Susan Maxwell for Radiation Safety Protection; David Parr for management of my grant and consumable orders; members of the McHugh lab for access and support with using the Typhoon; Dr Stan Ng (Ludwig Institute) for access to the Fujifilm; and Dr Petra Lukacik (Diamond Light Source) for assistance and advice for Native PAGE.
- To the BBSRC and Interdisciplinary Bioscience DTP for funding and for the invaluable support network of friends (in particular Amelia, Jai, Andrew and Alaa).
- To my housemates Barley, Zoe and Sophie for invaluable support especially through Covid lockdowns and during thesis writing, and to my life-long friendships away from Oxford who have been a huge source of comfort (especially Pip, Eleanor, Faye, and The Essex Girls).
- I would like to offer the biggest thank you to my parents - Alison and Peter - for their endless support in every sense of the word. To my siblings Liz and Owen for encouraging and inspiring me every day and to my grandmother Fiona for helping to collect leaves needed for my project.
- Finally, a huge thank you to Oscar for without his love, patience, and encouragement this would not have been possible.

Table of Contents

Abstract..... **3**

Acknowledgments..... **4**

Table of Contents..... **6**

List of Figures..... **10**

List of Tables..... **16**

Table of Abbreviations..... **18**

1.0 Introduction..... **19**

1.1 The role of N-glycosylation..... **19**

1.2 Lipid-linked oligosaccharide synthesis..... **22**

1.2.1 Enzymes of the early N-glycosylation pathway..... **24**

1.2.2 Conservation across the domains of life..... **27**

1.2.3 Multi-enzyme complexes amongst the early N-glycosylation pathway..... **34**

1.3 Polyisoprenol structure and function..... **36**

1.3.1 Biogenesis of polyisoprenols..... **38**

1.4 Congenital disorders of glycosylation..... **41**

1.4.1 Current challenges in the field of CDGs **42**

1.5 Project aims and hypotheses..... **45**

2.0 Materials and Methods..... **47**

2.1 Baculovirus-insect cell expression system..... **47**

2.1.1 Insect cell passage..... **47**

2.1.2 Ligation independent cloning..... **47**

2.1.3 Bacmid production..... **52**

2.1.4 Transfection and baculovirus amplification..... **53**

2.1.5 Small-scale test expression in insect cells..... **55**

2.1.6 Amplification of baculovirus to P2 stock..... **55**

2.1.7 MultiBac production..... **56**

2.2 Pathogenic variants of ALG13 and ALG14..... **57**

2.2.1 Site-directed mutagenesis..... **57**

2.2.2 RNA extraction..... **59**

2.3 Protein purification..... **60**

2.3.1 Insect cell infection and harvesting..... **60**
2.3.2 Homogenisation..... **60**
2.3.3 Sonication..... **61**
2.3.4 Membrane preparation **61**
2.3.5 Immobilised metal affinity chromatography..... **62**
2.3.6 Reverse purification..... **63**
2.3.7 Anti-FLAG affinity purification..... **63**
2.3.8 On-resin exchange into non-detergent membrane mimetics..... **63**
2.3.9 Size exclusion chromatography..... **65**
2.4 Protein characterisation..... **65**
2.4.1 SDS-PAGE..... **65**
2.4.2 2D Blue-Native/SDS-PAGE..... **65**
2.4.3 Western Blot..... **66**
2.4.4 TCA precipitation..... **67**
2.4.5 Tryptic digest MS/MS..... **67**
2.4.6 Intact denaturing MS..... **68**
2.4.7 Nano differential scanning fluorimetry..... **68**
2.4.8 BCA assay and densitometry..... **69**
2.4.9 SEC-MALS..... **70**
2.4.10 fSEC..... **72**
2.4.11 Negative Stain EM..... **72**
2.5 Synthesis of polyisoprenol and dolichol lipids..... **73**
2.6 Functional assessment of DPAGT1/ALG13/ALG14..... **76**
2.6.1 UMP/UDP-Glo assay..... **76**
2.6.2 Radioactive TLC assay..... **77**
2.7 Structural study..... **78**
2.7.1 AlphaFold Multimer using Google Collaboratory..... **78**
2.7.2 Crystallisation trials..... **79**
2.7.3 CryoEM..... **79**
3.0 Expression, purification and structural analysis of a DPAGT1/ ALG13/ ALG14 complex..... **82**
3.1 Introduction..... **82**

3.1.1 A multi-enzyme complex of DPAGT1, ALG13 and ALG14..... **82**
3.1.2 Expression and purification in insect cells..... **84**
3.1.3 Structural elucidation of DPAGT1/ALG13/ALG14..... **85**
3.2 Results..... **88**
3.2.1 Predicting the structure of human ALG13/ALG14..... **88**
3.2.2 Purification of ALG13 and ALG14..... **90**
3.2.2.1 ALG13/ALG14 Multi-bac expression and purification..... **96**
3.2.3 Characterisation of ALG13/ALG14..... **108**
3.2.4 DPAGT1/ALG13/ALG14 purification..... **115**
3.2.5 DPAGT1/ALG13/ALG14 complex formation..... **121**
3.2.6 CryoEM of His-DPAGT1/ALG13/ALG14..... **129**
3.2.7 CryoEM of DPAGT1-FLAG/ALG13/ALG14..... **135**
3.3 Discussion..... **147**
3.3.1 ALG13/ALG14..... **147**
3.3.2 DPAGT1/ALG13/ALG14..... **149**
4.0 Optimisation of a glycosyltransferase assay..... **154**
4.1 Introduction..... **154**
4.1.1 Kinetic study of bi-substrate enzymes..... **155**
4.1.2 Assays to monitor glycosyl transfer to a lipid acceptor... **157**
4.2 Results..... **161**
4.2.1 Promega Glo assay..... **161**
4.2.1.1 UMP release..... **162**
4.2.1.2 UDP release..... **170**
4.2.2 Radioactive TLC assay..... **172**
4.2.2.1 ALG13/ALG14 activity..... **178**
4.2.2.2 DPAGT1/ALG13/ALG14 activity..... **188**
4.3 Discussion..... **196**
4.3.1 UMP and UDP-Glo assay..... **197**
4.3.2 ^{14}C labelled UDP-GlcNAc assays..... **199**
5.0 Substrate specificity: testing a panel of native and non-native polyisoprenol and dolichol phosphates..... **203**

5.1 Introduction..... 203
5.1.1 Substrate specificity of polyisoprenol glycosyltransferases..... 204
5.1.2 Synthetic and semi-synthetic approaches to access polyisoprenols..... 207
5.2 Results..... 209
5.2.1 Accessing a panel of polyisoprenols and dolichols..... 209
5.2.1.1 Trial use of SRD5α3 for accessing dolichols..... 212
5.2.1.2 Characterisation of the lipid panel..... 214
5.2.2 Activity testing 219
5.2.2.1 Endpoint activity assay of DPAGT1 and ALG13/ALG14..... 219
5.2.2.2 Kinetic analysis with native and non-native lipids. 224
5.3 Discussion..... 229
6.0 Studying pathogenic variants of ALG13/ALG14..... 233
6.1 Introduction..... 233
6.2 Results..... 236
6.2.1 AlphaFold and AlphaMissense predictions of ALG13/ALG14 variants..... 236
6.2.2 Cloning of baculoviral expression vectors..... 242
6.2.3 Purification of pathogenic variants of ALG13/ALG14..... 245
6.2.4 Characterisation of pathogenic variants of ALG13/ALG14..... 253
6.3 Discussion..... 255
7.0 Discussion and Future Perspectives..... 259
7.1 Complex formation of human DPAGT1 and ALG13/ALG14..... 259
7.1.1 Single particle cryogenic electron microscopy..... 262
7.2 Functional study of glycosyltransferases..... 264
7.2.1 Insights into ALG13/ALG14 CDG mutants..... 266
7.3 Evolutionary consequences of eukaryotic dolichol use..... 267
7.4 Concluding remarks..... 269
Bibliography..... 270
Appendices..... 285

List of Figures

Figure 1.1 Schematic depicting the role of N-glycosylation in folding, quality control, localisation and function of proteins in the secretory pathway..... **21**

Figure 1.2 Asparagine linked glycosylation (ALG) enzymes 1-14 of the early N-glycosylation pathway..... **23**

Figure 1.3 Action of DPAGT1 and ALG13/ALG14 enzymes of the early N-glycosylation pathway..... **25**

Figure 1.4 Structure of the core eukaryotic N-glycan transferred to a nascent protein..... **26**

Figure 1.5 Parallels between early stages of eukaryotic and crenarchaeal N-glycosylation and bacterial peptidoglycan and lipopolysaccharide synthesis..... **28**

Figure 1.6 Comparison of the structure of DPAGT1 and MraY..... **30**

Figure 1.7 Comparison of ALG13/ALG14 homologs..... **32**

Figure 1.8 The ER translation-translocation machinery complex where the N-glycan is transferred..... **34**

Figure 1.9 Polyisoprenols utilised in glycan biosynthesis..... **36**

Figure 1.10 Biogenesis of polyisoprenol and dolichol phosphate..... **40**

Figure 1.11 Analysis of pathogenic DPAGT1 variants..... **44**

Figure 2.1 General plasmid maps of cloning vectors, as supplied by the CMD..... **49**

Figure 2.2 Infected Sf9 insect cells..... **54**

Figure 2.3 Acceptor (left) and donor (right) vectors combined during MultiBac production..... **56**

Figure 2.4 Calibration curve of refractive index area and concentration of OGNG/CHS/CDL (A) and DM/CHS (B) detergents used to calculate a dn/dC value by SEC-MALS..... **71**

Figure 2.5 Visualisation of cryoEM image processing work-flow..... **81**

Figure 3.1 Function of DPAGT1 and ALG13/ALG14 for lipid-linked oligosaccharide synthesis in the ER..... **83**

Figure 3.2 Structures of DPAGT1 as a monomer and homodimer..... **87**

Figure 3.3 Modelling the structure of human ALG13/ALG14…… **89**
Figure 3.4 Initial testing of ALG13 purification by IMAC and reverse IMAC…….. **91**
Figure 3.5 Wash buffer testing for ALG13 IMAC purification……... **92**
Figure 3.6 Buffer stability screen of ALG13 WT protein in 10 different buffer conditions…… **94**
Figure 3.7 Denaturing, intact mass spectrometry…… **95**
Figure 3.8 IMAC purification ALG13/His-ALG14 in DM or DM/CHS containing buffers……... **97**
Figure 3.9 IMAC purification His-ALG13/His-ALG14 in DM or DM/CHS containing buffers……... **98**
Figure 3.10 Precipitated proteins from the peak fractions from SEC of ALG13/ALG14……... **99**
Figure 3.11 Purification of ALG13/ALG14-His…… **101**
Figure 3.12 Purification of ALG14 and ALG13 with N-term 6xHis tag…….... **102**
Figure 3.13 Purification of ALG13 tagged and untagged with ALG14 protein from 4 L Sf9 cell culture…… **104**
Figure 3.14 Extended membrane preparation for purification of His-ALG13 /His-ALG14…… **106**
Figure 3.15 Purification from 6.6 L of Sf9 insect cells expressing ALG13/His-ALG14…… **107**
Figure 3.16 Intact denaturing MS of untagged ALG13…… **109**
Figure 3.17 Intact denaturing MS of His-ALG13…….. **110**
Figure 3.18 Intact denaturing MS of His-ALG14…….. **111**
Figure 3.19 ALG13/ALG14 are in a 1:1 complex in the DM/CHS micelle….. **112**
Figure 3.20 Thermostability of His-ALG13/His-ALG14…….. **113**
Figure 3.21 Freeze-thaw activity testing of ALG13/His-ALG14…….. **115**
Figure 3.22 AlphaFold-Multimer [21] predicted structure of DPAGT1 (green), ALG14 (pink) and ALG13 (blue) …….. **116**
Figure 3.23 Purification of His-DPAGT1/His-ALG13/His-ALG14 using DM/CHS or OGNG/CHS/CDL containing buffers…… **118**
Figure 3.24 Purification of His-DPAGT1/ALG13/His-ALG14 using DM/CHS or OGNG/CHS/CDL containing buffers…… **119**

Figure 3.25 Characterisation of the DPAGT1-His/ALG13/His-ALG14 protein complex…….. **120**
Figure 3.26 Analysis of DPAGT1-His/ALG13/ALG14-His by SEC-MALS… **122**
Figure 3.27 Pull down assay: ALG13/ALG14 interacts with DPAGT1-FLAG **123**
Figure 3.28 2D BN/SDS-PAGE of DPAGT1-FLAG/ALG13/His-ALG14….. **125**
Figure 3.29 Purification and freeze-thaw testing of His-DPAGT1/ALG13/His-ALG14 complex……. **127**
Figure 3.30 Negative stain electron microscopy images of His-DPAGT1/His-ALG13/His-ALG14 proteins at 100 and 50 µg/mL…… **128**
Figure 3.31 Determining the optimal protein concentration: screening and data collection of His-DPAGT1/ALG13/His-ALG14…… **130**
Figure 3.32 Grid screening and data collection of His-DPAGT1/ALG13/His-ALG14 in DM/CHS micelles with and without dolichol pyrophosphate GlcNAc……... **132**
Figure 3.33 Detergent and membrane mimetic screen for extraction and purification of DPAGT1/ALG13/ALG14……... **134**
Figure 3.34 Purification of DPAGT1-HisFLAG/ALG13/His-ALG14…….. **136**
Figure 3.35 Thermostability of DPAGT1-HisFLAG/ALG13/His-ALG14 before (red) and after (green) size exclusion chromatography….. **137**
Figure 3.36 Screen of 12 detergents for extraction and purification of DPAGT1-HisFLAG/ALG13/His-ALG14…….. **138**
Figure 3.37 Purification of DPAGT1-FLAG/ALG13/His-ALG14 in 3 different detergent conditions……. **140**
Figure 3.38 Summary of Atlas screen performed using eBIC Glacios cryo-electron microscope…… **142**
Figure 3.39 Purification and activity testing of DPAGT1-HisFLAG/ALG13/HisALG14…… **143**
Figure 3.40 Grid screening and 2D class averages of DPAGT1-HisFLAG/ALG13/His-ALG14……... **145**
Figure 3.41 Comparison of SEC profiles of ALG13/ALG14 constructs…….. **148**
Figure 3.42 Comparison of SEC profiles of DPAGT1/ALG13/ALG14 constructs…… **150**

Figure 3.43 2D class averages of (A) His-DPAGT1/ALG13/ALG14 and (B) DPAGT1-FLAG/ALG13/His-ALG14…… **152**
Figure 4.1 DPAGT1 and ALG13/ALG14 catalysed GT reactions…….. **154**
Figure 4.2 Schematic depicting the mechanism of UDP- and UMP-Glo assays…….. **161**
Figure 4.3 Standard curve of UMP in assay buffer detected using the UMP-Glo assay kit…… **162**
Figure 4.4 Quenching issues: DPAGT1 activity when addition of UMP-Glo kit reagents were used to stop the reaction……. **163**
Figure 4.5 Alternative methods of quenching DPAGT activity……... **165**
Figure 4.6 DPAGT1 titration…… **166**
Figure 4.7 Activity of DPAGT1 determined with respect to dolichol phosphate…… **167**
Figure 4.8 Activity of DPAGT1 determined with respect to UDP-GlcNAc… **168**
Figure 4.9 Catalytic mechanism of DPAGT1…….. **169**
Figure 4.10 UMP and UDP standard curve in the presence of a range of competing nucleotide phosphate concentrations…… **171**
Figure 4.11 Issues with the UDP-Glo kit……... **172**
Figure 4.12 Schematic depicting movement of lipid linked oligosaccharides by TLC…… **172**
Figure 4.13 TLC assay: standard curve UDP-[^{14}C]-GlcNAc……. **174**
Figure 4.14 Selecting a suitable solvent……. **175**
Figure 4.15 DPAGT1 time course…….. **176**
Figure 4.16 DPAGT1 ligand binding kinetics (N=2) …… **177**
Figure 4.17 Titration of ALG13/ALG14 from 2-0.063 µM……... **179**
Figure 4.18 Representative results of a dolichol pyrophosphate GlcNAc titration…… **181**
Figure 4.19 Representative results from a UDP-[^{14}C]-GlcNAc titration…….. **182**
Figure 4.20 Isothermal screening by nanoDSF of ALG13/ALG14, DPAGT1 and DPAGT1/ALG13/ALG14 complex……. **183**
Figure 4.21 ALG13/ALG14 titration from 2-0.6 µM when assayed at 25 °C... **184**
Figure 4.22 Early time points to define the linear phase of ALG13/ALG14…. **185**

Figure 4.23 Sugar donor kinetics with ALG13/ALG14……. **186**
Figure 4.24 Lipid acceptor kinetics with ALG13/ALG14…… **187**
Figure 4.25 Activity of DPAGT1 and the DPAGT1/ALG13/ALG14 complex……... **189**
Figure 4.26 Time course of DPAGT1/ALG13/ALG14……. **191**
Figure 4.27 Lipid acceptor kinetics with DPAGT1/ALG13/ALG14, N=3…… **194**
Figure 4.28 Sugar donor kinetics with DPAGT1/ALG13/ALG14…… **195**
Figure 5.1 Polyisoprenols utilised in glycan biosynthesis……... **203**
Figure 5.2 Effect of increasing number of cis isoprenoid units on affinity to GlcNAc transferases originating in three eukaryotic organisms: yeast [188], cow[190], rat [189]……...……. **206**
Figure 5.3 Schematic representing the structure of short (C55) and long (C90) polyisoprenoid lipids of dolichol and polyunsaturated nature……... **209**
Figure 5.4 Polyisoprenoid alcohols of human-like length were accessed from the leaves of the *Gingko biloba*…….. **210**
Figure 5.5 Hydrogenation of polyisoprenoid alcohols to dolichols……. **211**
Figure 5.6 Initial use of SRD5α3 to yield dolichol lipids…… **213**
Figure 5.7 Schematic depicting methodology to access polyisoprenols of various length and saturation state from leaves…….. **215**
Figure 5.8 Thermostability screen of DPAGT1/ALG13/ALG14 when bound to members of the native and non-native lipid panel…….. **217**
Figure 5.9 DPAGT1 endpoint activity with short (C55) and long (C90) polyisoprenol and dolichol lipids…… **220**
Figure 5.10 Summary of endpoint activity of ALG13/ALG14 with short (C55) and long (C90) polyisoprenol and dolichol lipids…… **221**
Figure 5.11 N=3 Endpoint activity screen of DPAGT1/ALG13/ALG14…….. **222**
Figure 5.12 DESI-MS of DPAGT1/ALG13/ALG14 assay plates……. **223**
Figure 5.13 Kinetic of DPAGT1/ALG13 with native and non-native lipid substrates (N=3 biological triplicates)…… **225**
Figure 5.14 Example TLC plates from titration of each of the native and non-native lipid phosphates with DPAGT1/ALG13/ALG14……..... **226**

Figure 5.15 Kinetic of DPAGT1/ALG13 with native and non-native lipid substrates (N=3 biological triplicates) **227**
Figure 5.16 Example TLC plates from titration of each of the native and non-native lipid pyrophosphate GlcNAc..... **228**
Figure 6.1 Location of missense variants of ALG13/ALG14 identified in patients with CDG..... **235**
Figure 6.2 Alignment of the predicted structure of pathogenic variants of ALG13 with that of wild type ALG13/ALG14..... **237**
Figure 6.3 Alignment of the predicted structure of pathogenic variants of ALG14 with that of wild type ALG13/ALG14..... **238**
Figure 6.4 AlphaMissense HotSpot analysis of ALG14 (A) and ALG13 isoform 1 (B) **239**
Figure 6.5 Figure PCR screen showing successful mutagenesis of ALG13/ALG14 MultiBac plasmids..... **242**
Figure 6.6 Sanger-sequencing showing successful introduction of pathogenic variants of ALG13..... **243**
Figure 6.7 Sanger-sequencing showing successful introduction of pathogenic variants of ALG14..... **244**
Figure 6.8 Representative SDS-PAGE gels showing progress of N=3 ALG13/AL14 mutant protein purifications..... **246**
Figure 6.9 Representative SDS-PAGE gels showing progress of N=3 ALG13/AL14 mutant protein purifications..... **247**
Figure 6.10 SEC profiles of 5 successfully expressed ALG13/ALG14 mutants (N=3)..... **248**
Figure 6.11 Protein detected in the fos-choline extracted cell lysate of ALG13/His-ALG14 expressing insect cells..... **250**
Figure 6.12 ALG13 (top) and ALG14 (bottom) gene amplified from RNA of Sf9 insect cells infected with MultiBac vectors encoding wild type or mutant versions of the proteins..... **252**
Figure 6.13 Relative activity of ALG13/ALG14 proteins bearing pathogenic mutations to wild-type (WT)..... **254**
Figure 6.14 Thermostability screen of ALG13/AL14 mutant protein stocks following storage at -80 °C..... **254**

List of Tables

Table 1.1	Summary of polyisoprenol structures across the domains of life...	**37**
Table 2.1	ALG13 and ALG14 specific primers for Ligation Independent Cloning.....	**49**
Table 2.2	Touchdown PCR using Herculase II Fusion DNA Polymerase....	**50**
Table 2.3	Sequencing primers.....	**52**
Table 2.4	MyTaq colony PCR screen.....	**52**
Table 2.5	Primers for introducing ALG14 patient mutations.....	**57**
Table 2.6	Primers used to introduce ALG13 patient mutations.....	**58**
Table 2.7	Q5 site directed mutagenesis: PCR protocol.....	**58**
Table 2.8	Critical micelle concentration (CMC) in water of detergents used during purification of DPAGT1, ALG13 and ALG14.....	**62**
Table 2.9	Antibodies used for Western Blot analysis of DPAGT1, ALG13 and ALG14.....	**67**
Table 2.10	Protein stock concentrations, μM, as measured by the BCA assay or nanodrop and adjusted for impurities using densitometry.....	**69**
Table 2.11	Parameters of the detergents and lipids (OGNG/CHS/CDL) used to determine the size of the protein complex by SEC-MALS.....	**71**
Table 2.12	Example of grids prepared for cryoEM containing DPAGT1-FLAG/ALG13/ALG14.....	**80**
Table 3.1	Matrix of buffer conditions screened for use in ALG13 purifications.....	**93**
Table 3.2	Summary of melting temperature and onset of precipitation of ALG13/ALG14.....	**114**
Table 3.3	Thermostability of DPAGT1-HisFLAG/ALG13/His-ALG14 before and after size exclusion chromatography (SEC).....	**137**
Table 4.1	Kinetic parameters of DPAGT1 compared to literature values, where n=3 technical replicates.....	**167**
Table 4.2	Steady state parameters of DPAGT1.....	**169**
Table 4.3	DPAGT1 kinetics by radioactive TLC (N=2).....	**178**
Table 4.4	Substrate kinetics with ALG13/ALG14.....	**188**
Table 4.5	DPAG1/ALG13/ALG14 kinetics N=3 biological repeats.....	**196**

Table 4.6 Summary of binding parameters measured using a TLC assay for DPAGT1 and ALG13/ALG14 alone and in complex…… **202**

Table 5.1 Thermostability screen of DPAGT1/ALG13/ALG14 when bound to members of the native and non-native lipid panel…….. **218**

Table 5.2 DPAG1/ALG13/ALG14 N=3 biological repeats……... **226**

Table 5.3 DPAG1/ALG13/ALG14 N=3 biological repeats……... **228**

Table 6.1 Global root mean square deviation (RMSD) of alignments of the predicted structures of pathogenic ALG13/ALG14 proteins to the wildtype……...……... **237**

Table 6.2 Summary of the predicted pathogenicity of ALG13 variants by AlphaMissense…… **240**

Table 6.3 Summary of the predicted pathogenicity of ALG14 variants by AlphaMissense…… **241**

Table 6.4 Summary of production of ALG13 and ALG14 mutant proteins, as screened on a DNA, RNA and protein level…….. **256**

Table of Abbreviations

ALG	Asparagine linked glycosylation
CHS	Cholesterol hemisuccinate
DHDDS	Dehydro dolichyl diphosphate synthase
Gal	D-galactose
Glc	D-glucose
DMAPP	Dimethylallyl pyrophosphate
Man	D-mannose
DPAGT1	Dolichol phosphate α-*N*-acetyl D-glucosaminyl-phosphotransferase
ER	Endoplasm reticulum
LacNAc	Galactose - *N*-acetyl D-glucosamine
GPI	Glycosyl phosphatidylinositol
IPP	Isopentyl pyrophosphate
kDa	Kilo Dalton
kb	Kilobase pair
LLO	Lipid linked oligosaccharide
LPS	Lipopolysaccharide
MS	Mass spectrometry
DESI-TOF	Matrix
MEP	Methylerythritol phosphate
MVA	Mevalonate pathway
GalNAc	*N*-acetyl D-galactosamine
GlcNAc	*N*-acetyl D-glucosamine
DM	n-Decyl-β-maltoside
DDM	n-Dodecyl-β-maltoside
OGNG	n-Octyl glucose neopentyl glycol
OST	Oligosaccharyl transferase
PAGE	Polyacrylamide gel electrophoresis
SDS	Sodium dodecyl sulphate
SRD5α3	Steroid 5 α-reductase 3
2D	Two dimensional
UDP	Uridine diphosphate
UMP	Uridine monophosphate
HMG-CoA	β-hydroxy β-methyl glutaryl coenzyme A

Chapter 1

Introduction

1.1 The role of N-glycosylation

Protein glycosylation is an essential modification that plays a role in many biological processes, including protein folding, stability and trafficking. It refers to the enzymatic addition of glycans to specific amino acid residues [1] and greatly diversifies the proteome. In humans, there are four main types of protein glycosylation: O-glycosylation, where sugar molecules are attached to oxygen atoms of serine, threonine and occasionally tyrosine residues [2]; glypiation of proteins at their carboxyl terminus to yield a GPI-anchor [3], N-glycosylation, which occurs when sugars are linked to the side-chain amide nitrogen of an asparagine residue via a glyosidic bond [4], and, rarely, C-mannosylation of specific tryptophan residues [5]. Defects in any of the steps involved in glycosylation or the synthesis of their substrates can lead to congenital disorders of glycosylation (CDG) [6, 7].

N-glycosylation is the most prevalent in the human proteome, estimated to occur on over 85% of proteins travelling through the secretory pathway [8]. This process begins in the endoplasmic reticulum (ER) and involves three phases: first, the synthesis of the lipid-linked oligosaccharide donor substrate (here referred to as the early N-glycosylation pathway); next, the *en bloc* transfer of a $(GlcNAc)_2(Man)_9(Glc)_3$ glycan to an asparagine residue in an N-X-S/T sequon of a nascent protein - where X can be any amino acid with the exception of proline; and finally, diversification and processing of the glycan in the ER and Golgi apparatus to yield the mature oligosaccharide [9].

These steps are depicted in Figure 1.1, along with notable functions of N-glycans. Addition of a large, polar oligosaccharide helps to drive protein folding by adding hydrophilicity to the local protein environment and also by enabling binding of chaperones [10]. Three glucose and up to six mannose residues of the N-glycan are trimmed whilst in the calnexin/calreticulin protein folding cycle, and these truncated forms are recognised by chaperones to either promote folding or mark them as terminally misfolded where they will be targeted for ER-associated degradation by the proteasome [10, 11]. In this way N-glycans serve as the first quality control checkpoint for ensuring functional, folded proteins.

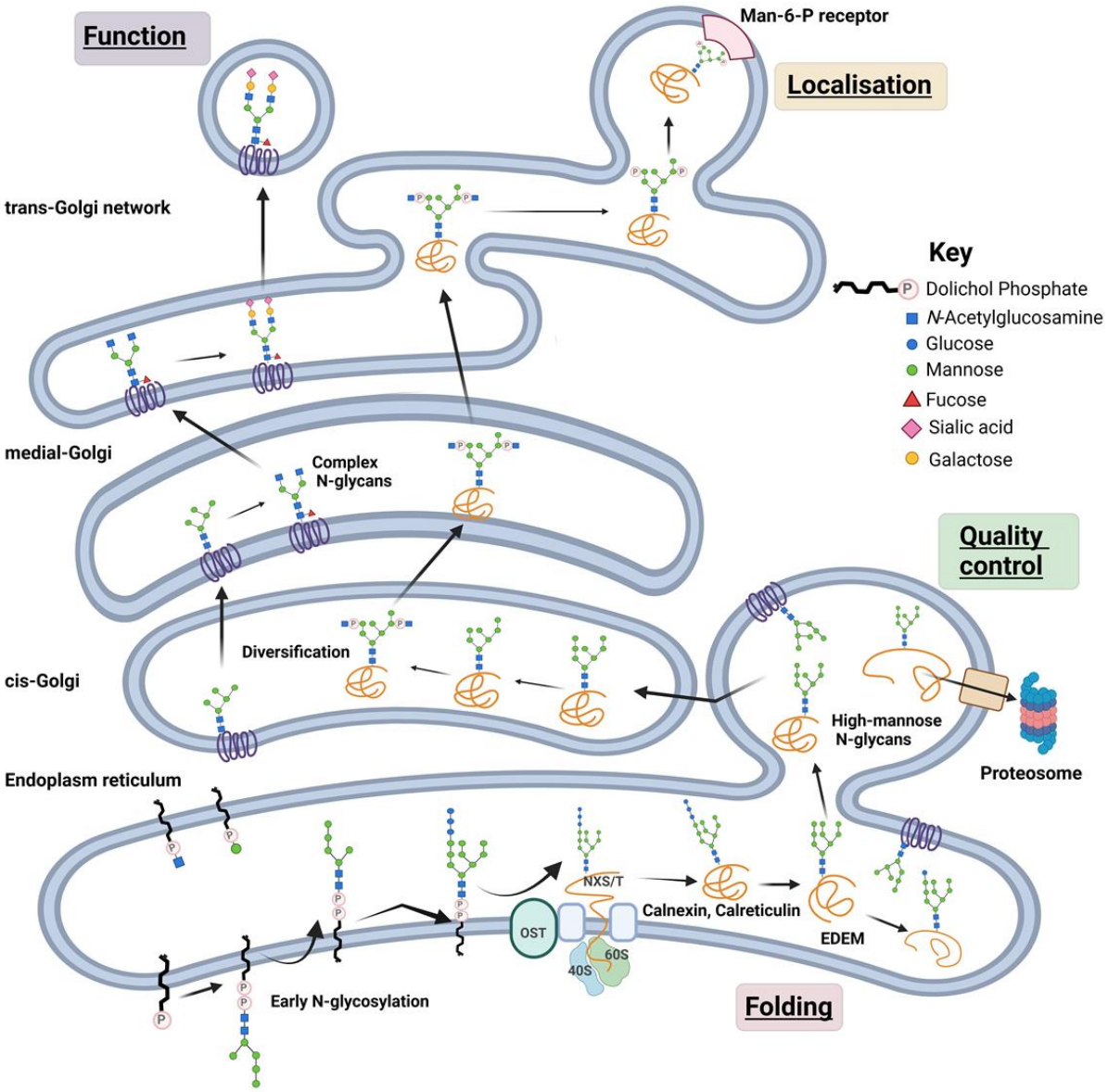


Figure 1.1 Schematic depicting the role of N-glycosylation in folding, quality control, localisation and function of proteins in the secretory pathway.

An elegant transport system for trafficking proteins along the secretory pathway is mediated by lectin recognition of N-glycans [10]. For example, binding of tetrameric ERGIC-53 to mannose residues guides glycoproteins from the ER to the ER-Golgi intermediate compartment in COP-coated vesicles [12, 13]. Recently, the roles of each of the six N-glycans on neuronal growth regulator 1 protein were deciphered, where, along with enabling folding, two N-glycans near the N-terminus were found to be crucial for ER to Golgi trafficking [14]. Viruses have long

been known to exploit this trafficking by N-glycans to allow infectivity. For instance, the N-glycan at position 260 on HIV protein gp120 protects it from lysosomal degradation and instead targets it to the endosome, making the modification crucial for its virulence [15]. Conversely, recognition of mannose-6-phosphate on the antennae of complex N-glycans on acid hydrolases by the mannose-6-phosphate receptor allows for the localisation of these enzymes as needed in the lysosome (as in Figure 1.1) [16, 17]. When in the more acidic (pH 5) environment of the lysosome, the affinity of the ligand is reduced allowing the release of the glycoprotein and action of the enzyme [18].

A diverse array of glycosidases and glycosyltransferases exist in the Golgi, which can add additional *N*-acetylglucosamine, galactose, fucose or sialic acid residues to the N-glycan core depending on the cell type, protein, or glycosylation site [17]. These N-glycans play a key role in receptor-ligand interactions, especially in enabling recognition of cells by the immune system [19]. For example, the N-glycan profile of receptors on B cells changes at different stages of differentiation, meaning the immune regulator galectin-9 binds to naïve and memory B cells but not germinal centre cells, as the latter have branched galactose and *N*-acetylglucosamine rather than linear chains on their N-glycan antennae [20]. Galectin-1 also recognises the N-glycans of glycoproteins CD7, CD43 and CD45 present on B and T cells, inducing apoptosis and downstream signalling [3, 21], again highlighting the fundamental role of N-glycosylation.

1.2 Lipid-linked oligosaccharide synthesis

The lipid-linked oligosaccharide (LLO) donor substrates needed for N-glycosylation are synthesised in the ER by asparagine-linked glycosylation (ALG) enzymes 1-14 (Figure 1.2). In the 1980s, Robbins *et al.* performed extensive genetic screening of temperature sensitive *Saccharomyces cerevisiae* mutants to identify the loci of *ALG* genes responsible for each step

of this early N-glycosylation pathway [22, 23]. By monitoring the accumulation of [^{3}H]-labelled glycans, the role of each silenced gene could be elucidated given the accumulated LLO represented its normal substrate. Since then, advances in techniques for accessing membrane proteins and their lipid-linked substrates have allowed for detailed *in vitro* study of many of the ALG enzymes involved, although to date much of this work has focused on the yeast model organism rather than those in higher eukaryotes.

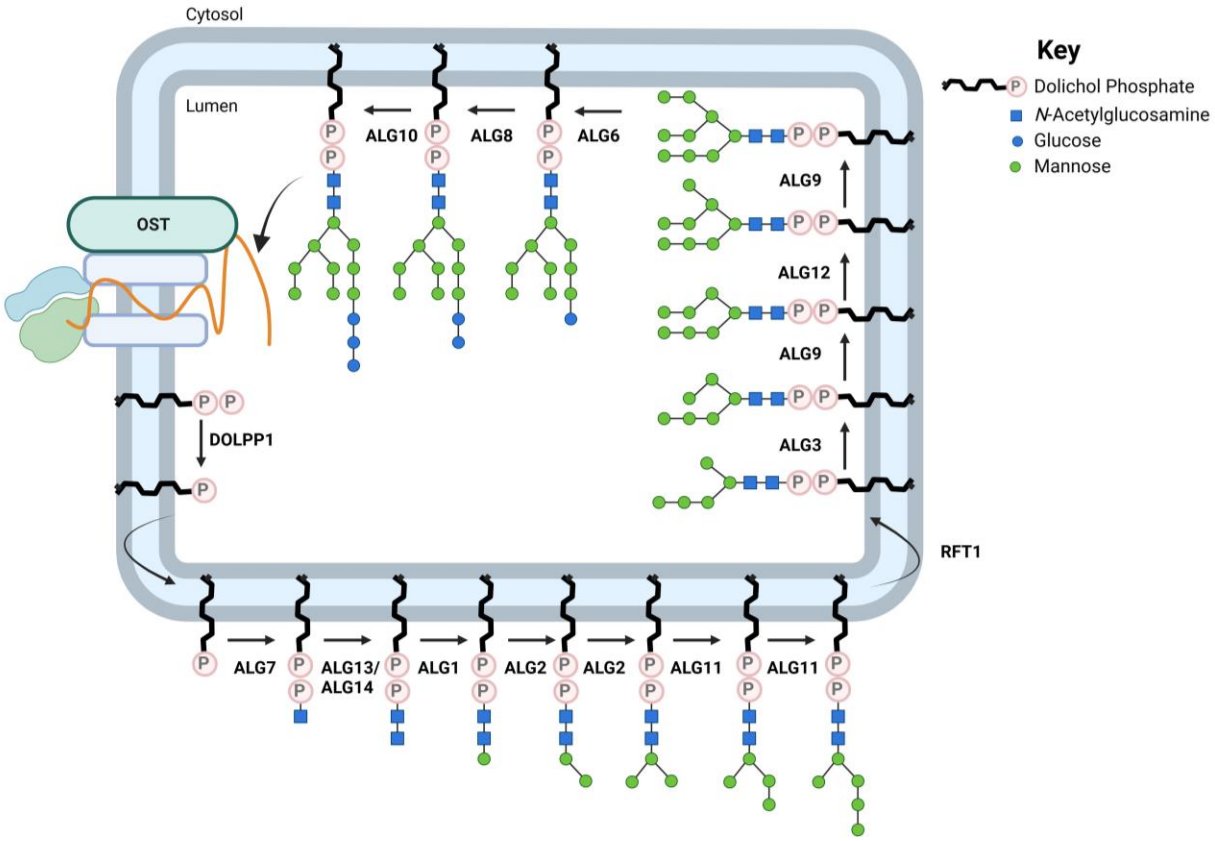


Figure 1.2 Asparagine linked glycosylation (ALG) enzymes 1-14 of the early N-glycosylation pathway, catalysing the synthesis of dolichol pyrophosphate-$(GlcNAc)_2(Man)_9(Glc)_3$. Human ALG7 is commonly known as DPAGT1.

1.2.1 Enzymes of the early N-glycosylation pathway

The early N-glycosylation pathway is initiated by dolichol-phosphate α-*N*-acetyl-glucosaminyl-phosphotransferase (DPAGT1), which catalyses the reaction between phosphorylated dolichol lipid and uridine diphosphate *N*-acetyl glucosamine (UDP-GlcNAc) to yield dolichol pyrophosphate GlcNAc (Figure 1.3) [24]. This enzyme is a member of the canonical polyprenyl-phosphate *N*-acetyl hexosamine 1-phosphate transferase (PNPT) enzyme family [25, 26]. Human DPAGT1 was previously studied by the Davis and Carpenter groups where structures, combined with mutagenesis and activity analysis, were used to elucidate the mechanism of catalysis [24] as discussed below.

The second step in lipid-linked oligosaccharide assembly involves the unusual heterodimeric complex ALG13/ALG14, which catalyses the transfer of a second GlcNAc from UDP-GlcNAc onto dolichol pyrophosphate GlcNAc yielding a chitobiose-linked dolichol (Figure 1.3). ALG14 is an integral ER membrane protein that recruits soluble ALG13 from the cytosol to the face of the ER membrane via an interaction at its C-terminus [27, 28]. Both ALG13 and ALG14 have been shown to be required to form the active glycosyltransferase [29]. In 2008, the Imperiali group characterised the structure of yeast ALG13 using NMR and found a UDP-GlcNAc binding site at its C-terminus [30]. These data suggest that ALG14 bears the lipid acceptor binding site, however, to date there has been no reported structure of the dimeric glycosyltransferase or of ALG14 alone.

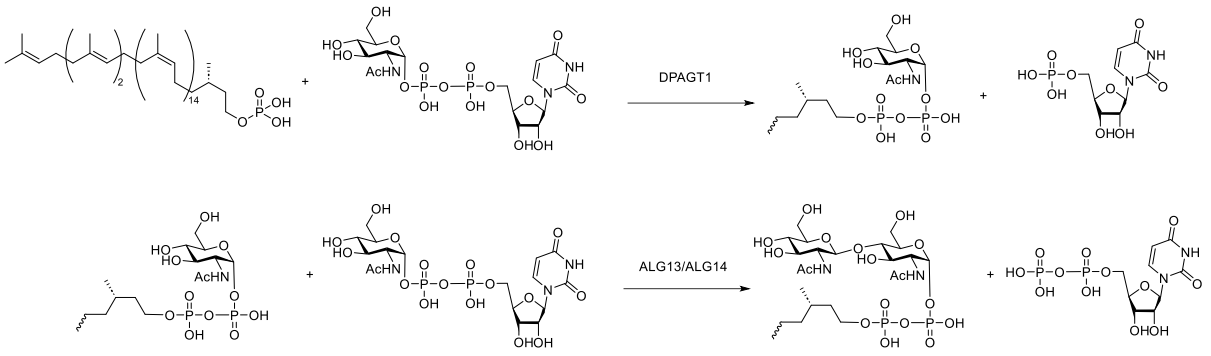


Figure 1.3 Action of DPAGT1 and ALG13/ALG14 enzymes of the early N-glycosylation pathway. Phospho-GlcNAc and GlcNAc are transferred from UDP-GlcNAc onto dolichol phosphate by DPAGT1 and ALG13/ALG14 respectively.

Following ALG13/ALG14 action, 5 branched mannose units are added by ALG1, 2 and 11 glycosyltransferases using GDP-mannose substrates. Detailed *in vitro* study of yeast ALG1 mannosyl transferase was achieved by Li *et al.* [31] who exploited the fact that ALG1 can utilise short, phytanol linked pyrophosphate $(GlcNAc)_2$ substrates as well as the native dolichol versions [32]. Using a liquid chromatography-mass spectrometry (LC-MS) assay they compared the activity of wild-type protein to those carrying missense mutations from patients with CDG [31]. Similarly, substrates with short, non-natural lipid moieties were used for studying action of downstream ALG2 and ALG11 mannosyl transferases, although it was found ALG2 required an acceptor with a lipid tail of at least four isoprene units [33]. Interestingly, ALG2 shows dual specificity as it is able to add mannose residues with both α-1,3 and α-1,6 glycosidic linkages [33-36], as depicted in Figure 1.4, whereas ALG11 decorates only the 1,3-linked sugar with two α-1,2 mannose [37]. There is therefore representation of both inverting and retaining glycosyltransferases in this early N-glycosylation pathway.

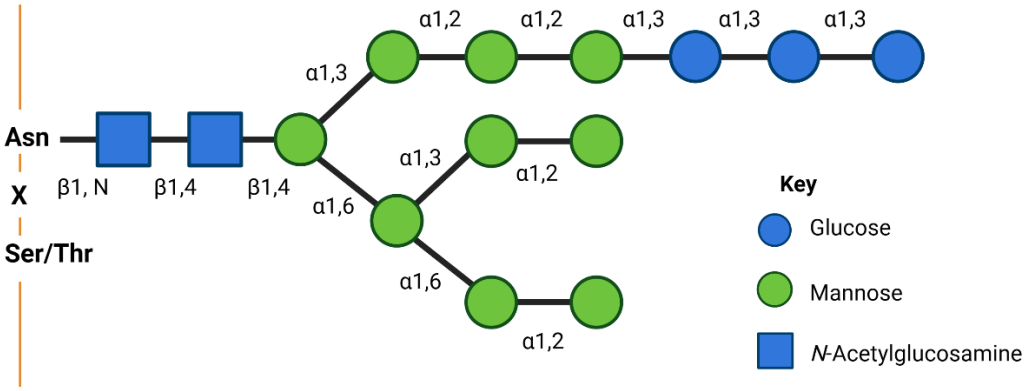


Figure 1.4 Structure of the core, eukaryotic N-glycan transferred to a nascent protein.

For final extension to the terminal N-glycan, the lipid-linked oligosaccharide requires movement from the cytosolic leaflet to the luminal side of the ER. The mechanism of movement of this large amphipathic molecule across the bilayer is still an active area of research, with several mechanisms proposed including spontaneous diffusion, action of flippases, or channel-mediated translocation [38, 39] . Similarly, the dolichol phosphate mannose and glucose donor molecules needed for luminal decoration of the N-glycan are synthesised on the cytoplasmic leaflet and therefore also require flipping across the ER membrane. Helenius *et al.* showed that yeast RFT1 is required for LLO flipping *in vivo* [40], although this activity has not been recapitulated when studied in liposomes, drawing into question whether RFT1 instead plays an accessory role [41, 42].

Once inside the lumen, ALG3, 9 and 12 mannosyl transferases act to provide the substrate for ALG6 which in turn catalyses the addition of the first glucose residue to the dolichol-linked glycan. The structure of yeast ALG6 in complex with a short, citronellyl-farnesyl linked sugar donor mimic was solved using cryogenic electron microscopy (cryoEM), utilising the binding of a Fab fragment to increase particle size [43]. The authors suggest a conserved architecture

amongst GT-C fold glycosyltransferases like ALG6 [43], with a common N-terminus containing seven transmembrane helices that interacts with the polyisoprenol donor substrate followed by a variable C-terminal domain. Finally, ALG8 and 10 add the terminal glucose residues, before oligosaccharide transferase (OST) acts to transfer the glycan onto a nascent protein emerging from the SEC61 translocon [44-46]. This makes N-glycosylation a co-translational modification in eukaryotes, where the Stt3 subunit of OST is homologous to the post-translational action of archaeal AglB and bacterial PglB enzymes [47].

1.2.2 Conservation across the domains of life

In 1976, the first non-eukaryotic glycoprotein was identified – a surface layer glycoprotein in the archaea *Halobacterium salinarum* [48]. This was later also found to be synthesised using sugars docked on phosphorylated dolichol [49]. Similarly, polyisoprenol phosphates in bacteria are essential for the synthesis of glycoconjugates such as lipopolysaccharide, capsular polysaccharide and glycoproteins [50]. In each case, intermediates are synthesised on the cytoplasmic leaflet of the lipid bilayer before translocating into the periplasm or the extracellular space in Gram-negative or Gram-positive bacteria respectively [51]. These building blocks are further polymerised during N-glycosylation and in O-antigen polysaccharide capsule synthesis, or transferred directly to a lipid A-core oligosaccharide [52]. There are therefore many parallels between eukaryotic and archaeal N-glycosylation and bacterial peptidoglycan and lipopolysaccharide synthesis [52], highlighted in Figure 1.5. In turn the glycosyltransferases responsible for these pathways are found to be highly conserved across the domains of life.

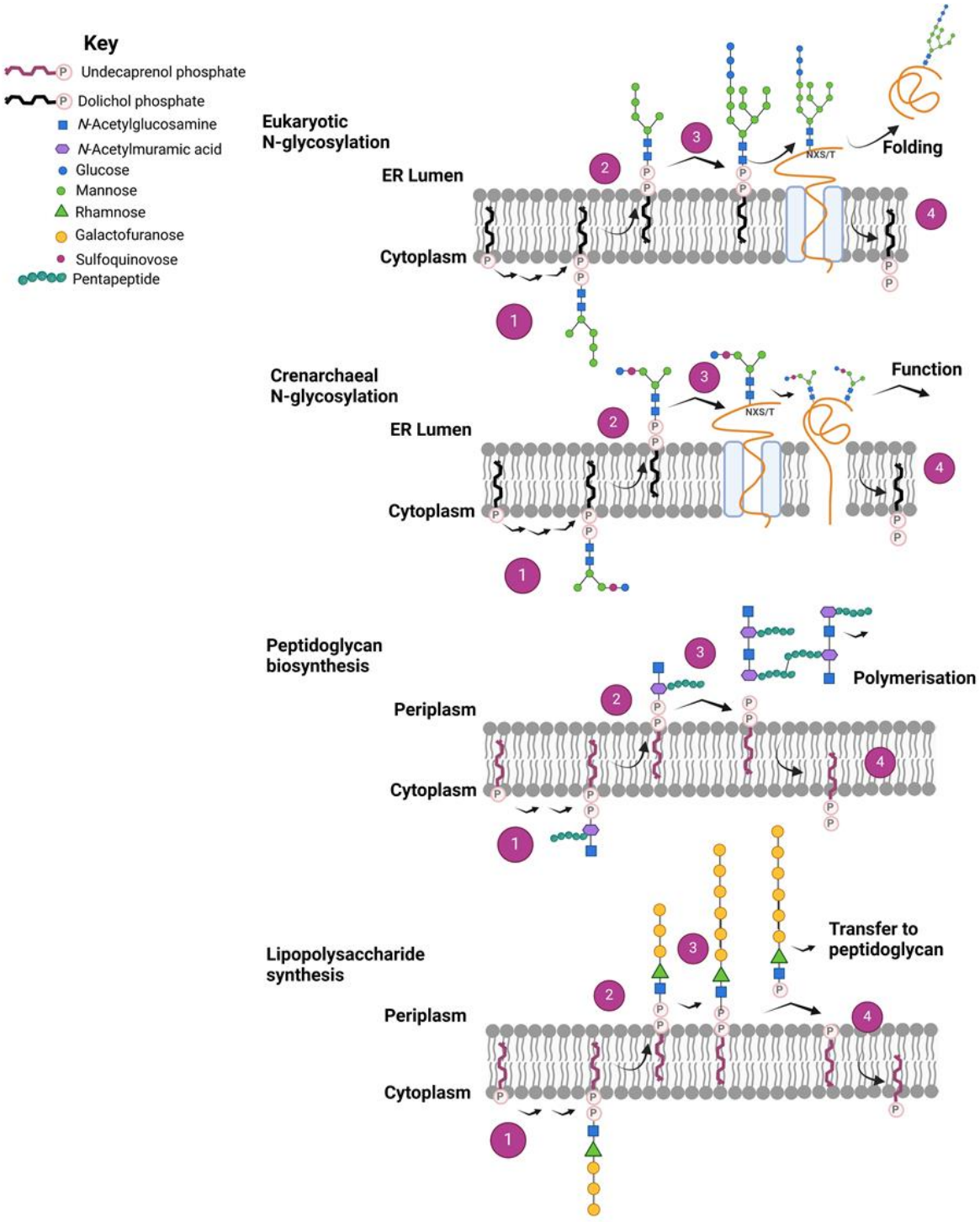


Figure 1.5 Parallels between the early stages of eukaryotic and crenarchaeal N-glycosylation and bacterial peptidoglycan and lipopolysaccharide synthesis. 1. Decoration of a polyisoprenol phosphate on the cytoplasmic leaflet. 2. Flipping of the lipid-linked oligosaccharide across the bilayer. 3. Transfer of the oligosaccharide for downstream use. 4. Recycling of the polyisoprenol phosphate to the cytoplasmic leaflet.

It is therefore likely that the basis of eukaryotic N-glycosylation was inherited from prokaryotic systems, and, given the common use of dolichol type lipids decorated with a chitobiose core, it has been proposed to be most likely from an archaea symbiosis event [52, 53]. Interestingly, Larkin *et al*. have reported an alternative pathway found in select archaea which use a monophosphate sugar linkage [54]. For example in the euryarchaeotic *Methanococcus volae,* the pathway is initiated by the action of AglK which transfers GlcNAc to dolichol phosphate, and AglJ in *Haloferax volcanii* which transfers a glucose molecule [45, 54]. These show closer sequence similarity to eukaryotic ALG5 which yields dolichol phospho-glucose than the initiating PNPT [45].

This PNPT enzyme family was first described by Mark Lehrman and includes eukaryotic DPAGT1 (ALG7) utilising UDP-GlcNAc and dolichol phosphate as substrates as well as the bacterial transferases MraY, WecA, and WbpL - which differ in their donor sugar specificity but all utilise undecaprenol phosphate as a lipid acceptor [25, 55] - and AGLH, which has been identified as an ancient homolog in the archaea *Sulfolobus acidocaldarius* [56]. WecA in *Escherichia coli* and the *Pseudomonas aeruginosa* homolog WbpL transfer GlcNAc to begin formation of lipopolysaccharide [57], whereas MraY transfers MurNAc linked to a pentapeptide to initiate peptidoglycan synthesis [58]. PNPTs typically have 10-11 transmembrane helices, and all require the divalent cations Mg^{2+} or Mn^{2+} for activity [55, 59]. A highly conserved DDxxD/N/E motif was previously predicted to be involved in metal ion coordination [55], but surprisingly the crystal structures of MraY [58] and DPAGT1[24] revealed different Mg^{2+} coordination mechanisms with none of the conserved aspartates coordinating the divalent cation in the eukaryotic protein. Furthermore, despite both being head to head homodimers, DPAGT1 and MraY have unique dimer interfaces which has been proposed to allow for cooperativity in the eukaryotic system [24, 60].

MraY translocase is potently inhibited by the lipid-nucleoside analogue tunicamycin from *Streptomyces lysosuperficus* and, given their similarity, it is not surprising that this natural product also inhibits DPAGT1 [61]. This renders it toxic to eukaryotes and therefore prevents its use as an antibiotic. The structure of DPAGT1 in complex with tunicamycin helped in the design of non-toxic derivatives [24], as it was found there was a gating loop in the human protein adjacent to the tunicamycin binding site that is absent in MraY (Figure 1.6). As described by Yoo *et al.*, this likely reflects the need for MraY to accommodate the pentapeptide on its lipid substrate [60]. Tunicamycin analogues containing two lipid chains were designed that selectively targeted *Mycobacterium tuberculosis* but not human DPAGT1 [24].

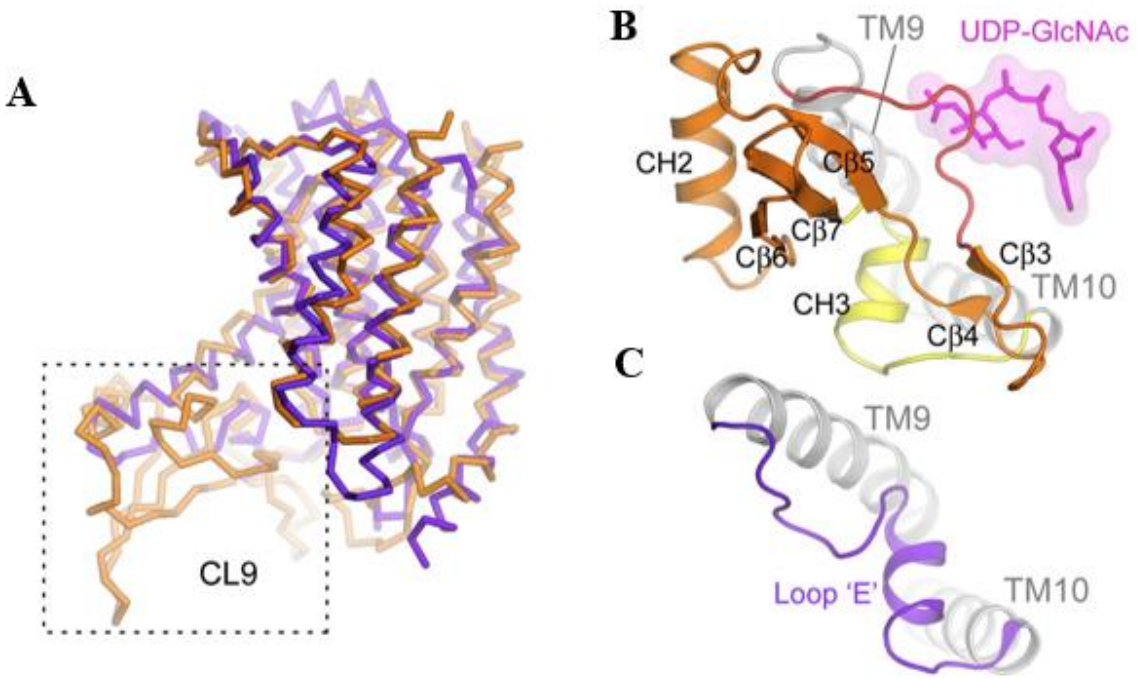


Figure 1.6 Comparison of the structure of DPAGT1 and MraY. A Ribbon representation of DPAGT1 (orange, PDB: 5LEV) and MraY (purple, PDB: 5CKR) highlighting a gating loop (CL9 domain, **B**) present in DPAGT1 where Loop 'E' is found in MraY (**C**). Figure adapted from Dong *et al.* [24] with permission under the Creative Commons license (http://creativecommons.org/licenses/by/4.0/).

WbaP from *Salmonella enterica* transfers galactose to undecaprenol phosphate but was not found to have high sequence similarities to the PNPT family enzymes and instead is a prototype for an evolutionarily unrelated polyprenyl-P hexose-1-P transferases (PHPT) class [25, 62]. PglC from *Campylobacter jejuni* is homologous to the C-terminal domain of wbaP, and uses the rare sugar bacillosamine [63, 64]. Both PNPT and PHPT glycosyltransferases initiate the first committed step for building complex glycans (whether that be O-antigens, lipopolysaccharide, peptidoglycan, or sugars for N-linked glycosylation) to yield a pyrophosphorylated lipid linked sugar.

While substrate specificity differs between eukaryotic and prokaryotic glycosylation pathways, the transfer of a GlcNAc using UDP-GlcNAc following initiation is universal. The identity of the inverting β1,4-GlcNAc transferase needed for this was the last of the ALG pathway to be elucidated in yeast owing to the unusual configuration as two distinct subunits [29]. These proteins are found to be highly conserved amongst eukaryotes, with the similar sequence identity of ALG13 and ALG14 proteins from human, yeast and rat shown in Figure 1.7. Using the EMBOSS Needle pairwise alignment tool [65], human ALG13 was found to have 26.5% identity with the yeast protein, and human ALG14 shares 30.4% identity with the yeast version. In humans, the *ALG14* gene is found on chromosome 1 (1p21.3) [66] whereas the *ALG13* gene is localised on the X chromosome at locus 23 (Xq23) [67]. Furthermore, in vertebrates, multiple isoforms of ALG13 exist [68]. In humans the long isoform 1 has unknown function whereas isoform 2 encodes the glycosyltransferase domain homologous to that found in yeast and plants (Figure 1.7B). It has been hypothesised that this separation of distinct subunits evolved to allow for increased regulation of the early N-glycosylation pathway in higher eukaryotes, with cytosolic ALG13 protein under tight control by proteolytic degradation [29, 69].

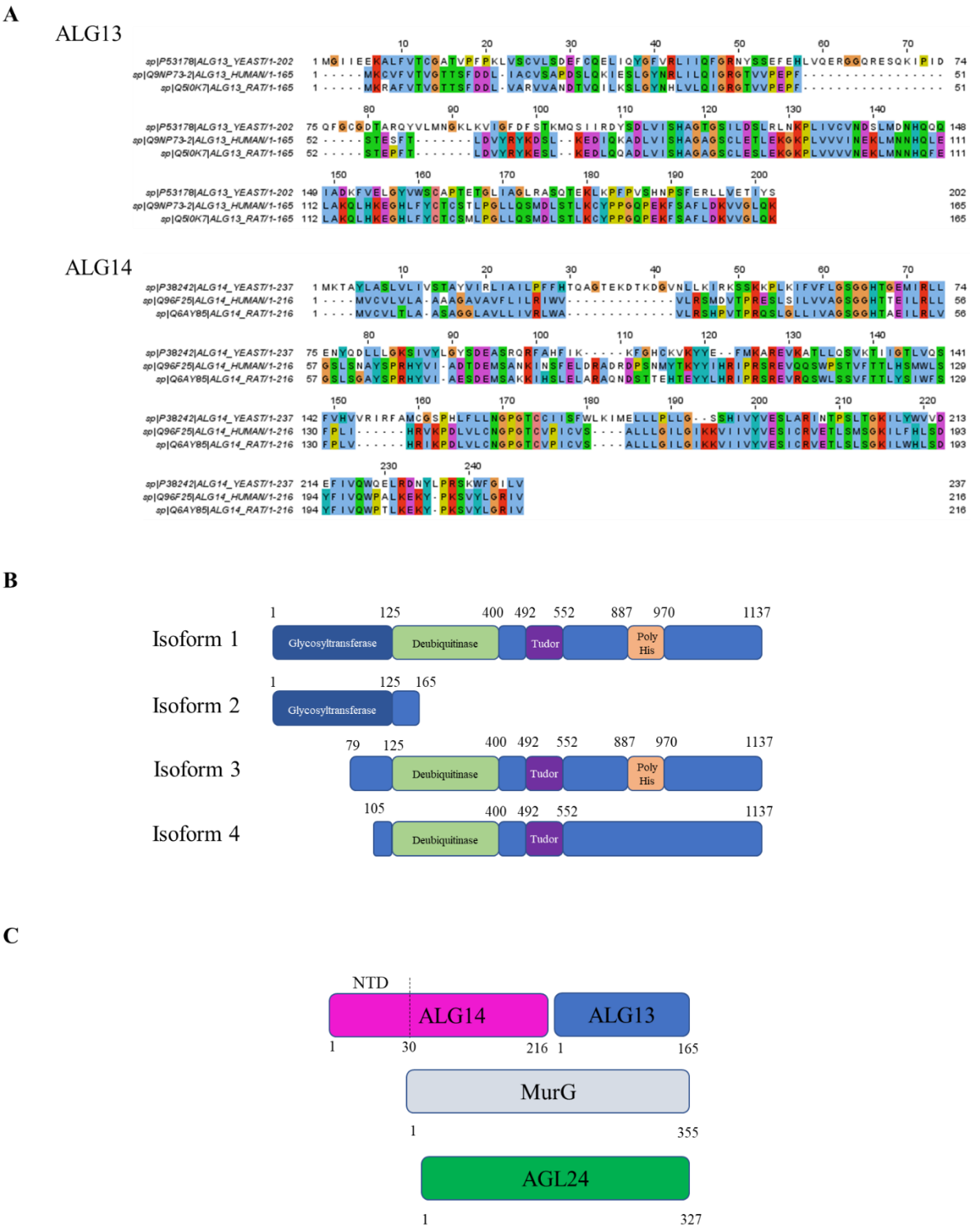


Figure 1.7 Comparison of ALG13/ALG14 homologs. A Clustal Omega multiple sequence alignment of ALG13 (top) and ALG14 (bottom) from yeast, human, and rat. Alignments were performed using EMBL-EBI [65] and visualised using Jalview [70], where residues were coloured as per Clustal Omega standards. UniProt identifiers were ALG13:

P53178 (Yeast), Q9NP73-2 (Human, isoform 2), Q510K7 (Rat), and ALG14: P38242 (Yeast), Q96F25 (Human), and Q6AY85 (Rat). **B** Structure of human ALG13 isoforms 1-4, labelled with known functional domains and by amino acid number from the start of isoform 1. Schematic adapted from that by Ng *et al.* [71]. **C** Schematic depicting the analogous nature of MurG, AGL24 and ALG13/ALG14, where ALG14 has an additional N-terminal domain (labelled as NTD) absent in bacterial MurG and archaeal AGL24.

Together the ALG13/ALG14 proteins are homologous to the bacterial GlcNAc-transferase MurG which transfers GlcNAc to MurNAc-pentapeptide pyrophosphate undecaprenol to yield lipid II [72]. Detailed structural information is available for both apo and UDP-GlcNAc bound MurG [72, 73], with an ALG14-like N-terminus that lacks a transmembrane domain (Figure 1.7C). Recently, an archaeal functional homolog to these proteins was identified, designated AGL24 [74]. The authors note that despite a low (16%) sequence identity, AGL24 is predicted to have a MurG-like GT-B fold [74] , and found that histidine residue at position 14 - conserved between the proteins across the three domains of life - was essential for the activity of this thermostable archaeal protein [74]. This again suggests a shared common origin of the enzymes in the early N-glycosylation pathway.

1.2.3 Multi-enzyme complexes amongst the early N-glycosylation pathway

Several enzymes of the early N-glycosylation pathway are predicted to function together in multi-enzyme complexes, although this remains an under-researched area particularly with regard to the organisation of the human proteins. Cryogenic electron tomography (CryoET) has allowed for detailed visualisation of the interactions of the OST complex with the ribosome and translocon which exists as a dynamic interplay between many proteins to ensure efficient protein production (Figure 1.8) [75, 76]. It is therefore likely that all members of the early N-glycosylation machinery are in close proximity to this complex for ready supply of N-glycans at a high local concentration.

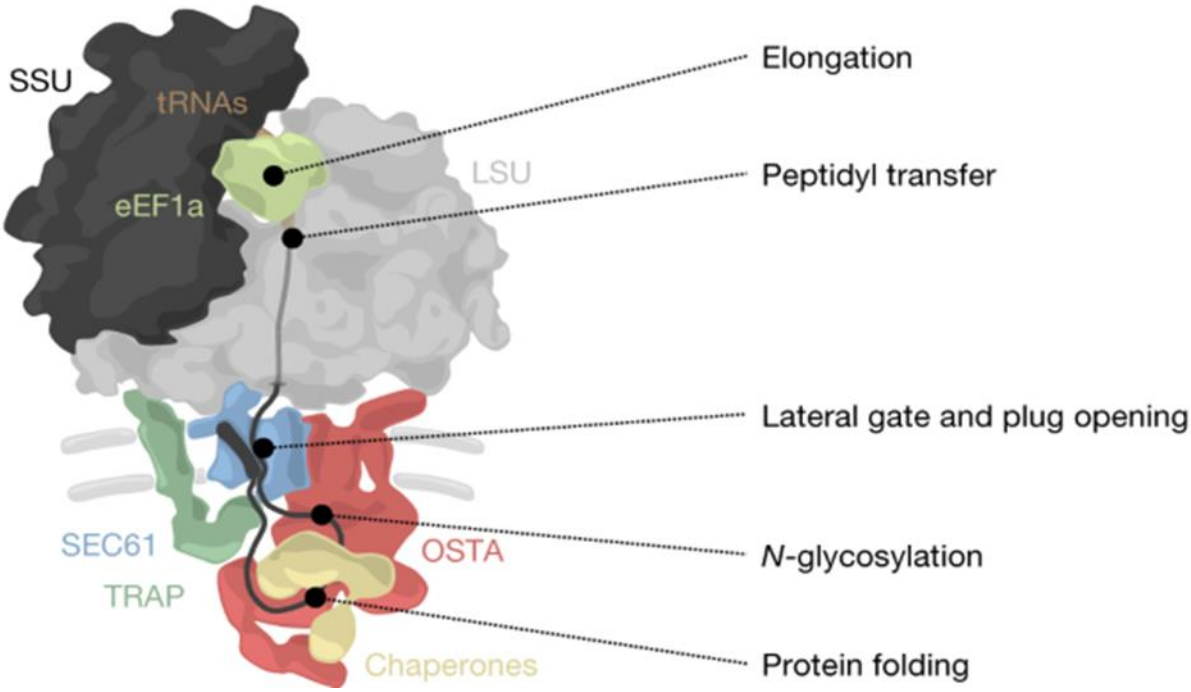


Figure 1.8 The ER translation-translocation machinery complex where the N-glycan is transferred. Figure adapted from Gemmer *et al.* [75] with permission under the Creative Commons license (http://creativecommons.org/licenses/by/4.0/).

The crystal structures of both human DPAGT1 and *E.coli* MraY were solved as homodimers [24, 60, 72], and in solution DPAGT1 has been detected as both monomers and dimers [24]. Given the large dimer interface and stability afforded by this dimerisation, it is assumed that this is also the case *in vivo* but this has not been confirmed experimentally [61]. Furthermore, ALG1 has long been known to form complexes with ALG2 or ALG11 proteins, as well as forming homodimers through interactions of its C-terminus [28]. Gao *et al.* suggest these are physiologically relevant complexes given mutant *ALG1* alleles that disrupted complex formation led to inactive mannosyl-transfer *in vivo* [28].

Noffz *et al.* were the first to show in yeast that ALG7 (DPAGT1) interacts with downstream ALG13/ALG14 [77], which was later shown to be facilitated through the interaction of the N-terminal domain of ALG14 [27]. Interestingly, the archaeal homolog AGL24 lacks a transmembrane region but has been found to be membrane associated *in vivo* and therefore it has been suggested to interact with upstream AGLH in a similar way [74]. Studying the interaction of the human initiator DPAGT1 with downstream ALG13/ALG14 would be particularly interesting as it will allow insight into the regulation and dynamics of the start of the pathway.

There are a number of processes governed by individual enzymes in bacteria that are organised as a complex in higher eukaryotes. This is classically exemplified by Type I and Type II fatty acid synthase (FAS) for which it is likely the eukaryotic multimeric assemblies evolved from the dissociated type II bacterial FAS enzymes [78, 79]. This arrangement may facilitate substrate channelling, where, for the early N-glycosylation pathway, the dolichol-linked oligosaccharide would be passed to subsequent enzymes without diffusion into the bilayer [80]. Lipids in synthetic membranes have been estimated to have diffusion coefficients of 10^{-8} cm^2 sec^{-1} meaning the rate of substrate diffusion in the ER bilayer would not be rate limiting given typical rates of enzyme catalysis [80]. However, it could be imagined that substrate channelling by enzyme

complexes may increase the overall rate of substrate production if it helps to alleviate the product inhibition seen for many of the ALG proteins [24, 81]. As discussed for the heterodimeric ALG13/ALG14, the use of complexes has also been suggested to provide an increased ability to respond to changes in requirements; being able to adjust the rates of many steps in the pathway through changes in regulation of a single subunit [82].

1.3 Polyisoprenol structure and function

As discussed, the polyisoprenol lipid dolichol serves as a scaffold for the stepwise assembly of the core N-glycan, $(GlcNAc)_2(Man)_9(Glc)_3$. The general structure of a polyisoprenol molecule is shown in Figure 1.7, where repeating isoprenoid units are arranged in *trans* and *cis* geometry. These lipids are found in a variety of chain lengths and saturation states across the three domains of life and are known as dolichols when the α-terminal unit is saturated. Polyisoprenols are non-bilayer forming lipids and instead their presence is associated with an increase in bilayer fluidity [83, 84], encouraging lipid movement across leaflets by promoting hexagonal phase II organisation and in turn aiding in membrane fusion events [85]. Their primary function however comes when phosphorylated where they act as sugar transfer molecules, anchoring oligosaccharides to membranes [45, 85].

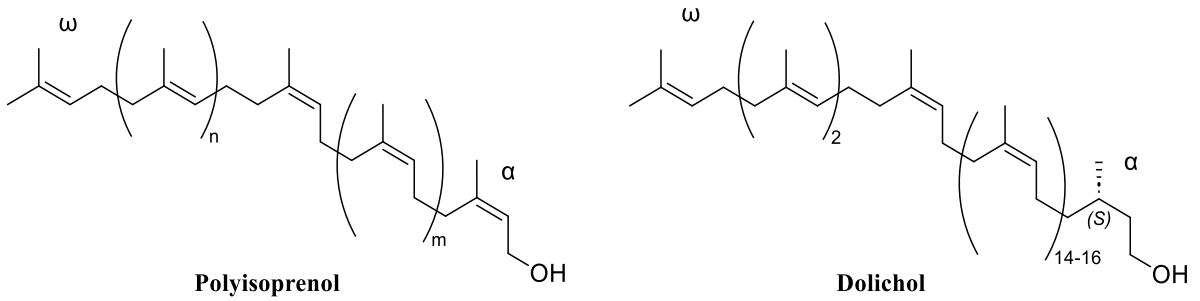


Figure 1.9 Polyisoprenols utilised in glycan biosynthesis (Left) General structure of polyisoprenols where n represents the number of *trans* configured isoprenoids and m the *cis*-configured units. The α and ω units are found in various saturation states across

nature. (Right) Dolichols are eukaryotic polyisoprenols with a characteristic saturated α-terminal isoprenoid unit.

Table 1.1 shows a comparison of polyisoprenol structures favoured in eukaryotes, bacteria and archaea cells. Typically, these polyisoprenols are represented in a distribution of chain lengths within membranes, rather than a single number of isoprenoid units being made [86]. In humans, dolichols possess 2 *trans*-configured isoprenoids (n, Figure 1.9) and between 12-17 *cis*-configured units (m, Figure 1.8). Similarly, *Saccharomyces cerevisiae* membranes contain dolichols of n= 2, m= 12-13 [86, 87]. Interestingly, a wide variety of polyisoprenols and dolichols have been isolated from plants, with either 2 and 3 *trans*-configured isoprenoids followed by anywhere from 6 to 37 *cis*-configured isoprenoid chains recorded [88]. Polyunsaturated lipids dominate in bacterial membranes, with the 2 *trans*, 7 *cis* C55 lipid known as undecaprenol being most prevalent [45, 89]. Archaea have been found to use dolichols with additional saturation at the ω-isoprene, as well as a variable number throughout the chains of C30-70 in length [45, 90, 91]. Studying the preference of glycosyl- and oligosaccharide-transferases for polyisoprenols of different saturation state and chain length may help us decipher the evolutionary history of the ability to utilise polyisoprenols for transfer of sugars across and within membranes [52].

Table 1.1 Summary of polyisoprenol structures across the domains of life [45, 84, 86, 87, 89, 90]

	trans	***cis***	**α- saturation**	**ω-saturation**
Eukaryotes	2-3	6-37	Yes	No
Archaea	2	2-10	Yes	Yes
Bacteria	2-3	6-9	No	No

1.3.1 Biogenesis of polyisoprenols

Saccharomyces cerevisiae has long been the workhorse model organism for studying eukaryotic dolichol synthesis, and it was in this yeast that the HMG-CoA reductase or mevalonate pathway (MVA) was identified. The MVA pathway yields mevalonate from acetyl-CoA and pyrophosphorylated mevalonate is decarboxylated to give the five carbon isopentyl pyrophosphate (IPP) which is isomerised to dimethylallyl pyrophosphate (DMAPP), together yielding geranyl pyrophosphate, responsible for the two-*trans* orientated ω-end of the molecule. Extension with subsequent IPP yields farnesyl pyrophosphate which is the starting point of polyisoprenol synthesis as well as of cholesterol and ubiquinone [85].

In the endoplasmic reticulum, elongation of farnesyl pyrophosphate to polyisoprenoid pyrophosphate is achieved by successive condensation of IPP molecules in *cis* orientations [85]. These reactions are catalysed by a dehydrodolichyl diphosphate synthase (DHDDS), yielding phosphorylated polyisoprenols in the cytosolic leaflet (Figure 1.10) [85]. Once dephosphorylated, steroid 5-α-reductase-3 (SRD5α3) catalyses the asymmetric reduction of the terminal isoprenol to give dolichol [92]. It is currently unknown which protein is responsible for this dephosphorylation in humans, and whether there is a single enzyme catalysing the removal of both phosphates rather than two separate proteins performing each dephosphorylation step. It has been theorised that dolichol pyrophosphate phosphatase (DOLPP1) catalyses the removal of pyrophosphate from both polyisoprenoid pyrophosphate during dolichol biosynthesis, and from dolichol pyrophosphate to recycle the monophosphorylated carrier following transfer of glycans to emerging proteins or other lipid carriers [93]. This would mean this phosphatase can utilise both terminally unsaturated and saturated polyisoprenol substrates. A geranylgeranyl reductase in halophilic archaeon is responsible for reducing the ω-prenol unit [94]. Ultimately, dolichol monophosphate is generated by dolichol kinase (DolK) in a CTP-dependent manner, which can react with phosphorylated sugars [85, 95].

In most bacteria and in the chloroplasts of plants, polyisoprenols are generated by sequential condensation of IPP from a pyrophosphorylated, all-*trans* oligoprene – typically of two isoprene units [96]. IPP and DMAPP are made via a non-mevalonate pathway using methylerthritol phosphate (MEP). Given these lipids are used in the polyunsaturated form, the sugar carrier molecule polyisoprenol phosphate can be generated by a single dephosphorylation (Figure 1.10) [97]. This means plants use both MVA and MEP pathways; a likely relic of endosymbiosis of cyanobacteria within the eukaryotic cell [87, 98].

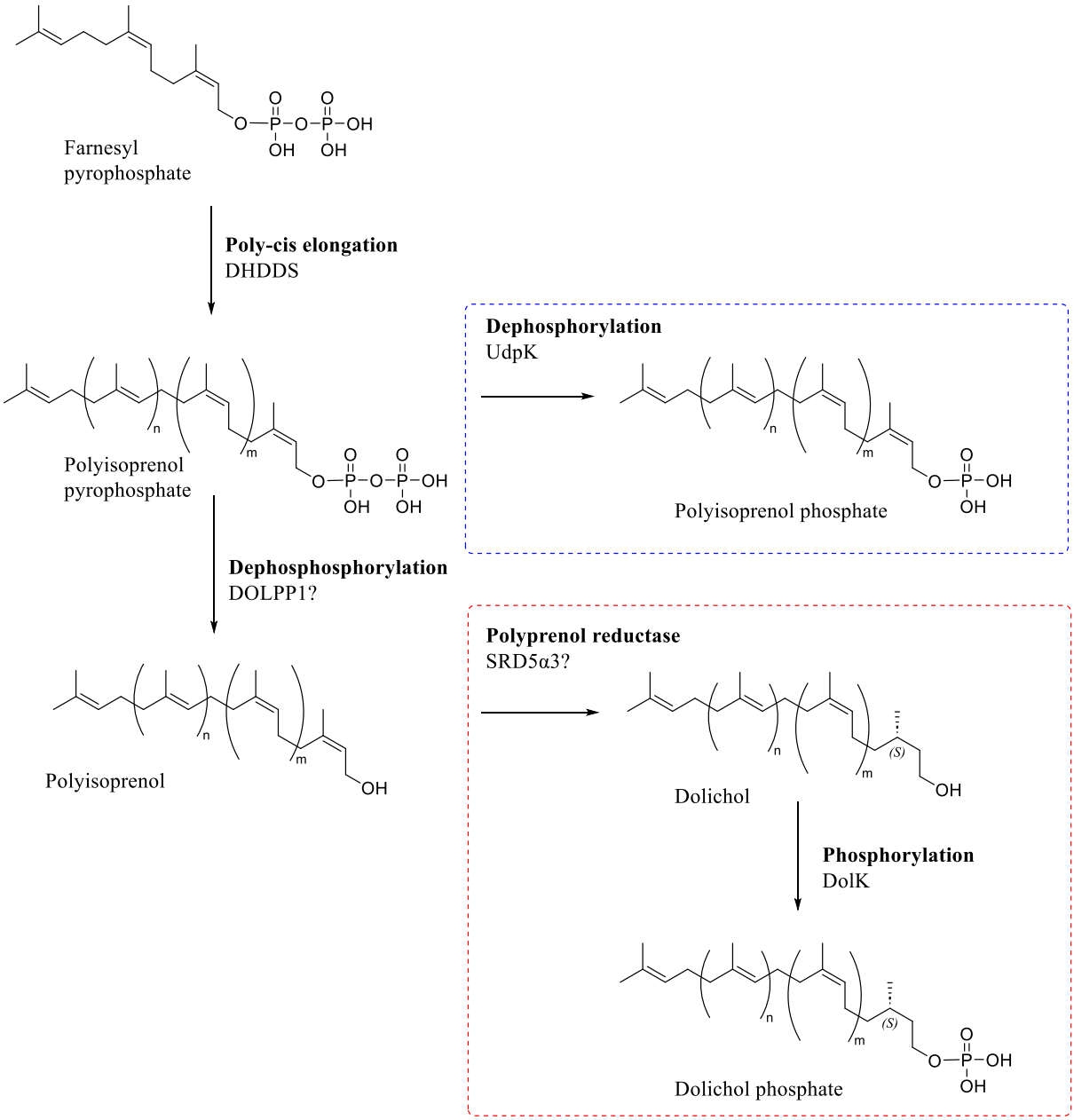


Figure 1.10 Biogenesis of polyisoprenol and dolichol phosphate. The blue box indicates the bacterial pathway, and orange that of dolichol-utilising eukaryotes. The enzymes responsible for each step are labelled, where DHDDS = dehydrodolichyl diphosphate synthase, UdpK = undecaprenol kinase, DOLPP1 = dolichyl pyrophosphate phosphatase 1, SRD5α3 = steroid 5-alpha reductase, and DolK = dolichol kinase. Presence of a question mark indicates unknown alternative enzymes that are likely to be involved in the process.

1.4 Congenital disorders of glycosylation

N-glycosylation is crucial to the normal function of healthy cells and when dysregulated is associated with Congenital Disorders of Glycosylation (CDG). By monitoring the impact of pathogenic variants of early N-glycosylation machinery, insight into the required regulation of the pathway in healthy cells can be gained. CDGs are a heterogeneous group of over 170 diseases, attributed to defects in any step involved in protein glycosylation including in the synthesis of dolichol and other required substrates [66, 99]. These are typically monogenic disorders with autosomal recessive inheritance although there are instances of both autosomal dominant and X-linked subtypes [99]. The first CDG was described in the 1980s by Jaeken *et al*. [100], but it has been in the last 20 years with the advent of next generation sequencing techniques that the majority of CDG subtypes have been identified.

Patients with CDGs present with a wide range of symptoms making diagnosis of these rare diseases challenging. A combination of clinical observations, biochemical analysis and genetic investigations is therefore needed [101]. Typical symptoms include developmental delay, hypotonia, intellectual disability, facial dysmorphisms and liver dysfunction [99, 102] , but interestingly, some forms of CDG primarily affect one organ. For example DPAGT1-, ALG2- and ALG14-CDG are associated with defects in the neuromuscular junction and result in myasthenic syndromes (CMS) [103, 104], whereas RFT1-CDG patients typically have deafness as well as a severe neurological disorder [105]. ALG13-CDG is characterised by severe infantile epilepsy, amongst other predominantly neurological symptoms [67].

Biochemical assessment usually takes the form of serum transferrin analysis to identify an abnormal glycosylation profile. Transferrin exists in various isoforms but typically contains two N-glycosylation sites which normally possess glycans with mature biantennary chains containing terminal sialic acid residues [102]. Depending on the CDG type, these glycans can

become truncated or missing and these abnormalities result in glycoproteins with altered isoelectric focusing (IEF) patterns resulting from the loss of the negatively charged sialic acid [102, 106]. A decrease in tetra-sialotransferrin and increase in di- and a-sialotransferrin bands measured by IEF are designated a type I pattern and are typically associated with CDGs with impaired N-glycosylation. A type II IEF pattern has the same features as type I but also shows an increase of tri- and mono-sialotransferrin [102]. This pattern is associated with CDGs with defects in diversification and processing of N-glycans [102]. Surprisingly, in some subtypes such as ALG13-CDG, a normal transferrin glycosylation pattern is observed which can delay diagnosis and has led some to question its classification as a glycosylation disorder [107, 108]. Genetic testing is the only definitive method for diagnosis of these CDGs, and this often requires whole-exome or whole genome sequencing for comprehensive assessment of gene variants.

1.4.1 Current challenges in the field of CDGs

There remains little mechanistic understanding of how most CDG mutations lead to their disease phenotype, with biochemical analysis often limited to monitoring protein expression levels to show a deficiency in a glycosylation enzyme. For example, the expression of ALG14 bearing CDG variant Pro65Leu was found to have reduced expression in HEK293 cells [103], and certain pathogenic variants of DPAGT1 were shown by Yuste-Checa *et al*. to lead to reduced localisation at the ER membrane [109]. In these cases the mutation leads to disruption in protein folding, stability or trafficking and hence lack of functional enzyme leads to disease. Other CDG mutations do not implicate expression levels but instead lead to disease through partial or complete loss of activity - for example those that implicate substrate binding or catalysis. These are often poorly characterised due to a lack of functional assays for enzymatic assessment.

PMM2-CDG is the most common type of CDG with over 700 cases globally, and is caused by a mutation in the phosphomannose mutase 2 gene that results in defective conversion of mannose-6-phosphate to mannose-1-phosphate during the synthesis of GDP-mannose and dolichol phosphate mannose [99]. Diagnosis of this CDG type is one of the few that can be confirmed by enzymatic assay, where the activity of PMM2 is assessed in patient leukocytes or fibroblasts using spectrophometric changes with the coupled reduction of NADP to NADPH [110, 111]. The most common missense mutation found in PMM2 is of an active site arginine residue replaced with histidine (R141H), which leads to protein with 0.4% activity of the wild-type, owing to the arginine binding the phosphate group of the substrate [112]. The mutant protein therefore has a 10-fold increase in K_M towards mannose-1-phosphate [113].

Although not routinely performed to aid diagnosis, detailed study of the activity of DPAGT1-CDG variants has also been reported. Figure 1.11 shows a comparison of activity of each of the heterozygous DPAGT1 mutations from patients with CDG and CMS by Dong *et al.*[24].This was achieved using a TLC based assay utilising [^{14}C]-labelled GlcNAc which was separated from the reaction once transferred to dolichol phosphate. It was found that patients typically had one allele with greater impact on DPAGT1 activity with the second having at least 50% the activity of the WT protein [24]. Those containing mutants closest to the substrate-binding sites (Figure 1.11B) were seen to have the most dramatic loss of activity as expected. Interestingly, both Val364Gly and Gly160Ser mutations were found to give much increased activity relative to the WT protein – and were later shown to cause abnormal RNA splicing – a finding that suggests there would be scope to therapeutically increase or modulate DPAGT1 activity [24].

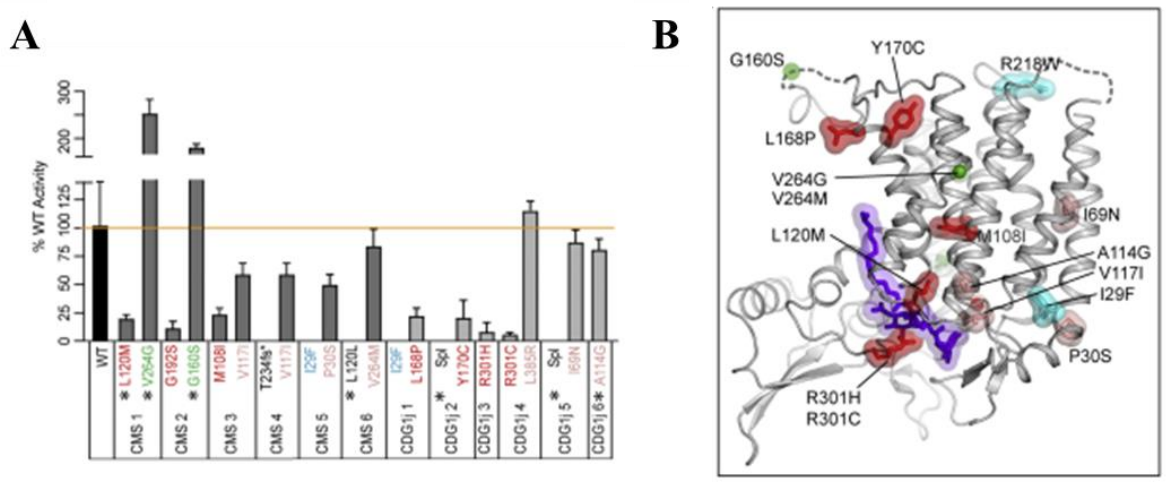


Figure 1.11 Analysis of pathogenic DPAGT1 variants. A Relative activity of CMS and CDG mutants to the WT (wild type) protein. The labelled variants are coloured red to indicate < 25% activity, pink for 50-100% activity and green for >WT. Those in cyan were confirmed unstable on purification. **B** The position of each mutation is mapped onto the tunicamycin (purple) bound DPAGT1 structure, coloured as in **A.** Figure adapted from Dong *et al.* Cell 2018 [24], with permission under the Creative Commons licence: https://creativecommons.org/licenses/by/4.0/.

Unfortunately, there remains a lack of targeted treatments for the majority of CDGs and instead supportive care is given to manage symptoms. This leaves a large unmet clinical need which will only be improved with increased molecular understanding of the functional and structural effects of each disease variant. For example, it is currently not clear why different CDG subtypes can have such different clinical phenotypes and severities. The striking difference in presentation of ALG13- and ALG14-CDG symptoms is particularly intriguing given the proteins act together to catalyse the same reaction. It is also not known why ALG13-CDG patients shows normal glycosylation patterns given its essential role in N-glycosylation. It is possible that pathogenic variants result in interruption of complex formation, or in gain of function elsewhere in the cell that lead to disease [108]. Further research into the functional effect of each ALG13/ALG14 variant is therefore needed to aid in the search for tailored therapeutics.

1.5 Project aims and hypotheses

The N-glycosylation pathway is a highly conserved and regulated process, essential for the proper functioning of proteins in cells. Owing to the difficulty of accessing both the membrane-associated enzymes and their lipid-linked substrates, gaps remain in our understanding of the molecular mechanisms involved in the first stages of this process in eukaryotes. Furthermore, studying the interactions of these enzymes may unveil potential multi-enzyme complexes which could be important targets for therapeutic interventions in glycosylation disorders. This thesis therefore aimed to utilise techniques in biochemistry, molecular biology and synthetic chemistry to:

1. *Express, purify and characterise the human ALG13/ALG14 heterodimer, and elucidate whether it forms a multi-enzyme complex with upstream DPAGT1*. Given the highly conserved nature of the eukaryotic N-glycosylation pathway, it was hypothesised that, like in yeast [27, 77], human ALG13/ALG14 would form a stable complex with DPAGT1. This trimeric complex - despite being small in size - is in the range solvable by cryoEM, of which initial efforts towards this endeavour are described in Chapter 3.

2. *Optimise a functional assay for study of GlcNAc transfer.* This study aims to assess the activity of ALG13/ALG14 alone and in complex using both a luminescence-based enzyme coupled assay as well as a radioactive thin layer chromatography assay, as detailed in Chapter 4.

3. *To assess the substrate tolerance of human DPAGT1 and ALG13/ALG14 enzymes.* Here, the activity of the human enzymes with native dolichol substrates and that with short, polyisoprenol equivalents used in bacteria, are compared to give insight into the evolutionary need for this long, α-saturated lipid, as discussed in Chapter 5.

4. *Characterise pathogenic variants of ALG13/ALG14 proteins identified in patients with Congenital Disorders of Glycosylation*. Mutagenesis of ALG13/ALG14 vectors allowed for the expression and purification of pathogenic variants of ALG13/ALG14 which were subsequently studied biochemically to understand the effect of these on the enzyme function, as shown in Chapter 6.

Chapter 2

Materials and Methods

2.1 Baculovirus-insect cell expression system

2.1.1 Insect cell passage

Sf9 insect cells from a clonal isolate derived from *Spodoptera frugiperda* (line IPLB-2Sf-21-AE) were purchased from ThermoFisher and cultured in Sf-900 II SFM medium supplemented with 10 µg/mL penicillin and streptomycin antibiotics (Gibco). Cells were incubated at 27 °C with 100 revolutions per minute (RPM) shaking in vented Erlenmeyer flasks, and passaged to maintain a confluence of between 0.8-6 x 10^6 cells/mL depending on requirements.

2.1.2 Ligation Independent Cloning

Ligation independent cloning (LIC) describes the technique of inserting a gene of interest into a plasmid without the need of a ligase enzyme, as first reported in 1990 [114]. This is achieved by introducing complementary single-stranded DNA overhangs on the insert and vector fragments using the exonuclease activity of T4 DNA polymerase. Nicks that remain in the DNA are tolerated and repaired by *E.coli* following transformation.

Plasmids containing the genomic sequence of target proteins DPAGT1, ALG13 and ALG14 were kindly provided by the Centre of Medicines Discovery as well as a range of baculovirus transfer vectors (pFastBac, pFB) containing sites for LIC that introduce various purification tags. Diagrams of those used are shown in Figure 2.1. These vectors contain polyhedron promoters and Tn7 sequences for *in vivo* recombination into bacmid DNA. Full-length sequences of the target proteins are shown in the appendices.

Polymerase Chain Reaction (PCR) was performed to amplify the gene of interest plus 5' and 3' extensions using primers detailed in Table 2.1. The primers also encode the necessary initiation and termination codons. A reaction mix containing Herculase II Fusion DNA PolymeraseII in 1x buffer (Agilent) with 0.6 µM forward and reverse primers, 0.25 ng/µL template DNA, and 0.3 mM dNTP mix in UltraPure water was subject to PCR using a Veriti thermo cycler (Applied Biosystems). A 'touchdown' PCR method was employed whereby successive rounds of amplification utilise decreasing annealing temperatures (Table 2.2). Samples of the PCR reaction were analysed on 1% agarose gel stained with ethidium bromide (UltraPure™, Thermo Fisher Scientific), alongside DNA Hyperladder (Bioline), to ensure presence of a single PCR product at the expected molecular weight (totalling the length of the target gene plus extensions regions). To remove the vector used as template DNA for the gene of interest, a Dpn1 treatment was performed. Dpn1 in CutSmart buffer was added to the PCR product and incubated for 1 hour at 37 °C. The PCR product was purified before downstream use using a QiaQuick kit (Quiagen) as per the manufacturers protocol.

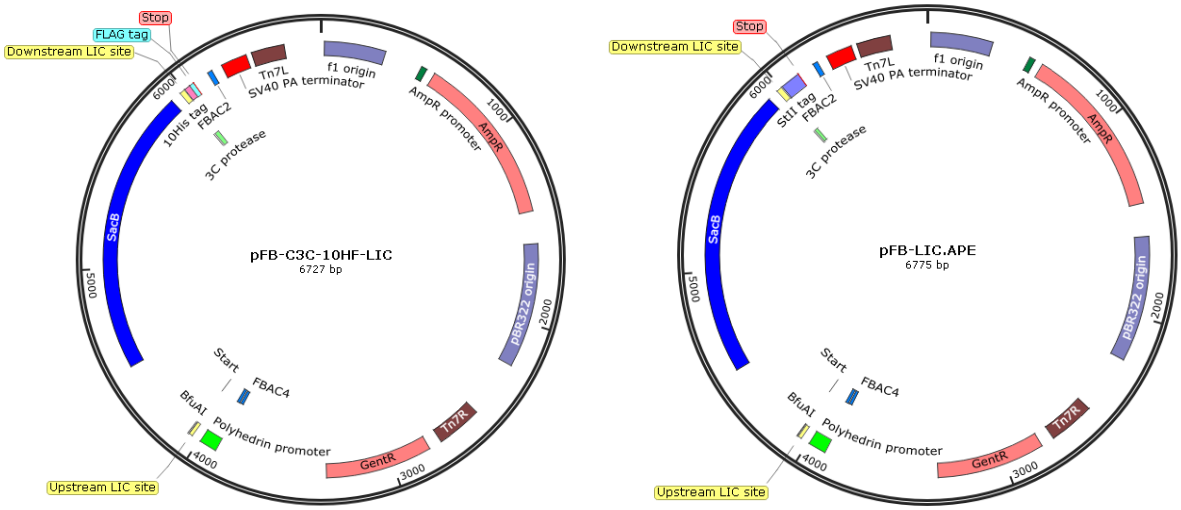


Figure 2.1 General plasmid maps of cloning vectors supplied by the CMD. (Left) pFB-C3C-10HF-LIC (Right) pFB-LIC.APE. The position and sequence of the tag, restriction site used for initial cloning and protease cleavage site are highlighted.

Table 2.1 ALG13 and ALG14 specific primers for Ligation Independent Cloning (LIC)

Name	Direction	Sequence
ALG13	5’	TACTTCCAATCCATGAAGTGCGTGTTTGTTACCGTA
	3’	TATCCACCTTTACTGTCATTTTTGTAATCCAACAACTT
ALG14	5’	TACTTCCAATCCATGGTGTGCGTTCTCGTTCTAGC
	3’	GATTGGAAGTAGAGGTTCTCTGCAACAATTCGCCCAAGGTA

Table 2.2 Touchdown PCR using Herculase II Fusion DNA Polymerase

	Temperature, °C	Time, min	Number of cycles
Initial denaturation	95	10	1
Amplification	95	0.5	5
	68	0.5	
	68	1.5	
	95	0.5	5
	60	0.5	
	68	1.5	
	95	0.5	5
	55	0.5	
	68	1.5	
	95	0.5	10
	50	0.5	
	68	1.5	
Final Extension	68	10	1
Cooling	4	∞	1

The required cloning vector was linearised by digestion with BseR1 restriction enzyme, and incubated at 37 °C for 3 hours. A sample of the digest mix was analysed using 1% agarose gel to ensure complete cutting, by the presence of two clean bands corresponding to the vector backbone and stuffer region. This vector was purified using QiaQuick kit (Quiagen) as per manufacturers protocol.

Both vector and insert were treated with T4 DNA polymerase to utilise its 3'exonuclease activity to reveal complementary overhangs. To the vector, 50 µL of digested plasmid was used with 1x T4 DNA Polymerase (NEB) in CutSmart buffer, 2.5 mM dGTP, 1 µL BSA (New England Biolabs, NEB), and 5 mM DTT and the reaction incubated for 30 minutes at 22 °C and 20 minutes at 75 °C. For the insert, materials were as before but using 2.5 mM dCTP.

Using a ratio of 1:3 treated insert: treated vector, samples were incubated for 2 hours at room temperature to allow for annealing. The plasmid was transformed into chemically competent Mach1 *E. coli* cells (prepared by Jing Wang) using heat shock. Briefly, 40 µL of cells and 4 µL plasmid were thawed on ice for 30 minutes, before heat-shock at 42 °C for 45 seconds, before moving back to ice for 30 seconds. 100 µL of SOC medium (Millipore) was added and cells allowed to outgrow for 1 hour at 37 °C. 100 µL of cells were plated onto LB agar containing 50 µg/mL ampicillin and 5% sucrose to select for those containing successfully ligated plasmids, and grown overnight at 37 °C.

A colony PCR experiment was performed to check for presence of the gene of interest. To a PCR mix containing 0.5 µM forward and reverse sequencing primers (see Table 2.3) and MyTaq DNA polymerase in 1x Buffer (Bioline) in UltraPure water, was dipped a single colony from the LB agar plate. PCR was performed as in Table 2.4, before the PCR reaction was analysed on a 1% agarose gel to confirm presence of the inserted gene at the expected molecular weight. The positive colony was used to inoculate 5 mL of LB broth containing selection antibiotics and grown overnight at 37 °C with shaking at 220 RPM. 1 mL of culture was used to prepare a 30% glycerol stock for long term storage at – 80 °C. From the remaining 4 mL culture, the plasmid was prepared using GeneJet miniprep (Thermofisher) as per the manufactures protocol.

Table 2.3 Sequencing primers

Name	Sequence
5' sequencing primer (FBAC1)	TATTCATACCGTCCCACCA
3' sequencing primer (FBAC2)	GGGAGGTTTTTTAAAGCAAGTAAA

Table 2.4 MyTaq colony PCR screen

	Temperature, °C	Time, min	Number of cycles
Initial denaturation	95	10	1
Amplification	95	0.33	25
	50	0.33	
	72	1.5	
Final extension	72	5	1
Cooling	4	∞	1

2.1.3 Bacmid production

Invitrogen's 'Bac-to-Bac' system was employed for site-specific *in vivo* transposition to produce bacmid DNA [115]. pFB vectors containing the gene of interest (described above) were transformed into competent DH10EMBacY *E.coli* (Life Technologies), using the heat shock method. This strain contains a parent bacmid encoding lacZ-mini-attTn7 which recombines with the pFB vector in the presence of a helper vector. Successful recombination can be identified using blue/white colony screening as the expression cassette interrupts the lacZ gene.

Briefly, screening was performed by plating 50 µL of outgrown and transformed DH10EmBacY onto selective LB agar plates containing 50 µg/mL kanamycin (for selection of

the bacmid), 7 µg/mL gentamycin (for selection of the pFB plasmid), 10 µg/mL tetracycline (for selection of the helper plasmid), and 40 µg/mL IPTG and 100 µg/mL blue-gal to monitor interruption of the lacZ gene. Plates were blocked from the light and incubated at 37 °C for 48 hours. White colonies were streaked to dilution on selective LB agar plates and incubated at 37 °C for 12 hours. 1 mL of LB agar containing 50 µg/mL kanamycin, 7 µg/mL gentamycin and 10 µg/mL tetracycline was inoculated with a white colony and grown overnight at 37 °C for 12 hours with 220 RPM shaking. 30% glycerol stocks were created from 120 µL of overnight culture, and the remaining culture was used to purify bacmid DNA using the 96-Well Plasmid and BAC Preparation Kit (Millipore), with washing and elution of bacmid DNA carried out in sterile buffer and in a biological safety cabinet.

This bacmid DNA was screened to check for the presence of the gene of interest using PCR with specific sequencing primers (as in Table 2.3). The bacmid DNA was diluted 10-fold in UltraPure water and incubated with MyTaq DNA polymerase (Bioline) as described above and in Table 2.4, before being analysed using 1% agarose gel to ensure a single band of the gene of insert plus 700 base pairs of the vector.

2.1.4 Transfection and baculovirus amplification

Bacmid DNA was transfected into Sf9 insect cells to initiate baculovirus production. P0 generation baculovirus was harvested and amplified by repeated rounds of infection to yield second generation (P2) virus for use in protein production.

Briefly, Sf9 cells were seeded at 0.2 x 10^6 cells/mL in Sf9 serum free media with 10 µg/mL penicillin and streptomycin into a 24-well plate (Costar 3524). 40 µL of transfection reagent (Insect GeneJuice, Sigma) in Sf9 serum-free media was added to each well of a 96 well sterile flat bottom microtiter plate. 2 µL recombinant bacmid DNA at 1 µg/µL was added to each well and the mixture incubated for 15 minutes. To the treated DNA, 160 µL serum-free media was

added. Media from the plated cells was removed by aspiration, and replaced with the media containing treated DNA. Plates were incubated for 4 hours at 27 °C in a humidified incubator, before addition of a further 400 µL serum-free media supplemented with 2% fetal bovine serum (FBS, Sigma). Cells were grown for 5 days before monitoring for virus production using a fluorescence microscope. Yellow-fluorescent protein (YFP) is encoded for in the backbone of the bacmid vector, and so yellow fluorescence indicates successful infection. An example of successfully infected insect cells is shown in Figure 2.2, taken using an Olympus fluorescence microscope at 10x magnification. The brightfield image shows the characteristic dark spot of infected Sf9 cells, with <90% of cells in the field of view expressing YFP. P0 generation virus was harvested by pelleting the cells with centrifugation at 2,500 RPM for 20 minutes. The virus-containing supernatant was stored away from light at 4 °C.

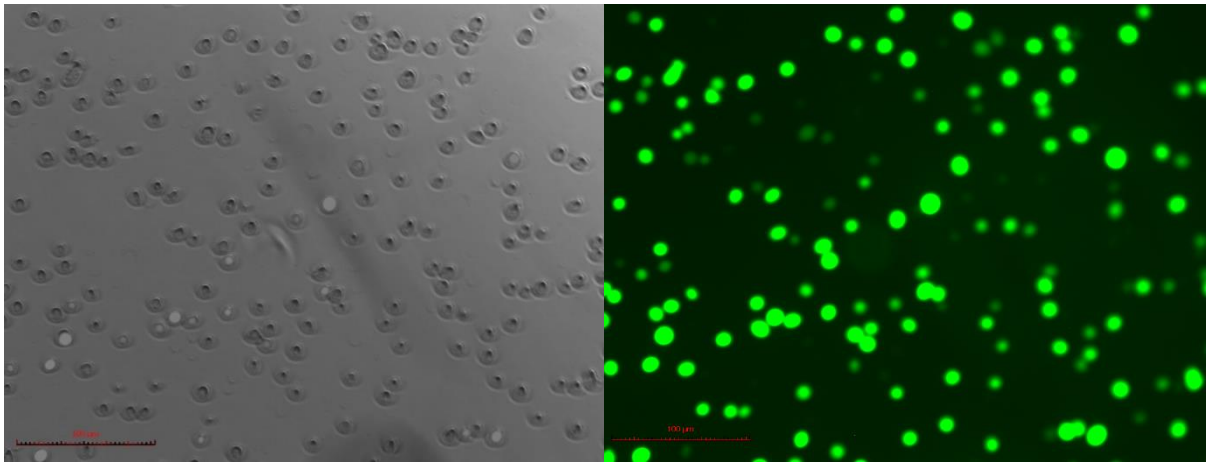


Figure 2.2 Infected Sf9 insect cells: Yellow fluorescent protein (YFP) expressing insect cells following baculovirus transfection. Left = brightfield, right = fluorescence of YFP following excitation at 450 nm. Scale bar shown at 100 µm.

2.1.5 Small-scale test expression in insect cells

The P0 generation baculovirus was amplified to P1 generation and tested on a small, 3 mL, scale for successful expression of the target proteins. Briefly, Sf9 cells at mid-log phase were seeded into a 24-deep well block with 3 mL per well in Sf9 serum-free media with 2% FBS and 50 µg/mL penicillin and streptomycin. 120 µL of P0 baculovirus was added and the cells incubated at 27 °C with 220 RPM shaking for 72 hours. Cells were pelleted by centrifugation at 900 x g for 20 minutes, with the supernatant stored as P1 generation baculovirus at 4 °C.

To check for expression of His-tagged target proteins, the cell pellet was resuspended in 1 mL EXB (see section 2.3 for details) containing 1% DM/CHS, and rotated at 4 °C for 1 hour to solubilise cell membranes. The cell debris was removed by centrifugation at 2,600 x g for 20 minutes at 4 °C, and the supernatant taken and bound to 100 µL of 50% Talon resin in EXB (equilibrated by 2x washes in MilliQ H_2O and 3x washes in extraction buffer, with pelleting using 700 x g for 2 minutes with slow deceleration) with rotation at 4 °C for 1 hour. The resin was moved to a 96-well filter bottomed plate, where it was washed in 2 x 800 µL EXB with 0.27% DM/CHS. All residual wash buffer was removed by 3- minute centrifugation at 500 x g at 4 °C, before bound proteins were eluted by addition of 50 µL EXB with 250 mM imidazole and 0.27% DM/CHS. The plate was incubated with shaking at room temperature for 10 minutes, before centrifugation at 500 x g for 3 minutes over a collection plate. 15 µL of the elution fraction was resuspended in SDS-sample buffer with reducing agent, and analysed by SDS-PAGE (see section 2.4).

2.1.6 Amplification of baculovirus to P2 stock

P2 generation baculovirus was generated by adding 250 µL P1 generation stock to 50 mL Sf9 cells at mid-log phase. The cells were incubated at 27 °C for 72 hours with 100 RPM shaking, before being pelleted by centrifugation at 900 x g for 20 minutes. The virus containing

supernatant was stored as P2 generation baculovirus at 4 °C for up to 6 months. P2 generation baculovirus was used to infect large scale cultures for expression and purification of target proteins for structural and functional assessment (see section 2.3).

2.1.7 MultiBac production

The MultiBac system was developed by Berger *et al.* [116] to allow for expression of heterologous complexes. Cloning of ALG13 and ALG14 into MultiBac vectors was performed by Jing Wang, using donor and acceptor plasmids as in Figure 2.3. These plasmids were combined at their respective LoxP sites using Cre-recombinase (NEB, used as per protocol) and selected for successful recombination by the combined antibiotic resistance donated from each vector following transformation into Mach1 *E.coli*. The resulting vector was recombined with bacmid DNA using Tn7 transposition inside DH10EmBacY, as in 2.1.3.

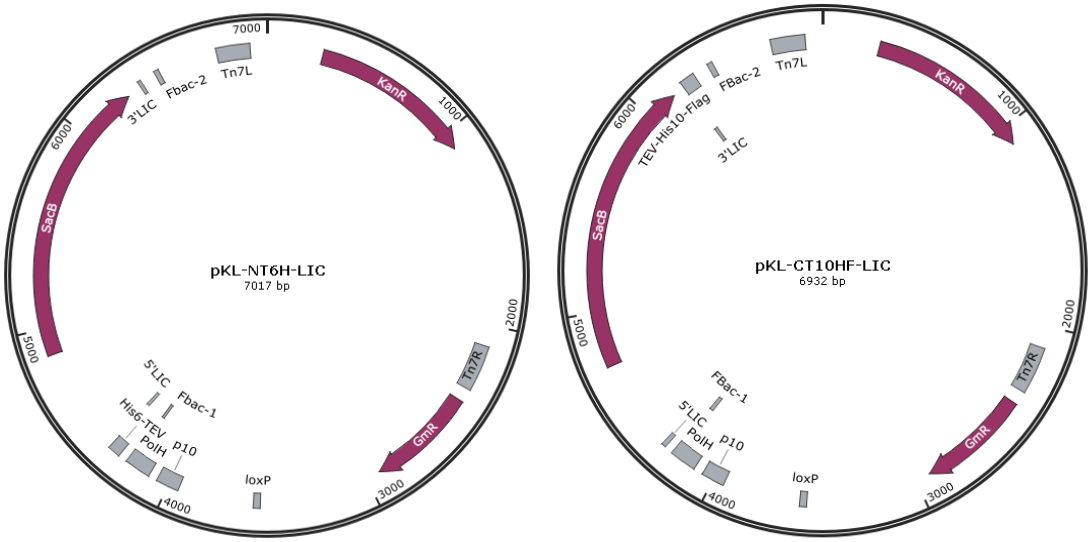


Figure 2.3 Acceptor (left) and donor (right) vectors combined during MultiBac production

2.2 Pathogenic variants of ALG13 and ALG14

2.2.1 Site-directed mutagenesis

Q5 Site Directed Mutagenesis kit (NEB) was used as per manufactures protocol to introduce mutations into ALG13 and ALG14. The primers used are shown in Tables 2.5 and 2.6, and the PCR reaction was performed as in Table 2.7. 1 µL of the PCR product was treated with kinase, Dpn1 and ligase (KDL) reagent for 5 minutes at room temperature, before heat-shock transformation into Mach1 competent *E. coli*. Cells were allowed to outgrow in 2xLB for 1 hour at 37 °C with shaking at 220 RPM, before spreading onto LB agar plates containing 50 µg/mL kanamycin and chloramphenicol and incubated overnight at 37 °C.

Colonies were picked and used to inoculate 3 mL LB broth with antibiotics and left to grow overnight at 37 °C with shaking at 220 RPM. The cells were pelleted by centrifugation at 5,000 RPM for 10 minutes, and plasmid DNA extracted using miniprep (Qiagen kit, standard protocol). DNA was sent for sequencing using gene specific sequencing primers to confirm introduction of the mutation.

Table 2.5 Primers for introducing ALG14 patient mutations, where the nucleotide corresponding to the point mutation is in lowercase.

Mutation introduced	Direction	Sequence
P65L	5’	TCCAAGAAGCtGGGAGGTTCA
	3’	ATTCGGTGAATGTAGTATTTGGTATAC
D74N	5’	TGCTGACACTaATGAAATGAG
	3’	ATGACATAATGTCTAGGTG
V141G	5’	CCAGATTTGGgGTTGTGTAAC
	3’	CTTCACCCTGTGAATTAG
R109Q	5’	CCAAGAAGCCaGGAGGTTCAG
	3’	AATTCGGTGAATGTAGTATTTG
I175N	5’	GTTGAAAGCAaCTGCCGTGTAG

	3’	GTAGACAATGATCACTTTCTTTATTC

Table 2.6 Primers used to introduce ALG13 patient mutations, with the nucleotide corresponding to the point mutation in lowercase.

Mutation introduced	Direction	Sequence
N107S	5’	TTGATGAACAgTCATCAGCTG
	3’	CTTTTCGTTTATAACCACTAC
I17N	5’	GACGACCTCAaTGCGTGTGTG
	3’	AAAGCTGGTGGTCCCTAC
K94E	5’	GGAAAAAGGAgAGCCACTCGTAG
	3’	AGAGTCTCCAAACAGCTTC
A81T	5’	TATTAGTCACaCAGGTGCAGG
	3’	ACAAGATCTGCTTTCTGAATG

Table 2.7 Q5 site directed mutagenesis: PCR protocol

	Temperature, °C	Time, seconds	Number of cycles
Denaturation	98	30	1
Amplification	98	10	25
	63	30	
	72	3	
Final Extension	72	120	1
Cooling	4	∞	1

2.2.2 RNA extraction

To test for expression of each of the ALG13/ALG14 pathogenic mutants an RNA screen was performed. 3 mL of Sf9 cells at 2 x 10^6 cells/mL in Sf9 serum-free media with 10 µg/mL penicillin and streptomycin were infected with 120 µL of P2 generation baculovirus encoding ALG13/ALG14 wild type or pathogenic mutants. The cells were incubated at 27 °C with 220 rpm shaking for 72 hours.

The insect cells were pelleted by centrifugation at 900 x g for 20 minutes, and supernatant removed by vacuum under sterile conditions. Total RNA was extracted from the cells using TRIzol reagent (Thermofisher) as per manufacturers protocol, and using a sterile and RNAase-free work station and equipment. Briefly, each cell pellet was lysed and homogenised in 750 µL TRIzol (containing phenol and guanidine isothiocyanate) to allow dissociation of nucleoproteins, before 150 µL chloroform was added and mixed. Following centrifugation at 12,000 x g at 4 °C, the upper, clear aqueous phase containing RNA was collected. RNA was precipitated by addition of 1:1 volume isopropanol, and collected by centrifugation at 12,000 x g at 4 °C for 20 minutes. The RNA was washed in 750 µL 75% ethanol, and collected by centrifugation at 7500 x g at 4 °C. After drying, the RNA was resuspended in 20 µL nuclease free water, with incubation at 55 °C for 10 minutes to completely dissolve the pellet.

DNA was removed from the samples using the TURBO DNA-free kit (Thermofisher), with incubation of each sample with DNase for 30 minutes at 37 °C before inactivation and removal of the DNase as per the kit manual. The RNA-only supernatant was taken forward and converted into cDNA using SuperScript IV Reverse Transcriptase (Invitrogen). Briefly, 1 µL of treated RNA was annealed to ALG13 or ALG14 gene specific 3' primers (see Table 2.3) in the presence of dNTP, by heating at 65 °C for 5 minutes, and then cooling on ice. SuperScript IV Reverse Transcriptase was added along with DTT and RNAse inhibitor. The reaction was

incubated for 10 minutes at 50 °C, and then stopped by inactivating the enzyme with heating at 80 °C for 10 minutes.

The cDNA created was amplified by touchdown PCR using the high-fidelity Hercules II Polymerase (Agilent, HercII) as described previously. Template cDNA was not diluted but used straight from the reverse transcription reaction mixture, and ALG13 or ALG14 forward and reverse primer pairs used are shown in Table 2.3. The resulting PCR product was analysed by electrophoresis using a 1% agarose gel, and DNA stained using ethidium bromide. Bands at the expected 498 bp and 753 bp were visualised for ALG13 and ALG14 respectively.

2.3 Protein purification

All purification steps were carried out at 4 °C unless stated otherwise.

2.3.1 Insect cell infection and harvesting

1 L cultures of Sf9 cells were infected with 5 mL of recombinant P2 baculovirus when at mid-log stage (2 x 10^6 cells/mL). The cells were grown for 65 h at 27 °C with 100 RPM shaking to allow maximum protein expression, before harvesting by centrifugation at 900 x g for 15 min at 16 °C. Cell pellets were washed in 15 mL phosphate buffered saline (PBS), flash frozen in liquid N_2 and stored at -80 °C.

2.3.2 Homogenisation

Cell pellets were thawed in 50 mL of extraction buffer per 1 L pellet. Extraction buffer (EXB) was 50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM TCEP, 5% glycerol and cOmplete EDTA-free protease inhibitor (Roche) unless stated otherwise. The cells were disrupted by 6x passes through a hand-held Dounce homogeniser followed by 2x passes through a C3 homogeniser (10-15,000 PSI). Clarification by 35,000 x g at 4 °C for 45 min in a JA25.5 rotor was performed to remove cell debris.

2.3.3 Sonication

To screen for expression of ALG13/ALG14 mutants (Chapter 6) cells were disrupted using sonication. Cell pellets from 50 mL of original Sf9 culture were resuspended in 2 mL EXB with cOmplete EDTA-free protease inhibitor (Roche), before exposing to sonication for 2 minutes at 20% power with pulses of 5 seconds on and 10 seconds off.

2.3.4 Membrane preparation and solubilisation

To improve the purity of DPAGT1, ALG13 and ALG14 preparations, a membrane preparation was employed. Cell pellets were thawed in EXB and homogenised as in 2.3.2 but without the centrifugation. Instead, the lysate was spun at 10,000 x g at 4 °C for 10 min in a JA.14 rotor to remove cell debris, before the supernatant was taken forward at 100,000 x g for 1 hour in a Ti 45 rotor. The membrane containing pellet was resuspended using 20 passes through a Dounce homogeniser in 20 mL EXB, before making to 50 mL in EXB with 1% (w/v) detergent (analytical grade from Anatrace, see Table 2.8). The sample was rotated at 4 °C for 1 hour to solubilise the membrane proteins. The insoluble fractions were removed by final centrifugation at 100,000 x g, 4 °C, Ti 45 for 1 hour, and the supernatant was taken forward for downstream purification.

During the extended membrane preparation as detailed in the Chapter 3, cell pellets were thawed in a hypotonic EXB (50 mM HEPES pH 7.5, 2 mM TCEP, 5% glycerol and Roche EDTA-free protease inhibitor) using 50 mL per 1 L pellet. The cells were disrupted by 6x passes through a hand-held Dounce homogeniser and 2x passes through a C3 homogeniser (10-15,000 PSI) as before, and the cell lysate subject to 100,000 x g at 4 °C for 1 h in a Ti 45 rotor. The membrane pellets were resuspended again in 50 mL hypotonic EXB using 20 passes through a Dounce homogeniser, before pelleting by ultracentrifugation as before. Two further

washes were performed using a hypertonic EXB, as above but with 1 M NaCl. The cell membranes were solubilised by resuspending pellets in the isotonic EXB with detergents as before, and any insoluble debris was removed by centrifugation at 100,000 x g, 4 °C, Ti 45 for 1 hour.

Table 2.8 Critical micelle concentration (CMC) in water of detergents used during purification of DPAGT1, ALG13 and ALG14 [117]

Name	CMC, %
n-Decyl-β-D-maltopyranoside (DM)	0.0087
n-Dodecyl-β-D-maltopyranoside (DDM)	0.087
Octyl glucose neopentyl glycol (OGNG)	0.058

2.3.5 Talon/Ni resin IMAC

6xHis-tagged target proteins were purified from clarified cell lysate using immobilised metal affinity chromatography (IMAC) with Co^{2+} charged resin (Talon, TakaraBio). Typically, 1 mL of equilibrated 50% resin was incubated per 1 L of original cell culture used. The samples were rotated at 4 °C for 1 hour to allow binding. The resin was then washed in 30x the column volume of wash buffer (WB) containing 50 mM HEPES pH 7.5, 2 mM TCEP, 150 mM NaCl, 5% glycerol, 10 mM imidazole pH 8.0, and detergents at 3 times their critical micelle concentration (CMC, as in Table 2.8) unless stated otherwise. Proteins were eluted in a buffer as in WB but with 250 mM imidazole buffer, using repeated application of 0.5 column volumes until all target protein was eluted. The fractions were inspected by absorbance at 280 nm using a Nanodrop spectrophotometer using the predicted extinction coefficient and molecular weight of the target protein by SATurn software [118]. Those containing the highest protein concentration

were loaded onto an equilibrated PD-10 column and exchanged into elution buffer without imidazole, and with reduced 20 mM HEPES pH 7.5 (referred to as GFB, gel filtration buffer).

2.3.6 Reverse purification

All target proteins contained TEV protease or 3C protease cleavage sites following their purification tag. Following elution and exchange into GFB, the sample was incubated overnight with 10:1 (w/w) protein: TEV protease (purified by Dr Yin Dong) or 3C protease (prepared by Jing Wang). Pre-equilibrated 50% Talon was added to re-bind impurities, and untagged-target protein was collected in the flow-through. The resin was washed in high imidazole elution buffer for analysis, and samples from each stage of the purification process were inspected by SDS-PAGE.

2.3.7 Anti-FLAG affinity purification

Anti-FLAG M2 affinity resin (Sigma) was used to purify DPAGT1 bearing a FLAG tag with sequence: DYKDDDDK. As the resin is formed by mouse IgG monoclonal antibody linked to agarose, reducing agent was not used in extraction or wash buffers so not to disrupt the disulphide bonds crucial for antibody stability. 0.5 mL of 50% equilibrated FLAG resin was used per 1 L of original cell culture. The proteins were bound by rotation for 1 hour, before the resin was pelleted by 750 x g centrifugation for 5 minutes. The flow through (supernatant) was stored for downstream analysis before the resin was loaded into a fritted column and washed in 30 times column volume of reducing agent-free EXB containing detergents at 3 x CMC. The target protein was eluted in wash buffer containing 150 ng/µL 3xFLAG peptide (Sigma) in batches of 0.5 column volume until all target protein had been removed.

2.4.8 Trial use of non-detergent membrane mimetics

The use of amphipol A8-35, nanodisc scaffold peptide (NSPr), membrane scaffold protein 1D1 and membrane scaffold protein 2N2 were trailed for use as alternatives to detergents for

solubilising DPAGT1-FLAG/ALG13/His-ALG14 (Chapter 3). The lipids phosphatidyl choline (PC), phosphatidylethanolamine (PE) and phosphatidyl glycerol (PG) (Laradon) were prepared in a 3:1:1 ratio. 25 mg of the lipid mixture was dried to a film in a round bottom flask under a stream of N_2 and dried *in vacuo*. 100 mM sodium cholate was made in EXB without detergent and used to dissolve the lipids by rotation at room temperature. The resulting lipid stock was used to encapsulate the target proteins with membrane scaffold proteins.

Briefly, the membranes from 1 L of DPAGT1-FLAG/ALG13/His-ALG14 expressing Sf9 cell culture were homogenised and used as in sections 2.3.1-3, with buffers without reducing agent. To the sample, 1 mL of 50% equilibrated anti-FLAG affinity M2 resin was added and bound for 1 hour by rotation at 4 °C. The resin was pelleted by 750 x g for 5 minutes, before loading into 4 fritted plastic columns. The columns were washed in 30 volumes of TCEP-free EXB with 3x CMC of DM/CHS. The expected yield of DPAGT1-FLAG/ALG13/His-ALG14 based on previous purifications was 33.5 µg per 250 mL of original culture, and this was used to determine quantities of membrane mimetics to use. The washed anti-FLAG resin with bound protein complex was resuspended in 500 µL of EXB with 3x CMC DM/CHS and the membrane mimetic. Amphipol or NSPr were used at 1: 4, protein complex: reagent, and for membrane scaffold proteins (MSP) 1D1 and 2N2, a ratio of 1: 50: 2500, protein complex: MSP: lipid stock. The resultant resin slurry was equilibrated with rotation at 4 °C overnight. BioBeads (BioRad) equilibrated in EXB without detergent were used to remove the DM/CHS from each sample by incubation with rotation for 2.5 hours.

The resin was washed in 20 column volumes of detergent-free EXB, before target proteins eluted using EXB with 150 ng/µL 3xFLAG peptide (as in section 2.3.5).

2.3.9 Size exclusion chromatography

Target proteins were further purified by size exclusion chromatography using the ÄKTA Pure system (Cytiva). Following affinity purification, the eluted sample was concentrated to 70% of the loop volume (typically 0.7 mL for use with a 1 mL loop) using a 50 kDa MWCO PES membrane concentrator (Vivaspin). The protein stocks were clarified by 21,000 x g for 20 minutes, before loading onto the size exclusion column. Superdex 200 increase 10/300, Superdex 75 increase 10/300 or Superose 6 increase 10/300 column (Cytvia) were used as in Chapter 3. Columns were equilibrated in degassed Gel Filtration Buffer (GFB) of 20 mM HEPES pH 7.5, 2 mM TCEP, 150 mM NaCl, 5% glycerol and 3x CMC detergent. The target proteins were eluted using a flow rate of 0.5 mL/min, collecting 500 µL fractions.

2.4 Protein characterisation

2.4.1 SDS-PAGE

Protein samples from DPAGT1, ALG13 and ALG14 purifications were mixed in 1x sodium dodecyl sulphate (SDS) loading dye (Thermofisher) containing DTT reducing agent, and incubated at room temperature for 5 minutes. SDS-Polyacrylamide gel electrophoresis (PAGE) was performed alongside SeeBlue Plus2 Pre-stained Protein ladder (unless otherwise stated). Pre-cast NuPAGE 4-12% Bis-Tris gels in MOPS running buffer (both Invitrogen) were used for the efficient separation of proteins in the size range of DPAGT1, ALG13 and ALG14. The gels were subject to 150 V for 1 hour, before staining with Coomassie solution (InstantBlue, AbCam) to visualise proteins. The washed gels were imaged using a BioRad Gel Doc and Image Lab software.

2.4.2 Blue-Native PAGE and 2D BN/SDS-PAGE

2D BN/SDS-PAGE was performed as described by Wittig *et al.* [119]. All gel tank equipment was washed thoroughly to remove traces of denaturing buffer and Native PAGE anode buffer

and dark and light blue cathode buffers (Invitrogen) were pre-cooled to 4 °C. DPAGT1 and ALG13/ALG14 in complex and alone were made to 1 mg/mL in GFB with 0.27% DM/CHS. Native, tris-glycine sample buffer and 5% G-250 sample additive (Invitrogen) were added to make a 1x solution. The protein samples were loaded alongside NativeMark unstained protein standard onto Native PAGE 4-16% Bis-Tris gels (Invitrogen) and run at 4 °C in anode and dark blue cathode buffers using 150 V for 1 hour. The dark blue cathode buffer was then replaced with light blue, and the gel ran for a further 30 minutes at 150 V. The gels were washed in MilliQ and, to visualise the ladder, stained in Coomassie solution as before. The lanes of interest were excised using a scalpel, and submerged in SDS-Sample buffer with reducing agent overnight with agitation. The denatured gels were placed onto the wide well of a 2D NuPAGE 4-12% Bis-Tris protein gel (Invitrogen), and ran in MOPS running buffer at 150 V for 1 hour.

2.4.3 Western Blot

For Western blot, proteins were separated by SDS-PAGE alongside MagicMark XP Western Protein Standard (Invitrogen). The proteins were transferred from the unstained gel to a nitrocellulose membrane using electrophoresis with a Power Blotter in 1x transfer buffer (ThermoFisher). The membrane was blocked by incubating with 5% milk in 1x Tris buffered saline with 0.1% Tween 20 (TBS/Tween) for 1 hour with rocking, before 3x 5-minute washes in milk-free TBS/Tween. The membrane was labelled with primary or secondary antibodies conjugated with horseradish peroxidase (as in Table 2.9) at 1:1000 in 1% milk TBS/Tween, and rocked for 1 hour at room temperature before washing three times in TBS/Tween. The blots were developed by application of enzyme linked chemiluminescence (ECL) substrate (Amersham), and imaged using a ChemiDoc (BioRad).

Table 2.9 Antibodies used for Western Blot analysis of DPAGT1, ALG13 and ALG14

Antibody name	Supplier	Product number	Host Species/Isotype
Anti-ALG13 monoclonal	ProteinTech	67283-1-Ig	Mouse/ IgG1
Anti-6xHis tag monoclonal HRP	ThermoFisher	MA1-21315-HRP	Mouse/IgG2b
Anti-mouse IgG HRP	Dako	P0447	Goat

2.4.4 TCA precipitation

In order to detect and analyse proteins at very low concentration, trichloroacetic acid (TCA) precipitation was used. Briefly, into 0.5 mL protein samples, 125 µg/mL Na-deoxycholate (4.25 µL of a 2% solution) was added, samples vortexed, and incubated at room temperature for 15 minutes. In a fume hood, TCA was added to a final concentration of 6% (w/v), and samples vortexed. Tubes were centrifuged at 12,000 x g at 4 °C for 30 min. The supernatant was carefully removed and pellets washed in acetone at -20 °C to remove excess TCA, with centrifugation at 12000 x g for 5 min at 4 °C. The pellets were solubilised in SDS-PAGE buffer, and used as in 2.4.1.

2.4.5 Preparing samples for tryptic digest MS/MS

Coomassie stained-gel bands were excised following SDS-PAGE, and stored in a buffer of 0.5 mM TCEP in 20% MeOH in deionised water. Samples were sent for analysis by tryptic digest MS/MS, as performed by Dr Rod Chalk, CMD.

2.4.6 Intact denaturing MS

Intact denaturing MS was assisted by Dr Tiago Moreira, CMD. Protein at 1 mg/mL was diluted 1:50 in 0.1% (v/v) formic acid (50 µL final volume), before reversed-phase chromatography was performed in-line prior to MS using an Agilent 1100 HPLC system (Agilent Technologies Inc., Palo Alto, CA, USA). 50 µL was injected onto a 2.1 mm × 12.5 mm Zorbax 5µm 300SB-C3 guard column housed in a column oven set at 40 °C. The solvent system used consisted of 0.1% (v/v) formic acid in LC-MS grade water (Millipore, solvent A) and 0.1% (v/v) formic acid in methanol (LC-MS grade, Chromasolv, solvent B). Chromatography was performed starting in 90% A and 10% B and a flow rate of 1.0 mL/min. After 15 seconds at 10% B, a linear gradient from 10% B to 80% B was applied over 45 seconds, followed by 80% B to 95% B over 3 seconds. Elution continued at 95% B for 1 min 12 seconds followed by equilibration at initial conditions for a further 45 seconds. Protein intact mass was determined using an MSD-TOF electrospray ionization orthogonal time-of-flight mass spectrometer (Agilent Technologies Inc., Palo Alto, CA, USA). The mass spectrometer was operated in positive ion, 2 GHz detector mode. Source parameters were drying gas 350 °C, flow 12 L/min, nebulizer 60 psi, capillary 4000 V. Fragmentor was 250 V, collision energy 0 V and data acquired from 100-3200 m/z. Data analysis was performed using Masshunter Qualitative Analysis B0.7 proprietary software and deconvolution performed using the Maximum Entropy algorithm.

2.4.7 Nano-differential scanning fluorimetry

The thermostability of DPAGT1 and ALG13/ALG14 was measured by nano-differential scanning fluorimetry (DSF) using a Nanotemper Prometheus. Samples were taken into capillary tubes and loaded onto the Prometheus in triplicate and alongside buffer only controls. A discovery scan was performed to select for an appropriate excitation power, before the fluorescence monitored when the temperature was increased by 1 °C /min from 20 to 95 °C.

The ratio of fluorescence at 330 nm: 350 nm and its first derivation were used to monitor protein unfolding as tryptophan residues became exposed to the solvent. Furthermore, a proprietary scattering wavelength was used to monitor precipitation of the proteins with temperature.

2.4.8 BCA assay and densitometry

In an attempt to accurately determine the concentration of DPAGT1 and ALG13/ALG14 proteins, results from the Nanodrop spectrophotometer were compared to results of a BCA protein assay (ThermoFisher). The assay was performed as per the manufacture's procedure, where proteins were serially diluted and compared to known concentrations of bovine serum albumin (BSA) standards. The assay quantifies proteins based on a change in absorbance at 564 nm of bicinchoninic acid (BCA) induced when proteins in an alkaline environment reduce Cu^{2+} to Cu^{1+}.

Given the high levels of impurities remaining in the protein stocks, densitometry was performed on Coomassie stained gels to calculate the ratio of target protein: impurities. This was done using ImageJ software and was used to adjust the concentrations measured by the BCA assay and Nanodrop, as in the example shown in Table 2.10.

Table 2.10 Protein stock concentrations, µM, as measured by the BCA assay or nanodrop and adjusted for impurities using densitometry.

	BCA assay	Adjusted	Nanodrop	Adjusted
ALG13/ALG14	11.3	2.24	4.78	0.95
DPAGT1	5.53	1.02	5.4	1.00
DPAGT1/ALG13/ALG14	5.06	1.33	3.5	0.92

2.4.9 SEC-MALS

In order to measure the size of protein particles in detergent micelles, analytic SEC-multi angle light scattering (MALS) was performed using a Malvern OmniSEC with multi detectors of low angle and right-angle light scattering. An equilibrated Superdex 200 increase column (using degassed GFB with OGNG/CHS/CDL or DM/CHS at 3x CMC) was used to separate protein samples at 1 mg/mL. The samples were cleared by 20 minutes at 150,000 x g to remove any precipitant that would confound analysis. 160 μL of the resulting supernatant was used and compared to 1 mg/mL bovine serum albumin (BSA) standards in the same buffer. OmniSEC version 4.7 software was used to analyse the size of particles using a copolymer method.

In order to distinguish the proportion of the light scattering contributed by protein and that from the detergent micelle, a calibration curve of both DM/CHS and OGNG/CHS/CDL was created. Detergents were titrated from 0- 2% in buffer containing 3x CMC of the detergent and were run in triplicate. By plotting the refractive index area of the peaks produced against concentration, a calibration curve of actual concentration of detergent was generated using a trend-line forced through (x,y) (0,0) (Figure 2.4 A). By inputting these actual concentrations from this trendline into OmniSEC software for each peak, the calculated dn/dC was found to be 0.194 for OGNG/CHS/CDL. This was consistent across the detergent concentrations in the titration. The same analysis was performed to calculate the dn/dC value for DM/CHS as 0.293 (Figure 2.4 B). The parameters used to analyse the DPAGT1/ALG13/ALG14 complex were based on an assumed extinction coefficient and molecular weight of a 1:1:1 complex, and the dn/dC of the detergent as measured by a calibration curve (Table 2.11). Similarly, SEC-MALS data of ALG13-His/ALG14-His in DM/CHS was used (as shown in Chapter 3).

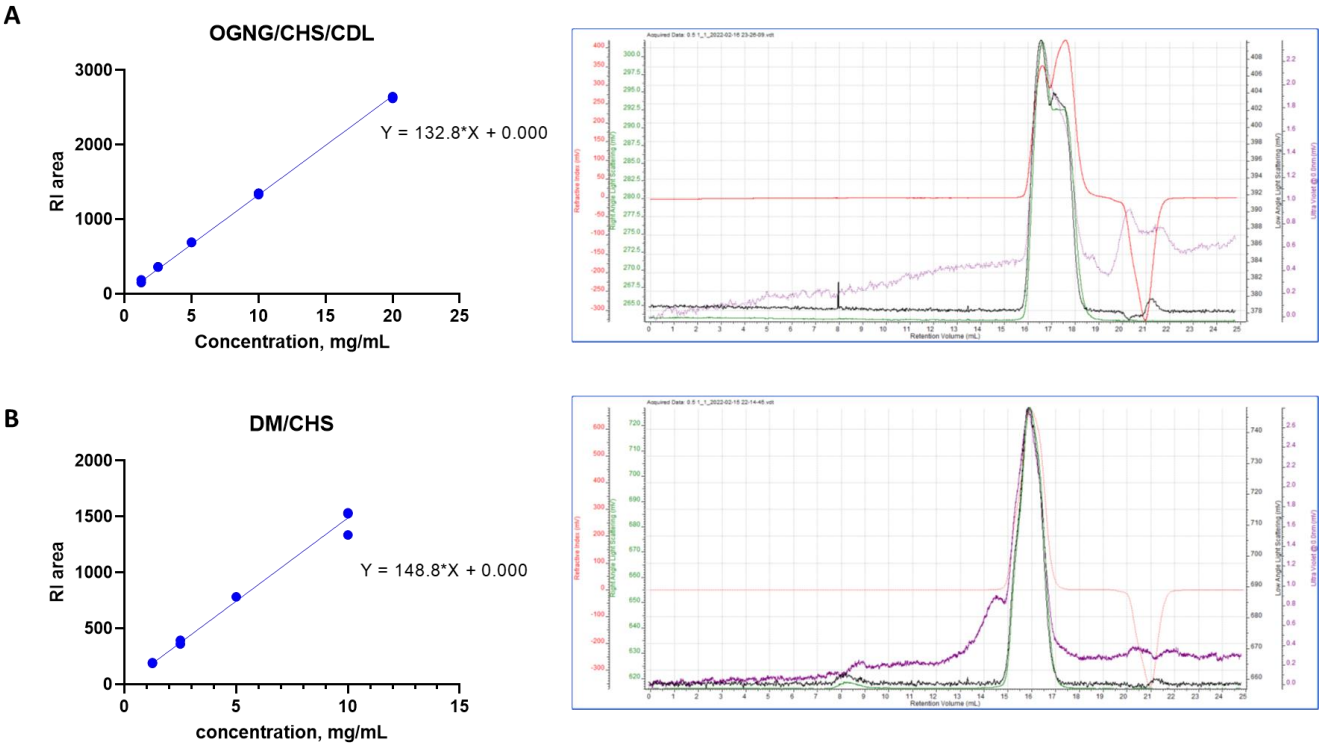


Figure 2.4 Calibration curve of refractive index area and concentration of OGNG/CHS/CDL (A) and DM/CHS (B) detergents used to calculate a dn/dC value by SEC-MALS. Multidetector spectra from Malvern SEC-MALS shown for a representative 0.5% (5 mg/mL) detergent standard in gel filtration buffer with 3x CMC detergent.

Table 2.11 Parameters of the detergents and lipids used to determine the size of the protein complex by SEC-MALS

	Differential refractive index, dn/dC	**dA/dC**
OGNG/CHS/CDL	0.1940	-
DPAGT1-His/ALG13/ ALG14-His	0.185 (Assumed)	0.8601 (ε / Mw = 80220/93272.72)

DM/CHS	0.2933	-
ALG13/His-ALG14	0.185 (Assumed)	0.9763 (ε / Mw = 46465/47595.39)

2.4.10 fSEC

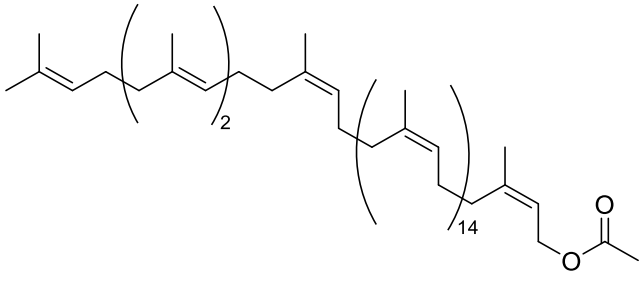
Fluorescence SEC (fSEC) was performed using a Shimadzu HPLC with a temperature-controlled sample chamber to allow for the high-throughput analysis of pathogenic ALG13 and ALG14 proteins (Chapter 6) and during the screening of detergents for extraction and purification (Chapter 3). A 3 mL Superdex 200 column in GFB was used with a flow rate of 0.5 mL/min for 40 minutes per sample, and tryptophan residues were monitored by excitation at 295 nm and emission at 350 nm.

2.4.11 Negative stain electron microscopy (EM)

Protein stocks for analysis by negative stain electron microscopy were diluted in GFB both with and without detergent to give samples at 100 µg/mL and 50 µg/mL. Carbon films on 300 copper mesh grids (Agar Scientific) were used that had been subject to Glow discharge using the Quorum GloQube (exposed to plasma at 30 mA for 30 seconds in negative polarity). In a radioisotope laboratory, 5 µL of sample was added onto these grids using a pipette. After 30 seconds the grids were blotted onto filter paper before dabbing 2x 10 seconds in water with blotting in between, before 2x 20 seconds in uranyl acetate with blotting in between. Images of the grids were taken on a JEOL 2100 200 kV electron microscope (Diamond Light Source) with a lanthanide hexaboride electron source by Dr Peter Harrison.

2.5 Synthesis of polyisoprenol and dolichol lipids

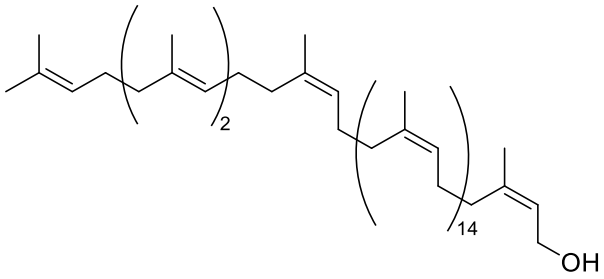
All solvents and reagents were supplied by Sigma Aldrich unless stated otherwise.



1

Polyisoprenoid acetates were extracted from the leaves of dried *Gingko biloba* (purchased from ByWholeFoodsOnline.co.uk) as described by Ibata *et al.* [120]. Briefly, 50-100 g of leaves were powdered by flash freezing in liquid nitrogen and crushed with a pestle and mortar. The lipids from crushed leaves were extracted in 1 L round bottom flask using a mixture of acetone and *n*-hexane (500 mL, 1:1, v/v) stirring under argon at room temperature for 5 days. The mixture was filtered over Celite in a plugged funnel, and concentrated *in vacuo*. Silica gel flash chromatography was used to purify polyisoprenoid acetates using petroleum ether: diethyl ether (19:1, v/v) eluent. Fractions were collected to yield 887.1 mg.

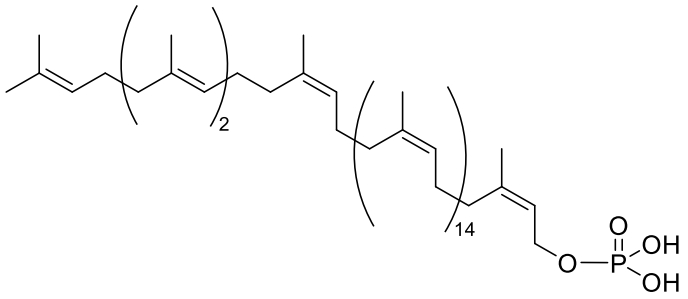
^{1}H NMR (601 MHz, $CDCl_3$) δ 5.43 – 5.26 (m, 4H), 5.19 – 5.06 (m, 13H), 4.59 (d, *J*= 7.1 Hz, 1H), 4.56 (d, *J*= 7.3 Hz, 2H), 1.76 (d, *J*= 1.5 Hz, 3H), 1.68 (d, *J*= 1.5 Hz, 38H), 1.62 – 1.58 (m, 11H).



2

Polyisoprenoid acetates were converted into 262.7 mg of polyisoprenoid alcohols using Zemplan deacetylation. Briefly, acetates (**1**) were dried under vacuum and backfilled with argon. Material was dissolved in 2 mL dry toluene and dry MeOH (1:1 v/v). 1 eq of NaOMe (25% in MeOH) was added and stirred at room temperature for 1 hour. The reaction was quenched by addition of Dowex NH_4^+ and stirred for 5 minutes. Dowex was filtered off and solvent removed by rotary evaporation. Si column flash purification was able to separate two very similar length alcohols, using an eluent of 15% diethyl ether in petroleum ether.

^{1}H NMR (601 MHz, $CDCl_3$) δ 5.45 (td, *J*= 7.2, 1.5 Hz, 1H), 5.17 – 5.07 (m, 18H), 4.09 (t, *J*= 6.4 Hz, 2H), 2.13 – 1.99 (m, 65H), 1.75 (d, *J*= 1.4 Hz, 3H), 1.70 – 1.65 (m, 42H), 1.62 – 1.58 (m, 9H), 1.10 (t, *J*= 5.6 Hz, 1H).



3

5-ethylthiotetrazole (29.85 mg, 0.229 mmol, 4.41 eq) and phosphoramidite (41.4 µL, 0.127 mmol, 2.83 eq) were dissolved in 1 mL DCM, with vigorous stirring. Polyisoprenoid alcohols (**2**, 64.6 mg, 0.052 mmol) were added in 1 mL DCM and stirred at room temperature under argon for 2 hours. TLC in 20% ethyl acetate in petroleum ether (with p-anisaldehyde staining) was used to confirm completion. Material was worked up by diluting in DCM and washed in 1x $NaHCO_3$, 1x H_2O and 1x NaCl, before drying in anhydrous sodium sulphate, filtering and concentrating under vacuum. The following day, material was dissolved in 1.02 mL tetrahydrofuran (to give 0.05 M) and oxidised by addition of 0.113 mL, 30% H_2O_2 at -78 °C, before warming to room temperature and stirring for 2 hours. The reaction was cooled to 0 °C and 1 mL of ice cold Na_2SO_3 was added dropwise to quench the H_2O_2. 15 mL of ethyl acetate was used to dilute reaction, before washing in 1x $NaHCO_3$, 1x H_2O and 1x NaCl before drying in anhydrous sodium sulphate, filtering and concentrating under vacuum. Material was deprotected by dissolving in 0.92 mL dry MeOH and dropwise addition of 0.037 mL (0.68 mmol, 13.16 eq) NaOMe (25% in MeOH) and stirring at room temperature overnight. The reaction was diluted in 10 mL DCM: MeOH (1:1) and neutralised by addition of Dowex NH_4^+, which was filtered off using cotton wool. TLC in $CHCl_3$: MeOH: H_2O: NH_4^+ revealed starting material remained, and so 100 µL NaOMe in 0.92 mL MeOH and 1 mL toluene was added. The reaction was monitored at 1 hour where material remained on the baseline and so reaction was left to stir overnight in attempt to react all starting material. Silica flash column

chromatography was performed in $CHCl_3$: MeOH (9:1) in a gradient to 100% EtOAc : MeOH : H_2O : NH_4^+ (4 : 2 : 1 + 1%). 29 mg of product, **3**, was collected and analysed.

^{31}P (162 MHz, MeOD) δ 2.59 – -1.52 (m).

^1H NMR (400 MHz, MeOD) δ 5.25 (d, *J*= 9.4 Hz, 2H), 5.03 (d, *J*= 6.7 Hz, 14H), 1.96 (d, *J*= 5.5 Hz, 51H), 1.75 (s, 1H), 1.59 (s, 35H), 1.52 (d, *J*= 5.2 Hz, 9H), 1.17 (s, 17H), 0.85 – 0.71 (m, 3H)

2.6 Functional assay

2.6.1 UMP/UDP-Glo assay

The UMP- and UDP-Glo assay kits (Promega) use relative luminescence from luciferase to monitor UMP or UDP release of phospho-/glycosyltransferases [121]. They were tested for use in studying DPAGT1 and ALG13/ALG14 action. All reactions were performed in a buffer of 20 mM HEPES pH 7.5, 200 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$, 0.27% DM and 0.027% CHS, unless stated otherwise.

DPAGT1 purified by Dr Yin Dong, was studied over 25-minutes in 10 µL total reactions. (2x) 1 µM DPAGT1 was incubated for 1 hour on ice with (2x) 40 µM dolichol phosphate to allow for lipid substrate uptake into detergent micelles. 5 µL of this enzyme and substrate stock was added into wells of a PCR plate. Reactions were sequentially initiated by addition of 5 µL Ultrapure UDP-GlcNAc (to give 1x concentrations). Use of a negative time course allowed reactions to be terminated by incubating the at 95 °C for 5 minutes to inactivate all DPAGT1. The plate was allowed to reach room temperature and briefly spun to collect all reaction mix. 5 µL of the reaction was mixed into 5 µL of assay detection reagent in a 384 well plate (white ProxiPlate plate, Perkin Elmer). The plate was briefly spun to ensure reagents were in the bottom of each well before being incubated at room temperature for 1 hour. Luminescence was

read using a Clariostar Microplate reader (BMG). Alongside all assays, an UMP titration was tested to produce a standard curve and enable conversion of relative luminescence units (RLU) into UMP concentration, as in Chapter 4.

Kinetic parameters were obtained by taking time points every 2 minutes up to 20 minutes with 1.5 μM DPAGT1. The concentration of one substrate was varied between 0- 100 μM whilst the second substrate was fixed at 50 μM. Rates of the reactions were plot against varied substrate concentration and modelled by the Michaelis-Menten equation (Chapter 4, eq 1) in GraphPad Prism.

For determination of enzyme mechanism, substrate matrix experiments were performed using sub-saturating to saturating concentration of the fixed substrate. 1.5 μM DPAGT1 was assessed as before with 0-100 μM varied substrate at 1.25, 2.5, 5, 10, 20, 40 or 80 μM fixed substrate. The resulting data were modelled by the ping-pong and ternary complex equations (Chapter 4, eq 2 and 3). The goodness of fit of both models were compared using in-built Akaike's Information Criterion (AICc) assessment.

2.6.2 Radioactive TLC assay

For endpoint activity study, 2 μM His-DPAGT1, ALG13/His-ALG14 or His-DPAGT1/ALG13/His-ALG14 were incubated at 25 °C for 1 hour with 100 μM UDP-[^{14}C]-GlcNAc (American Radiolabelled Chemicals) and 100 μM lipid phosphate or pyrophosphate GlcNAc (see Chapter 5 for details) using a 10 μL assay volume in thin-walled PCR tubes. Assay buffer was used as for the Promega Glo assays (GFB supplemented with 10 mM $MgCl_2$, 0.27% DM and 0.027% CHS). Using a 1:1 by volume addition of methanol the reactions were quenched. 3x 1 μL of each reaction were spot onto aluminium backed thin layer chromatography plates (Merk MilliPore). A mobile phase of ethyl acetate: isopropanol: water (3:2:1) was used to separate lipid species, and the samples ran for 30 minutes. The plates were

dried in a fume hood before wrapping in plastic film-wrap and exposing to a BAS-IP MS Phosphor screen (Cytiva) for 3 days in a light-sealed cassette. The screen was read for phosphorescence using a Typhoon laser scanner (Cytiva) with laser excitation at 635 nm and image collected at a pixel size of 50 μm. Where stated, phosphorescence was read using an FLA-5100 laser scanner (FujiFilm).

ImageJ software was used to quantify the total counts of each lane of the TLC plate by densitometry. The amount of upper lipid linked-[^{14}C]-GlcNAc was adjusted for total radioactivity loaded using the counts of both the lipid-[^{14}C]-GlcNAc spot and baseline UDP-[^{14}C]-GlcNAc. The data were analysed using in-built GraphPad Prism multiple t tests where the null hypothesis was rejected if $P < 0.05$.

Following identification of the linear reaction phase, the rate of reaction of 0.25 μM ALG13/ALG14 with 0-200 μM of one substrate at 100 μM of the other was taken by amount of product after 1 minute. For DPAGT1, 1.5 μM enzyme was used and the rate of reactions were determined by amount of lipid linked [^{14}C]-GlcNAc after 20 minutes. These rates were plot against varied substrate concentration as before, and modelled by the Michaelis-Menten equation (Chapter 4, eq 1) in GraphPad Prism to obtain apparent K_M and V_{max}.

2.7 Structural study

2.7.1 AlphaFold Multimer using Google Collaboratory

The structure of ALG13, ALG14 and DPAGT1 alone and when in complex were modelled using AlphaFold Multimer [122, 123]. Full-length sequences of each protein as wild-type or containing pathogenic mutations were input on a local installation of ColabFold [124] using Google Collaboratory servers. Generated structures were aligned using PyMOL [125] to evaluate the root mean square deviation (RMSD).

2.7.2 Vapour diffusion-controlled crystallisation trials

250 µL of 10 mg/mL freshly purified ALG13/ALG14 protein was dispensed into plates containing a matrix of crystallisation conditions using MemGold2 plates (Molecular Dimensions [126]). The plates were stored at 4 °C or 20 °C in humidified environments, and were monitored for signs of crystallisation under the microscope.

2.7.3 Single particle cryogenic electron microscopy (cryoEM)

To gain structural information about the DPAGT1/ALG14/ALG14 complex, single particle cryogenic electron microscopy (cryoEM) was explored. Grid preparation and data collection for cryoEM were performed with Dr Peter Harrison (Diamond Light Source).

Protein samples destined for study by cryoEM were spun for 20 minutes at 4 °C and 120,000 x g to remove aggregation. Quantifoil R 1.2 / 1.3 Au 300 mesh grids were glow discharged using a Quorum GloQube Glow discharge system at 30 mA for 60 seconds, and samples were vitrified using a Vitrobot Mark IV cryo-plunger (ThermoFisher) at 4 °C and 100 % humidity. 3 µL of target protein was applied to the grid and after a 10 second incubation, grids were blotted onto filter paper for various times and forces before being plunged into liquid ethane at -188 °C to form vitreous ice. Blot force and time, as well as concentration of protein applied, were varied to give a range of ice thickness and particle density to visualise when screening.

For protein bound to lipid substrates, DPAGT1/ALG13/ALG14 was incubated on ice for 30 minutes with a 1:1 volume of 500 µM lipid (either DPAGT1 substrate C90-Dol-P, ALG13/ALG14 substrate, C90-Dol-PP-GlcNAc or ALG13/ALG14 product, C55-Poly-PP-$(GlcNAc)_2$. This gave final protein concentrations of 1 mg/mL protein and 250 µM lipid, giving a 24-fold molar excess of substrate. An example is shown in Table 2.12 for DPAGT1-FLAG/ALG13/ALG14. Microscope grids were clipped ready for loading.

Table 2.12 Example of grids prepared for cryoEM containing DPAGT1-FLAG/ALG13/ALG14. In this case, the 2 mg/mL apo protein sample blotted with force 2 and blot time 3 (row 3) was selected for large scale data collection following screening.

[DPAGT1-FLAG/ ALG13/ALG14], mg/mL	[Lipid], µM	Blot force	Blot time
1	-	2	3
1	-	2	3
2	-	2	3
2	-	2	3
1	C90-Dol-P, 250 µM	2	3
1	C90-Dol-P, 250 µM	2	3
1	C90-Dol-PP-GlcNAc, 250 µM	2	3
1	C90-Dol-PP-GlcNAc, 250 µM	2	3
1	C55-Poly-PP-$(GlcNAc)_2$, 267 µM	2	3
1	C55-Poly-PP-$(GlcNAc)_2$, 267 µM	2	3
1	C90-Dol-PP-GlcNAc, 250 µM	3	3
1	C55-Poly-PP-$(GlcNAc)_2$, 267 µM	3	3

Data were collected at the electron Bio-Imaging Centre (eBIC) at Diamond Light Source, using a 200 kV Glacios cryo transmission electron microscope (Thermofisher) with a Falcon IV detector under proposal nt33941. Thermofisher EPU software was used to perform an initial atlas screening at 135x magnification to allow selection of the grid squares with best particle coverage, intact squares and ice thickness for full data collection. A magnification of 120,000

x (1.192 Å px^{-1}) was used with a dose of 3.86 e px^{-1} sec^{-1} applied over 13.43 seconds to give a total dose of 60 e $Å^{-2}$, unless stated otherwise.

Figure 2.5 shows a visualisation of the typical single particle cryoEM processing workflow. The collected movies were corrected for motion using MotionCorr2 [127] using the eBIC auto processing pipeline in Relion 3.1 [128]. Motion corrected micrographs were then imported into CryoSPARC v3.3.2 (later CryoSPARC v4.4.0)[129] and contrast transfer function (CTF) correction was performed using CTFFind4 [130]. Initially, particles were picked using the Laplacian of Gaussian blob picker function, and filtered based on CTF fit resolution. These were extracted from the micrographs and averaged into 50 2D classes. Following selection of the best-looking classes likely containing protein of interest, the particles from this subset were extracted and averaged to successively fewer 2D classes. Iterative selection of the best-looking classes from previous rounds of classification was performed.

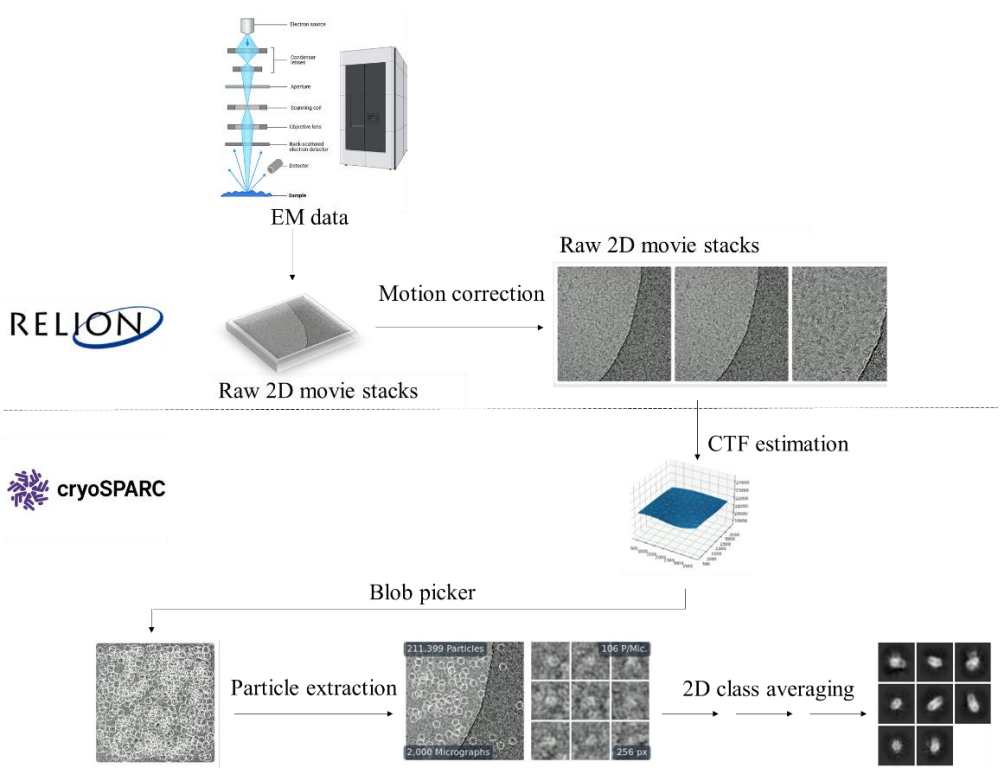


Figure 2.5 Visualisation of cryoEM image processing work-flow, first using Relion by the auto-processing pipeline at eBIC [128], and then using cryoSPARC [129].

Chapter 3

Expression, purification and structural analysis of a DPAGT1, ALG13 and ALG14 complex

3.1 Introduction

3.1.1 A multi-enzyme complex of DPAGT1, ALG13 and ALG14

Lipid-linked oligosaccharide assembly in the ER is initiated by dolichol-phosphate α-*N*-acetyl-glucosaminyl-phosphotransferase (DPAGT1), which catalyses the reaction between phosphorylated dolichol lipid carrier and uridine diphosphate *N*-acetyl glucosamine (UDP-GlcNAc) to yield dolichol pyrophosphate (PP) GlcNAc [44]. The second step in lipid-linked oligosaccharide assembly involves the unusual heterodimeric complex ALG13/ALG14, which catalyses the transfer of a second GlcNAc onto dolichol-PP-GlcNAc from UDP-GlcNAc yielding a chitobiose-linked dolichol (Figure 3.1).

Interestingly, it has been shown that in yeast DPAGT1 and ALG13/ALG14 form a multi-glycosyltransferase complex, reported by Noffz *et al.* to be 2:2:2 in nature [77]. The Gao group have shown the N-terminal transmembrane domain of ALG14 is likely to interact with DPAGT1 [27, 131], which is unsurprising given DPAGT1 is integral to the ER membrane with 10 spanning helices [44]. This formation of multi-order glycosyltransferase complexes is also seen in other domains of life. MraY and MurG perform analogous steps to DPAGT1 and ALG13/ALG14 during lipid II production in gram-negative bacteria, and these too have been found to form a complex [132]. MurG has since also been shown to act as a scaffold for a large, Mur enzyme complex [133], suggesting the potential that ALG13/ALG14 might also. It is

therefore of key interest to elucidate whether the human ALG proteins are organised in higher order complexes in the ER membrane, which might have profound effects on the kinetics of dolichol-linked oligosaccharide synthesis.

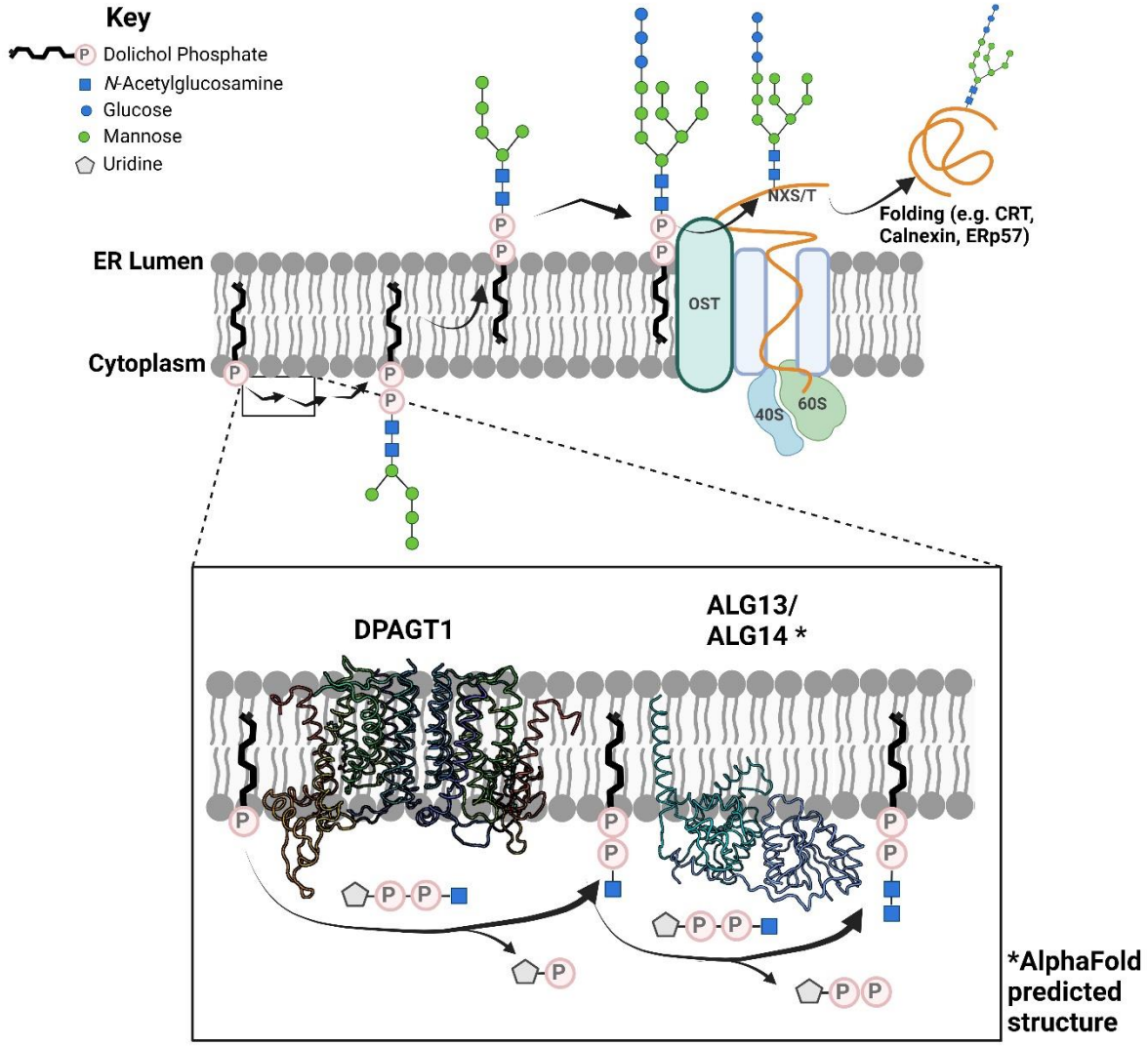


Figure 3.1 Function of DPAGT1 and ALG13/ALG14 for lipid-linked oligosaccharide synthesis in the ER.

3.1.2 Expression and purification in insect cells

To explore this phenomenon, the expression and purification of human ALG13 and ALG14 proteins alone and in complex with DPAGT1, were optimised. Previously, the Dong group had cloned full length *ALG13* into a baculovirus vector to allow expression of the recombinant protein in insect cells. This long established technique utilises an adapted baculoviral genome to hijack the expression system of insect cells, allowing high yields of recombinant protein to be produced [115]. The gene of interest is first cloned into a donor plasmid vector (e.g. pFastBac) which is then transformed into an adapted strain of *E.coli* known as DH10bac [134]. This contains a modified baculovirus genome as well as a helper plasmid encoding proteins required for Tn7 transposition [134]. Once recombined, the bacmid containing the gene of interest can be isolated and transfected into insect cells to produce baculovirus progeny, which are subsequently used to infect and induce expression of the target gene under the control of a polyhedron promoter.

Insect cell expression in this way has a number of advantages - in particular for the production of membrane-associated or integral membrane proteins like DPAGT1 and ALG13/ALG14. It strikes the balance between an affordable method for large scale production - important given typically low yields achieved for membrane proteins - and expression of high-quality protein with many of the simple post-translational modifications found in human cells that those expressed in bacteria would lack [135]. This is especially relevant for production of multi-protein complexes given their assembly can often be reliant on protein modifications such as removal of signalling peptides, glycosylation, disulphide bond formation or use of chaperone proteins. Additionally, the stoichiometry of each subunit in a multi-protein complex is often vital for its assembly. Co-expression of the three proteins therefore needed to be optimised to ensure this balance was achieved. Berger *et al.* developed a MultiBac system to address this which allows many subunits to be expressed from the same baculoviral vector – unifying the stoichiometry of expression [136].

Many factors require optimisation to access pure stocks of isolated membrane protein for structural and functional study. Firstly, the constructs themselves; can the full-length proteins be stably produced or is it necessary to truncate disordered regions, or focus attention on specific subunits? The addition of N- or C-terminal purification tags is normally required, and the type of tag as well as the position in the protein with any linkers or protease cleavage sites requires careful testing to ensure normal folding and function of the resulting proteins is maintained. To ensure stability of the isolated proteins, the conditions of all buffers (pH, salt concentration and composition, crowding agents, reducing agents etc.) require tailoring to meet the needs for each specific protein. For maintenance of native and active membrane proteins, it is also necessary to use detergents to stabilise the proteins once removed from their lipid bilayer [137]. The best performing detergent for extraction and solubilisation of membrane proteins will depend on the size and shape of the target as well as its intended downstream use, and this also requires careful screening to select the appropriate chemistry of detergent that does not disrupt the proteins' function.

3.1.3 Structural elucidation of DPAGT1/ALG13/ALG14

Evaluating the 3D structure of a protein allows for better understanding of its function, and it was therefore pertinent to determine the structure of human ALG13/ALG14 alone or in complex with DPAGT1. As discussed, the structure of DPAGT1 was solved by X-ray crystallography in 2018, with the sugar donor UDP-GlcNAc and the potent natural product inhibitor tunicamycin bound in the cytoplasmic facing active site [24, 60]. The structure of the monomer and homodimer are shown in Figure 3.2, where 10 membrane spanning helices can be seen. X-ray crystallography remains a powerful method for structural determination, and in particular is the technique of choice for small proteins such as ALG13 and ALG14 which have a combined molecular weight of just 42.3 kDa. X-ray crystallography does however require large amounts of high purity protein, and also relies on the protein crystallising which is

sometimes not possible for flexible complexes [138]. NMR has previously been utilised to access the structure of yeast ALG13 – exploiting the fact that this is the small and soluble subunit of the bipartite ALG13/ALG14 making it amenable for study by NMR [30, 139]. Others have employed the advancements in resolution that can now be achieved using single particle cryogenic electron microscopy (cryoEM) to elucidate the structure of ALG proteins. For example, ALG6 was recently solved by Bloch *et al*. who used an antigen binding fragment (Fab) to add mass and stability to the glycosyltransferase to gain high resolution structures [140]. Structures for small (under 100 kDa) proteins remain difficult to resolve using this methodology, although 77 unique membrane proteins of this size have now been resolved to better than 5 Å using cryoEM [141].

The recent revolution in our ability to predict protein structure from sequence alone using computational structural modelling has also allowed insight into the ALG enzymes. Homology modelling was used by Mitusińska *et al.* to assess human ALG13, using the reported structures for yeast ALG13 and bacterial MurG [142]. This model predicts the human protein to be most similar to the Raman *et al.* yeast NMR structure of ALG13 [139], and to contain five parallel β-sheets and eight α-helices [142]. The structure of human ALG14 has been modelled by the machine learning system AlphaFold with high predicted accuracy, as in Figure 3.2B [123]. The only model of ALG13 present in the AlphaFold database is for the long isoform 1- currently with unknown biological function- and it is of very low confidence for the majority of the protein. Crucially, predictive methods are often not accurate for multi-enzyme complexes, and by nature are limited in their ability to predict novel folds meaning it remains necessary to confirm findings experimentally. It is therefore of interest to access high quality DPAGT1 and ALG13/ALG14 proteins to advance our understanding of the early N-glycosylation pathway in humans.

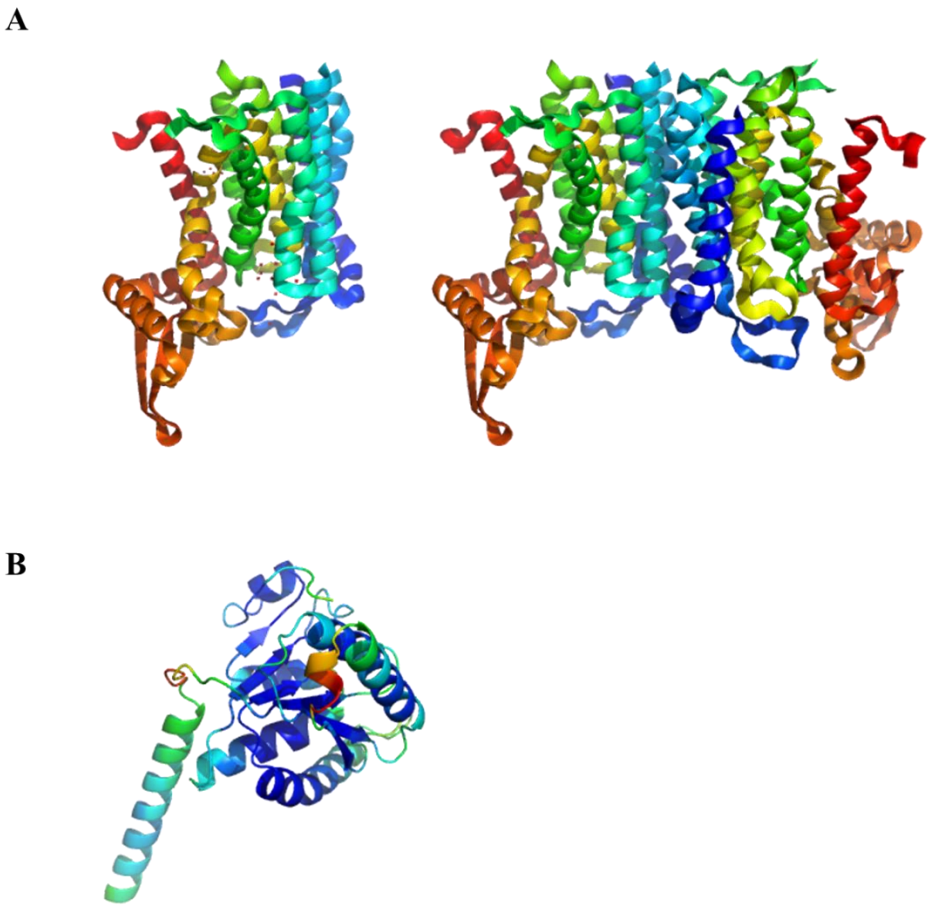


Figure 3.2 A Structures of DPAGT1 as a monomer and homodimer. Left, PDB 5LEV, right PDB 6FM9, coloured by chain. **B** Predicted structure of ALG14, AF-Q96F25-F1, coloured by AlphaFold per-residue model confidence score (pLDDT) where blue dark blue is very high and red is very low confidence.

3.2 Results

3.2.1 Predicting the structure of human ALG13/ALG14

The structure of ALG13 in complex with ALG14 was modelled using AlphaFold Multimer [122] given there is no published structure – experimental or predicted – currently available for this heterodimeric enzyme. This was executed on Google Colab graphics processing unit (GPU) and found to be of high confidence with pLDDT > 90 for the majority of the protein (Figure 3.3A). This prediction suggests the C-terminal helix of ALG13 interacts with the globular, C-terminal domain of ALG14 and the N-terminal helix of ALG14 is seen to protrude away from the rest of the proteins. This has previously been modelled as the transmembrane spanning region [143]. The predicted cytosolic structure resembles that of the *Pseudomonas aeruginosa* MurG, as shown in Figure 3.3B, with a global RMSD of just 1.399 Å. ALG13 and ALG14 have 11.7% sequence identity, yet they are predicted to adopt very similar Rossman fold globular domains. This prediction therefore corroborates the experimentally determined topology of ALG14 with an N-terminal ALG14 transmembrane domain and C-terminus that interacts with ALG13 [27]. Interestingly, ER localisation in yeast ALG14 was found even in the absence of the transmembrane spanning domain suggesting the globular C-terminus also mediates ER association [143].

This model was also aligned to the experimentally determined yeast ALG13 structure (Figure 3.3C), with a global RMSD of 3.588 Å. The terminal helix of yeast ALG13 is seen to fold with the rest of the globular Rossman fold [30], whereas this is predicted to protrude out and interact with ALG14 in the AlphaFold model- as highlighted in the box in Figure 3.2B. This is also absent in the homology model of ALG13 presented by Mitusińska *et al.* which is unsurprising given it was built on this yeast structure [142]. It is perhaps likely that this is also the induced

configuration of the yeast helix when bound to ALG14, as the structure was solved when ALG13 was in isolation.

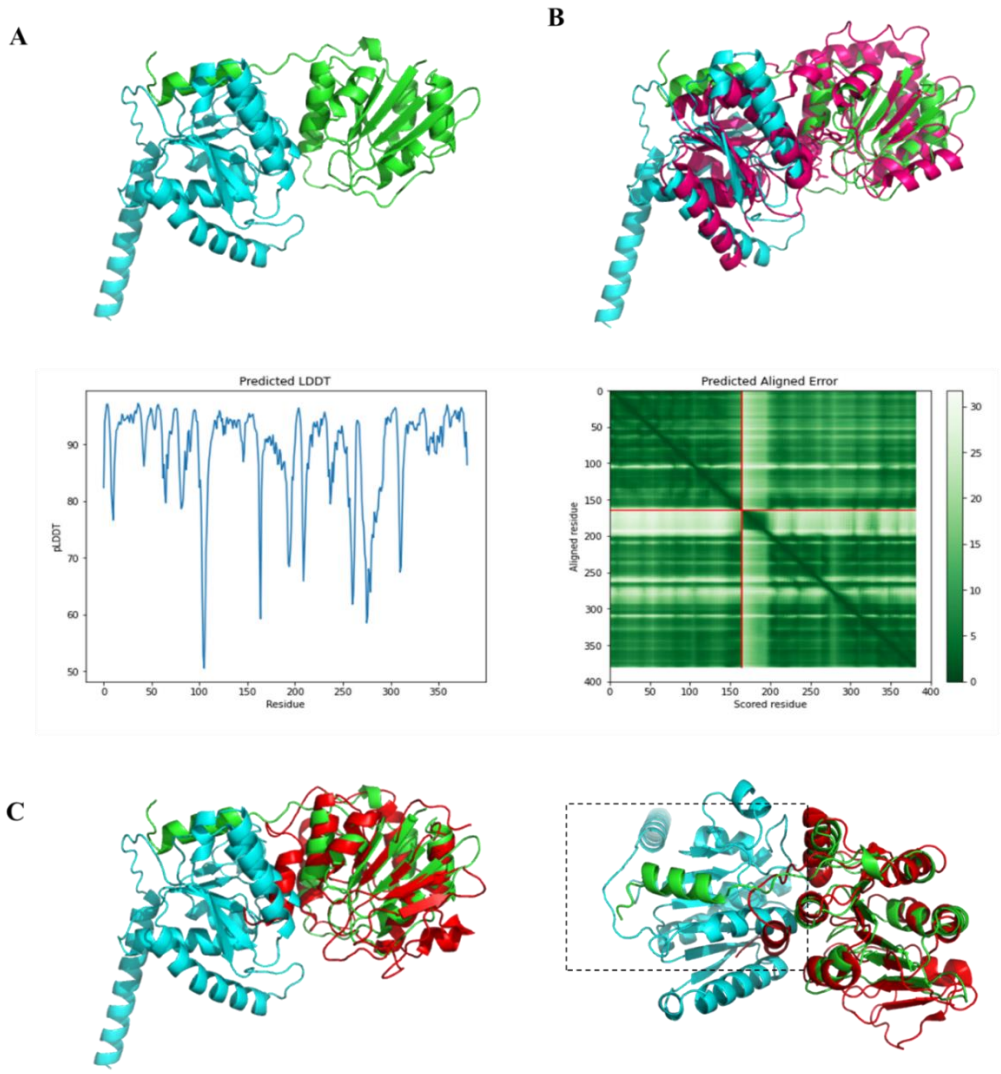


Figure 3.3 Modelling the structure of human ALG13/ALG14. **A** AlphaFold-Multimer predicted structure of ALG13 (green) and ALG14 (light blue), with predicted LDDT and alignment error plots generated by AlphaFold Multimer below. **B** Alignment of the predicted human ALG13/ALG14 model with the structure of *Pseudomonas aeruginosa* MurG [144] (magenta, PDB 3S2U) **C** Alignment of the predicted human ALG13/ALG14 model with a yeast NMR structure of ALG13 [30] (red, PDB 2JZC) showing side and top view.

3.2.2 Purification of ALG13 and ALG14

To corroborate these findings experimentally, extensive method development has been undertaken to access wild type (WT) ALG13 and ALG14 protein. Initially, attempts to access soluble, cytoplasmic ALG13 were made so that comparisons to protein containing pathogenic mutations could be drawn. Full length, human ALG13 with a TEV cleavable N-terminal 6xHis tag was cloned into a baculovirus vector (cloning and baculovirus generation performed by Jing Wang), to allow over-expression in Sf9 insect cells.

To start, a small-scale test expression and purification was performed, using 800 mL of Sf9 insect cells in mid-log phase infected with the ALG13-encoding baculovirus. Following incubation for 72 hours to allow maximum levels of protein production, the cells were harvested by centrifugation and immobilised metal affinity chromatography (IMAC) was performed. Co^{2+} resin (Talon) was found to bind more target protein than Ni-NTA resin (data not shown), and so this was used in all purifications. Following IMAC, the sample was incubated with TEV protease in an attempt to cleave the 6xHis purification tag which would allow for a reverse purification to remove impurities and the 6xHis tagged TEV protease. Samples from across the purification pipeline were analysed by Western blot and probed for the presence of the 6xHis tag which would remain on the target before tag cleavage (Figure 3.4). Excitingly, His-ALG13 at its expected molecular weight of 20.7 kDa was visible before TEV-protease cleavage. It was clear however that the majority of target protein remained bound to the Talon resin after TEV-protease incubation suggesting cleavage of the tag was not successful and rendering the reverse purification ineffective. Additionally, ALG13 was lost during washing of the resin, with protein seen in the 40 mM imidazole wash elution (lane 4, Figure 3.4). Although attempts were made to equalise the loading of the blot to the volume of original sample, the wash fraction was 30x the volume of the elution fractions and therefore this represents a large amount of target protein being lost.

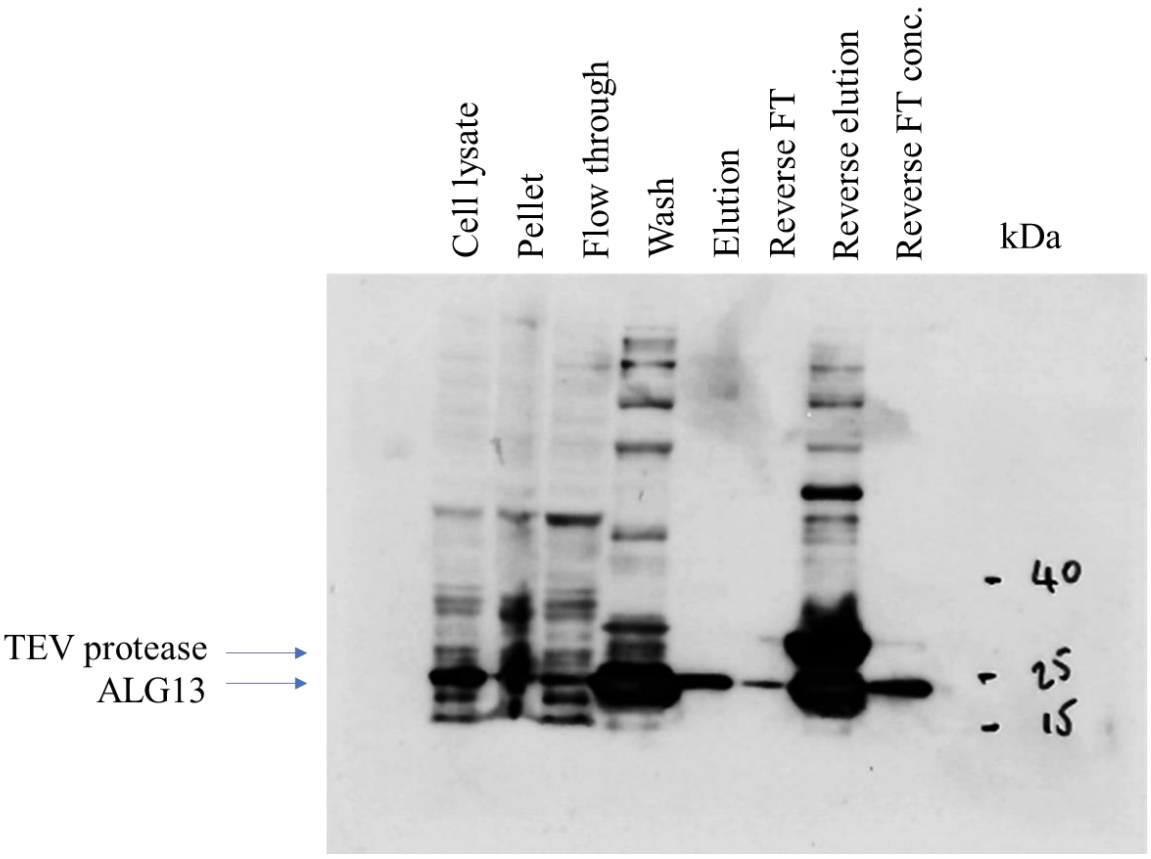


Figure 3.4 Initial testing of ALG13 purification by IMAC and reverse IMAC. Western blot probing for the 6xHis purification tag of samples across the purification pipeline. 1 ° Ab mouse anti-His (CPA9028) at 1:1000 and 2 ° goat/rabbit anti-mouse IgG-HRP at 1:1000 were used, and visualised by exposing X-ray film to the blotted membrane for 5 seconds. The position of key protein standards in the ladder were marked onto the film before exposure. Reverse FT = flow through from Talon resin following incubation with TEV protease to remove the 6xHis tag, and that following concentration labelled as Reverse FT conc.

In order to reduce the amount of ALG13 lost during IMAC, a range of lower imidazole wash buffers were tested for use. Figure 3.5 shows representative results from a 2 L scale purification of ALG13 protein. Bands of the expected molecular weight (20 kDa) were seen in elution fractions 3-6 and were excised and confirmed to be the target protein by tryptic digest MS/MS (Performed by Dr Rod Chalk, Centre for Medicines Discovery, data not shown). Lost target

protein was seen in wash buffers containing 20- and 30 mM imidazole (lanes 3 and 8 respectively) but not 10 mM, and so this was selected as an appropriate wash buffer for subsequent purifications.

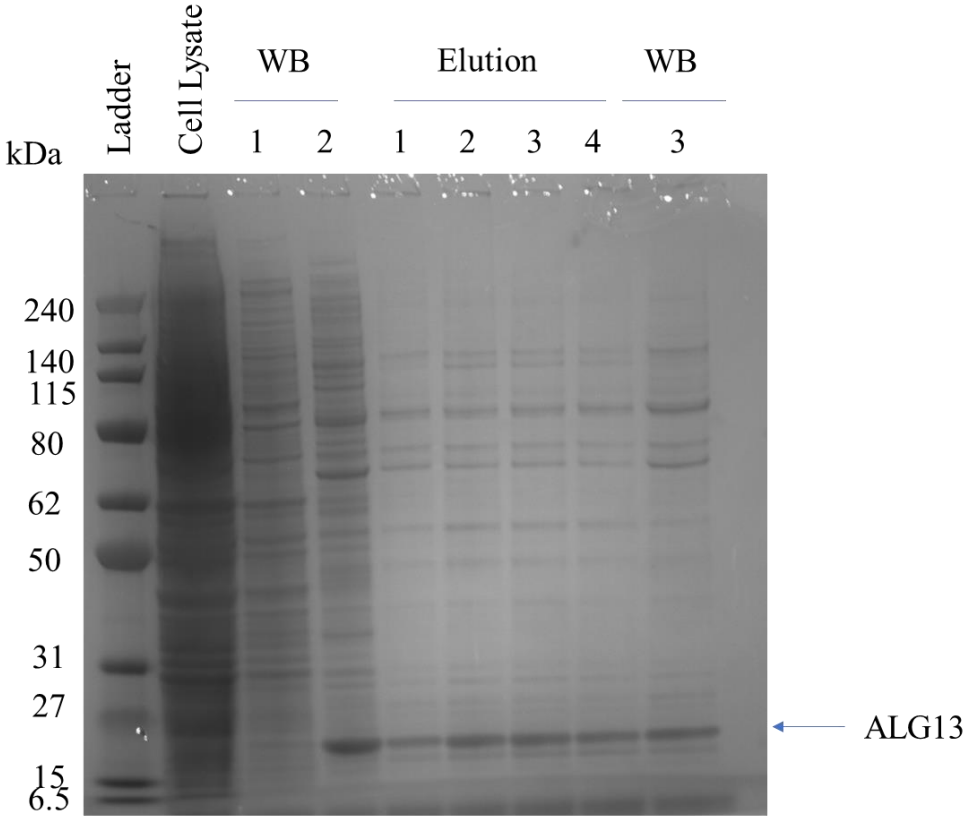


Figure 3.5 Wash buffer testing for ALG13 IMAC purification. Coomassie stained SDS-PAGE of ALG13 protein IMAC with three wash buffers containing 10, 20 or 30 mM (wash buffer (WB) 1, 2 and 3 respectively). Elution fractions from IMAC contained ALG13 (identity confirmed by tryptic digest MS/MS).

Next, a screen was performed to optimise the buffers used during purification of ALG13. A range of different buffers spanning pH 7-8, as well as those with and without reducing agent, glycerol, and various salts were tested, as outlined in Table 3.1. Using protein eluted from IMAC in a base buffer of 150 mM NaCl, HEPES pH 7.5 and 0.5 M TCEP, ALG13 was exchanged into 9 different buffers and level of precipitation was monitored following 3 days

at 4 °C (Figure 3.6). ALG13 appeared remarkably stable across the conditions, with little protein seen to precipitate over the 3-day period. There was no protein pellet at all when ALG13 was stored in buffer containing 500 mM NaCl, as well as in that containing 2 mM TCEP. Buffers with high salt and reducing agent were therefore used in future purifications. The added stability in these conditions is perhaps unsurprising given ALG13 has six putative surface exposed cysteines, which could lead to aggregation when non-reduced.

Table 3.1 Matrix of buffer conditions screened for use in ALG13 purifications. Condition number 1 represents the standard buffer that was used for extraction and initial purification of the protein using IMAC, before protein was exchanged into 1 mL of conditions 2 – 10. The condition changed from buffer 1 is highlighted in blue.

	Buffer	**pH**	**NaCl, mM**	**TCEP, mM**	**% Glycerol**	**$MgCl_2$, mM**
1	HEPES	7.5	150	0.5	0	0
2	TRIS	7.5	150	0.5	0	0
3	TRIS	8	150	0.5	0	0
4	PBS	7.5	150	0.5	0	0
5	HEPES	7	150	0.5	0	0
6	HEPES	7.5	500	0.5	0	0
7	HEPES	7.5	150	1	0	0
8	HEPES	7.5	150	2	0	0
9	HEPES	7.5	150	0.5	5	0
10	HEPES	7.5	150	0.5	0	5

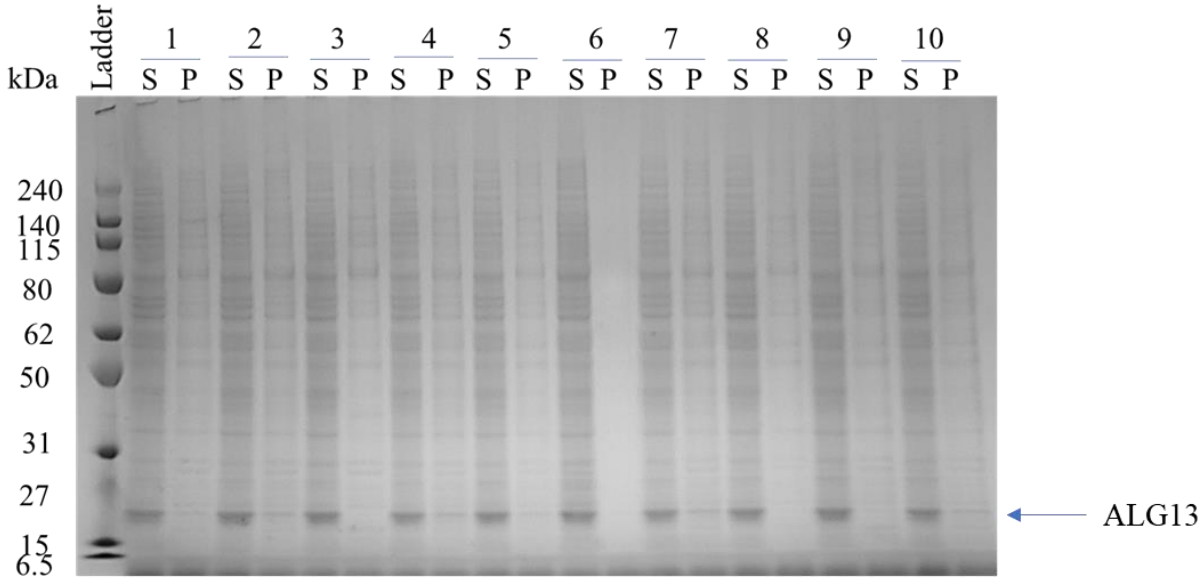


Figure 3.6 Buffer stability screen of ALG13 WT protein in 10 different buffer conditions. Numbers correspond to buffer conditions in Table 3.1. 0.5 mL of ALG13 was exchanged into 1 mL of each buffer condition using equilibrated G25 columns. After 3 days at 4 °C, samples were centrifuged at 21,000 RPM at 4°C for 10 min (JLA 25.5) to pellet any precipitated protein. The supernatant was removed and pellets dissolved in 1 mL of 8 M urea. SDS-PAGE was performed to test relative protein levels in the supernatant and pellet of each buffer S = supernatant, P = pellet. ALG13 molecular weight of 20.7 kDa, shown in lowest intense band.

Ultimately, further attempts at purifying this ALG13 construct were stopped when it was discovered using denaturing, intact mass spectrometry that the expressed ALG13 was of a truncated form (Figure 3.7). The experimentally measured mass of 17,523.3 Da maps to a peptide of predicted 17,524.2 Da of ALG13 without the N-terminal 6xHis tag, as well as loss of 11 amino acids from the C-terminus. Given the full-length clone was confirmed during sequencing, this cleavage appears to occur following expression in Sf9 insect cells. Purification of this construct was therefore not pursued further, and instead cloning was undertaken to introduce alternative purification tags with the aim of improving stability.

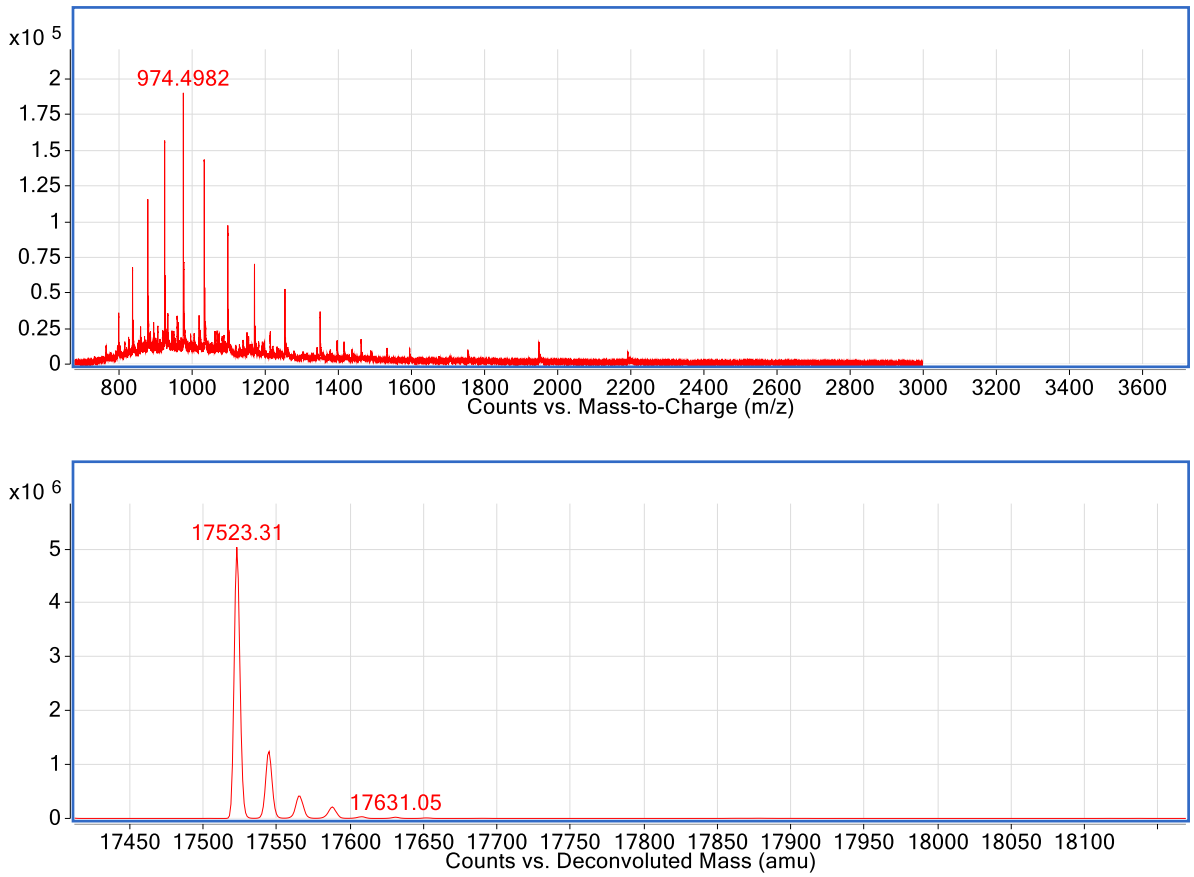


Figure 3.7 Denaturing, intact mass spectrometry. ALG13 was analysed by LC/MS-TOF and found to be truncated. The ion series between 800-3600 (m/z) is shown above with corresponding deconvoluted mass of protein below. The experimentally determined mass matches to a peptide whose sequence is shown in blue with the remaining full length, tagged ALG13 in black (figure created using Paws software).

3.2.2.1 ALG13/ALG14 MultiBac expression and purification

As the N-terminally 6xHis tagged ALG13 was purified in a truncated form, cloning was performed to access alternatively tagged constructs. Additionally, given the active form of ALG13 is in complex with ALG14, it was decided to design these on vectors co-expressing ALG14 using the MultiBac system [116] . This would enable functional study but also hopefully improve the stability of ALG13 and enable purification. Full length ALG13 and ALG14 with various combinations of purification tags were cloned into MultiBac vectors (work carried out by Jing Wang). It has previously been reported that the C-terminus of ALG14 is important for its interaction with ALG13 [27, 145], hence the N-terminus was selected for tagging. Out of these, two were found to successfully express both proteins. These encoded ALG14 bearing an N-terminal 6xHis-tag, with either untagged or N-terminal 6xHis tagged ALG13. These MultiBac vectors were both taken forward to optimise purification of ALG13/ALG14.

ALG14 contains a transmembrane domain, tethering the ALG13/ALG14 complex to the ER membrane. It was therefore necessary to solubilise the proteins using detergent. Initially, two detergent conditions were compared for use in ALG13/ALG14 extraction and purification; n-Decyl-β-D-Maltopyranoside (DM) or DM supplemented with cholesterol hemisuccinate (DM/CHS). Figure 3.8 and 3.9 summarise findings for each detergent during the purification of untagged ALG13 with His-ALG14, and His-ALG13/His-ALG14 respectively. Overall, both detergents performed equally at solubilising the proteins, where untagged ALG13, His-ALG13 and His-ALG14 were seen at their expected molecular weights of 18.2, 20.7 and 26.6 kDa respectively. During size exclusion chromatography (SEC) using a Superdex 200 Increase column (Cytiva), there were broad and multiple peaks eluting – in particular for the double 6xHis tagged construct - suggesting aggregation was occurring.

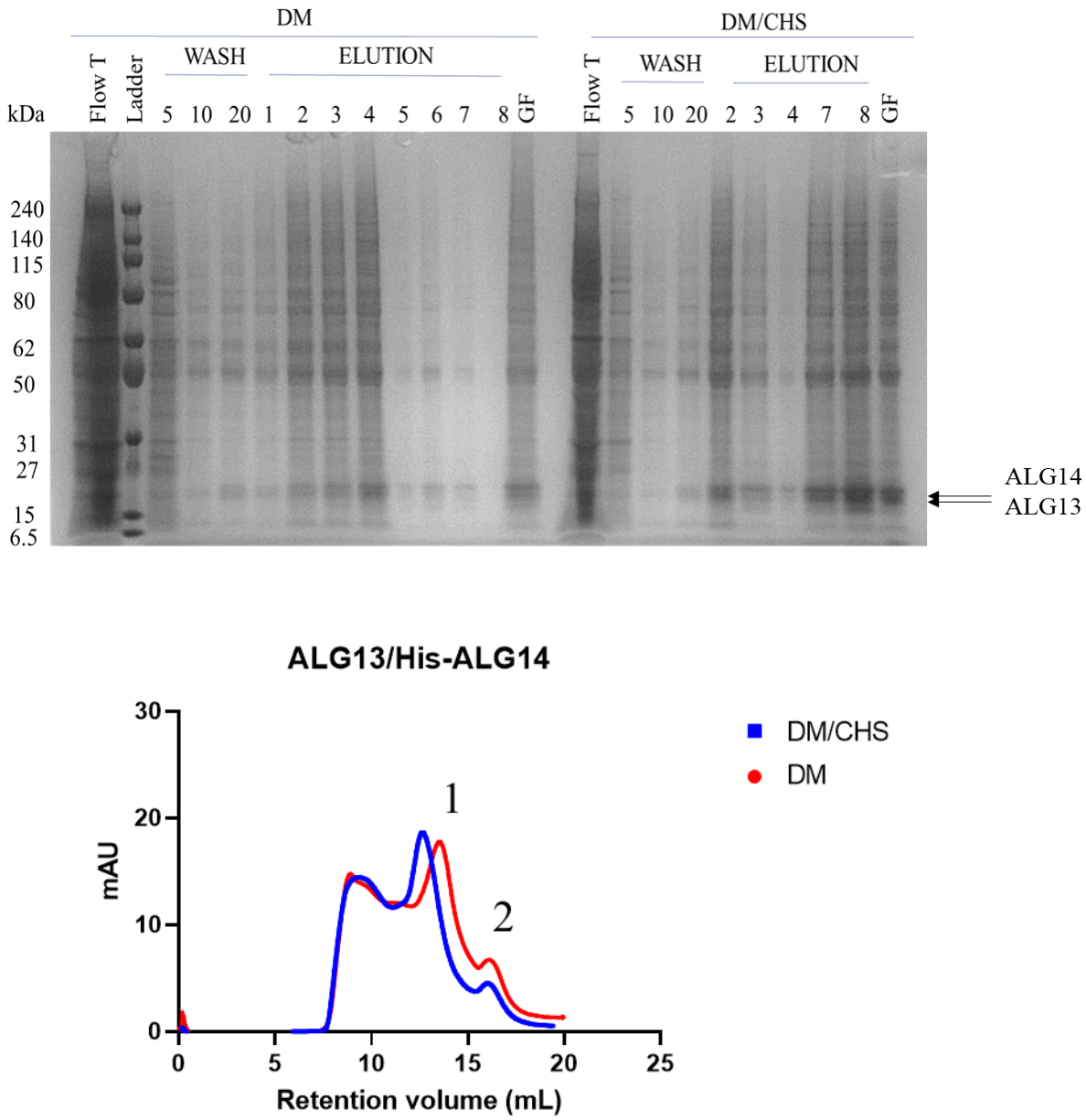
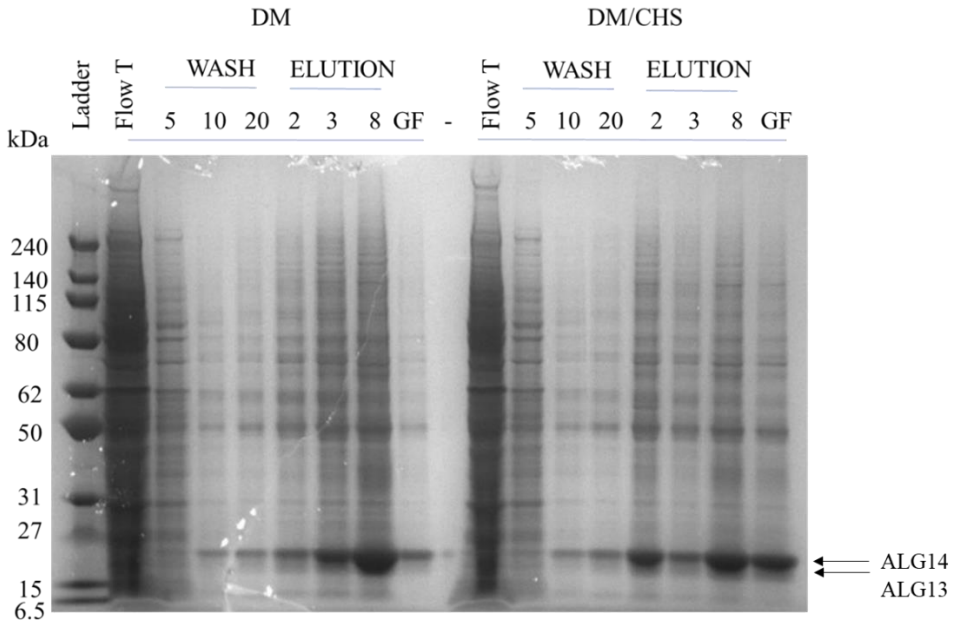


Figure 3.8 IMAC purification ALG13/His-ALG14 in DM or DM/CHS containing buffers. Samples from the purification of 2 L baculovirus infected Sf9 cells by IMAC are shown by Coomassie stained SDS-PAGE. Flow T = flow through after Talon resin binding, Wash = buffers containing either 5, 10 or 20 mM imidazole. GF = combined IMAC fractions buffer exchanged into gel filtration buffer. Absorbance at 280 nm of proteins eluting from a Superdex 200 Increase column are shown below, with peaks labelled as in lanes of Figure 3.10.



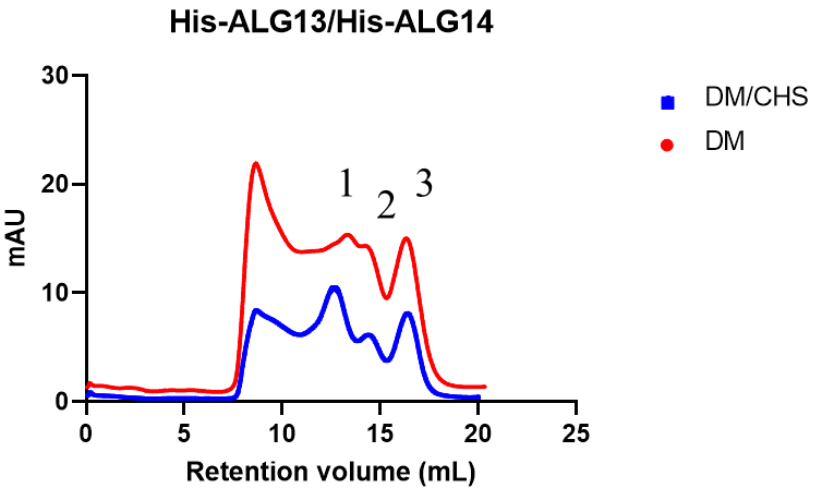


Figure 3.9 IMAC purification His-ALG13/His-ALG14 in DM or DM/CHS containing buffers. Samples from the purification of 2 L baculovirus infected Sf9 cells by IMAC are shown by Coomassie stained SDS-PAGE. Flow T = flow through after Talon resin binding, Wash = buffers containing either 5, 10 or 20 mM imidazole. GF = combined IMAC fractions buffer exchanged into gel filtration buffer. Absorbance at 280 nm of proteins eluting from a Superdex 200 Increase column are shown below, with peaks labelled as in lanes of Figure 3.10.

Due to the low concentration of protein in each SEC fraction, total protein from each of the SEC peaks was precipitated using trichloroacetic acid (TCA) before analysis by Coomassie-stained SDS-PAGE (Figure 3.10). ALG13 and 14 have very similar molecular weights: His-tagged ALG14 at 26760.38 Da and ALG13 at 20834.99 Da, making identification on the gels difficult. Putative ALG13/ALG1414 containing bands were excised and submitted for tryptic digest MS/MS (Dr Rod Chalk, CMD). Promisingly, the presence of ALG13 and ALG14 in both constructs and in both detergents was detected. It was decided to use DM/CHS as the detergent system for extraction and purification in all downstream ALG13/ALG14 preparations given cholesterol is abundant in human, ER membranes.

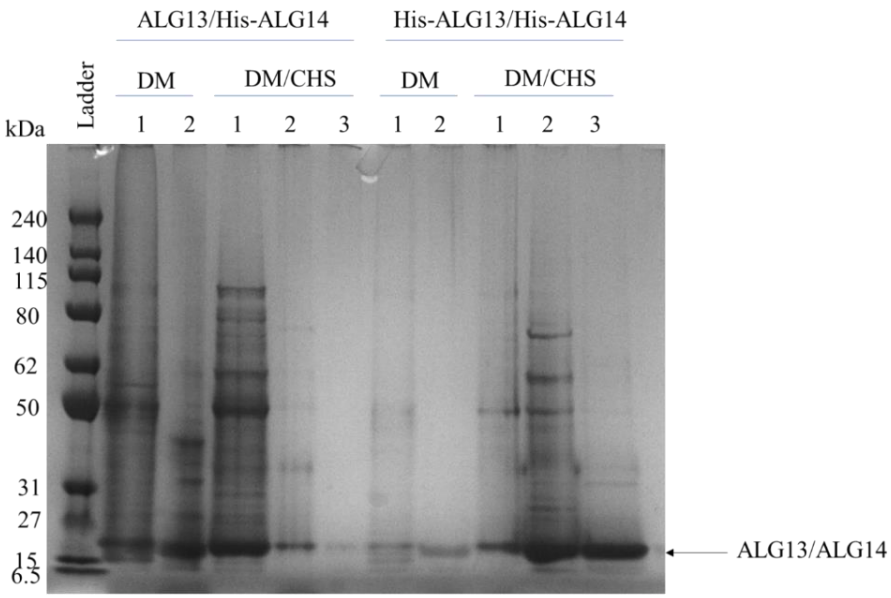


Figure 3.10 Precipitated proteins from the peak fractions from SEC of ALG13/ALG14. ALG13 untagged (left) or 6xHis tagged (right) was co-expressed with His-ALG14 and purified in buffers containing DM or DM/CHS, as shown in Figures 3.8 and 3.9 respectively. Proteins from the each of the peaks seen eluting from SEC were precipitated using TCA. Peak number corresponding to those labelled in Figures 3.8 and 3.9 are shown above each lane. ALG13 and ALG14 bands labelled, as confirmed by tryptic digest MS/MS.

In an attempt to improve the purity of ALG13/ALG14 proteins loaded onto SEC, a reverse purification step following TEV protease cleavage of the 6xHis tag was trialled. Additionally, an imidazole gradient from 100-250 mM was applied to elute proteins bound to the Talon resin during IMAC, to again try and reduce background impurities. Both MultiBac vectors were trialled as before - untagged ALG13 or 6xHis tagged ALG13 with 6xHis tagged ALG14 - as shown in Figures 3.11 and 3.12. The reverse purification was not successful, with target proteins continuing to bind to Talon resin following TEV protease incubation. This is perhaps not surprising given the inherent 'sticky-ness' ALG13 displayed to Talon resin when a truncated, non-tagged version was expressed. Additionally, the imidazole gradient was not successful in reducing the levels of impurities, with ALG13 and ALG14 seen in all fractions across the 100-250 mM gradient. All elution fractions were therefore pooled and concentrated before downstream purification by SEC. Again, a number of peaks were seen for both ALG13/ALG14 constructs. A TCA precipitation was performed prior to analysis given low amounts of protein in each SEC fraction, and the target proteins were seen in the peak following the void in both cases.

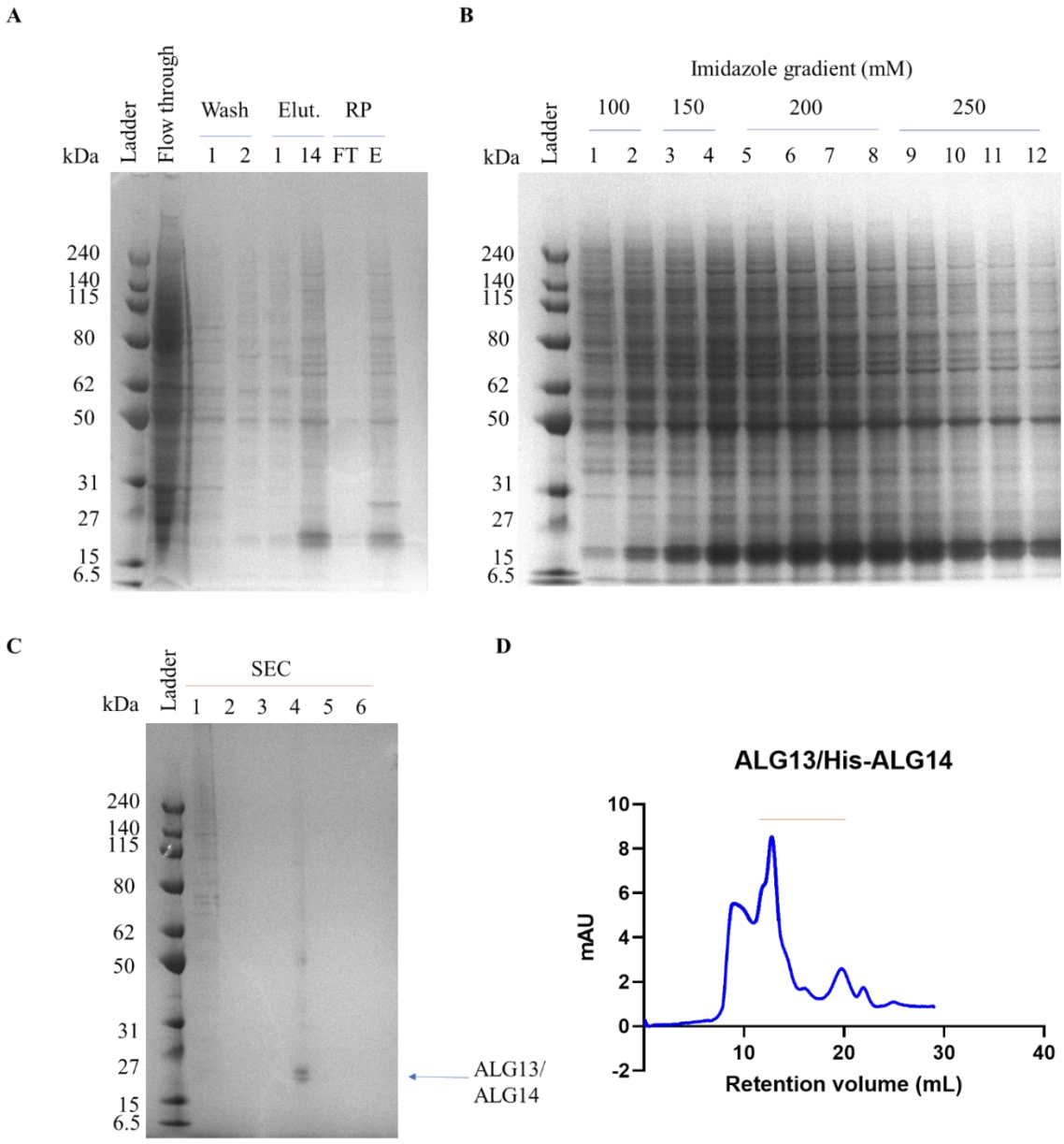


Figure 3.11 Purification of ALG13/ALG14-His. ALG14-His and ALG13 were purified from 1 L baculovirus infected Sf9 cells. **A** Samples from across IMAC, where FT = flow through after Talon resin binding, Wash = protein content following washing twice in 10 mM imidazole, Elut. = elution fractions 1 and 14 corresponding to those from an imidazole gradient shown in **B**. RP = reverse purification following overnight incubation with TEV protease, with E = protein eluted using 250 mM imidazole. **C** TCA precipitated proteins from elution fractions from SEC using a Superdex 200 Increase 10/300 column, where the orange line corresponds to that above the UV spectra at 280 nm, **D**.

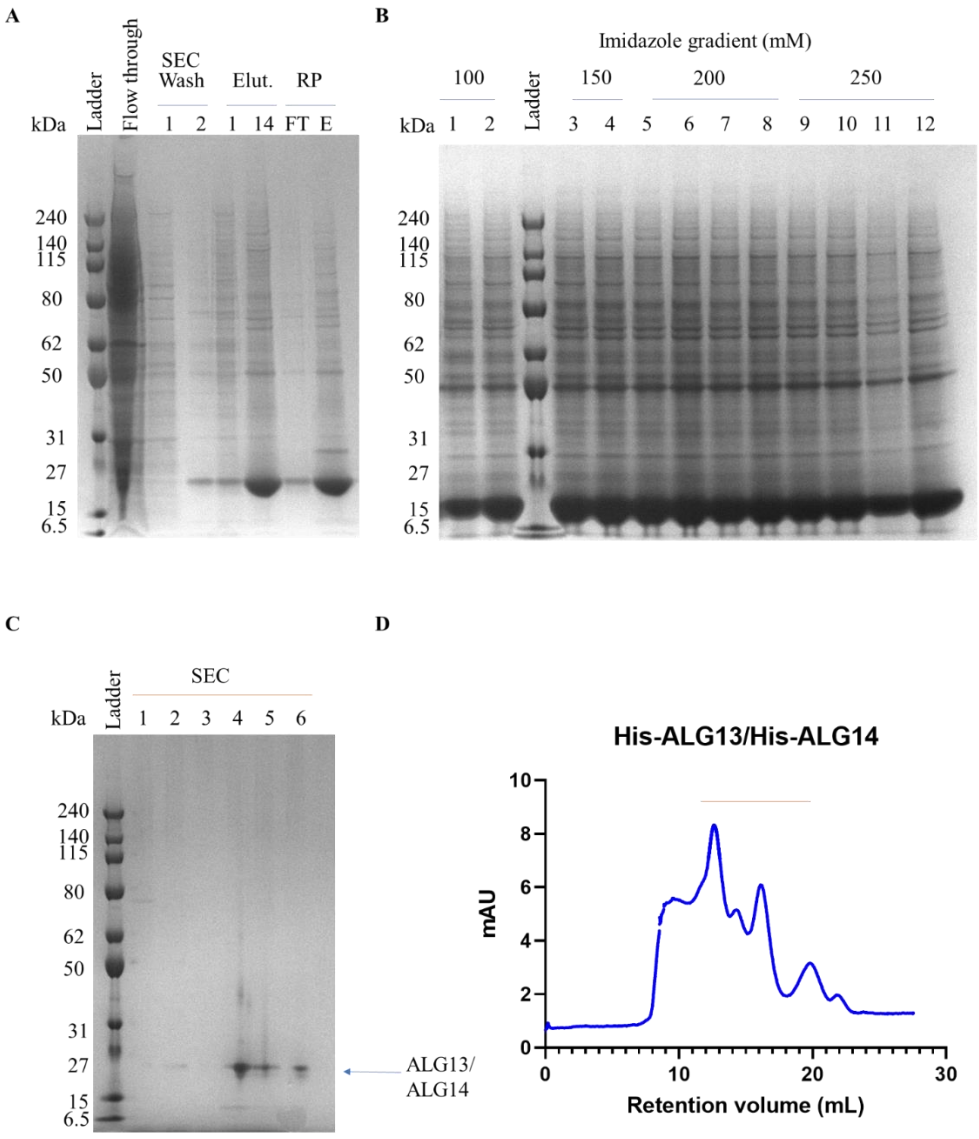


Figure 3.12 Purification of ALG14 and ALG13 with N-term 6xHis tag. ALG14-His and ALG13-His were purified from 1 L baculovirus infected Sf9 cells. **A** Samples from across IMAC, where FT = flow through after Talon resin binding, Wash = protein content following washing twice in 10 mM imidazole, Elut. = elution fractions 1 and 14 corresponding to those from an imidazole gradient shown in **B**. RP = reverse purification following overnight incubation with TEV protease, with E = protein eluted using 250 mM imidazole. **C** Precipitated proteins from elution fractions from SEC using a Superdex 200 Increase 10/300 column, where the orange line corresponds to that above the UV spectra at 280 nm, **D**.

The most significant protocol change that improved the purity of ALG13/ALG14 preparation was the introduction of a membrane extraction to exclude all soluble components of the insect cells before IMAC. Example purifications from 4 L of insect cells for each construct, are shown in Figure 3.13. Insect cell membranes were prepared by ultracentrifugation of the homogenised cell lysate at 100,000 x g for 1 hour, and the supernatant containing soluble proteins was discarded. The membrane containing pellet was washed in a buffer containing high salt (1 M NaCl) which helped remove large amounts of membrane-associated impurities (Figure 3.13A). Once the membranes were solubilised in DM/CHS, the proteins were subject to IMAC as before but with an additional high salt wash. Bound proteins were eluted from the resin in 250 mM imidazole (Figure 3.13B) and peak fractions (6-8 for ALG13/His-ALG14, and 5-8 for His-ALG13/His-ALG14) were pooled and loaded onto a Superdex 200 Increase column as before. These additional steps greatly improved the profiles following size exclusion chromatography, with more protein seen in the second peak than the void (Figure 3.13**C**), although significant overlap of ALG13/ALG14 and the void peak remained. 0.16 and 0.08 mg His-ALG13/His-ALG14 and ALG13/His-ALG14 respectively were prepared for downstream characterisation, giving 39.51 and 19.08 ng per original L of Sf9.

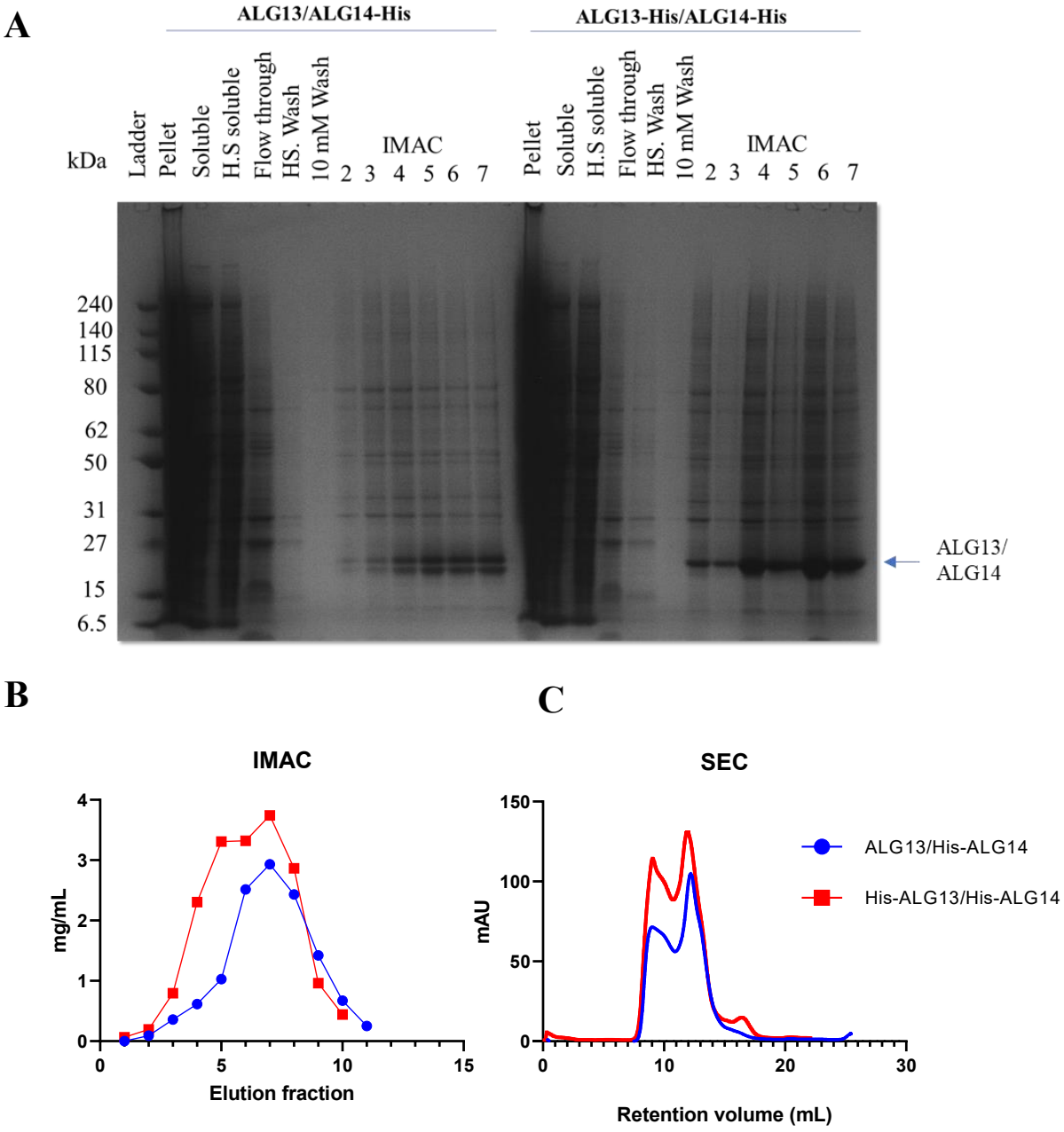


Figure 3.13 Purification of ALG13 tagged and untagged with ALG14 protein from 4 L Sf9 cell culture. A SDS-PAGE of protein content in samples taken across the membrane preparation and IMAC. **B** Protein content across elution fractions from IMAC and SEC for each construct, using a Superdex 200 Increase column.

To further improve the purity and yield of ALG13/ALG14, an extended membrane preparation was trialled. During this protocol, membranes were prepared from cell lysate using a hypotonic extraction buffer, washed twice in hypotonic buffer and then twice in a hypertonic buffer as before. The proteins were then exchanged into an isotonic buffer for further purification by IMAC and SEC. A Superose 6 Increase 10/300 column was used to see if better resolution could be achieved during size exclusion chromatography than with Superdex 200. The results of large-scale purification of the double-His tagged construct and that with untagged ALG13 are shown in Figures 3.14 and 3.15 respectively. Excitingly, purity was hugely improved with a single tight peak seen in the size exclusion spectra (Figure 3.154), with just a small leading shoulder remaining in the spectra. Peak fractions were pooled and concentrated, giving overall yields of 3.12 mg for His-ALG13/His-ALG14 (0.63 mg/ L of Sf9 cells) and 1.88 mg for ALG13/His-ALG13 (0.29 mg/ L of Sf9 cells).

The ALG13/His-ALG14 sample was concentrated to 10 mg/mL and dispensed into plates containing a matrix of crystallisation conditions using MemGold2 (Molecular Dimensions). Plates were incubated at 20 °C and 4 °C and monitored for signs of crystalline protein. Unfortunately, crystallisation of ALG13/His-ALG14 did not occur.

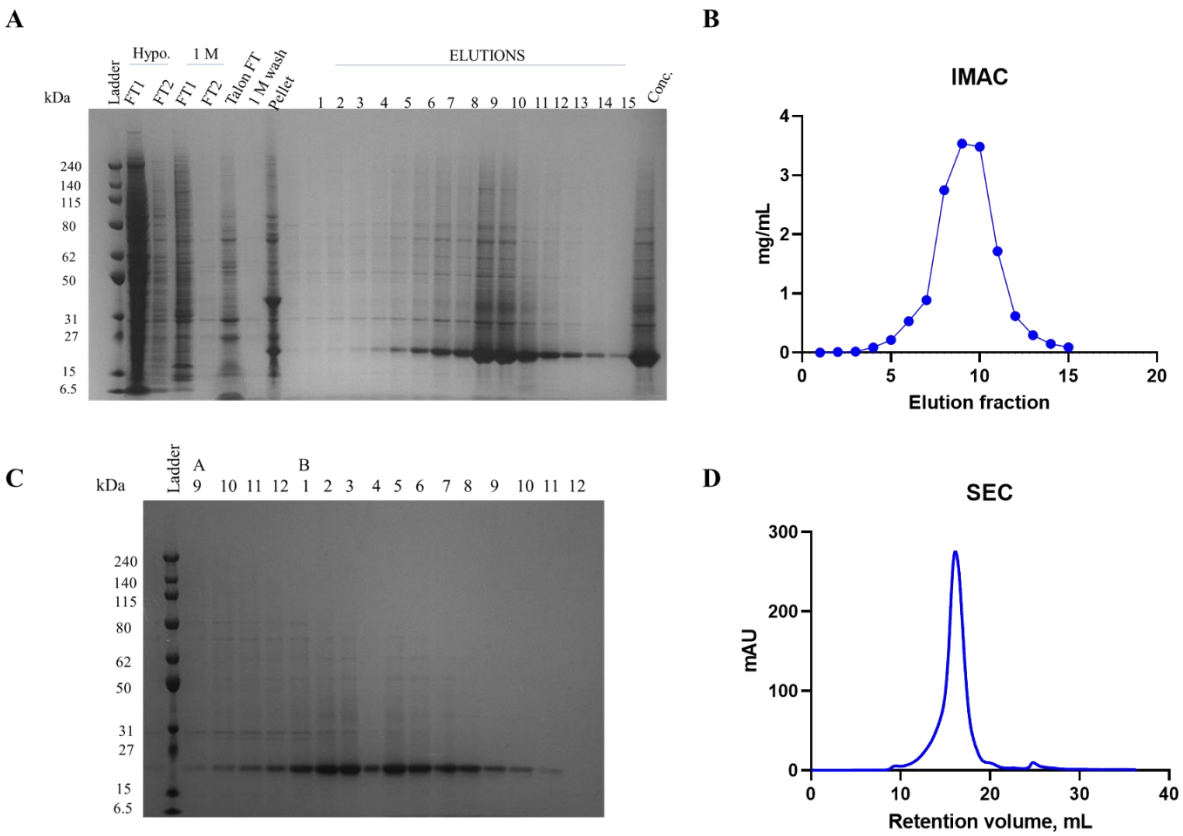


Figure 3.14 Extended membrane preparation for purification of His-ALG13 /His-ALG14. A SDS-PAGE of protein content in samples taken across the membrane preparation and IMAC, where Hypo.= hypotonic extraction buffer, with 0 M NaCl, 1M = 1 M NaCl, hypertonic extraction buffer, FT = flow through, Conc. = concentrated protein from pooled elution fractions 8-11 before loading onto size exclusion column. **B** Protein content across elution fractions from IMAC. **C** SDS-PAGE of His-ALG13/His-ALG14 of elution fractions from Superose 6 Increase 10/300 column, **D**. N.b Low volume of B4 loaded compared to other lanes and hence impurities not visible.

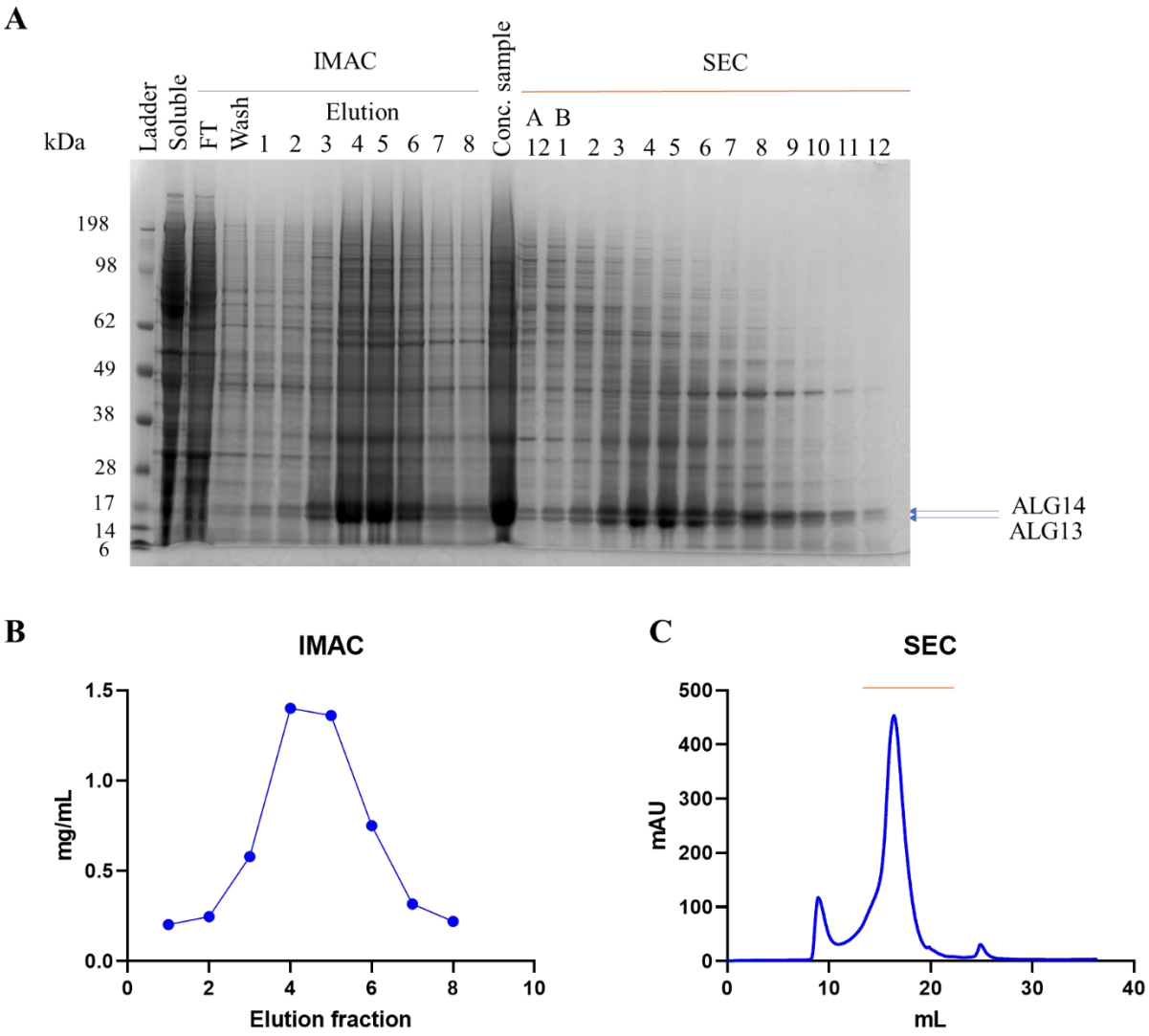
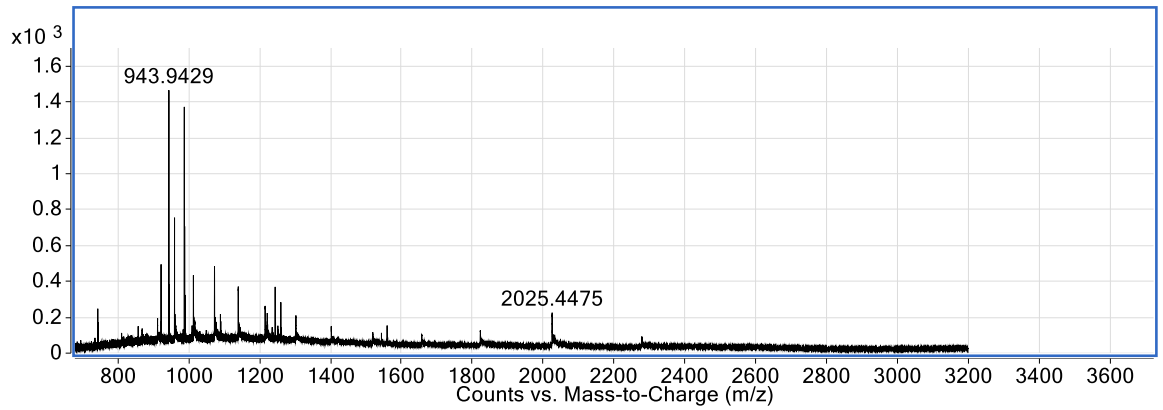


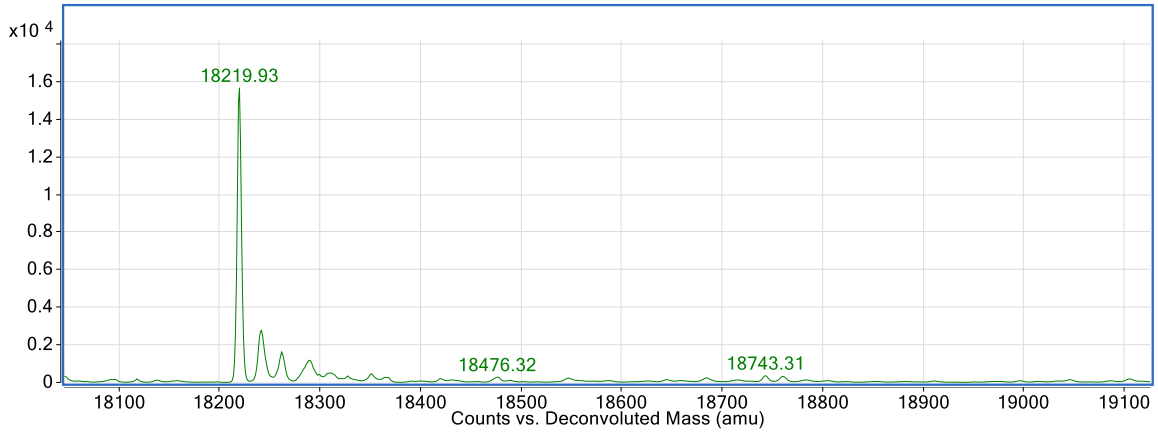
Figure 3.15 Purification from 6.6 L of Sf9 insect cells expressing ALG13/His-ALG14. **A** Purification progress from membrane preparation, IMAC and SEC, where Soluble = supernatant from discarded following membrane preparation, FT = flow through following binding to Talon resin, Conc. sample = elution fractions 3-6 pooled and concentrated to 1.2 mL at 2.4 mg/mL (as shown in **B)** that was loaded onto the Superose 6 Increase column where SEC elution across the peak shown, corresponding to the trace in panel **C**.

3.2.3 Characterisation of ALG13/ALG14

Both ALG13/ALG14-His and ALG13-His/ALG14-His proteins were extensively analysed to confirm the identity, protein stability, purity and functional activity of stocks before downstream use in structural and mechanistic studies. In each case, characterisation was carried out on freshly prepared protein.

To confirm that full length ALG13 and ALG14 had been successfully purified, both ALG13/His-ALG14 and His-ALG13/His-ALG14 samples were analysed by denaturing intact mass spectrometry using liquid chromatography (LC) and electrospray ionisation time of flight (ESI-TOF). ALG13 in its untagged form was identified at 18,219.9 Da (Figure 3.16) and with an N-terminal 6xHis for the second construct at 20,740.7 Da (Figure 3.17) despite low signal of protein. 6xHis tagged ALG14 was identified in both samples at 26,625.5 Da (Figure 3.18). Interestingly, these masses match to oxidised forms of ALG13 and ALG14, where all cysteine pairs are cystines forming disulphide bonds.

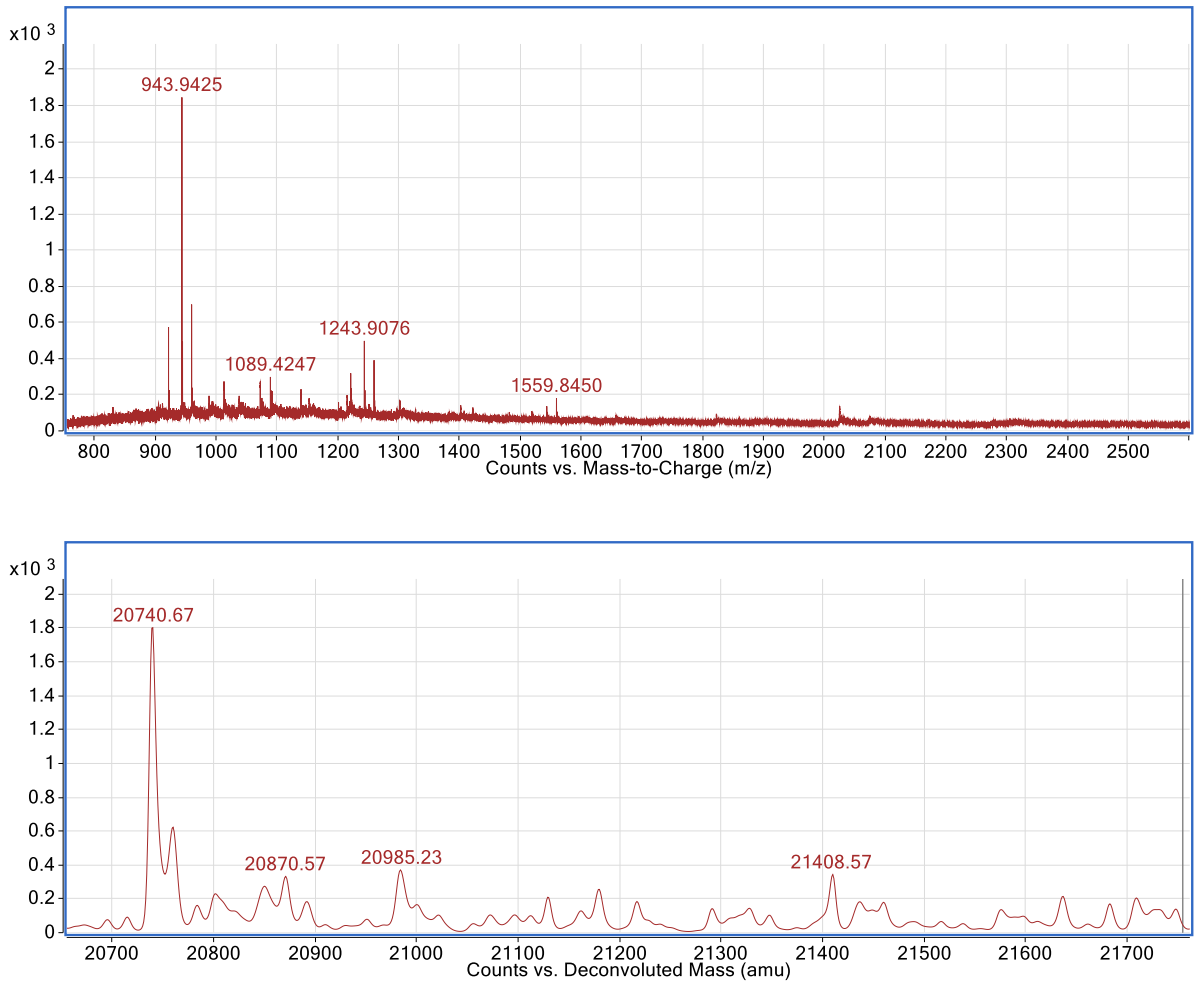




M[1-165]K = 18219.1

1	M	K	C	V	F	V	T	V	G	T	T	S	F	D	D	L	I	A	C	V	S	A	P	D	S	L	Q	K	I	E	30
31	S	L	G	Y	N	R	L	I	L	Q	I	G	R	G	T	V	V	P	E	P	F	S	T	E	S	F	T	L	D	V	60
61	Y	R	Y	K	D	S	L	K	E	D	I	Q	K	A	D	L	V	I	S	H	A	G	A	G	S	C	L	E	T	L	90
91	E	K	G	K	P	L	V	V	V	I	N	E	K	L	M	N	N	H	Q	L	E	L	A	K	Q	L	H	K	E	G	120
121	H	L	F	Y	C	T	C	S	T	L	P	G	L	L	Q	S	M	D	L	S	T	L	K	C	Y	P	P	G	Q	P	150
151	E	K	F	S	A	F	L	D	K	V	V	G	L	Q	K																165

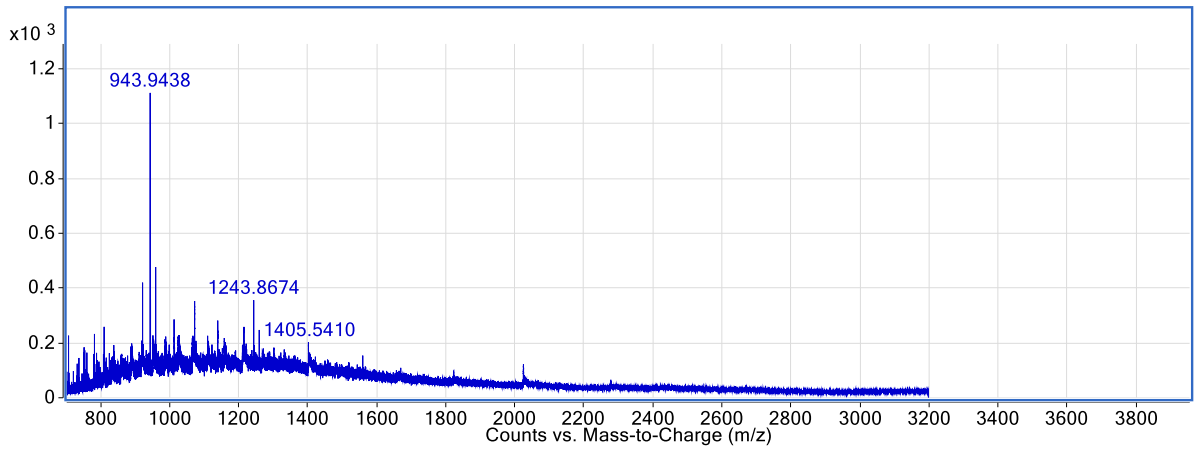
Figure 3.16 Intact denaturing MS of untagged ALG13. (Top) The ion series between 800-3600 (*m/z*) with corresponding deconvoluted mass of protein (middle). The experimentally determined mass of 18,219.9 Da matches that of full-length protein in its oxidised form, with sequence shown below.

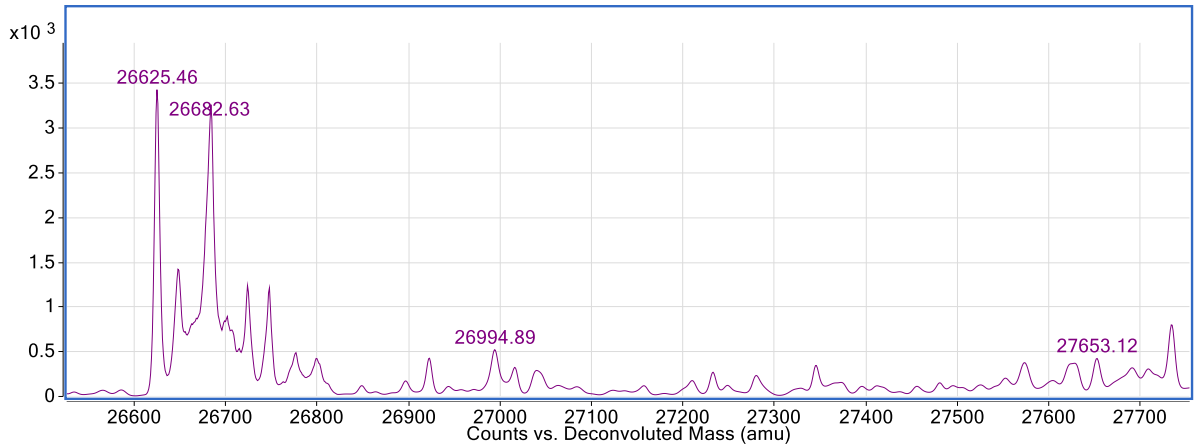


M[1-187]Q = 20742.7

1	M	G	H	H	H	H	H	H	S	S	G	V	D	L	G	T	E	N	L	Y	F	Q	S	M	K	C	V	F	V	T	30
31	V	G	T	T	S	F	D	D	L	I	A	C	V	S	A	P	D	S	L	Q	K	I	E	S	L	G	Y	N	R	L	60
61	I	L	Q	I	G	R	G	T	V	V	P	E	P	F	S	T	E	S	F	T	L	D	V	Y	R	Y	K	D	S	L	90
91	K	E	D	I	Q	K	A	D	L	V	I	S	H	A	G	A	G	S	C	L	E	T	L	E	K	G	K	P	L	V	120
121	V	V	I	N	E	K	L	M	N	N	H	Q	L	E	L	A	K	Q	L	H	K	E	G	H	L	F	Y	C	T	C	150
151	S	T	L	P	G	L	L	Q	S	M	D	L	S	T	L	K	C	Y	P	P	G	Q	P	E	K	F	S	A	F	L	180
181	D	K	V	V	G	L	Q	K																							188

Figure 3.17 Intact denaturing MS of His-ALG13. (Top) The ion series between 800-3600 (*m/z*) with corresponding deconvoluted mass of protein (middle). The experimentally determined mass of 20,740.7 Da matches that of full-length protein in its oxidised form, with sequence shown below.





G[2-239]V = 26624.0

1	M	G	H	H	H	H	H	H	S	S	G	V	D	L	G	T	E	N	L	Y	F	Q	S	M	V	C	V	L	V	L	30
31	A	A	A	A	G	A	V	A	V	F	L	I	L	R	I	W	V	V	L	R	S	M	D	V	T	P	R	E	S	L	60
61	S	I	L	V	V	A	G	S	G	G	H	T	T	E	I	L	R	L	L	G	S	L	S	N	A	Y	S	P	R	H	90
91	Y	V	I	A	D	T	D	E	M	S	A	N	K	I	N	S	F	E	L	D	R	A	D	R	D	P	S	N	M	Y	120
121	T	K	Y	Y	I	H	R	I	P	R	S	R	E	V	Q	Q	S	W	P	S	T	V	F	T	T	L	H	S	M	W	150
151	L	S	F	P	L	I	H	R	V	K	P	D	L	V	L	C	N	G	P	G	T	C	V	P	I	C	V	S	A	L	180
181	L	L	G	I	L	G	I	K	K	V	I	I	V	Y	V	E	S	I	C	R	V	E	T	L	S	M	S	G	K	I	210
211	L	F	H	L	S	D	Y	F	I	V	Q	W	P	A	L	K	E	K	Y	P	K	S	V	Y	L	G	R	I	V		239

Figure 3.18 Intact denaturing MS of His-ALG14. (Top) The ion series between 800-3600 (*m/z*) with corresponding deconvoluted mass of protein (middle). The experimentally determined mass of 26,625.6 Da corresponds to full length ALG14 in an oxidised form, with loss of the initiator methionine before the 6xHis tag. Mass of 26,682.6 Da corresponds to protein with Co^{2+} resin bound.

Next, to confirm that ALG13 and ALG14 were in complex rather than co-purified, proteins were subject to analytical SEC-multi angle light scattering (MALS) to determine the molecular weight of the protein in detergent micelles. Only the His-ALG13/His-ALG14 sample was analysed in this way (Figure 3.19). The protein stock was cleared by ultracentrifugation at 150,000 x g, 4 °C for 40 minutes to remove any precipitants that might confound results. Using the dN/dC value calculated for buffer containing DM/CHS (see methods section for details), the first peak seen in the UV trace after the void (data in red, Figure 3.19) was found to contain matter at 71,354 Da, with protein content contributing 45,381 Da. This confirms the protein exists in a 1:1 His-ALG13/His-ALG14 complex, given the theoretical mass of 47,566.1 Da. The second peak corresponded to free detergent micelles of DM/CHS, and was found to be 40,807 Da which matches the 40 kDa size reported for DM only micelles [117].

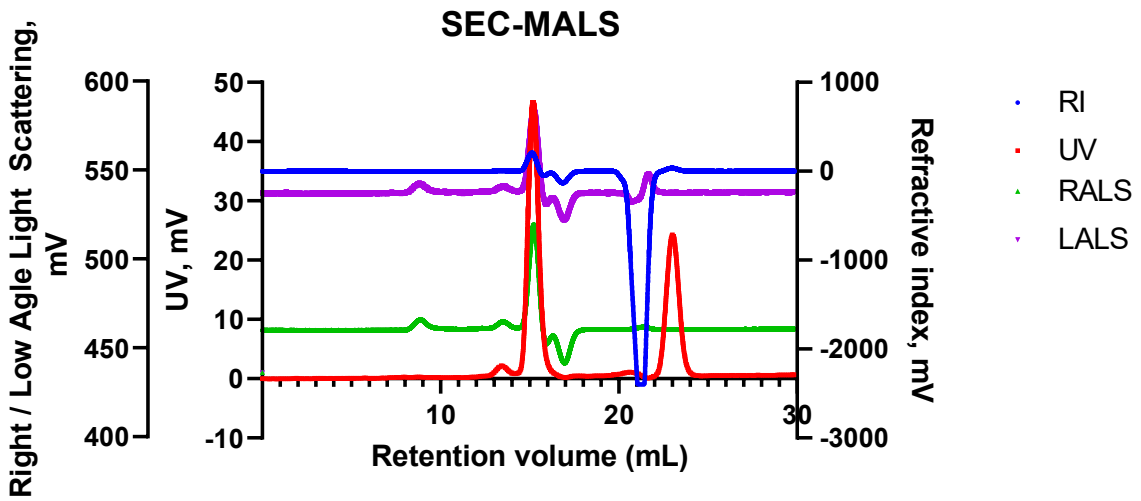


Figure 3.19 ALG13/ALG14 are in a 1:1 complex in the DM/CHS micelle. Light scattering of ALG13/ALG14 in DM/CHS was used to determine the absolute mass of the particles in solution, where RI = refractive index, UV = ultra violet, RALS = right angled light scattering, LALS = low angled light scattering.

To further the characterisation of ALG13/ALG14, the thermostability of the proteins were measured by nano-differential scattering fluorimetry (nano-DSF). The ratio of emission by tryptophan residues at 350 and 330 nm, was used to measure protein unfolding when residues become exposed to the solvent. Each sample was taken into capillary tubes in triplicate, and the temperature increased from 20 °C to 95 °C by 1 degree per minute. Melting temperatures of 47.4 °C for ALG13/His-ALG14 and 52.0 °C for His-ALG13/His-ALG14 were recorded (Figure 3.20, Table 3.2). Changes in scattering with respect to the buffer control was used as an estimate of the temperature that leads protein precipitation. For His-ALG13/His-ALG14 this was 48.5 °C, and 44.7 °C for ALG13/His-ALG14. The untagged ALG13 construct had lower unfolding and precipitation temperatures, suggesting it is a less stable complex and more at risk of denaturation under stress than that bearing the 6xHis tag.

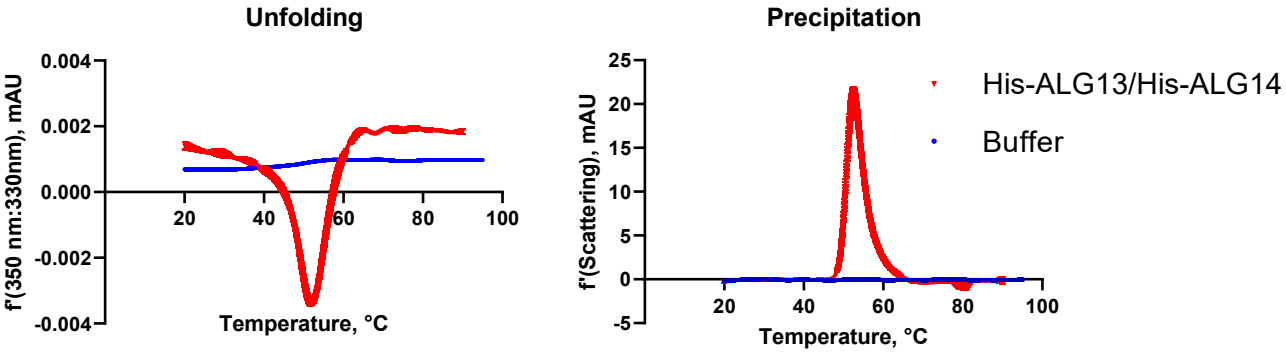


Figure 3.20 Thermostability of His-ALG13/His-ALG14. (Left) Protein unfolding shown by the first derivative of the ratio of fluorescence 350 nm: 330 nm ratio when samples were increased by 2 °C per minute from 20 to 95 °C, used to determine the melting temperature. (Right) The first derivative of the scattering mAU (at Prometheus proprietary wavelengths), where changes with respect to buffer controls are used to estimate the onset of protein precipitation.

Table 3.2 Summary of melting temperature and onset of precipitation of ALG13/ALG14

Sample	Melting temperature, T_M, °C	Precipitation onset, °C
ALG13/His-ALG14	47.4	44.7
His-ALG13/His-ALG14	52.0	48.5

Finally, the enzymatic activity of ALG13/His-ALG14 was tested to identify whether the proteins were correctly folded and in a complex. Given the low thermostability of ALG13/ALG14, this was done both before and after snap freezing in liquid N_2 to ensure long term storage of protein aliquots at -80 °C was possible. Figure 3.21 shows the activity of ALG13/His-ALG14, where samples straight from purification by SEC were either kept on ice or snap frozen in liquid N_2 and thawed immediately. Both aliquots were tested for activity by incubating with 50 µM dolichol pyrophosphate GlcNAc (supplied by Dr Andrew Giltrap) and 100 µM UDP-[^{14}C]-GlcNAc for 30 minutes. End-point activity was measured using TLC as described in Chapter 5. Excitingly, the enzyme was found to be active both fresh and after freeze-thawing, meaning it could be stored and was in an active conformation when solubilised in DM/CHS. This also indicated ALG13 and ALG14 were in complex given neither protein bares enzymatic activity when alone [29] . An enzymes specific activity can be used as a measure of purity, where µmole of product produced per minute is reported per mg of enzyme stock used. The freshly purified stock had a specific activity of 0.073 µmole product min^{-1} mg^{-1} , and the frozen stock at 0.071 µmole product min^{-1} mg^{-1}.

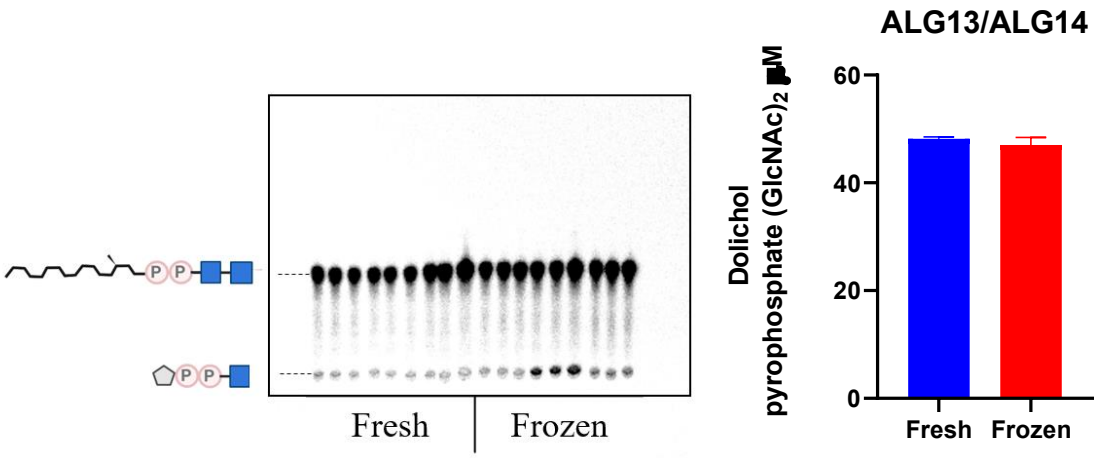


Figure 3.21 Freeze-thaw activity testing of ALG13/His-ALG14, where radioactivity detected on the TLC plate showing GlcNAc transfer from soluble UDP-[^{14}C]-GlcNAc to dolichol pyrophosphate GlcNAc is on the left. The activity represented in µM of product formed after 1 hour is on the right.

3.2.4 DPAGT1/ALG13/ALG14 purification

As discussed, it has been suggested that in yeast ALG7 (the DPAGT1 homolog) forms a multi-glycosyltransferase complex with ALG13/ALG14 [27, 103, 131]. To test whether this was analogous for the human proteins, purifications of DPAGT1/ALG13/ALG14 in complex were attempted. Adding mass to the small ALG13/ALG14 protein complex would also make structural elucidation by cryo-EM amenable.

AlphaFold-Multimer was also used to predict the structure of the protein complex formed between DPAGT1, ALG13 and ALG14 (Figure 3.22) [122]. The modelled DPAGT1 structure is consistent with that published from crystallography data as expected given this would be available in the training set for the model [44]. However, the membrane orientation of ALG13/ALG14 is likely incorrect as the protein complex is known to act on the cytoplasm

leaflet. Therefore, experimental determination of the structure of the three proteins in complex is needed.

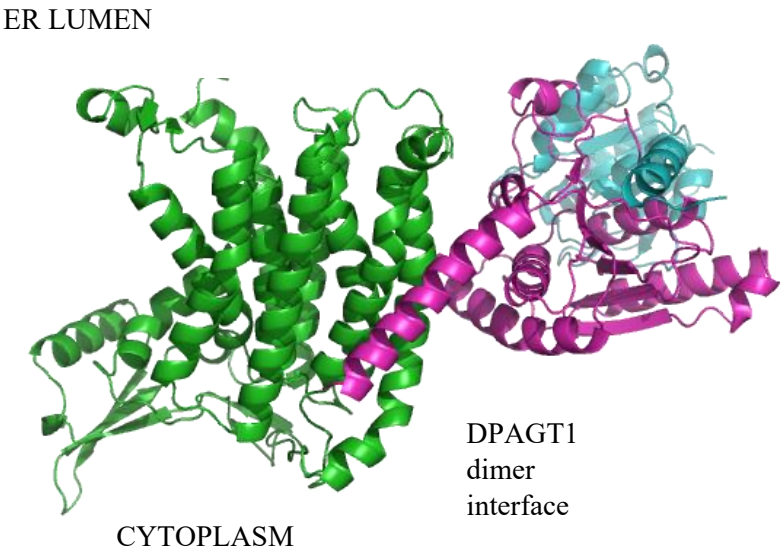


Figure 3.22 AlphaFold-Multimer [122] predicted structure of DPAGT1 (green), ALG14 (pink) and ALG13 (blue). Labels indicate likely orientation in the ER membrane based on experimentally determined DPAGT1 structure [44].

Initially, both ALG13/ALG14 constructs characterised in section 3.2.1 were tested for expression with DPAGT1. The cDNA sequence encoding full-length DPAGT1 with an N-terminal 6xHis tag (His-DPAGT1), was cloned into a pFB-LIC Bse expression vector by Jing Wang. Baculovirus were produced by transformation of vectors into DH10Bac cells (see Materials and Methods). To test the optimal ratios of expression vector needed to successfully co-purify the complex, a small-scale screen was performed whereby 3 mL Sf9 cells at 2 x 10^6 cells/mL were infected with 15 µL P2 virus' encoding ALG13/ALG14 and DPAGT1 in volume ratios of 4:1, 2:1, 1:1, 1:2, 1:4. It was clear that using a 4:1 ratio of ALG13/ALG14 P2:

DPAGT1 P2 virus gave the highest intensity expression of all three proteins (data not shown), and so this was used for co-expressing the proteins.

Using the results of this test expression, the buffer tolerance of the complex was tested. The same buffer used for previous ALG13/ALG14 purifications containing DM/CHS as detergents and 50 mM HEPES pH 7.5, 2 mM TCEP, 5 % glycerol and 500 mM NaCl, was compared to the reported buffer conditions used for DPAGT1 structural study [44] , containing octyl glucose neopentyl glycol (OGNG) with CHS and also cardiolipin (CDL) in 50 mM HEPES pH 7.5, 5 mM $MgCl_2$, 2 mM TCEP, 5 % glycerol and 200 mM NaCl.

The same extended membrane protocol as described above for ALG13/ALG14 alone was used to prepare the membrane fractions of 1 L Sf9 expressing DPAGT1-His and either ALG13/His-ALG14 or His-ALG13/His-ALG14. For both constructs, purification was performed in buffers containing either DM/CHS or OGNG/CHS/CDL. The complexes were subject to IMAC with Talon resin and SEC using a Superose 6 Increase 10/300, as summarised in Figures 3.23 and 3.24. Very similar spectra were seen for each construct in the two buffer conditions, suggesting either detergents would be suitable for purification of the complex. DPAGT1, ALG13 and ALG14 at their expected molecular weights were seen by SDS-PAGE, and ALG13/ALG14 in both constructs was confirmed using tryptic digest MS/MS. Presence of DPAGT1 was not confirmed using this method, perhaps due to the protein having only a short soluble region which is predicted to yield just 3 tryptic peptides. The highest yield of complex was achieved with untagged ALG13 and this was therefore used in all later preparations, especially given this leaves ALG13 in its native form.

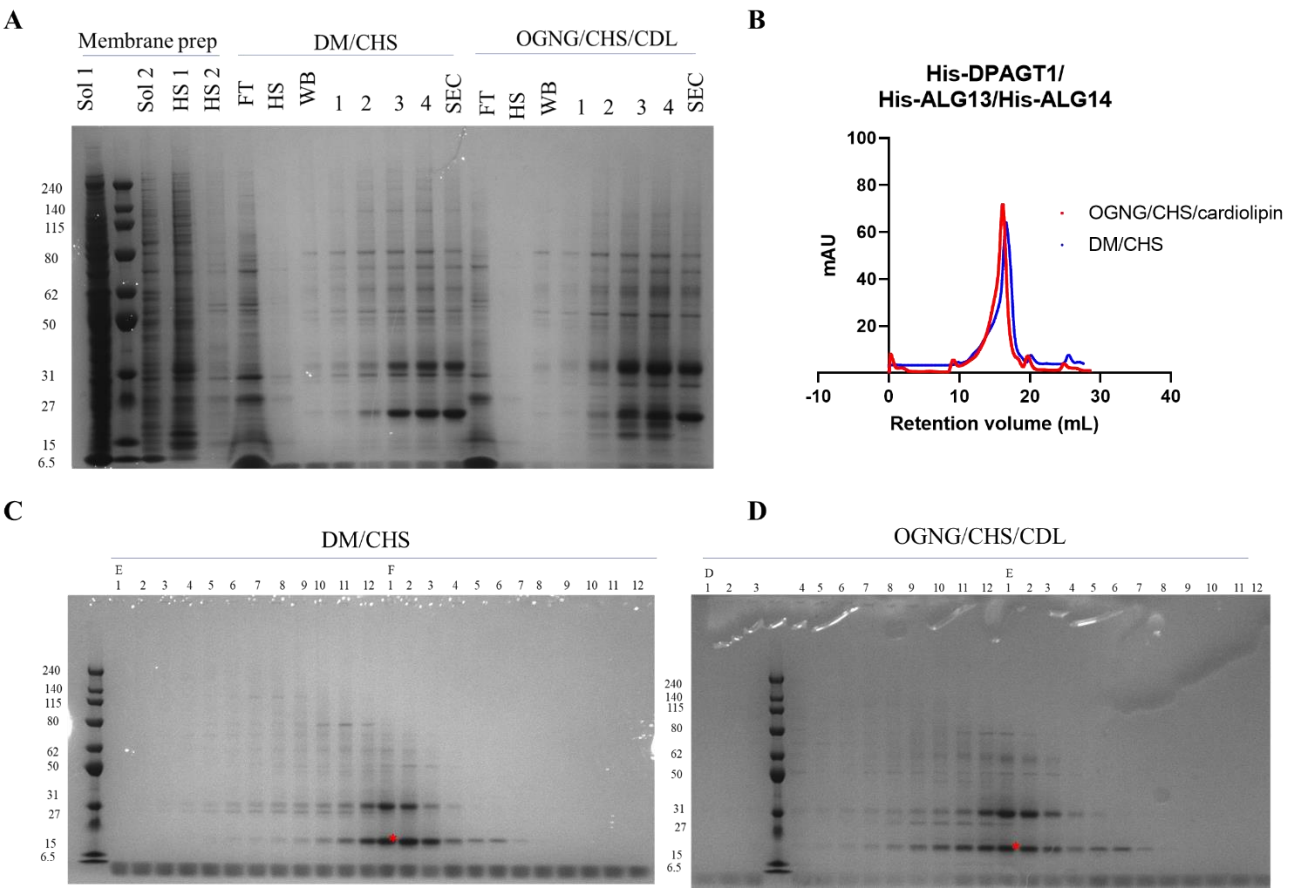


Figure 3.23 Purification of His-DPAGT1/His-ALG13/His-ALG14 using DM/CHS or OGNG/CHS/CDL containing buffers. A SDS-PAGE showing progress across the membrane preparation and IMAC, where Sol 1 and 2 represent wash fractions from hypotonic buffer, and HS 1 and 2, those with hypertonic buffer. FT = flow through from Talon resin, WB = wash buffer, SEC = concentrated samples loaded onto Superose 6 Increase column, with corresponding UV trace in panel **B,** and elution fractions analysed by SDS-PAGE in **C** and **D.** Putative His-DPAGT (48.7 kDa), His-ALG13 (20.8 kDa), and His-ALG14 (26.8 kDa) bands were excised and analysed by tryptic digest MSMS Those labelled with red asterisk where confirmed to be His-ALG13 and ALG14-His.

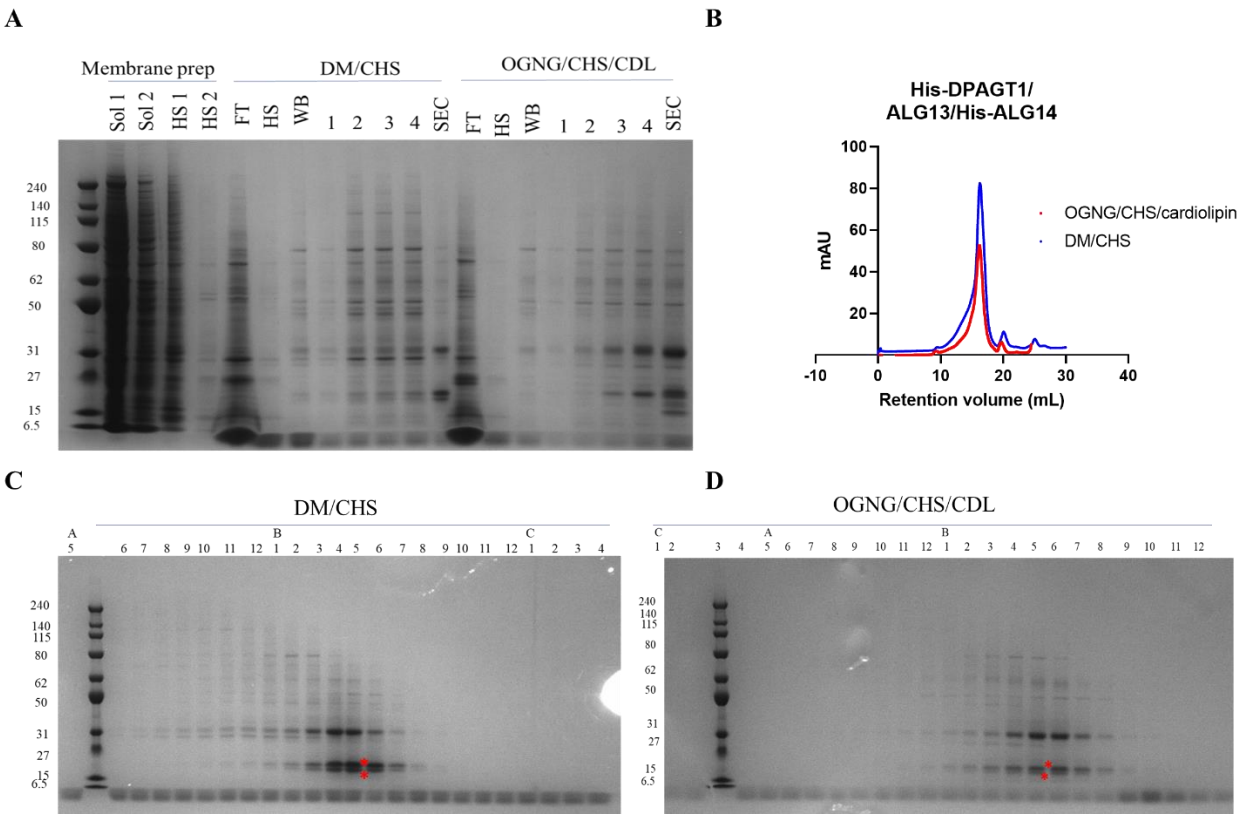


Figure 3.24 Purification of His-DPAGT1/ALG13/His-ALG14 using DM/CHS or OGNG/CHS/CDL containing buffers. A SDS-PAGE showing progress across the membrane preparation and IMAC, where Sol 1 and 2 represent wash fractions from hypotonic buffer, and HS 1 and 2, those with hypertonic buffer. FT = flow through from Talon resin, WB = wash buffer, SEC = concentrated samples loaded onto Superose 6 Increase column, with corresponding UV trace in panel **B,** and elution fractions analysed by SDS-PAGE in **C** and **D.** Putative His-DPAGT (48.7 kDa), ALG13 (18.2 kDa), and His-ALG14 (26.8 kDa) bands were analysed by tryptic digest MSMS, and those labelled with red asterisk were confirmed to be ALG13 and ALG14-His.

Attempts at denaturing intact mass spectrometry of the DPAGT1/ALG13/ALG14 complex were not successful (data not shown), and given tryptic MS/MS was only able to detect ALG13/ALG14 and not DPAGT1, the complex was analysed by Western blot to confirm protein identity. Samples were run alongside preparations of ALG13/ALG14 and DPAGT1

alone. The Coomassie blue stained gel showed three intense bands in the complex at the same molecular weight as those in samples of the enzymes alone (Figure 3.25). DPAGT1 and ALG14 were identified by their 6xHis tags, and ALG13 using a protein specific antibody. Background ALG13/ALG14 - likely endogenous from Sf9 insect cells – was also detected in the DPAGT1 only sample.

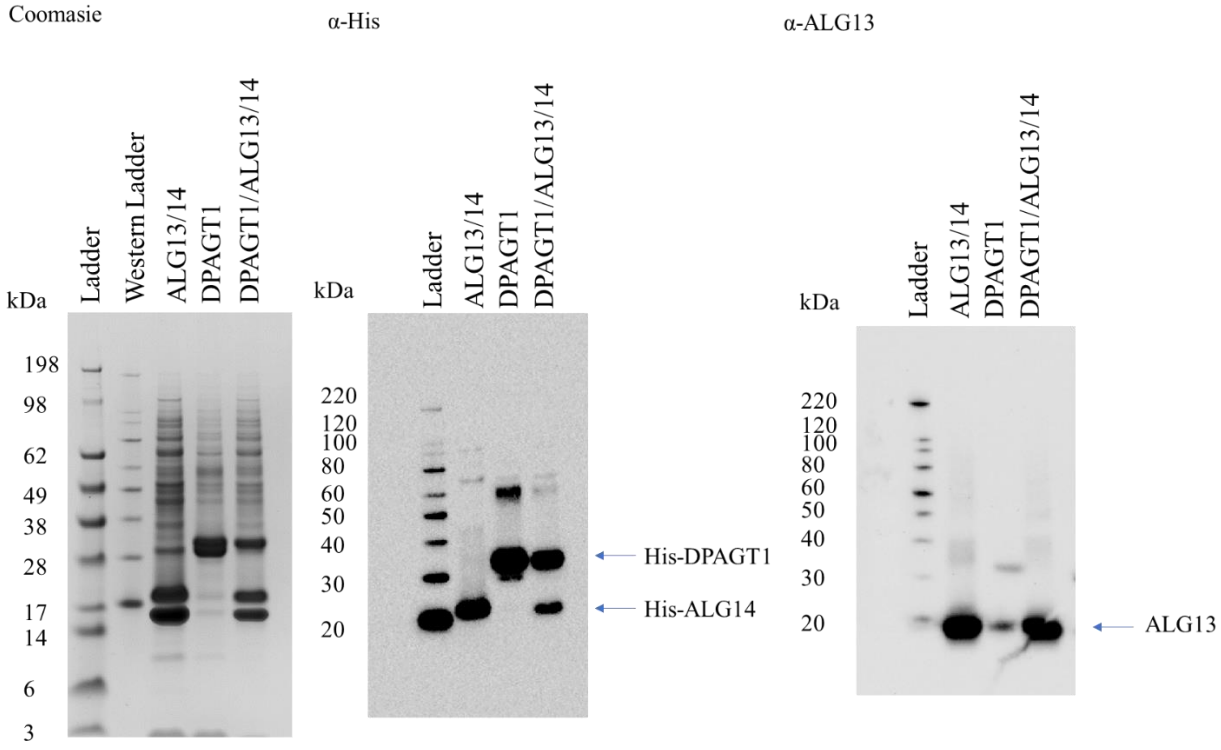


Figure 3.25 Characterisation of the DPAGT1-His/ALG13/His-ALG14 protein complex. Samples of ALG13/ALG14 and protein alone and in complex were separated by SDS-PAGE. Left = Coomassie blue stained gel, middle = anti-His tag Western blot, right = anti-ALG13 Western blot, confirming identity of all three enzymes.

3.2.5 DPAGT1/ALG13/ALG14 complex formation

Following the successful purification of DPAGT1/ALG13/ALG14, efforts were made to elucidate whether DPAGT1 was forming a stable complex with ALG13/ALG14 or simply eluting at the same position during chromatography. First, analytical SEC-MALS was performed to determine the absolute mass of particles in solution. His-DPAGT1/ ALG13/His-ALG14 complex was purified as before from 6 L of infected Sf9 insect cells using DM/CHS detergent for solubilisation (Figure 3.26). In this preparation, there was a shoulder on the right of the peak on SEC, and so two fractions were collected (labelled as i and ii in Figure 3.26), and these were analysed separately by SEC-MALS (Figure 3.26C). Using co-polymer analysis, the samples i and ii were found to contain particles of 139,980 Da and 135,637 Da, of which protein contributed 130,374 Da and 126,546 Da respectively. These are both larger than the theoretical Mw of a 1:1:1 His-DPAGT1: ALG13: His-ALG14 complex at 96,258 Da. However, DPAGT1 has been reported to readily form homo-dimers [44], which have a theoretical mass of 97,398 Da and hence the sample could represent a 2:1:1, DPAGT1:ALG13:ALG14, with theoretical mass of 144,957 Da. However, these results may have been confounded by impurities in the protein sample, and so alternative methods of analysis were sought.

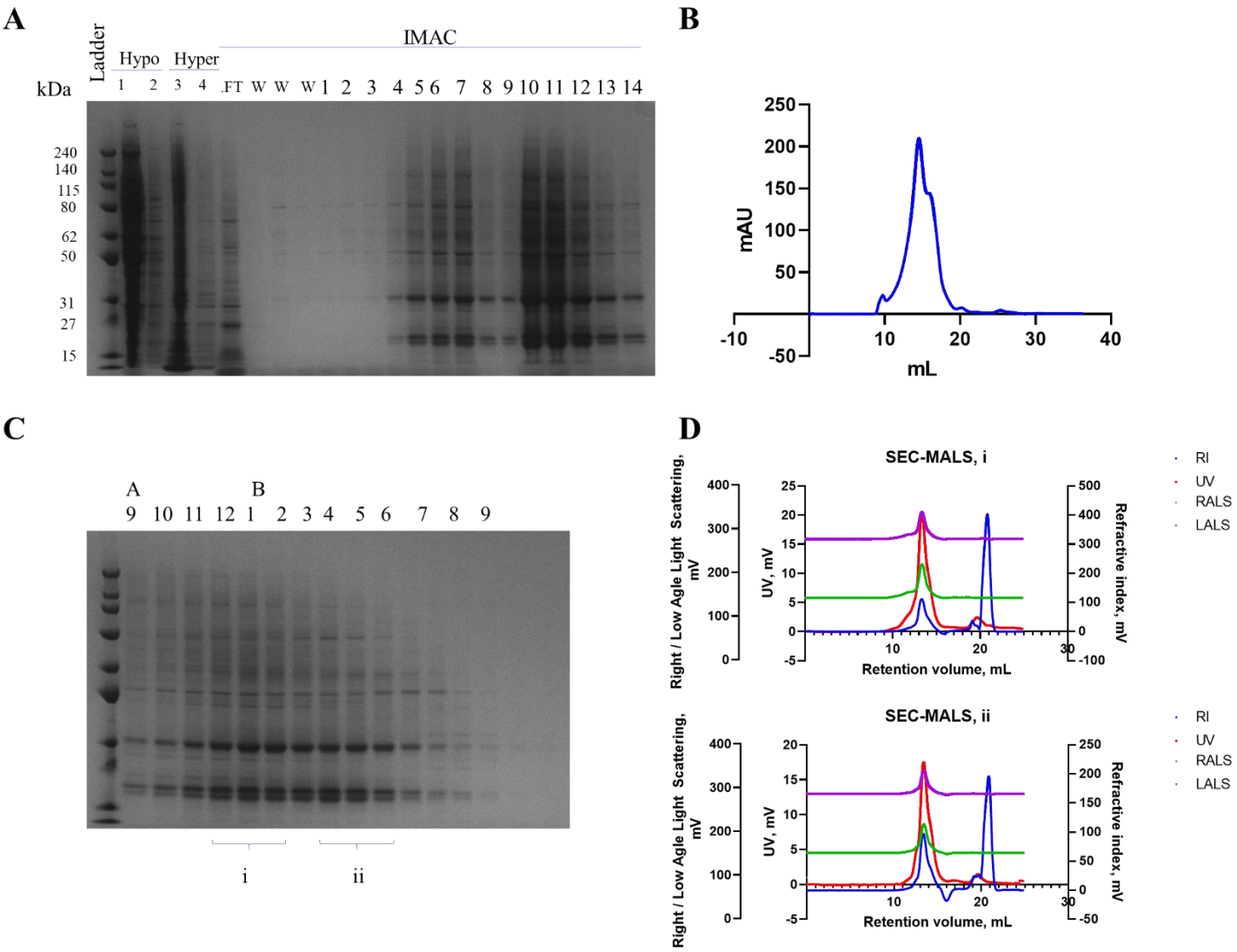


Figure 3.26 Analysis of DPAGT1-His/ALG13/ALG14-His by SEC-MALS. A Membrane preparation and IMAC fractions, with those in lanes 8-11 pooled for further purification by size exclusion chromatography (**B and C**). Fractions from the two peaks present were pooled to give samples i and ii, which were analysed by SEC-MALS shown in **D.**

In all purifications reported previously, both DPAGT1 and ALG13/ALG14 contained 6xHis tag affinity tags and therefore in theory could be independently binding to the Co^{2+} resin during IMAC rather than being in complex. To test this, pull-down assays were performed using proteins expressed with unique purification tags. DPAGT1 was cloned to contain both C-terminal FLAG and 10xHis tags (see Materials and Methods). Promisingly, ALG13 and ALG14 were pulled down using FLAG resin when co-expressed with DPAGT1 and not when expressed alone (Figure 3.27) confirming the proteins must be interacting.

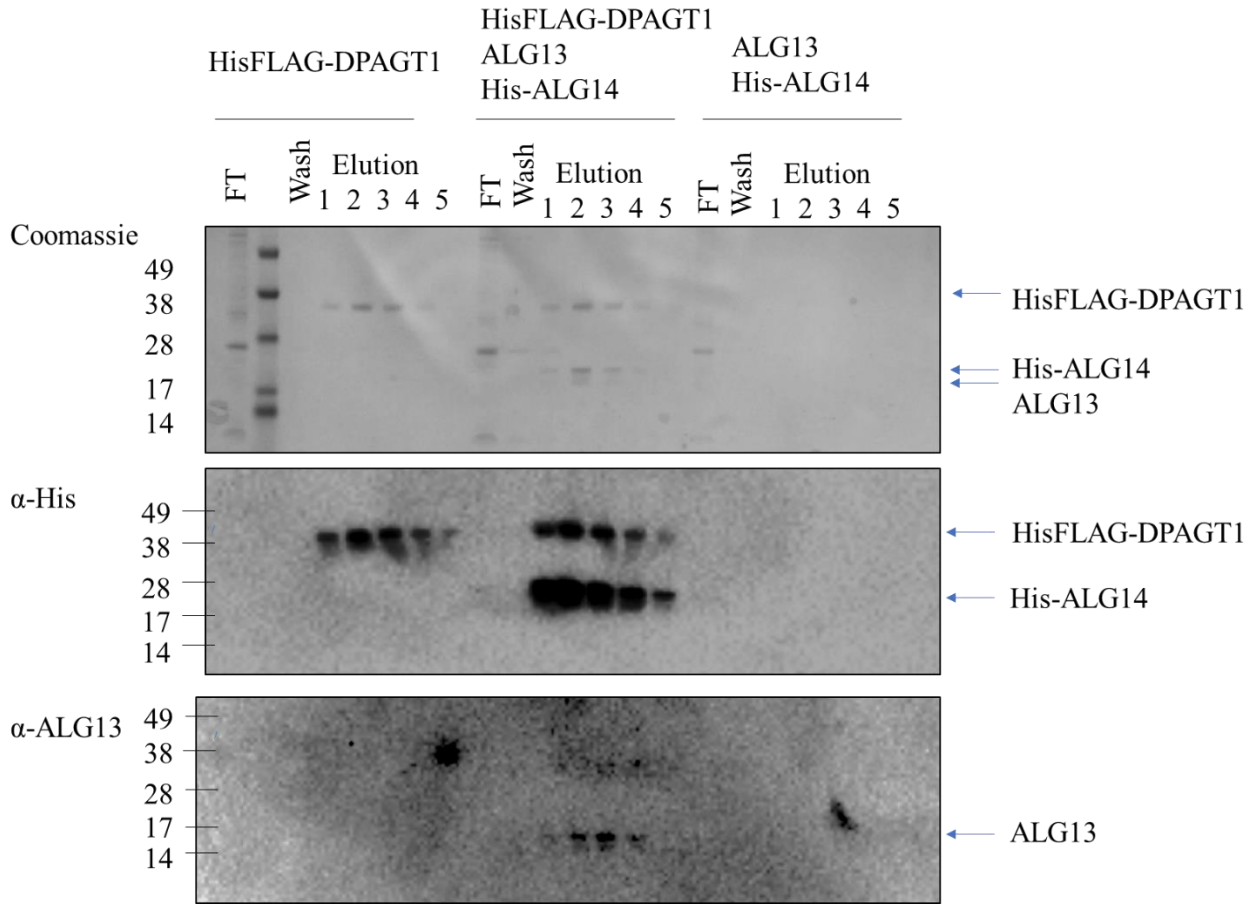


Figure 3.27 Pull down assay: ALG13/ALG14 interacts with DPAGT1-FLAG. Elution fractions from FLAG resin incubated with cell lysate of cells expressing ALG13/His-ALG14 alone or with DPAGT1-HisFLAG. The Coomassie stained gel is shown above Western blots probing for the 6xHis tags of DPAGT1 and ALG14, and for ALG13.

Additional confirmation of a DPAGT1/ALG13/ALG14 complex came in the form of native PAGE. This technique utilises conditions where proteins remain folded and in their 'native' charge under biological conditions, meaning they can be separated based on their overall size and charge. Estimation or prior knowledge of the protein's isoelectric point (pI) is used to adjust the pH of the buffer to allow electrophoretic migration of the native protein; in most cases towards the positive electrode. Following difficulties getting the solubilised membrane protein complex to enter native gels (data not shown), blue native (BN) PAGE was explored. This method utilises the negative charge of Coomassie blue G250 dye to coat the protein without denaturing or interrupting protein-protein interactions, allowing proteins to be separated based on their folded size [146]. This has been reported to be advantageous for membrane proteins as G250 Coomassie dye binds to hydrophobic regions which increases the chance of a protein entering the gel matrix and migrating [119, 146].

It was clear from BN-PAGE that DPAGT1 alone forms many higher oligomeric states in the conditions of these gels, with a ladder effect of the protein seen (lane 1, Figure 3.28A). ALG13/ALG14 ran as two intense bands, likely corresponding to free protein and the complex of the two at the expected 45 kDa (Figure 3.28 B). Promisingly, the DPAGT1-FLAG/ALG13/His-ALG14 complex ran as a single, smeary band by BN-PAGE (Figure 3.28A), disrupting the oligomerisation of DPAGT1 seen when analysed alone. To confirm the presence of all three subunits in this band on BN-PAGE, the gel lanes were excised, denatured and reduced in SDS sample buffer, and run in a second dimension under denaturing conditions whereby the three subunits appeared at their respective molecular weights (Figure 3.28B-D). This again provides indication of a DPAGT1/ALG13/ALG14 complex.

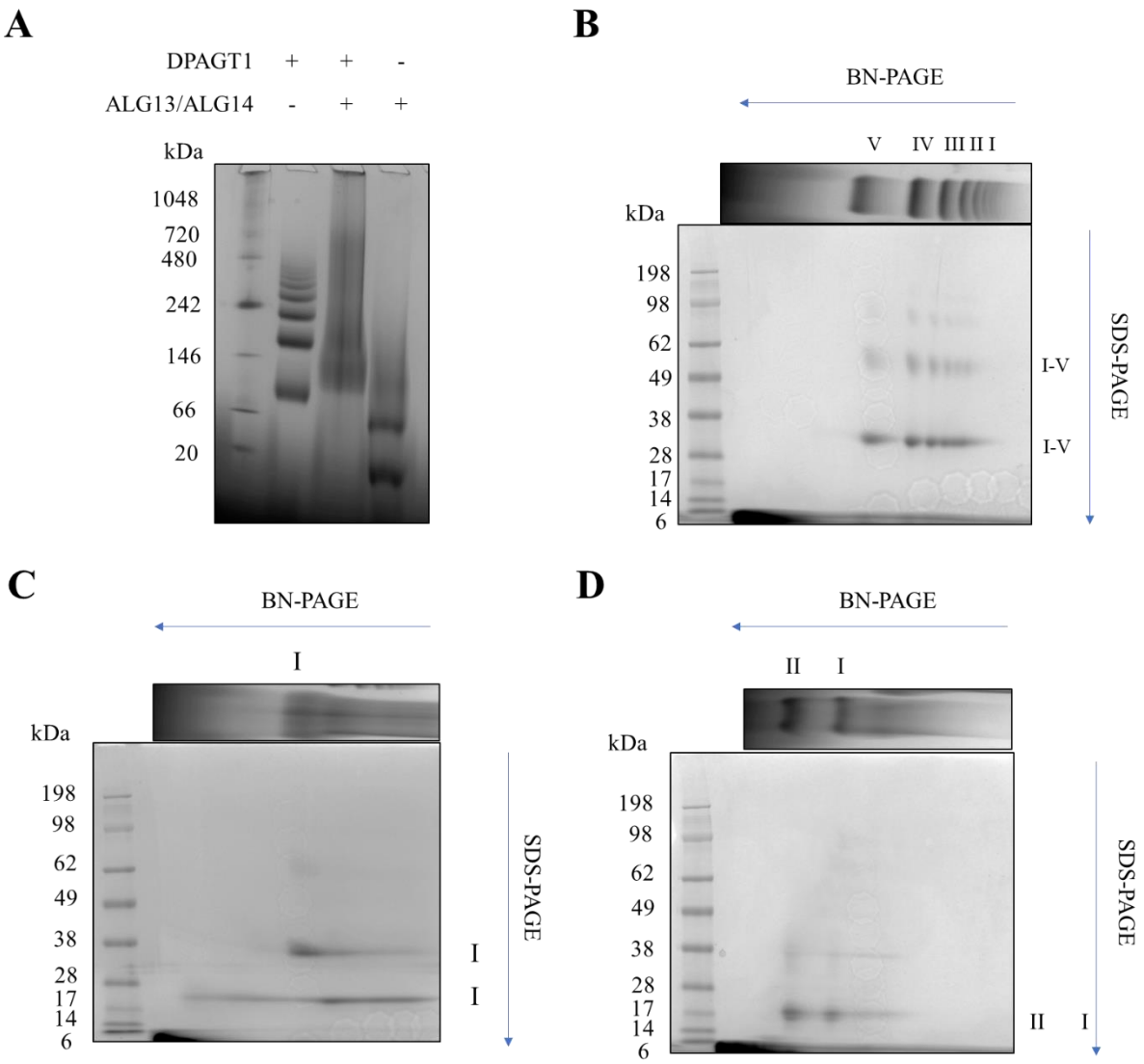


Figure 3.28 2D BN/SDS-PAGE of DPAGT1-FLAG/ALG13/His-ALG14. A 1 mg/mL samples of DPAGT1 and ALG13/ALG14 alone and in complex were coated in Coomassie G-250 dye and separated by size under native conditions**,** where DPAGT1 (46 kDa) is in lane 1, DPAGT1-FLAG/ALG13/His-ALG14 (expected Mw of 89 kDa) in lane 2 and ALG13/ALG14 in lane 3 (together ALG13/His-ALG14 are 45 kDa and alone ALG13 is 18 kDa ALG14 is 27 kDa). **B-D** show results when DPAGT1, DPAGT1-FLAG/ALG13/His-ALG14, and ALG13/His-ALG14 lanes of the BN gel were reduced and separated in a second dimension under denaturing conditions.

As with stocks of ALG13/ALG14 alone, the activity of both components of the DPAGT1/ALG13/ALG14 complex were tested to ensure proteins were folded in their functional conformation. His-DPAGT1/ALG13/His-ALG14 was purified as before with solubilisation in DM/CHS, and the complex was screened for activity when freshly purified and following freeze/thaw (Figure 3.29). The enzyme complex was assayed for 1 hour as described in Chapter 5, using ^{14}C-labelled sugar donor and monitored by TLC. The fresh and frozen samples of the complex were tested with both DPAGT1 substrate dolichol phosphate, and ALG13/ALG14 substrate dolichol pyrophosphate GlcNAc. In this way, both enzyme components of the complex were confirmed to be active with their native substrates, and transfer of the DPAGT1 product to ALG13/ALG14 occurred, with two spots visible when dolichol phosphate was supplied. Additionally, there was no significant difference in activity before and after freezing allowing the complex to be stored at -80 °C for further functional study. As for ALG13/ALG14 alone, the specific activity of the fresh stock was measured. When the stock was supplied with dolichol pyrophosphate GlcNAc, the specific activity of the fresh stock was 0.029 μmole product min^{-1} mg^{-1} , and following freeze-thaw at 0.030 μmole product min^{-1} mg^{-1}. This stock therefore showed a slightly lower specific activity than when ALG13/ALG14 was expressed alone, indicated lower purity was achieved.

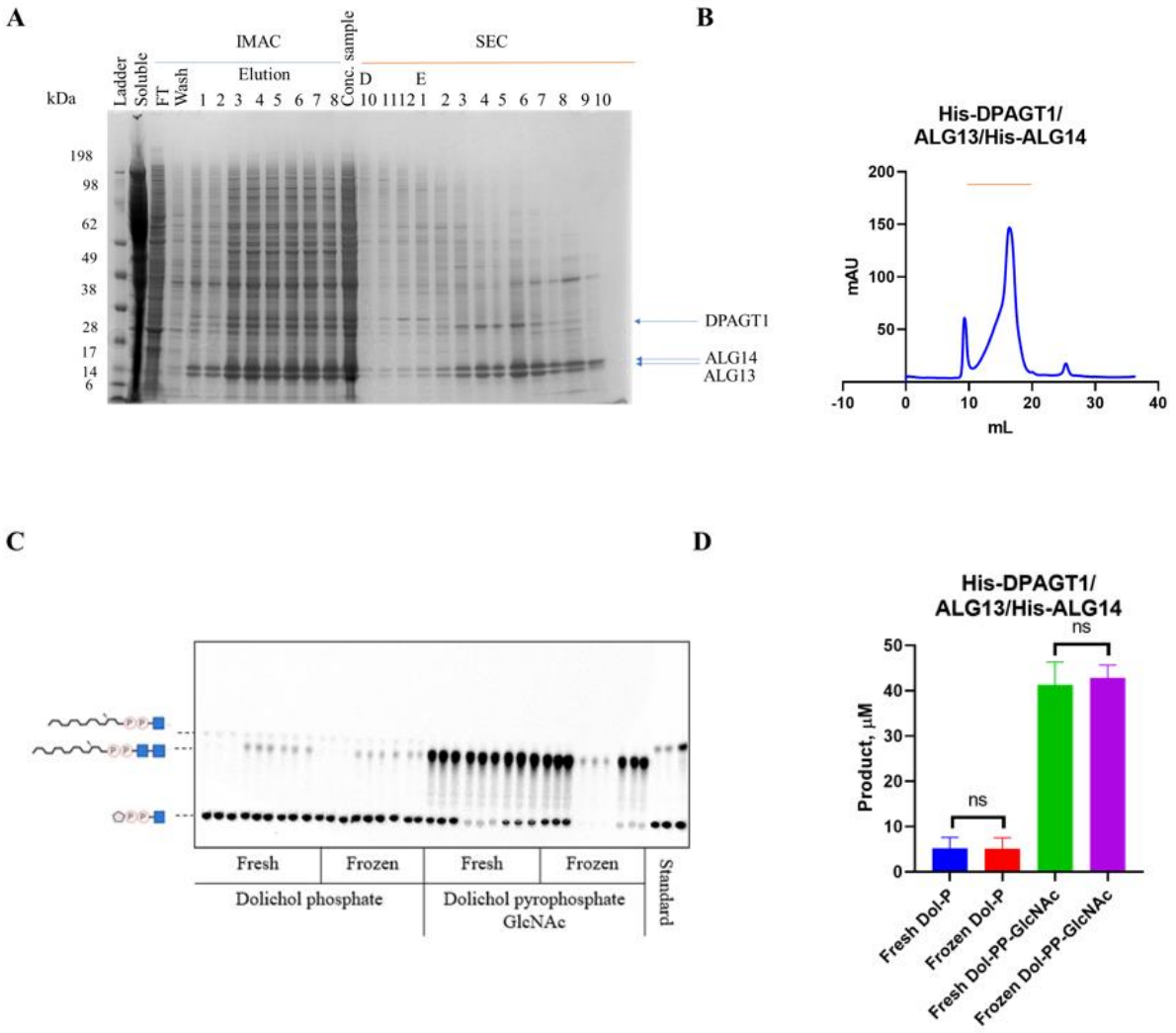


Figure 3.29 Purification and freeze-thaw testing of His-DPAGT1/ALG13/His-ALG14 complex. A Purification progress from membrane preparation, IMAC and SEC, where Soluble = supernatant from discarded following membrane preparation, FT = flow through following binding to Talon resin, Conc. sample = elution fractions 3-6 pooled and loaded onto the Superose 6 Increase column (SEC elution across the peak shown, corresponding to the trace, **B**. Freeze thaw activity testing of ALG13/ALG14, where **C** = TLC plate showing GlcNAc transfer from soluble UDP-[^{14}C]-GlcNAc to dolichol phosphate and dolichol pyrophosphate GlcNAc, alongside a DPAGT1 only standard (right hand three lanes). **D** Activity represented in µM of product formed after 1 hour.

Finally, to observe the distribution of particles in the purified stock, His-DPAGT1/His-ALG13/His-ALG14 was imaged using negative stain electron microscopy (EM). The proteins were was diluted in gel filtration buffer without detergent to give samples at 100 µg/mL and 50 µg/mL. 5 µL of each sample was loaded onto a glow discharged copper grid with a continuous carbon film and stained using uranyl acetate (work assisted by Dr Peter Harrison). Figure 3.30 shows representative EM images. Promisingly, there was minimal aggregation seen and a good distribution of even particles that were found to be of approximately 12 nm in diameter, suggesting structural elucidation by cryogenic EM (cryoEM) might be amenable.

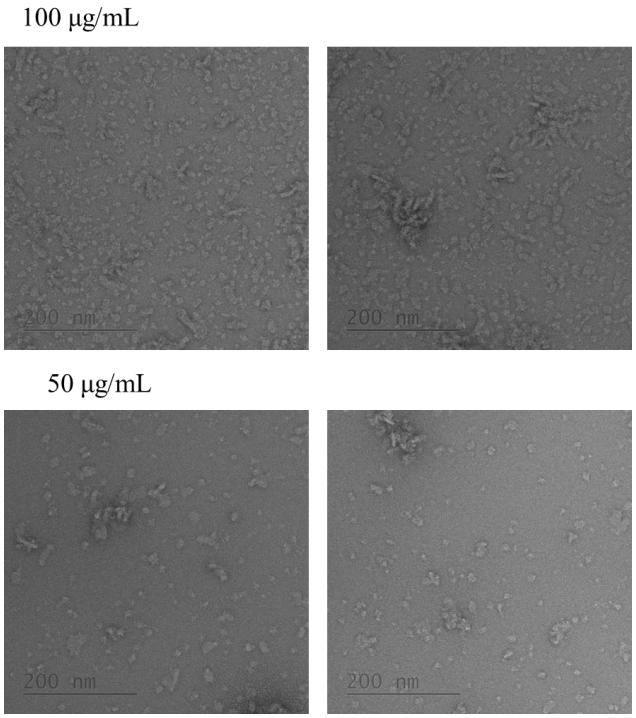


Figure 3.30 Negative stain electron microscopy images of His-DPAGT1/His-ALG13/His-ALG14 proteins at 100 and 50 µg/mL. Proteins were stained using uranyl acetate to visualise particles. The scale bar shown represents 200 nm.

3.2.6 Cryogenic electron microscopy to elucidate the structure of the DPAGT1/ALG13/ALG14 complex

Following the exciting finding that DPAGT1/ALG13/ALG14 exists in complex, it was decided to pursue structural elucidation by cryoEM. Initially, frozen protein from the preparation shown in Figure 3.22 and purified in OGNG/CHS/CDL buffer was used, as this was of high purity and in the same detergent previously used for X-ray crystallography of DPAGT1 [44]. To determine the optimal concentration of protein to apply to the grids, His-DPAGT1/ALG13/His-ALG14 was clarified by ultracentrifugation for 20 minutes at 150,000 x g, before diluting to 3, 2, 1.5, 1, and 0.5 mg/mL using detergent free gel filtration buffer. Samples were vitrified using a Vitrobot with negatively glow discharged Quantifoil carbon on gold grids at two different blotting conditions per protein concentration. These 10 grids were screened using a 200 kV Glacios TEM with a Falcon IV detector at eBIC Diamond Light Source (performed with Dr Peter Harrison).

Figure 3.31 shows representative images of a grid containing 0.5 mg/mL protein at progressively higher magnification, allowing a grid square, hole and then the final data collection image to be visualised. In all grids it was difficult to detect particles – likely due to the small size of the complex – with a high background of what appeared to be empty detergent micelles in the sample. It was difficult to elucidate at this stage whether stable complex was present but not visible, and so the grid containing protein at the highest protein concentration of 3 mg/mL was selected and a data set from 10 grid squares was collected. Processing was performed in CryoSPARC 3.0 [129] by Peter Harrison, where 253,051 particles were extracted from the data set and grouped based on their 2D shape into 50 classes. 43 of the best classes are shown in Figure 3.30, B. Sadly, only featureless particles at poor resolution were seen, and so further processing was not performed on this data set. This might have been due to frozen

samples being used rather than freshly prepared complex, or due to a high background of OGNG/CHS/CDL in the sample following concentration.

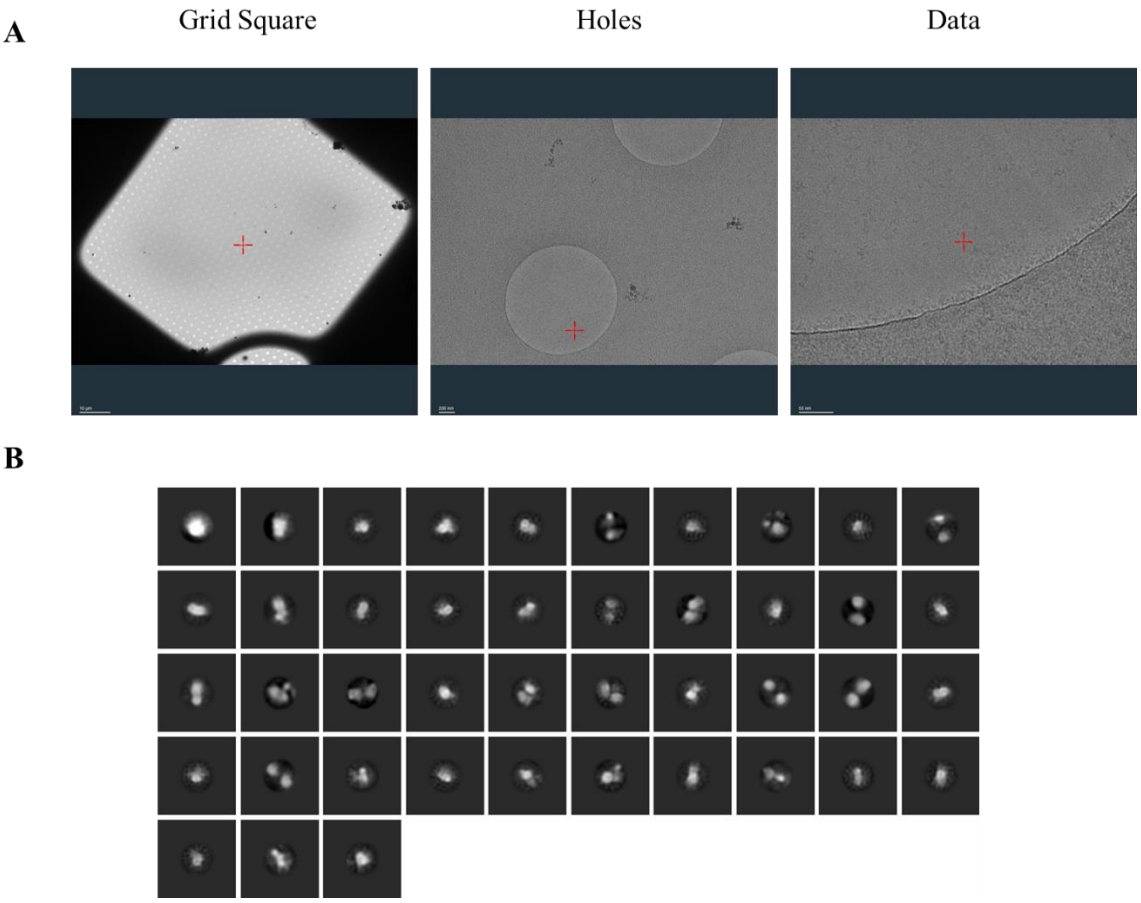


Figure 3.31 Determining the optimal protein concentration: screening and data collection of His-DPAGT1/ALG13/His-ALG14. **A** Representative images showing a grid square, holes and data collection from 0.5 mg/mL His-DPAGT1/ALG13/His-ALG14 in OGNG/CHS/CDL on gold on carbon grids, where scale bars from the left to right image are 10 µm, 200 nm, and 50 nm. **B** 253,051 particles were grouped into 43 2D classes using in CryoSPARC 3.0 [129]. eBIC session nr27436-19.

Therefore, during the next attempt at cryoEM, fresh His-DPAGT1/ALG13/His-ALG14 was used (purification shown in Figure 3.29), which was solubilised in DM/CHS. As well as of the apo complex, grids were prepared that contained proteins bound to lipid substrates in the hopes of gaining the structure of the ternary complex. Additionally, binding of lipid substrates were

found to dramatically increase the stability of the proteins (see Chapter 6 for thermostability data) which is favourable for structural analysis. Briefly, 3 mg/mL His-DPAGT1/ALG13/His-ALG14 was incubated on ice for 30 minutes with 250 μM lipid (an 8-fold molar excess) of either dolichol phosphate, dolichol pyrophosphate GlcNAc, or assay buffer with 3x the critical micelle concentration of DM/CHS.

Grids were screened as before using a Glacios TEM at eBIC, and the best quality grids were selected for data collection based on number of intact grid squares, ice thickness, and visibility of uniform particles. Representative grid squares, holes and data collection images are shown for each in Figure 3.32A. For the apo His-DPAGT1/ALG13/His-ALG14 complex, 8,039 movies were collected and for that bound to dolichol pyrophosphate GlcNAc, 4,500 movies. These were motion corrected to micrographs using Relion as previously, before downstream analysis was performed in CryoSPARC v4.4.0 [129] (see Materials and Methods for the detailed EM processing workflow).

For the dolichol pyrophosphate GlcNAc bound sample, 4,621,805 particles were extracted from the micrographs, of which 874,972 were averaged into 10 2D classes (Figure 3.32B). Poor resolution was achieved of this sample, and so further processing was not pursued. 761,335 particles were extracted of the apo complex, of which 178,549 were grouped into 10 2D classes, as in Figure 3.32B. Further processing of this data set is ongoing, as promising density was observed at the 2D class average stage – with what appears to be a transmembrane domain in a detergent micelle. Unfortunately, there was no obvious ALG13/ALG14 protruding from the micelle in these 2D classes, perhaps suggesting the complex has dissociated and the density is of DPAGT1.

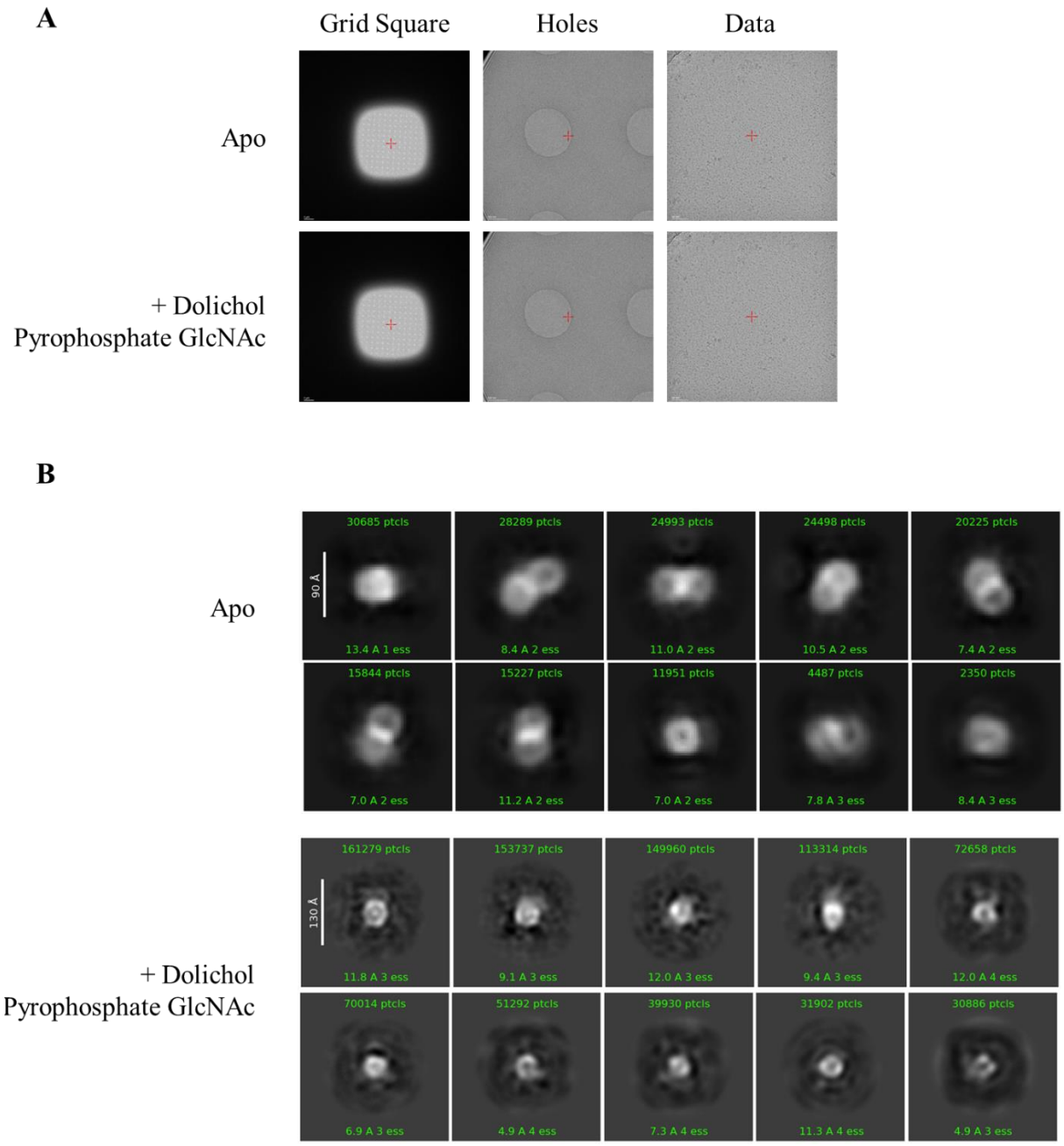


Figure 3.32 Grid screening and data collection of His-DPAGT1/ALG13/His-ALG14 in DM/CHS micelles with and without dolichol pyrophosphate GlcNAc. A Representative grid square, holes and data collection of 3 mg/mL His-DPAGT1/ALG13/His-ALG14 alone (Apo) or with 250 µm dolichol pyrophosphate GlcNAc, where scale bars from the left to right images are 5 µm, 500 nm, and 50 nm. **B** 2D class averages for the apo and dolichol pyrophosphate GlcNAc bound samples were obtained using CryoSPARC v4.4.0 [129]. eBIC session nt33941-13.

Given the high background signal of detergent in these samples, it was decided to screen 8 other detergents as well as non-detergent alternatives for use in solubilisation of the DPAGT1/ALG13/ALG14 for cryoEM. This screen was carried out in collaboration with the Membrane Protein Laboratory at Diamond Light Source and was performed using just 50 mL of Sf9 cells expressing His-DPAGT1/ALG13/His-ALG14. Cell membranes were prepared by sonication and ultracentrifugation and each of the detergents were tested for use in extraction and purification of the membrane complex (Figure 3.33). Non-detergent alternatives included amphipols (A8-35, and the non-ionic equivalent, NAPol, nanodisc scaffold peptide (NSPr) and the polymer styrene-maleic acid (SMA). For these mimetics, membrane solubilisation was achieved using DDM, before on-resin exchange was performed into the respective reconstitution agent in detergent-free buffer.

All detergents were able to extract DPAGT1 and ALG13/ALG14, with DDM/CHS and FC12 giving non-specific solubilisation with a large number of impurities seen by SDS-PAGE (Figure 3.33). DPAGT1 was found to stain with higher intensity than ALG13/ALG14 – something that was not seen in previous rounds of expression and purification – which could be due to the larger size of DPAGT1 compared to ALG13/ALG14 which would result in increased staining by Coomassie dye, rather than a higher expression level. Tryptophan fluorescence enables a more specific profile than UV 280 nm, which was especially useful on this small scale. The purified proteins were therefore analysed using an fSEC, with excitation at 295 nm to monitor emission at 350 nm (Figure 3.33B). These results suggest LMNG and LMNG/CHS lead to the highest purity with a good yield.

Amphipols A8-35 and NAPol were not successful in solubilising the protein complex, but promisingly all three proteins were detected following exchange into NSPr. Results of the SMA extraction were inconclusive given low molecular weight straining across the gel lane. Despite a lower yield, these data suggest promise for use of a membrane mimetic in stabilising the His-

DPAGT1/ALG13/His-ALG14 complex, which may aid future efforts at structural elucidation by cryoEM.

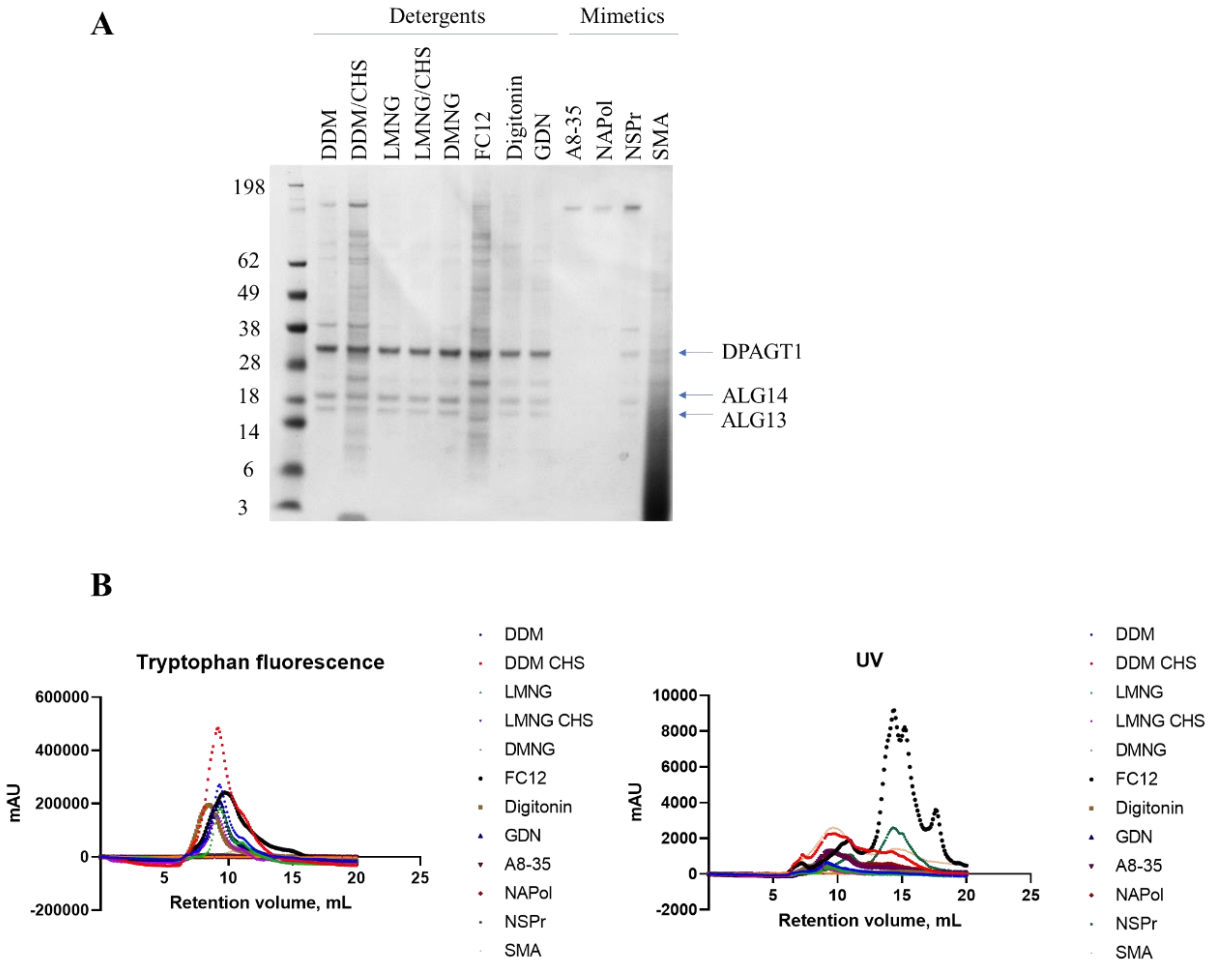


Figure 3.33 Detergent and membrane mimetic screen for extraction and purification of DPAGT1/ALG13/ALG14. A 3 mL of Sf9 cells over-expressing His-DPAGT1/ALG13/His-ALG14 were purified by IMAC with Ni^{2+} resin using 8 different detergents and 4 non-detergent membrane mimetics, and protein in the elution fraction analysed by SDS-PAGE. **B** Spectra showing tryptophan fluorescence (Ex:295 nm, Em:350 nm, left) and UV at 280 nm (right) of proteins by analytical SEC.

3.2.7 Use of DPAGT1-FLAG construct for structural elucidation by cryoEM

As shown above, a C-terminally FLAG tagged DPAGT1 was used to perform pull down experiments that revealed ALG13/ALG14 stably interacts with DPAGT1. Anti-FLAG affinity chromatography was found to give much improved purity over IMAC with Talon resin, and it was therefore decided to explore use of this complex for structural elucidation by cryoEM.

To start, DPAGT1-HisFLAG/ALG13/His-ALG14 were purified from the membrane fraction of 4 L of Sf9 cells using anti-FLAG affinity resin and SEC (Figure 3.34). Proteins were eluted from the anti-FLAG resin using 3xFLAG peptide, and found to contain all three proteins as confirmed by tryptic digest MS/MS (Dr Rod Chalk, CMD). Those containing the highest protein concentration were pooled to give an overall yield of 0.54 mg (i.e. 0.134 mg/ L original cells) from anti-FLAG affinity chromatography. The concentrated sample was further purified using a Superose 6 Increase 30/100 column. Three peaks were seen, which were found to contain DPAGT1 with ALG13/ALG14 followed by ALG13/ALG14 and then free ALG14. Unfortunately, the stoichiometry of the three proteins across the peaks differed – indicating the complex had fallen apart during SEC.

To explore this theory, the thermostability of proteins eluted from the anti-FLAG resin was compared to that after using nano-DSF (Figure 3.35). Results indicated a destabilisation following SEC of 2.53 °C for onset of precipitation and 1.45 °C for T_m (Table 3.3, Figure 3.35), which may be due to loss of complex formation. These data suggest a transient interaction of ALG13/ALG14 with FLAG tagged DPAGT1 that is lost during dilution of the proteins by SEC. Attempts to stabilise the complex were therefore undertaken.

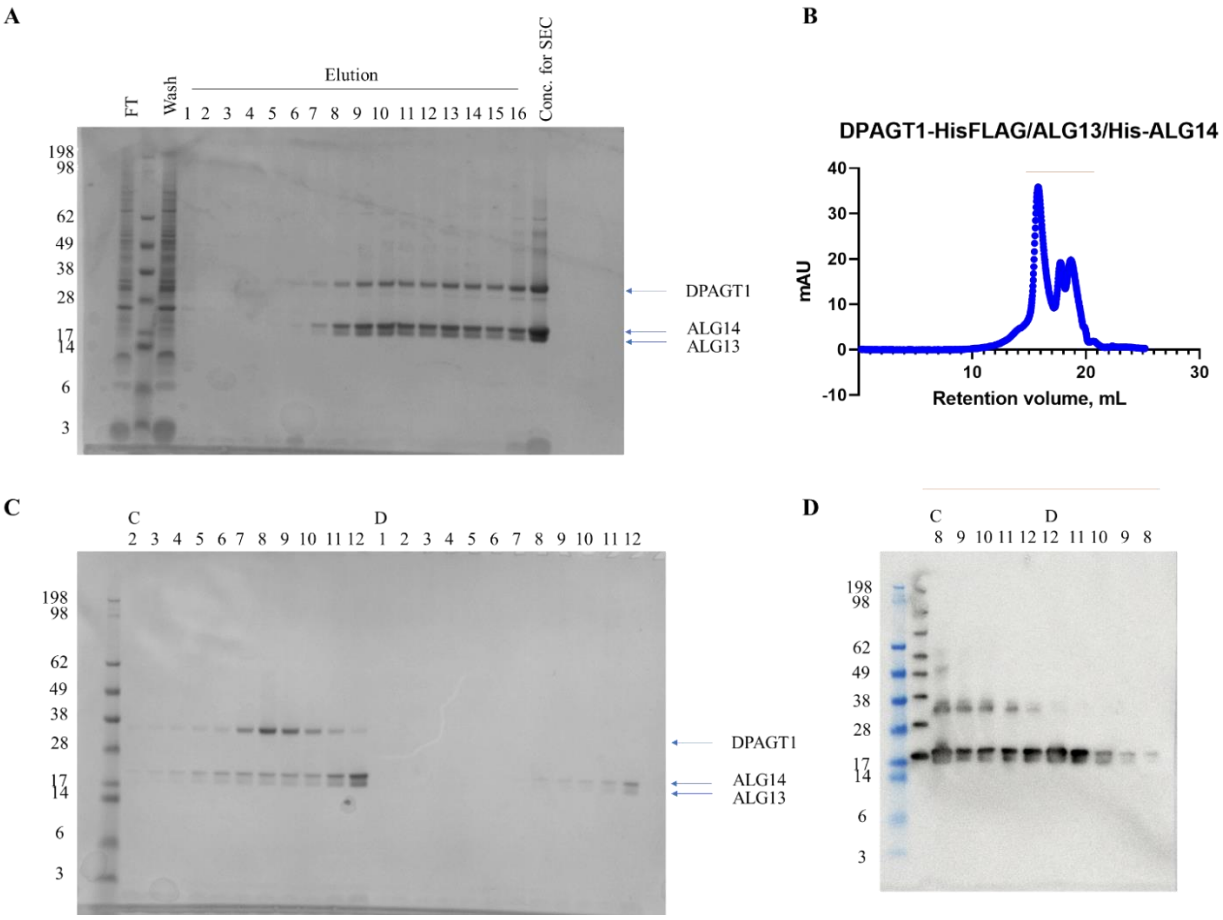


Figure 3.34 Purification of DPAGT1-HisFLAG/ALG13/His-ALG14. **A** Anti-FLAG affinity chromatography where FT = flow through following binding to resin, Wash = flow through of wash buffer, Conc. for SEC = pooled sample of elution fractions concentrated using ready for additional purification by Size Exclusion Chromatography. **B** UV 280 nm spectra of elution from a Superose 6 Increase column, with samples from each peak run on SDS-PAGE (**C**, Coomassie stained, **D**, Anti-6xHis tag Western blot). Protein identities are labelled, as confirmed by MS/MS for DPAGT1 and ALG13.

Table 3.3 Thermostability of DPAGT1-HisFLAG/ALG13/His-ALG14 before and after size exclusion chromatography (SEC)

Sample	Melting temperature, T_M, °C	Precipitation onset, °C
Before SEC	42.95	42.70
After SEC	41.50	40.17

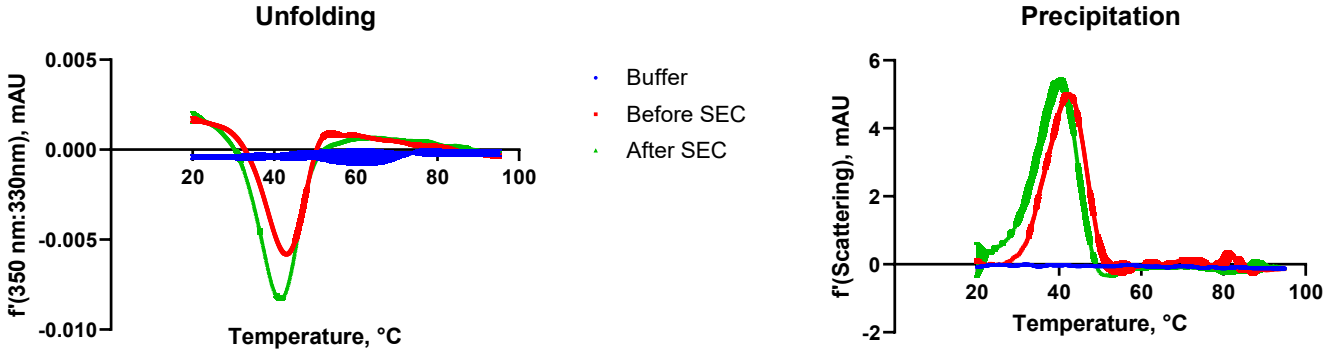


Figure 3.35 Thermostability of DPAGT1-HisFLAG/ALG13/His-ALG14 before (red) and after (green) size exclusion chromatography. (Left) Protein unfolding shown by the first derivative of the ratio of fluorescence 350 nm: 330 nm ratio when samples were increased by 1 °C per minute from 20 to 95 °C, as used to determine the melting temperature. (Right) The first derivative of the scattering mAU (at Prometheus proprietary wavelengths), where changes with respect to buffer controls are used to estimate the onset of protein precipitation.

Firstly, it was explored whether an alternative detergent system may be beneficial for stabilising the protein complex. As for the His-DPAGT1 complex, a high throughput detergent screen was carried out to select the best performing conditions for extraction and solubilisation. This time, detergents were supplemented with CHS and CDL, as this was previously shown to aid the stability of DPAGT1 alone [44]. The extraction and purification of DPAGT1-

HisFLAG/ALG13/His-ALG14 by these detergents were compared to that of Fos-choline 12 as this is a harsh detergent and so will solubilise a broad range of membrane proteins (Figure 3.36). When DDM was compared to that supplemented with CHS and CDL, there was an increase in band intensity on the SDS-PAGE for DPAGT1, as well as a larger peak signal following fSEC. The sample using DM/CHS/CDL showed intense bands on SDS-PAGE as well as a uniform peak profile by fSEC. Interestingly, C12E9/CHS/CDL showed the highest absorbance on fSEC, and so these were taken forward for medium scale testing.

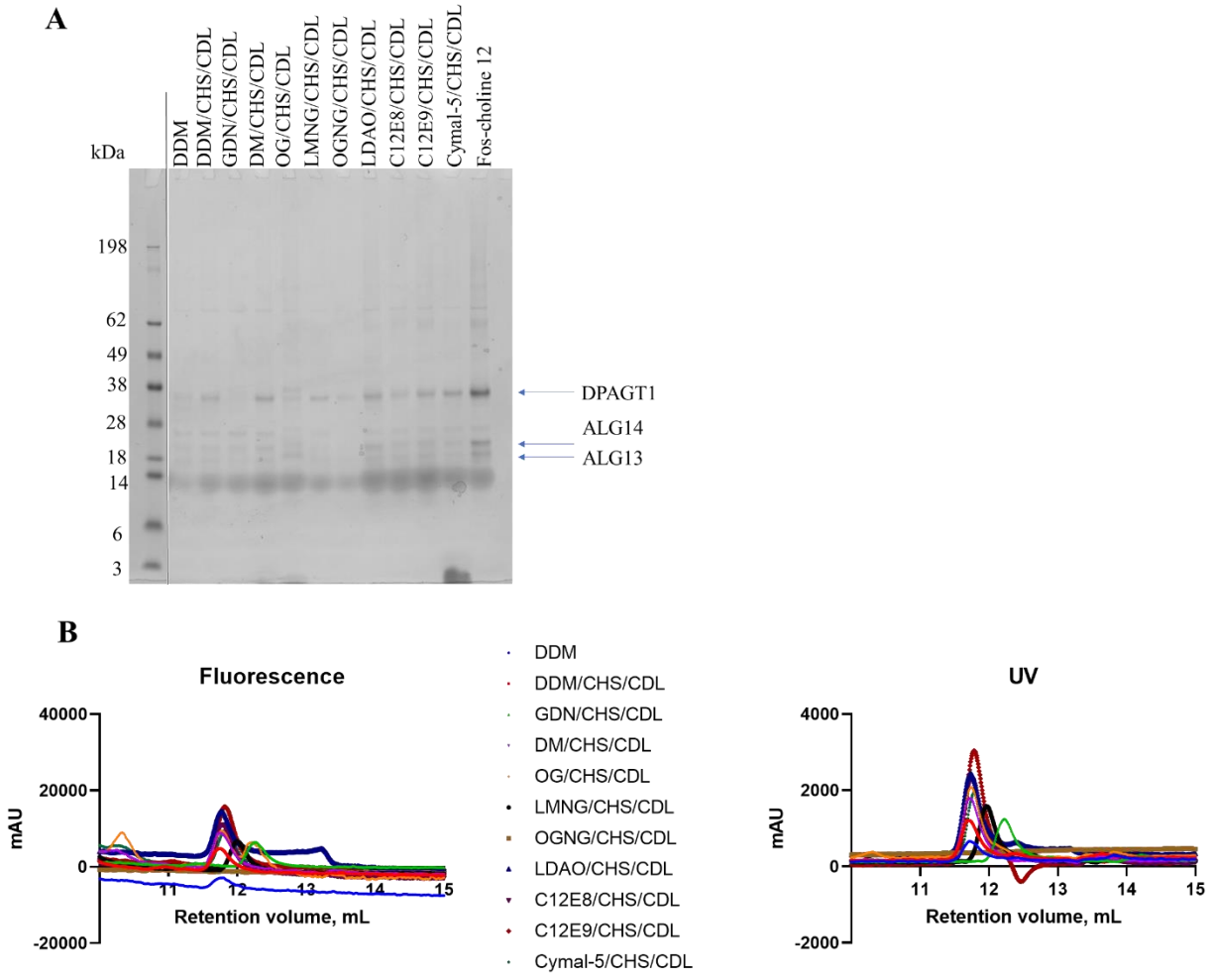


Figure 3.36 Screen of 12 detergents for extraction and purification of DPAGT1-HisFLAG/ALG13/His-ALG14. A 12 different detergents were used in the extraction and purification of DPAGT1-HisFLAG/ALG13/His-ALG14 from 3 mL of Sf9 cells

using IMAC with Ni^{2+} resin. Protein in the elution fraction was analysed by SDS-PAGE. **B** Spectra showing tryptophan fluorescence (Ex:295 nm, Em:350 nm, left) and UV at 280 nm (right) of proteins by analytical SEC. Screen performed with Monika Olejnik, Research Technician, MPL.

These three best performing detergents – DM, OGNG and C12E9 - were used in a medium scale purification test, where the clarified membranes from 3 L of Sf9 insect cells expressing DPAGT1-HisFLAG/ALG13/His-ALG14 were used for each condition. Anti-FLAG affinity chromatography was comparable for all three detergents (Figure 3.37), with intense bands for all three components seen with few contaminants. However, as before, when samples were analysed by fSEC all detergent conditions contained a population of protein in complex followed by one of free ALG13/ALG14. This was determined as the band intensity of ALG13/ALG14 is seen to peak twice. These results again suggest a weak interaction of DPAGT1-HisFLAG and ALG13/His-ALG14.

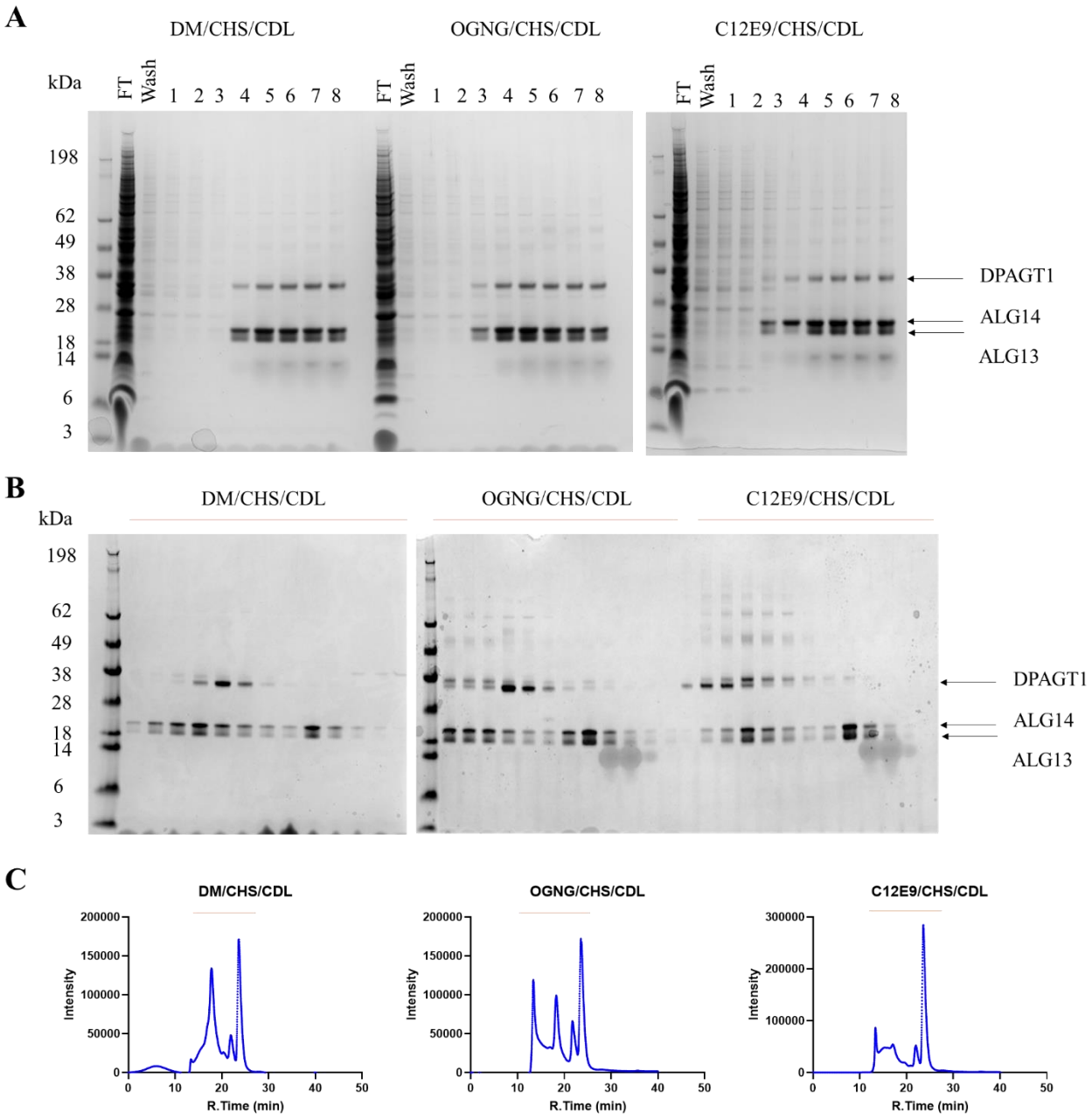


Figure 3.37 Purification of DPAGT1-FLAG/ALG13/His-ALG14 in 3 different detergent conditions. **A** The progress of anti-FLAG affinity chromatography for each of three detergents, before downstream analysis by fSEC. Labels 1-8 correspond to order of elution fractions from the anti-FLAG affinity resin. UV trace across the elution (**C**) is shown below corresponding fractions analysed by SDS-PAGE, **B**.

It was therefore decided to purify DPAGT1-FLAG/ALG13/His-ALG14 using only anti-FLAG affinity chromatography without the SEC polishing step, given a high level of purity was already achieved with this method. A range of native and non-native lipid substrates were also added to the samples in an attempt to stabilise the complex. To the eluted protein, 250 μM nerol pyrophosphate GlcNAc, undecaprenol pyrophosphate GlcNAc or dolichol phosphate were added. These samples were applied to glow discharged gold grids and screened by cryoEM (Figure 3.38). Unfortunately, this revealed the prepared samples were too dilute to be amenable for structural elucidation so these were not progressed further. Furthermore, a high background of detergent micelles was noted. This was perhaps unsurprising given the complex was purified from just 2 L of cells across each of the four substrate conditions, meaning the samples were subject to repeat rounds of concentration before applying to the grids which also concentrates the detergent.

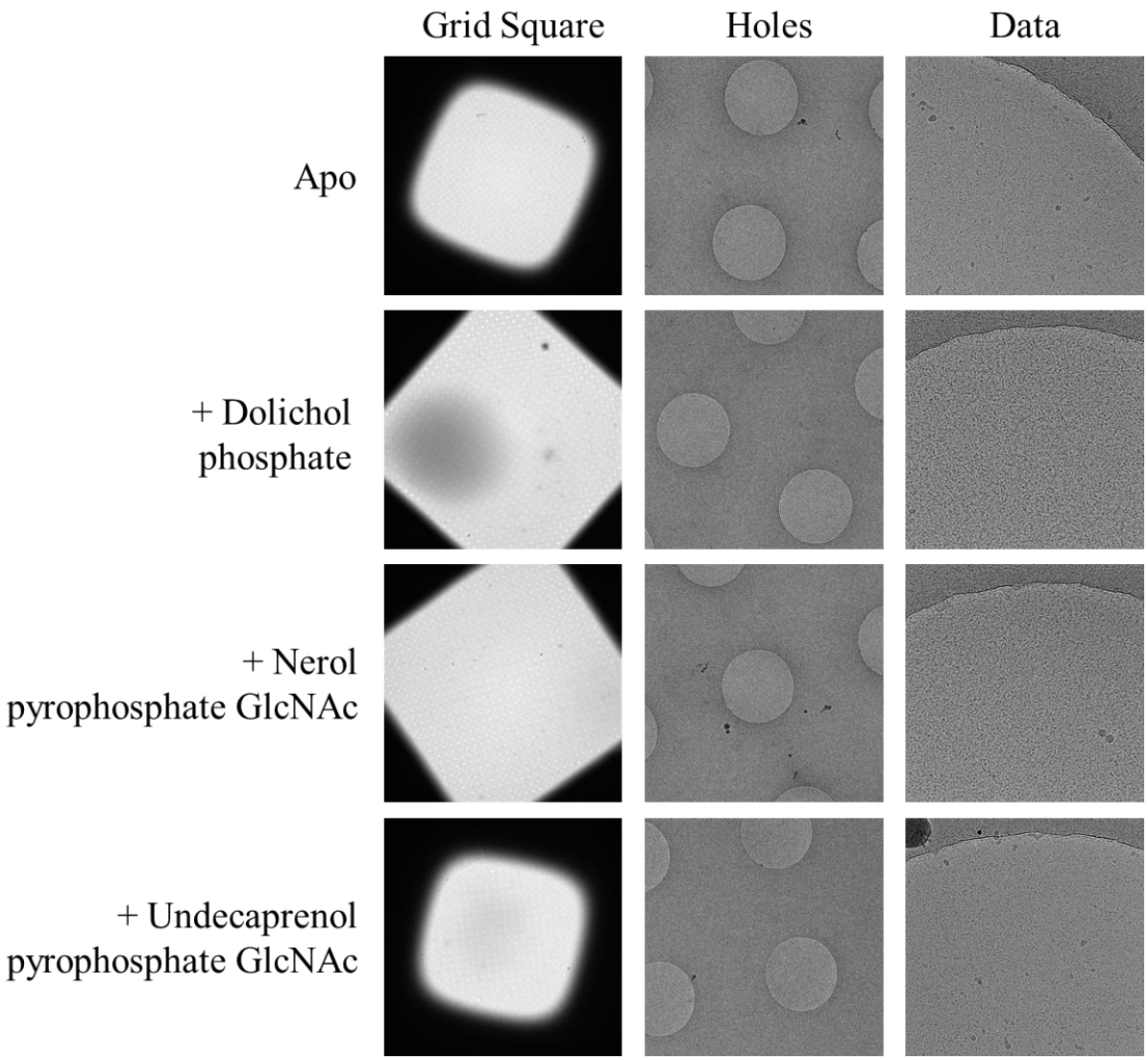


Figure 3.38 Summary of Atlas screen performed using eBIC Glacios cryo-electron microscope. Grids contained DPAGT1/ALG13/ALG14 plus three different lipid substrates (undecaprenol-PP-GlcNAc, nerol-PP-GlcNAc or dolichol phosphate). eBIC session nt29493-99.

In an attempt to improve the yield of the protein complex and to reduce the detergent background in samples for cryoEM, a 12 L scale expression of the C-term FLAG tagged DPAGT1 with ALG13/ALG14 was performed, this time using buffer containing 2x the CMC of DM/CHS/CDL (rather than 3x used in previous attempts). Peak fractions from anti-FLAG

affinity chromatography (Figure 3.39) were pooled and concentrated as minimally as possible to reduce detergent accumulation; a 3-fold decrease in volume was used to yield protein at 2 mg/mL. Unfortunately, very low expression of ALG13 protein was detected in this preparation- an issue that was later identified to be due to the ALG13/His-ALG14 baculovirus. The prepared DPAGT1-HisFLAG/ALG13/His-ALG14 was however tested for activity by incubating for 1 hour with both DPAGT1 substrate, dolichol phosphate, and with ALG13/ALG14 substrate dolichol pyrophosphate GlcNAc and labelled UDP-[^{14}C]-GlcNAc, as in Chapter 5. The complex was found to be active with both substrates despite this low expression, with a specific activity when supplied with dolichol pyrophosphate GlcNAc of 0.0033 µmole product min^{-1} mg^{-1}. This was ten-fold lower than that of the stock with the N-terminally 6xHis tagged DPAGT1.

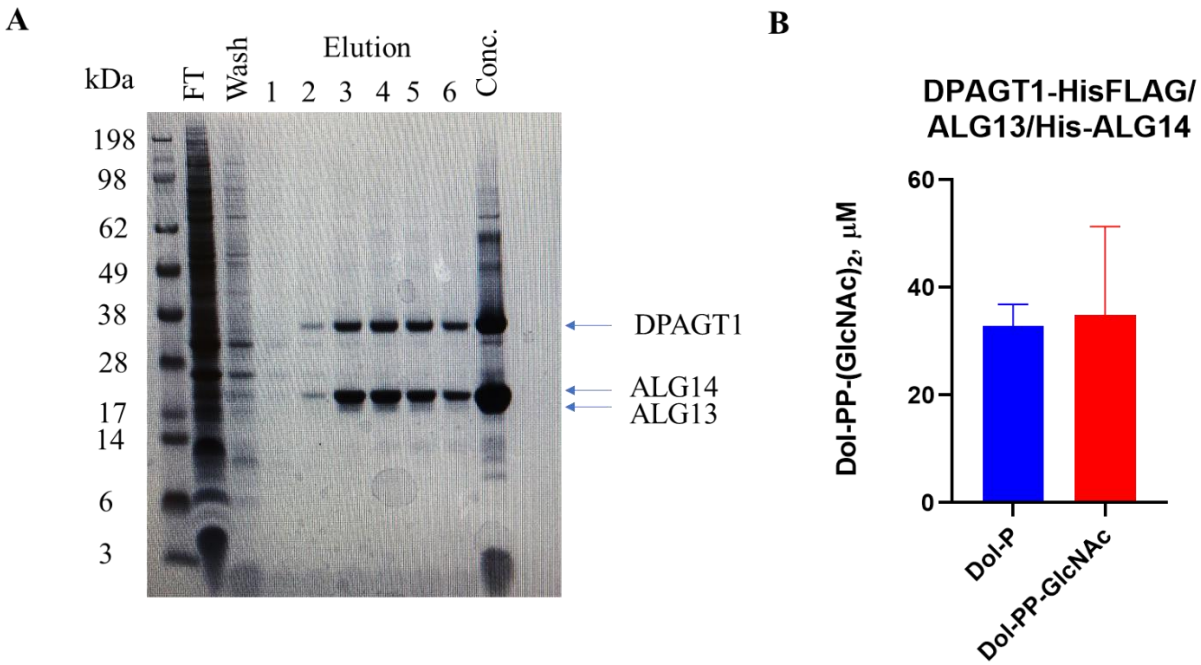


Figure 3.39 Purification and activity testing of DPAGT1-HisFLAG/ALG13/HisALG14, where FT = flow through from anti-FLAG affinity resin, wash = wash flow through of resin with 30 column volumes of extraction buffer. Elution fraction 1 was 1 mL, and remaining fractions eluted in 1.5 mL fractions (0.4 and 0.5

times column volume respectively). Conc. = concentrated sample of the pooled elution fractions 3-5, that was used for downstream structural analysis. (**B**) Activity of the protein stock was tested using endpoint activity. 2 μM DPAGT1-HisFLAG/ALG13/His-ALG14 was incubated for 1 hour at 37 °C with 50 μM UDP-[^{14}C]-GlcNAc and 100 μM dolichol-phosphate (Dol-P) or dolichol pyrophosphate GlcNAc (Dol-PP-GlcNAc), and transfer of radioactive GlcNAc monitored by TLC.

Given the activity, protein was incubated with a range of native and non-native lipid acceptor species to explore substrate and product bound conformations of the DPAGT1-HisFLAG/ALG13/His-ALG14 complex by cryoEM. Samples were incubated with 250 μM dolichol phosphate, dolichol pyrophosphate GlcNAc, or undecaprenol pyrophosphate $(GlcNAc)_2$. Following an initial atlas screen of all grids, the best grid was selected based on number of intact grid squares, ice thickness, and visibility of uniform particles of likely size and shape of the target proteins. This was found to be a grid containing apo DPAGT1-FLAG/ALG13/ALG14 protein (Figure 3.40A), and 5,785 movies were collected.

Processing of the data set was performed as for the His-DPAGT1 containing complex. Particle picking was performed using cryoSPARC in-built blob picking function on a subset of randomly sampled 2,000 micrographs, before 1,234,068 particles were extracted from the full set of micrographs and 288,474 selected for 2D classification. These 11 classes are shown in Figure 3.40B, with a detergent micelle visible and what appears to be a transmembrane domain, likely of DPAGT1. The 2D classes are similar to those for the His-DPAGT1 complex, giving confidence in the particle extraction and processing. These data were corroborated by processing of the dataset by Dr Peter Harrison in Relion (Figure 3.4C), where 189,500 particles were extracted and grouped into 50 classes. As before, there is little evidence of association of ALG13/ALG14 which would be expected to protrude from the micelle given ALG14's predicted structure of a single pass transmembrane domain and soluble Rossman fold that

interacts with the globular, soluble ALG13. This is likely to be due to the low ALG13 expression seen in the preparation.

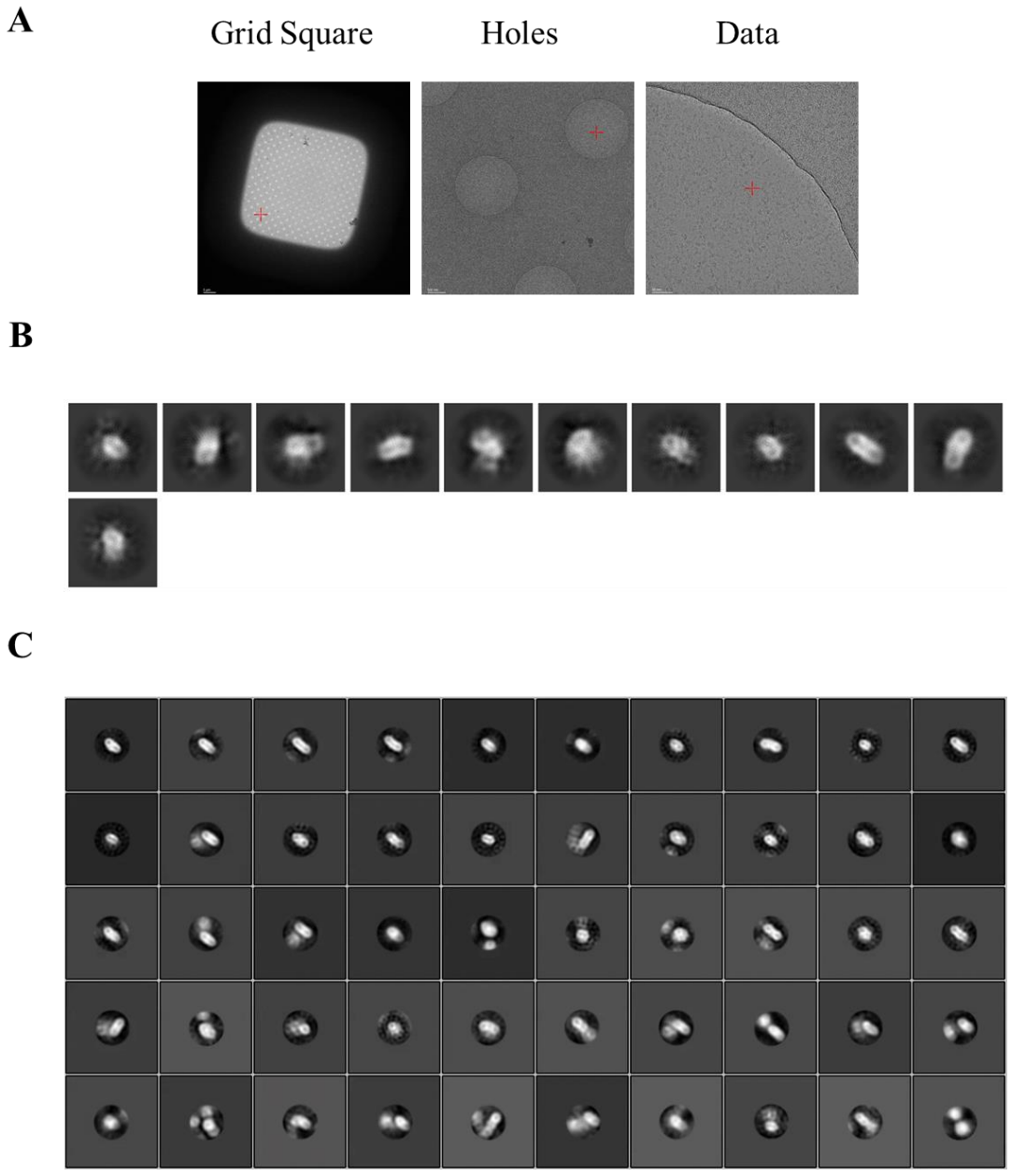


Figure 3.40 Grid screening and 2D class averages of DPAGT1-HisFLAG/ALG13/His-ALG14. A Representative grid square, holes and data collection of 3 mg/mL DPAGT1-HisFLAG/ALG13/His-ALG14, where scale bars from the left to right images are 5 µm, 500 nm, and 50 nm. **B** 1,234,068 particles were extracted from micrographs in CryoSPARC version v3.3.2, of which 288,474 were split into 11 2D

classes. **C** 189,500 particles were grouped into 50 2D classes using Relion 4.0 (Dr Peter Harrison). eBIC session nt33941-9.

3.3 Discussion

3.3.1 ALG13/ALG14

To gain structural and functional insights into the integral ER membrane proteins ALG13 and ALG14, their expression and purification were optimised. Initial attempts at purification of ALG13 alone were terminated when it was discovered using intact denaturing mass spectrometry that the expressed protein was truncated at both N and C termini (Figure 3.7). The loss of the purification tag explains results of purification attempts: TEV protease cleavage was not possible and the protein easily washes off from Talon resin during IMAC. The untagged, truncated protein likely has inherent binding to Talon resin, making washing very difficult. This finding corroborates that reported by Averbeck *et al.* who found yeast ALG13 protein to be unstable and, when alone, readily degraded by the proteasome in a ubiquitin-independent manner [69]. The authors suggest this could be a mechanism of eukaryotic regulation of glycosylation levels by controlling levels of ALG13 available in the cytosol. Instead, ALG13 was co-expressed with ALG14 using a MultiBac vector which allowed for protection of the ALG13 protein and stochiometric production of the functional enzyme heterodimer.

Two constructs were successfully purified: N-terminally 6xHis tagged ALG14 with either untagged ALG13 or that bearing an N-term 6xHis tag. To maintain the option of using either of these constructs for downstream functional assessment, they were both extensively characterised. As ALG14 is integral to the ER membrane, a detergent was required to solubilise the protein complex. DM has a hydrophilic maltose head group and a single alkyl chain forms its hydrophobic region. This was originally selected as a starting point for ALG13/ALG14 solubilisation because it forms relatively small micelles (40 nm, with a Mw of 40 kDa [117]) and

it is mild and non-ionic in nature ensuring the proteins would not be denatured. Additionally, DM has been widely reported to be successful in purification of other membrane proteins used in structural analysis; it has been used to purify 13.9% of all membrane protein structures determined by crystallography [117]. DM was compared alone or when supplemented with 1/10th by weight of the soluble cholesterol mimic, cholesterol hemiscuccinate (CHS). This was because the non-bilayer forming lipid is abundant in ER membranes and has precedent for stabilisation of purified human proteins [147], as well as aiding in their function [148]. This supplemented DM/CHS detergent system proved efficient at extraction and purification (Figures 3.8-3.10) and was therefore used in all subsequent ALG13/ALG14 purifications.

Attempts were made to increase the purity of ALG13/ALG14 achieved using IMAC with Co^{2+} resin, given the target proteins were seen to be readily lost during washing of the resin and high levels of contaminants remained in elution fractions (Figures 3.11-3.13). Ultimately, an extended membrane preparation was employed which greatly improved sample purity and allowed for protein with a mono-disperse SEC peak to be produced (Figures 3.14-3.15). This method was first described by Ray Stevens, and involves the heavier, membrane-containing cell compartments being washed twice in both hypotonic and hypertonic extraction buffers to thoroughly clear all soluble and membrane-associated proteins by ionic disruption [149]. Both constructs (with untagged or 6xHis tagged ALG13) were successfully purified using this method and very little difference was seen between their SEC profiles, as compared in Figure 3.41.

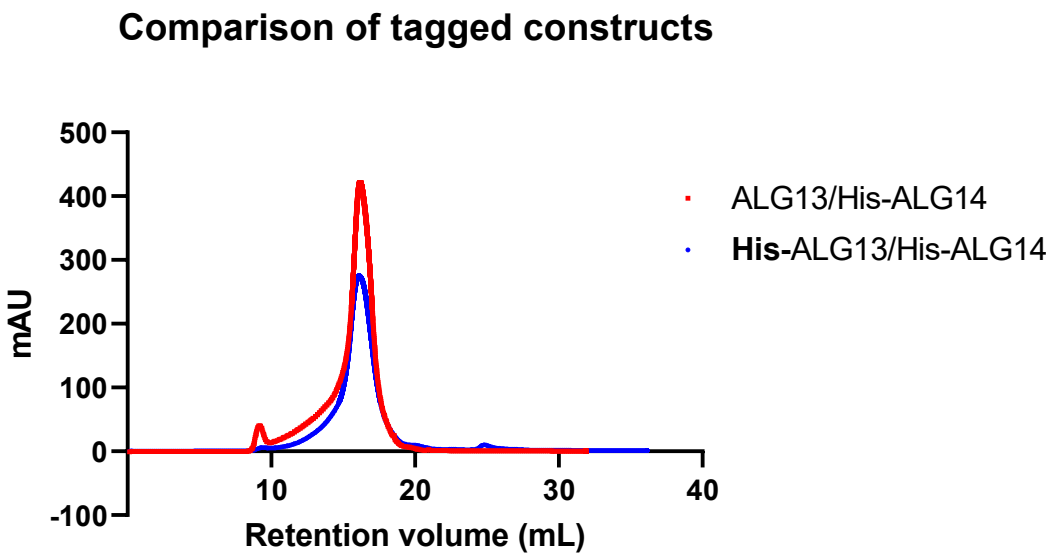


Figure 3.41 Comparison of SEC profiles of ALG13/ALG14 constructs. (Red) ALG13/His-ALG14 was purified from 5 L of cell culture. (Blue) N-terminally His tagged ALG13 with His-ALG14 was purified from 6 L of cell culture. Both protein preparations were solubilised in DM/CHS buffer and separated using a Superpose 6 Increase 30/100 column.

Thorough analyses to ensure high quality ALG13/ALG14 had been produced were necessary to ensure productive downstream use in functional and structural studies; especially important given issues of truncation previously seen with ALG13 expressed alone. Using SEC-MALS it was definitively shown for the first time that human ALG13 and ALG14 form a 1:1 complex, and intact denaturing mass spectrometry was able to confirm that full-length proteins had been purified when both ALG13/ALG14 vectors were used (Figures 3.16-3.19). The intact masses of all target proteins unexpectedly matched those expected for proteins in their oxidised form when disulphide bonds are formed between all cystine residues. This was surprising given proteins were solubilised in buffer containing 2 mM of the reducing agent TCEP, perhaps only explained if there was an issue with the reducing agent used or if oxidation occurred during buffer exchange during the LC stage.

3.3.2 DPAGT1/ALG13/ALG14

Building on the methodology developed to access ALG13/ALG14, evidence was obtained in favour of the early N-glycosylation enzymes forming a complex with DPAGT1, using pull-down and 2D blue native/ SDS-PAGE. To characterise this complex, it was decided to focus efforts using the untagged ALG13/His-ALG14 MultiBac vector as this left ALG13 in its native, untagged form. SEC-MALS was able to determine particle size in the detergent micelles when co-expressed with His-DPAGT1, and a predicted 130 kDa of protein was found to contribute to signal in the detergent micelle. Given the theoretical Mw of a 1:1:1 His-DPAGT1: ALG13: His-ALG14 complex is 96,258 Da, this may suggest homo-dimeric DPAGT1 interacts with ALG13/ALG14 which would have a theoretical mass of 144,957 Da. Discrepancies between theoretical and experimental masses are likely due to the presence of contaminant in the protein preparations, and perhaps due to the assumptions made as to the molecular weight and extinction coefficient of the proteins during co-polymer analysis.

Interestingly, when the FLAG tagged DPAGT1 was used in the DPAGT1/ALG13/ALG14 complex, despite allowing pull-down by affinity chromatography, free ALG13/ALG14 eluted during SEC following that containing DPAGT1. This suggested the complex was weak or transient - perhaps disrupted by the dilution experienced during SEC. This was not the case when co-expressed with His-DPAGT1, where a single peak with a broad leading shoulder after the void was monitored, as demonstrated in Figure 3.42. This may suggest that the C-terminus of DPAGT1 is implicated in binding to the downstream enzymes, or it might be that the highly charged nature of the FLAG tag was destabilising.

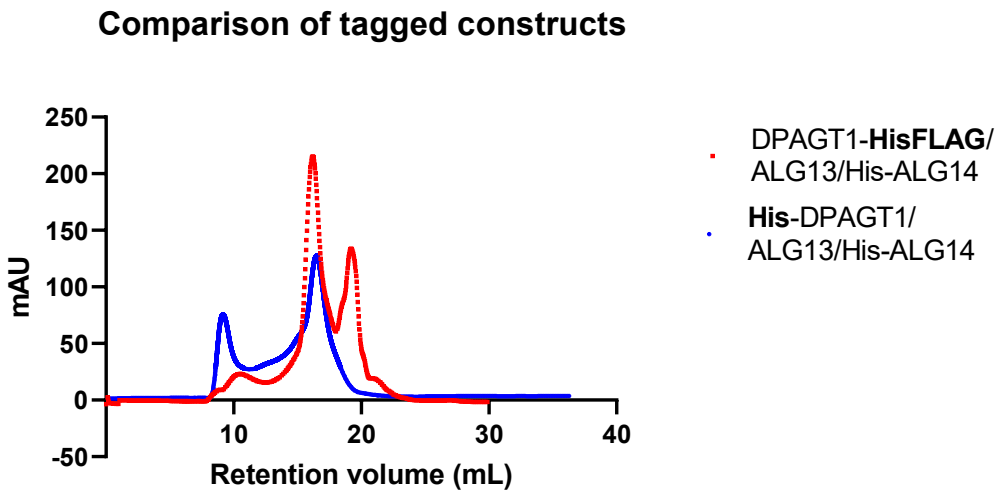


Figure 3.42 Comparison of SEC profiles of DPAGT1/ALG13/ALG14 constructs. (Red) DPAGT1 bearing a C-terminal FLAG tag along with ALG13/ALG14 was purified from 12 L of cell culture. (Blue) N-terminally His tagged DPAGT1 co-expressed with ALG13/ALG14 was purified from 6 L of cell culture. Both protein preparations were solubilised in OGNG/CHS/CDL buffer and separated using a Superpose 6 Increase 30/100 column.

SEC-small-angle X-ray scattering was used by Watanabe *et al.* for resolving ERGIC-53 in complex with an example glycosylated cargo protein glycoprotein [13], and could be a useful technique to apply to define the size of DPAGT1/ALG13/ALG14. Furthermore, the complex would be an excellent candidate for crosslinking-mass spectrometry. *N*-hydroxysuccinimide (NHS) esters that non-specifically cross link amine groups at the N-terminus or in lysine residues are commonly used for this [150], as well as maleic acid imides that react with the sulfhydryl group of free cysteines in a protein, and zero-length cross linkers like 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) which cross-link aspartates or glutamates to nearby lysines [151]. Bioconjugation in this way coupled with tryptic digest mass spectrometry can identify residues that are close in 3D space when the protein is folded (as opposed to residues

near each other in primary sequence) and as such could help elucidate where the protein-protein interface of DPAGT1 and ALG13/ALG14 lies [151].

With the added mass of monomeric or dimeric DPAGT1 bound to ALG13/ALG14, structural elucidation by single particle cryogenic electron microscopy became amenable. This technique offers the ability to capture the breadth of conformations that a protein exists in when in solution - and hence insight into its function - as the protein remains flexible until it is vitrified in liquid ethane [152]. This is in contrast to the typically single protein conformation captured using X-ray crystallography – that being the most stable conformation but not necessarily a functional one [152].

During all screening attempts, it was difficult to visualise particles of DPAGT1/ALG13/ALG14 due to the small size of this membrane protein complex even at 120,000x magnification (1.192 Å px^{-1}). Additionally, a high background of empty detergent micelles was present in the samples – especially in those containing lipid substrates given these were solubilised in buffer containing 1% detergent. Once diluted with the addition to the protein sample, DM was at 0.5%. As DM is a small detergent it has a high CMC, meaning this was 5 times CMC. It has been suggested that the optimal detergent concentration for membrane proteins studied by cryoEM is 0.2–0.4% [153] which would put this higher than the suitable range. High levels of free detergent can lead to artefacts on the grids making detergents with low CMC's like GDN or LMNG more attractive options [153]. These were both found to be suitable for extraction and solubilisation of His-DPAGT1/ALG13/His-ALG14 (Figure 3.33) and so it would be exciting to trial data collection in these detergents during future endeavours.

Despite this, small data sets were collected from grids containing His-DPAGT1/ALG13/His-ALG14 in DM/CHS both in the apo form and when bound to dolichol pyrophosphate GlcNAc (Figure 3.32). Promising signs of density inside a detergent micelle were identified at the 2D

class average stage, with what appears to be mostly side views of particles. This density is likely to be DPAGT1 as it known to be largely integral to the ER membrane. There was no obvious sign of ALG13/ALG14 protruding from the micelle, which again suggests the proteins had dissociated and as such might be only weakly interacting. Perhaps DPAGT1 and ALG13/ALG14 dissociated when exposed to the air-water interface during application to the EM grids. D'Imprima *et al.* report that 90% of complexes adsorbed to this interface become denatured during sample preparation [154], although presence of detergent would be expected to protect the membrane proteins. Given the high levels of purity achieved when purifying the DPAGT1-FLAG/ALG13/His-ALG14 construct, iterative attempts were made to stabilise this complex for cyroEM (Figures 3.34-3.37). The final processed 2D class averages look remarkably similar to that of the His-DPAGT1/ALG13/ALG14 (Figure 3.43), again suggestive of loss of ALG13/ALG14 in this sample too.

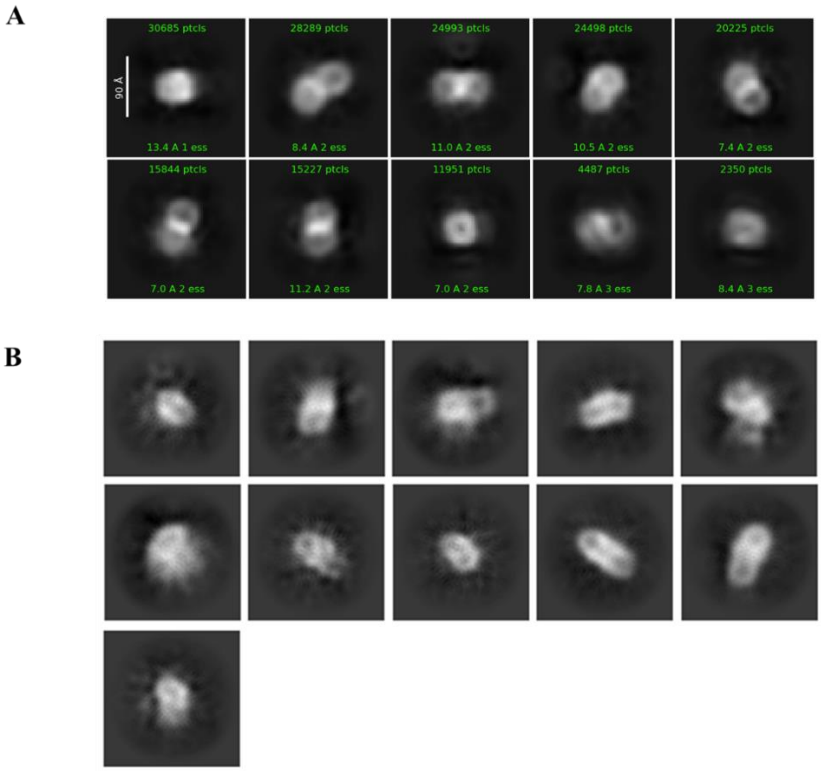


Figure 3.43 2D class averages of (A) His-DPAGT1/ALG13/ALG14 and (B) DPAGT1-FLAG/ALG13/His-ALG14

Whilst data processing continues from the latest collection of His-DPAGT1/ALG13/His-ALG14 both alone and bound to dolichol pyrophosphate GlcNAc, any future efforts towards a structure of DPAGT1/ALG13/ALG14 should focus on use of membrane mimetics rather than detergents. In particular, trialling the use of lipid nanodiscs where membrane proteins are protected by lipids packaged with membrane scaffold proteins based on apolipoprotein A1 [155], as NSPr (nanodisc scaffold peptide) showed promise for recapitulating all three components of the complex (Figure 3.35). Proteins inside the nanodisc remain freely available to move as in a native lipid bilayer but are soluble in detergent free buffer [156] . It will be important to test whether packaging by NSPr leads to aggregation of DPAGT1/ALG13/ALG14 when detergent is removed, for example by analysing the sample with SEC and negative stain EM, rather than just by SDS-PAGE as done so far. Additionally, crosslinking proteins before analysis by cryoEM could be explored, to protect the subunits from dissociating during SEC and grid preparation. Although, this could lock the protein complex into artificial or limited conformations [156].

In summary, the purification of ALG13/ALG14 and DPAGT1/ALG13/ALG14 proteins from the membranes of insect cells has been described, which successfully allowed for the functional characterisation of both steps in the early N-glycosylation pathway- as shown in Chapters 4-6. For the first time, evidence has been provided of a human multi-glycosyltransferase complex between these ER enzymes, which allowed progress to be made towards a structure by cryoEM.

Chapter 4

Optimisation of a glycosyltransferase assay

4.1 Introduction

DPAGT1 and ALG13/ALG14 are endoplasmic reticulum (ER) membrane bound glycosyltransferases that initiate synthesis of the lipid-linked oligosaccharide substrate used in N-glycosylation. DPAGT1 catalyses the transfer of a phosphorylated GlcNAc from UDP-GlcNAc onto dolichol phosphate, before downstream ALG13/ALG14 acts to catalyse the transfer of a second GlcNAc forming the chitobiose core of the N-glycosylation polysaccharide (as shown in Figure 4.1). In order to better understand the kinetics of these glycosyltransferases and to complement our efforts for structural characterisation, the development of a functional activity assay was undertaken.

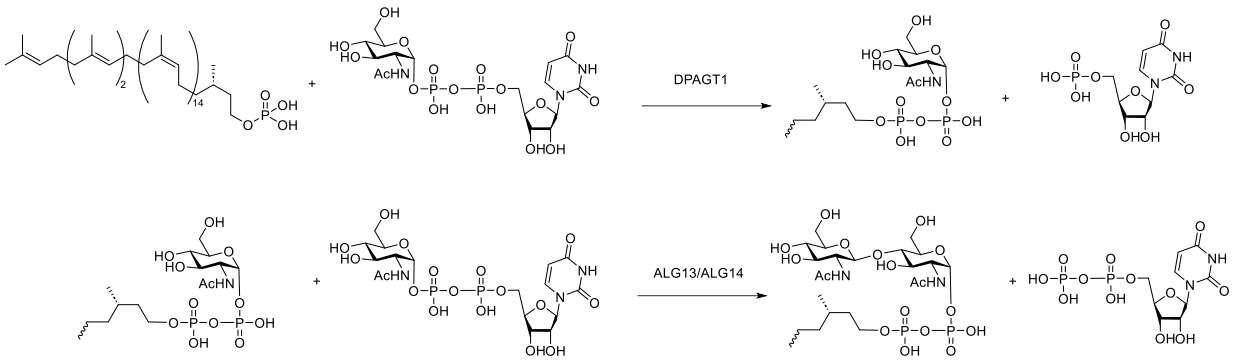
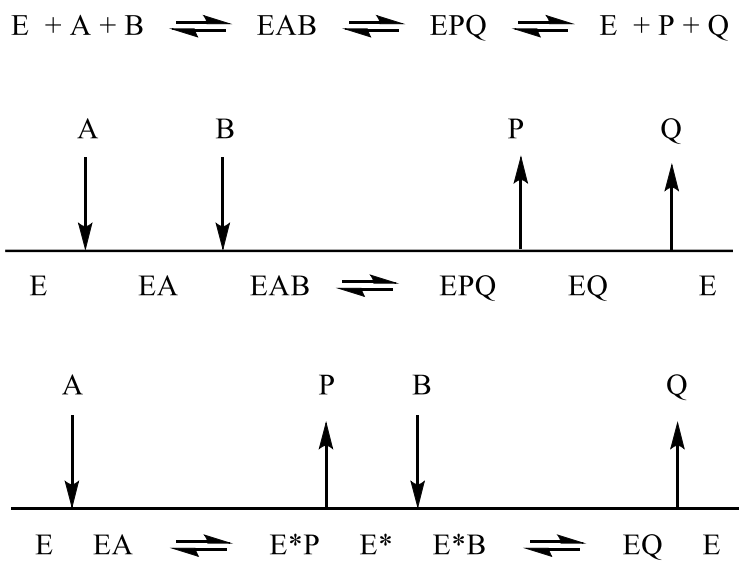


Figure 4.1 DPAGT1 and ALG13/ALG14 catalysed GT reactions. Phospho-GlcNAc and GlcNAc are transferred from UDP-GlcNAc onto dolichol phosphate by DPAGT1 and ALG13/ALG14 respectively.

4.1.1 Kinetic study of bi-substrate enzymes

Bi-substrate bi-product (bi-bi) enzymes convert two substrates (A and B) into two products (P and Q), as shown in scheme 1. These reactions are governed by a complex set of rate equations and so are studied when the concentration of one substrate is held constant and the other is varied, giving pseudo-single substrate conditions that enable insight into the enzymes catalytic mechanism [157].



Scheme 1

Scheme 2

Scheme 3

where E = enzyme, E* = modified enzyme, A and B = substrates, P and Q = products.

These bi-bi enzymes utilise either ternary complex formation (scheme 2) or ping-pong mechanisms (scheme 3). In some cases, the 'EAB' ternary complex requires a specific order of substrate binding for catalysis to take place i.e. substrate A must bind before B, whereas for other enzymes the binding can be in a random order with either A or B binding to the unbound enzyme. In the case of the ping pong mechanism, A binds and is converted into product P leaving a modified E* form of the enzyme. This is then in the correct conformation for substrate B to bind and be converted into product Q to give the original form of the enzyme, E.

By performing substrate matrix experiments (varying substrate A at various, fixed concentrations of B and *vice versa*) and comparing the fit of the data to the alternative bi-bi kinetic equations governing ternary complex or ping-pong mechanisms (equations 2 and 3) using non-linear regression modelling, the most plausible mechanism can be elucidated [157, 158]. Additionally, if the first product, P, is observed when the enzyme is incubated with only substrate A, then a ping pong mechanism must be occurring.

Michaelis-Menten:

$$v = \frac{V_{max}[S]}{[S]+K_M} \qquad \text{Equation 1}$$

Ternary complex:

$$v = \frac{V_{\max}[A][B]}{K_{ia}K_M^B+K_M^B[A]+K_M^A[B]+[A][B]} \qquad \text{Equation 2}$$

Ping-Pong:

$$v = \frac{V_{\max}[A][B]}{K_M^B[A]+K_M^A[B]+[A][B]} \qquad \text{Equation 3}$$

where v is the initial rate, V_{max} is the maximum velocity, [A] is the concentration of the varied substrate, [B] is the concentration of the fixed substrate, K_M^A and K_M^B are the Michaelis constants of the varied and fixed substrates respectively, and K_i^a is the dissociation constant of the varied substrate.

In the case of the bi-bi phosphoglycosyl transferase (PGT) DPAGT1, nucleophilic attack of the β-phosphate of the UDP-GlcNAc by the oxyanion of dolichol phosphate results in the transfer of sugar-1-phosphate to yield a dolichol pyrophosphate linked GlcNAc [44, 159, 160]. Structural and functional studies of DPAGT1 provide strong evidence for the formation of a ternary complex, and no covalent intermediate or half reactions have been observed making a ping-pong mechanism unlikely [44]. Perhaps unsurprisingly given their similarity in structure and function,

the bacterial PGTs MraY and WecA that are essential for cell wall synthesis, are also predicted to follow this "one-step" mechanism [161-163] with the enzymes stabilising a sugar donor and lipid acceptor ternary complex.

PglC is another bacterial PGT found in *Campylobacter* that initiates the protein glycosylation pathway. It was found to have membrane architecture distinct from the polytopic phospho-glycosyltransferases mentioned previously, containing a single re-entrant helix compared to the 10 transmembrane helices of DPAGT1 and MraY [164]. Interestingly, evidence supports PglC utilising a ping-pong mechanism whereby nucleophilic attack by a key aspartic acid residue onto the sugar donor forms a covalent phosphosugar-enzyme intermediate, before binding and attack by the lipid phosphate [165]. This may be the case for other PglC-like PGTs - WbaP, PglB, and WcaJ - although detailed analysis has not been reported to date.

ALG13/ALG14 is also predicted to be a bi-bi enzyme, catalysing the conversion of UDP-GlcNAc and dolichol pyrophosphate GlcNAc into UDP and dolichol pyrophosphate $(GlcNAc)_2$, as shown in Figure 4.1. To date, the catalytic mechanism of this enzyme has not been studied, although MurG - a bacterial homolog of ALG13/ALG14 - has been shown to catalyse GlcNAc transfer during lipid II synthesis by formation of a ternary complex [166].

Accessing the full length human DPAGT1 in complex with ALG13/ALG14 likely better recapitulates the native arrangement in the ER. To compare the reaction dynamics of the enzymes alone and in complex, a suitable assay to monitor both steps in the synthesis of dolichol pyrophosphate chitobiose by DPAGT1/ALG13/ALG14 was investigated.

4.1.2 Assays to monitor glycosyl transfer to a lipid acceptor

The difference in polarity of nucleotide diphosphate sugar donors and lipid phosphate acceptor substrates has long been exploited when studying glycosyltransferases with lipid acceptor molecules [159], with separation achieved by thin layer or high performance liquid

chromatography [44, 160, 161, 167-169], or by extraction with hydrophobic polymer beads [170]. To detect the transfer, it has been common to use radioactively labelled nucleotide diphosphate (NDP) sugars, given this allows highly sensitive and quantitative recordings. The use of radioactivity is particularly useful for these assays given the need for detergents for solubilisation of membrane or membrane-associated proteins and lipid acceptor molecules, which often co-elute with the lipid linked oligosaccharides during chromatography and therefore lead to high background noise.

In 1982, Sharma *et al.* studied the activity in yeast membranes of the first three enzymatic steps in the N-glycosylation pathway (now known to be catalysed by DPAGT1, ALG13/ALG14 and ALG1,2 and 11) using UDP-[^{14}C]-GlcNAc and thin layer chromatography (TLC). K_M values for the sugar donor of 15, 10 and 7 μM were reported for the three enzymatic steps respectively, and 180, 40 and 17 μM with respect to the lipid acceptor [171]. This was the first study to identify reliance of divalent cations in these early N-glycosylation enzymes, as well as to report product inhibition. Activity of ALG13/ALG14 from mung bean seedlings was studied using UDP-[^{3}H]-GlcNAc [172], as was that in pig liver microsomes by Tai *et al.* in 2001, with extraction and scintillation counting of the separated organic and aqueous phases used to show specific activity of ALG13/ALG14 for UDP-GlcNAc over modified versions of the sugar donor when the C2-acetamido group was replaced by fluoro, ethoxy or trifluoroacetamido functional groups [173].

Similarly, UDP-[^{14}C]-GlcNAc was used for study of DPAGT1 with separation of dolichol products by TLC [44]. A K_M of 4.5 μM and k_{cat} of 0.21 min^{-1} towards UDP-GlcNAc, and 36.3 μM and 0.20 min^{-1} for dolichol phosphate was recorded. Despite enabling accurate recording of reaction rates and ligand binding, this assay format does have a number of disadvantages. Namely, it is labour intensive to run each stopped reaction by TLC; activation of a phosphor screen to visualise the radioactive TLC requires a long incubation time (3-5 days) before results

can be read; radioactive substrates are expensive to synthesise or purchase; and the use of radioactive material poses significant safety risks and therefore requires tight control over use and release of radioactivity. This makes it a low throughput choice.

The Gao group overcame some of these issues with an assay for study of ALG2 whereby mild acid cleavage of the sugar moiety from short, non-natural phytanyl (C20) polyprenol acceptors enabled separation and detection by LC-MS [34, 35]. This method was applied to study products of the ALG1, 2 and 11 complex, as well as downstream ALG3, 9 and 12 [31, 174]. Recently, the Gao group assayed ALG13/ALG14 in the same way, with heat and acid cleavage to remove the chitobiose moiety from the dolichol pyrophosphate product [68].

Others have used fluorescently labelled substrates to assay glycosyl transfer to a lipid acceptor molecule, for example both MurG and MraY have been shown to be able to accept a fluorescently modified version of their lipid-linked pentapeptide substrate [166, 175-177]. This has allowed for the development of continuous, FRET-based assays. For MurG study, the UDP-GlcNAc sugar donor was decorated with a donor fluorophore (indole-3-acetic acid) and transfer of energy was monitored when the sugar was moved to a dansylated lipid-peptide FRET acceptor [176]. Inhibitors of MraY activity were screened using a FRET-assay whereby BODIPY in the pentapeptide transferred energy to an acceptor lipid (Lissamine rhodamine B dipalmitoyl phosphatidylethanolamine) embedded in detergent micelles along with the undecaprenol substrate and MraY protein [178]. Increase in proximity of the BODIPY containing undecaprenol product to the reporter FRET lipid following transfer was used to monitor activity. This could prove useful for study of other lipopolysaccharide synthesis enzymes with pentapeptide substrates.

Coupling the transfer of glycans onto lipid molecules to other enzyme catalysed reactions has also be used to monitor activity. The Locher group utilised the promiscuous activity of an

oligosaccaryltransferase (OST) from *Trypanosoma brucei* to transfer glycan products of cytoplasmic facing (ALG1,2 and 11) and luminal (ALG3, 9, 12 and 6) mannosyl and glycosyl transferases from their dolichol anchor to a fluorescently labelled peptide which was amenable for detection via tricine gel electrophoresis [33, 43, 179]. This assay was not used for kinetic study, rather end point activity screening of the enzymes. A continuous, coupled spectrophotometric system however, was used to monitor the kinetics of phosphoglycosyl transfer in *Streptomyces* onto a membrane bound lipid acceptor, using nucleotide monophosphate kinases to yield ADP from the released NMP which could be coupled to NADH oxidation by the well-established pyruvate kinase and lactate dehydrogenase enzyme system [180].

In 2016, the Imperiali group demonstrated use of an alternative enzyme coupled system developed by Promega, known as UMP-Glo, for study of the bacterial PGTs PglC and WecA [121, 165]. UMP released by action of these phosphosugar transferases was converted into a luminescence signal, again removing the need for a fluorescent or radiolabelled sugar substrate. The assay was found to give highly comparable measurements to those recorded by use of UDP-[^{3}H]-diNAcBAc. Despite also being discontinuous, this assay greatly improves throughput compared to the radioactive TLC alternatives, and so it was decided to test use of this technology for study of the human PGT, DPAGT1. Similarly, Promega's UDP-Glo assay was explored for study of downstream ALG13/ALG14, especially interesting given there has been no reported use of the kit for study of a glycosyltransferase of this type with a lipid acceptor molecule. In this way, a glycosyltransferase assay was optimised that allowed insight into the comparative rates of the enzymes alone and in complex.

4.2 Results

4.2.1 Promega Glo assay

As discussed, the rate of incorporation of radioactive isotopes has long been the method of choice for studying phosphoglycosyl transferases, but Promega's UMP-Glo assay kit has shown promise in providing a safe and fast alternative method for monitoring PGT activity [181]. The reagents in the kit are added to terminate the PGT reaction and contain a proprietary enzyme that converts UMP (or CMP) into adenosine triphosphate (ATP). This ATP can be used by luciferase to produce bioluminescence (Figure 4.2). Luminescence can be read using a plate reader and is proportional to the UMP produced and therefore the activity of the PGT. Promega also offers variations of this assay technology specific to other nucleotide sugars; UDP, GDP, ADP and ATP-Glo kits are all available. The UMP-Glo and UDP-Glo were therefore selected for study of DPAGT1 and ALG13/ALG14 respectively.

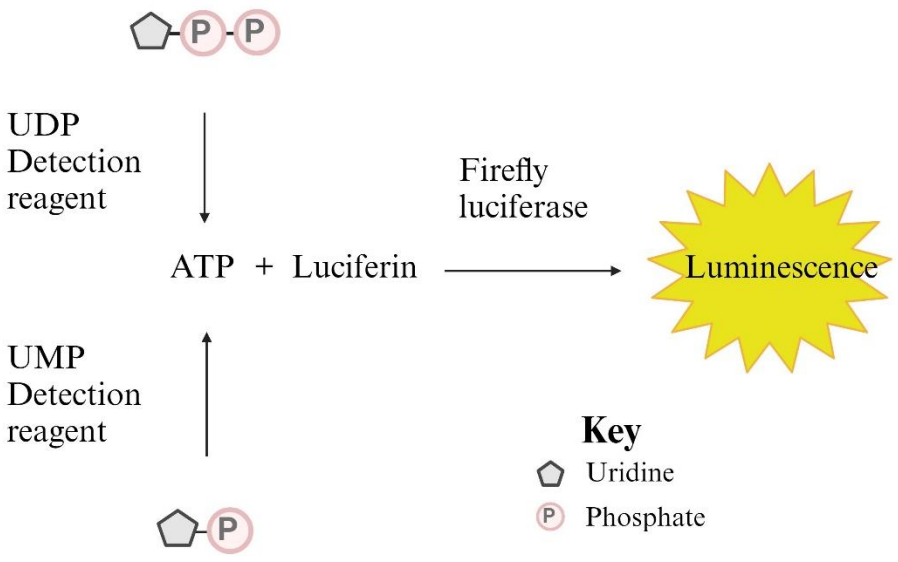


Figure 4.2 Schematic depicting the mechanism of UDP- and UMP-Glo assays.

4.2.1.1 UMP release

UMP-Glo assay development was performed using the same buffer previously used for the study of DPAGT1 [44] and with 0.27% n-Decyl-β-d-maltopyranoside (DM) supplemented with 0.027% cholesterol hemisuccinate. The assay was used in a 384 well format, with a final reaction volume of just 5 µL, making this a high-throughput option. Initially, a 6-point 1:2 dilution series of UMP from 50 µM was performed in assay buffer to test the tolerance of the UMP-detection reagents for the pH, detergent and salt composition of the assay buffer. The plate was read for luminescence in a CLARIOstar (BMG) plate reader (Figure 4.3), and a linear increase in luminescence was recorded as expected. Running a standard curve alongside all subsequent DPAGT1 assays enabled conversion of relative luminescence units (RLU) into UMP, µM.

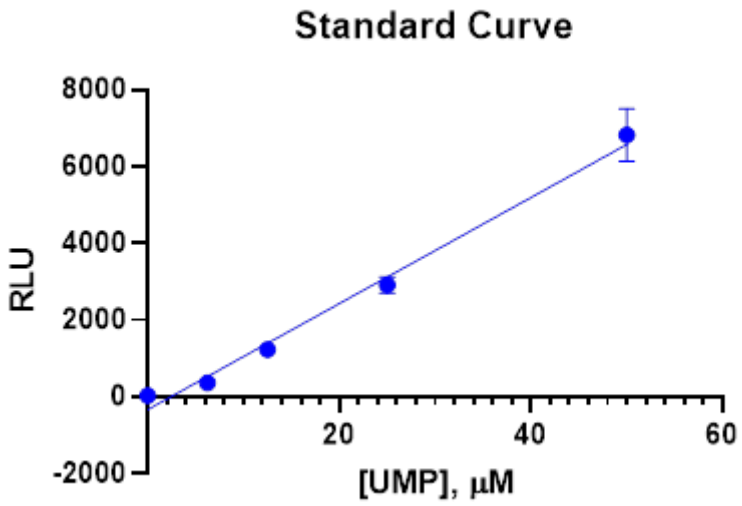


Figure 4.3 Standard curve of UMP in assay buffer detected using the UMP-Glo assay kit. RLU = relative luminescence units, as measured using the CLARIOStar plate reader. Linear regression line (blue) with $R^2 = 0.99$. N=3 with standard deviation shown.

To test the UMP-Glo in a PGT reaction, full length, untagged DPAGT1 protein that had been expressed and purified by Dr Yin Dong was used. 1 µM DPAGT1 was incubated at room temperature, with 15 µM Ultrapure UDP-GlcNAc (Promega) and 100, 50, 25 or 0 µM dolichol phosphate (C95-Dolichyl monophosphate, Larodan) in 5 µl reactions. The reaction was stopped at 2-minute intervals up to 20 minutes by addition and mixing of 1:1 by volume UMP-Glo detection reagents, and compared to control wells lacking either substrate or DPAGT1 (Figure 4.4). A zero minute time point was included whereby addition of UMP-Glo detection reagents was performed before the reaction was initiated with dolichol phosphate substrate - in theory to inactivate the enzyme before it could act. In wells containing all reaction components, luminescence was detected even at 0 minutes and there was an apparent decrease in product production over the time course. This indicated the Glo detection reagents were not stopping the reaction as suggested by the manufacturers, and instead DPAGT1 in the early time points was receiving the same opportunity to transfer as those at the end.

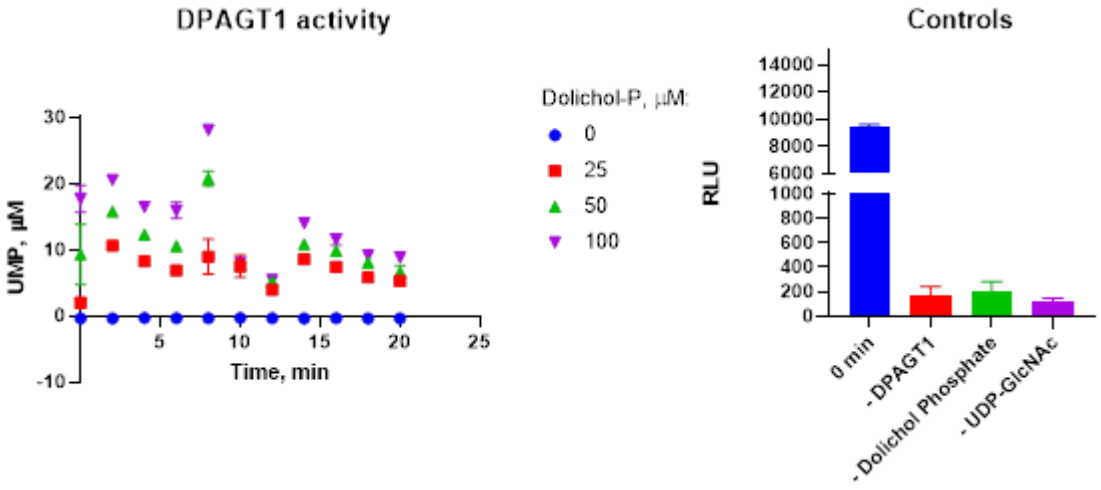


Figure 4.4 Quenching issues: DPAGT1 activity when addition of UMP-Glo kit reagents were used to stop the reaction. (Left) Apparent UMP released over 20 minutes by DPAGT1 at a range of dolichol phosphate concentrations. (Right) Relative

luminescence units (RLU) produced by negative control wells either at 0 minutes or without each reaction component.

Alternative methods of quenching the assay were therefore explored, namely denaturation of the protein by heat or lowering of the pH. It was important to test whether these conditions would interfere with the detection reagents of the UMP-Glo kit given these are also enzymes. Quenching by addition of 0.5% trifluoracetic acid (TFA) to the reactions led to a higher level of luminescence than with the Glo reagent quench (Figure 4.5), perhaps suggesting this addition did not sufficiently lower pH to denature the proteins. Fortunately, heating the reaction to 95 °C for 5 minutes proved an effective way of stopping the activity. Heating in this way was shown to have no effect on the UMP standard curve (Figure 4.5). The assay format was therefore changed, whereby a 'reverse' time course was performed by starting the reactions at sequential time points, so that they could be stopped at the same time by heating the assay plate to 95 °C in a thermocycler. The plate was then allowed to cool to room temperature before the UMP-Glo detection reagents were added. 0-minute time points were generated by addition of UDP-GlcNAc to control wells following heat inactivation.

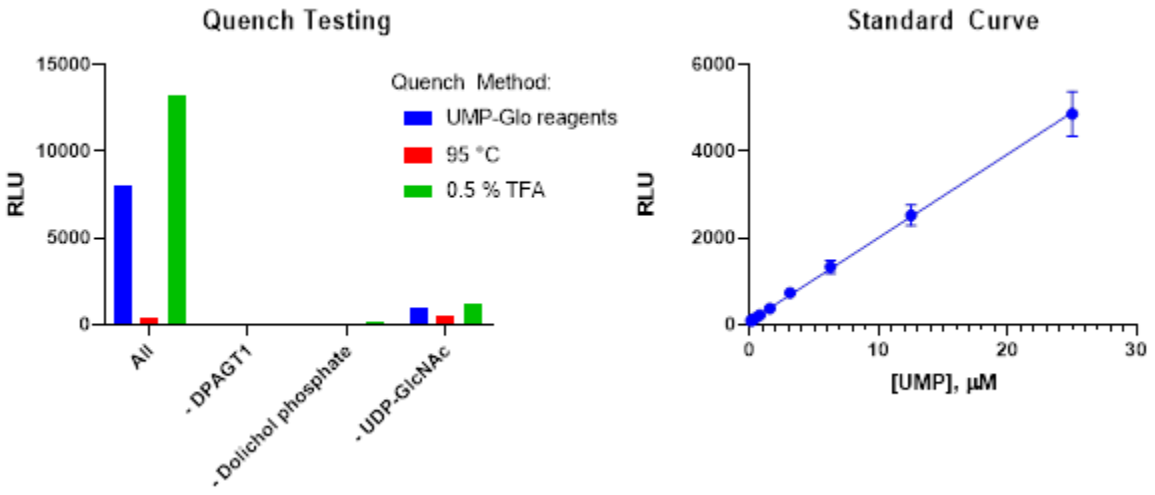


Figure 4.5 Alternative methods of quenching DPAGT activity. (Left) Relative luminescence units (RLU) detected when using the Glo reagents (blue), heat inactivation (red) or addition of 0.5% TFA (green) to terminate DPAGT1 activity, where All = all components of the assay present and subsequent control reactions without enzyme or substrates. (Right) Standard curve generated by 0-25 µM UMP that experienced the same heat treatment as assay wells, with R^2= 0.99.

Other changes were introduced to improve reproducibility of the assay. All reactions were performed at 37 °C in a heat block to control for temperature, and a brief centrifugation of the assay plate was used following heat inactivation to ensure any condensation on the lid was collected. Additionally, 10 µL PGT reactions were used rather than 5 µL which reduced variability (data not shown). Finally, the sugar donor was used to initiate reactions rather than dolichol phosphate, which was instead pre-mixed on ice with the enzyme before use in the assay.

1, 0.5 or 0.25 µM DPAGT1 with 40 µM dolichol phosphate and 40 µM UDP-GlcNAc was studied over a 20-minute time course (Figure 4.6). Activity correlated with increasing enzyme concentration as expected, and the reaction progress remained linear in this time frame. An 18-

minute time course using 1 µM DPAGT1 was therefore selected for use in determining the kinetic parameters of each of the substrates.

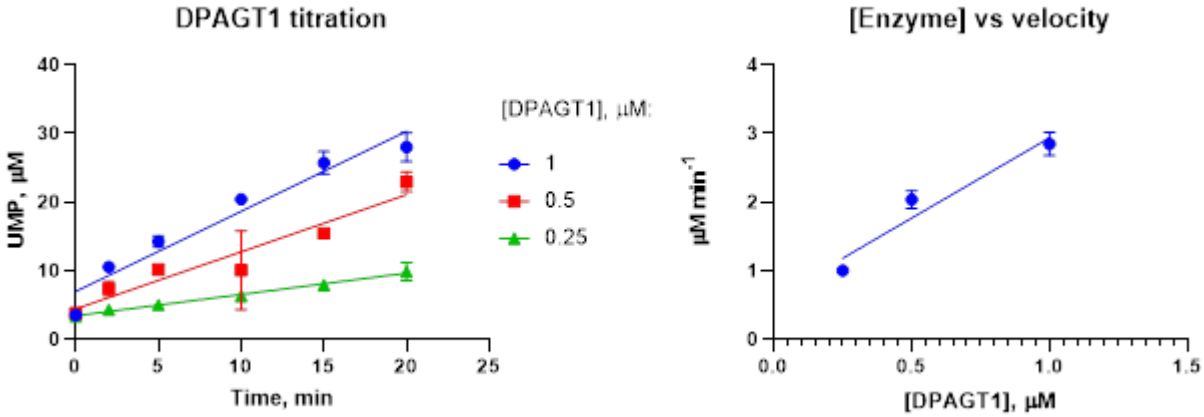


Figure 4.6 DPAGT1 titration: (Left) Activity of 1, 0.5 and 0.25 µM DPAGT1 with 40 µM UDP-GlcNAc and dolichol phosphate over a 20-minute time course, as detected using the UMP-Glo kit (Promega). R^2 of linear regression lines were 0.93, 0.83 and 0.94 for 1, 0.5 and 0.25 µM DPAGT1 respectively. (Right) Rate of DPAGT1 activity linearly increased with concentration.

Dolichol phosphate was varied from 0-50 µM, at fixed, saturating concentration (100 µM) of UDP-GlcNAc, and *vice versa*. Initial rates were plotted against substrate concentration and used to determine K_M and k_{cat} (Figures 4.7 and 4.8). Kinetic parameters found using the UMP-Glo assay kit were similar to those in the literature (Table 4.1), with apparent K_M' values (± SE) of 11.88 ± 4.92 and 10.03 ± 5.12 µM towards dolichol phosphate and UDP-GlcNAc respectively, and k_{cat} of 0.53 ± 0.08 and 0.46 ± 0.07 min^{-1}. This assay therefore provides a robust, high-throughput alternative for DPAGT1 activity study.

Table 4.1 Kinetic parameters of DPAGT1 compared to literature values, where n=3 technical replicates.

	UMP-Glo assay		Literature [44]	
	K_M, µM ± SE	k_{cat}, min^{-1} ± SE	K_M, µM ± SE	k_{cat}, min^{-1} ± SE
Dolichol phosphate	11.88 ± 4.92	0.53 ± 0.08	36.30 ± 7.20	0.20 ± 0.012
UDP-GlcNAc	10.03 ± 5.12	0.46 ± 0.07	4.50 ± 0.80	0.21 ± 0.007

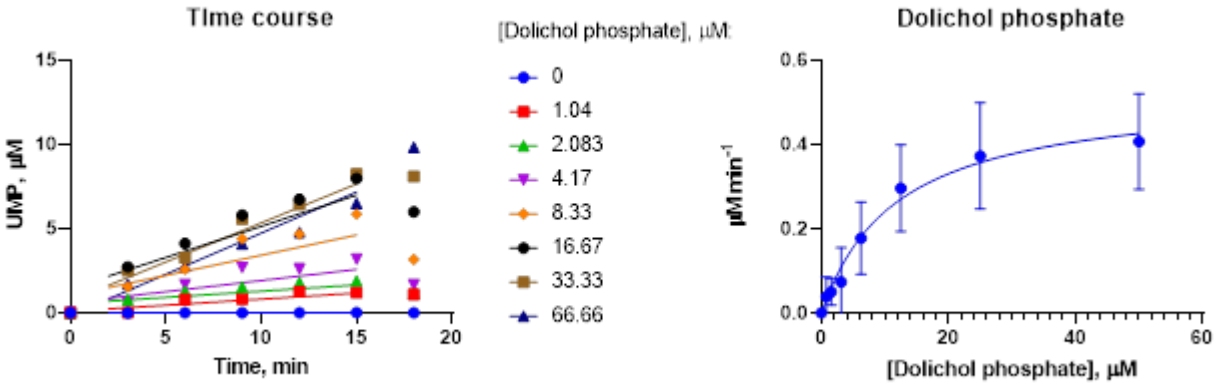


Figure 4.7 Activity of DPAGT1 determined with respect to dolichol phosphate. Representative initial rate data is shown (Left), with Michaelis-Menten plot from n=3 experimental replicates. Error bars represent 95% confidence intervals.

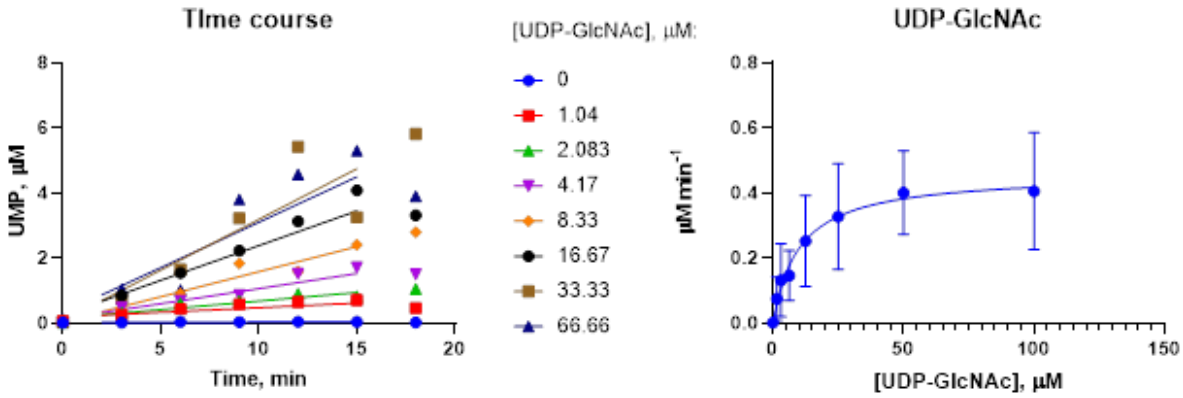


Figure 4.8 Activity of DPAGT1 determined with respect to UDP-GlcNAc. Representative initial rate data is shown (left), with Michaelis-Menten plot from n=3 experimental replicates. Error bars represent 95% confidence intervals.

Additional substrate matrix experiments were performed using various fixed concentrations of the alternative substrate. The data were modelled using the ternary bi-bi and ping pong equations in GraphPad Prism (equations 2 and 3, section 4.1.1) in an attempt to ascertain the catalytic mechanism. These results support DPAGT1 following a ternary bi-bi mechanism when compared to the ping-ping model using Akaike's Information Criteria (AICc) (Figure 4.9), with a strong preference when with respect to varied dolichol phosphate, and a moderate preferential fit with respect to varied UDP-GlcNAc. The probability the ternary bi-bi model was correct was >99.99% when dolichol phosphate was varied, and 65.56% with varied UDP-GlcNAc. Steady state parameters obtained using the ternary bi-bi model are shown in Table 4.2, albeit with large standard errors with respect to UDP-GlcNAc. These data suggest a covalent intermediate is not formed, but rather DPAGT1 stabilises a UDP-GlcNAc and dolichol phosphate bound complex to allow catalysis in a one-step nucleophilic attack.

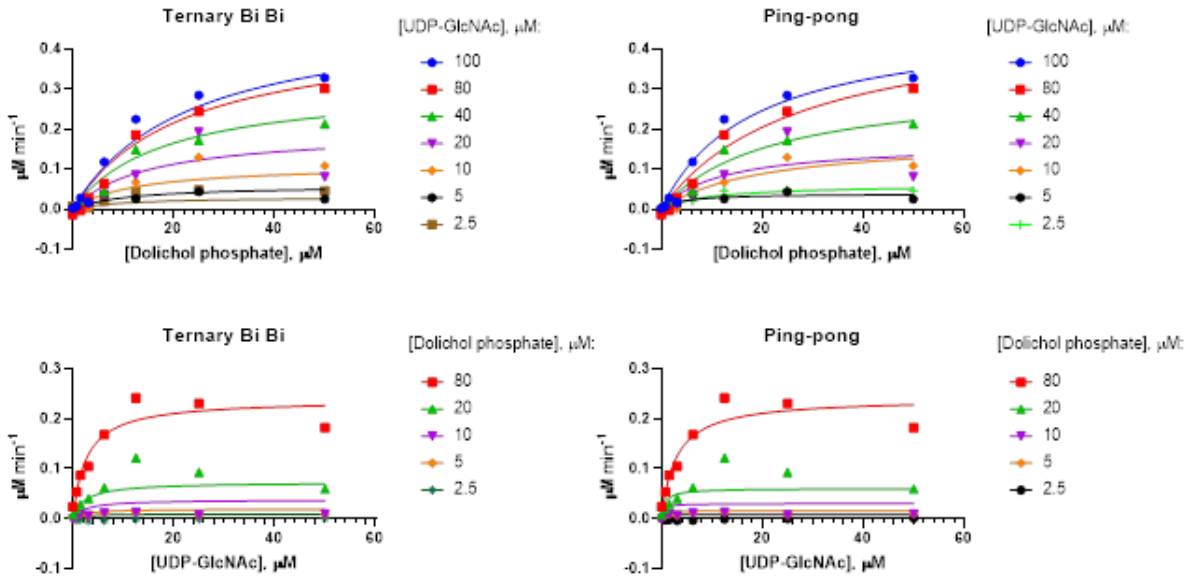


Figure 4.9 Catalytic mechanism of DPAGT1. Initial rates were fit to the ternary bi-bi (left, equation 2) or ping pong model (right, equation 3) when dolichol phosphate was varied from 0-50 μM at 2.5, 5, 10, 20, 40, 80 or 100 μM UDP-GlcNAc, or *vice versa* when UDP-GlcNAc was varied from 0-50 μM at 2.5, 5, 10, 20, or 80 μM dolichol phosphate. The goodness of fit of each model was compared using AIC in GraphPad Prism. N.b data shown is N=1.

Table 4.2. Steady state parameters of DPAGT1

	Varied substrate	**Fixed substrate**	**k_{cat}, min^{-1} ± SE**	***K_d* (varied), μM ± SE**	***K_M* (varied), μM ± SE**	***K_M* (fixed), μM ± SE**
Ternary bi-bi	UDP-GlcNAc	Dolichol phosphate	0.98 ± 0.82	2.35 ± 2.34	3.62 ± 8.91	249.60 ± 266.6
	Dolichol phosphate	UDP-GlcNAc	0.79 ± 0.23	9.53 ± 0.23	30.92 ± 15.73	60.81 ± 33.06

4.2.1.2 UDP-release

Following the successful use of the UMP-Glo kit, UDP-Glo was tested for use in studying ALG13/ALG14 activity where UDP is released from UDP-GlcNAc following transfer of the sugar to a dolichol pyrophosphate GlcNAc. In theory, when DPAGT1 and ALG13/ALG14 are studied together in complex, both GlcNAc transfers catalysed by this complex could be monitored simultaneously by exploiting the fact that DPAGT1 releases UMP, whereas ALG13/ALG14 liberates the diphosphate nucleotide. In a 10 µL reaction, 5 µL could be taken into UMP-Glo reagents and 5 µL into the UDP-Glo reagents, and in this way, quantification of UMP and UDP of each reaction could be measured at the same time.

This assay design relies on the UDP/UMP kits being selective for their respective uridine phosphate given a mixture of both will be present in each reaction time point. Therefore, standard curves were generated in the presence of a range of UDP and UMP concentrations (Figure 4.10). As before, an assay buffer of 20 mM HEPES pH 7.5, 200 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$, 0.27% DM, and 0.7 % CHS was used. UDP and UMP were titrated in an 8-point two-fold dilution series from 25 and 50 µM respectively, in the presence of 50, 25, 12.5 or 0 µM of the competing nucleotide phosphate molecule (i.e. UDP for testing of the UMP-Glo kit, and UMP for UDP-Glo kit). Both assay kits were found to be highly selective for their intended target, with similar luminescence readings for each concentration of competing nucleotide phosphate. This gave promise for use of the kits to study the dynamics of an DPAGT1/ALG13/ALG14 complex.

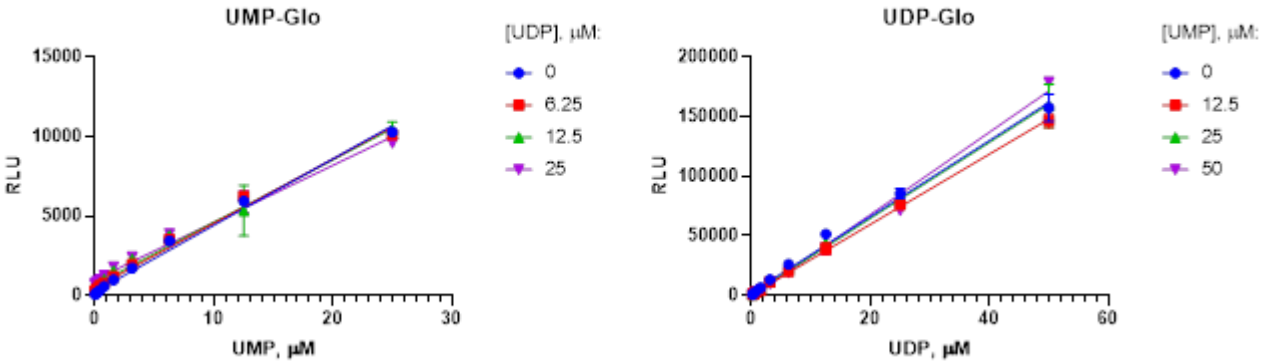


Figure 4.10 UMP and UDP standard curve in the presence of a range of competing nucleotide phosphate concentrations. (Left) The UMP-Glo kit (Promega) was used to detect UMP concentration (0-25 µM), in the presence of 0-25 µM UDP. (Right) The UDP-Glo kit (Promega) was used to detect UDP concentration (0-50 µM), in the presence of 0-50 µM UMP. RLU = relative luminescence units.

Initially, the endpoint activity of DPAGT1 and ALG13/ALG14 alone and in complex was tested alongside no enzyme and no UDP-GlcNAc controls. 1 µM of each enzyme was incubated at 37 °C with 50 µM UDP-GlcNAc and 100 µM undecaprenol pyrophosphate GlcNAc (a close mimic of the native product of DPAGT1, and substrate of ALG13/ALG14, synthesised by Dr Andrew Giltrap) or dolichol phosphate (Laradon). As with the UMP-Glo assay optimised previously, 10 µL reactions were run in triplicate and quenched using heat inactivation. Once cooled to room temperature, the reactions were split into 2 x 5 µL and added to 5 µL of UDP or UMP-Glo detection reagents. Negative control reactions without protein, UDP-GlcNAc or lipid acceptor substrate were run alongside.

Unfortunately, it became clear that for wells detected by the UDP-Glo kit there was a high background signal with similar levels of luminescence produced in each of the enzyme conditions as well as in the no enzyme control when assayed with ALG13/ALG14 substrate (Figure 4.11). This disappeared without UDP-GlcNAc suggesting that thermal degradation of

the UDP-GlcNAc stock to UDP was leading to high luminescence. As this background was not detected by the UMP-Glo assay, the decomposition of UDP-GlcNAc when heated must occur at the phosphate-GlcNAc bond and not between the phosphates. Ultimately this meant signal from ALG13/ALG14 activity could not be detected above background (Figure 4.11).

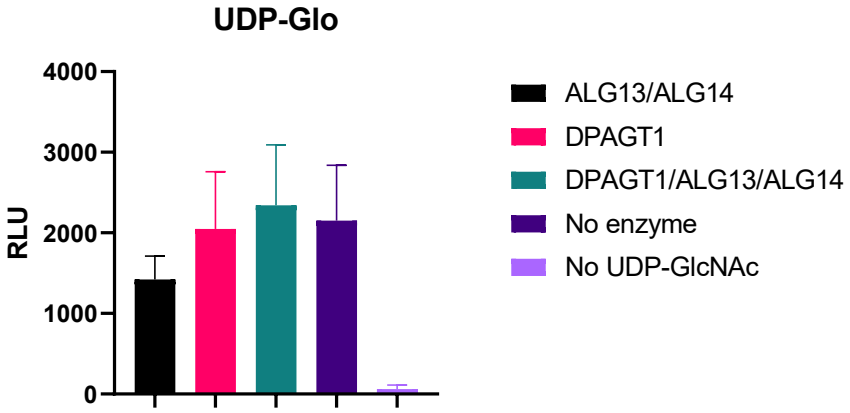


Figure 4.11 Issues with the UDP-Glo kit: endpoint UDP activity of ALG13/ALG14 and DPAGT1 alone and in complex with undecaprenol pyrophosphate GlcNAc. RLU = relative luminescence units.

4.2.2 Radioactive thin layer chromatography assay

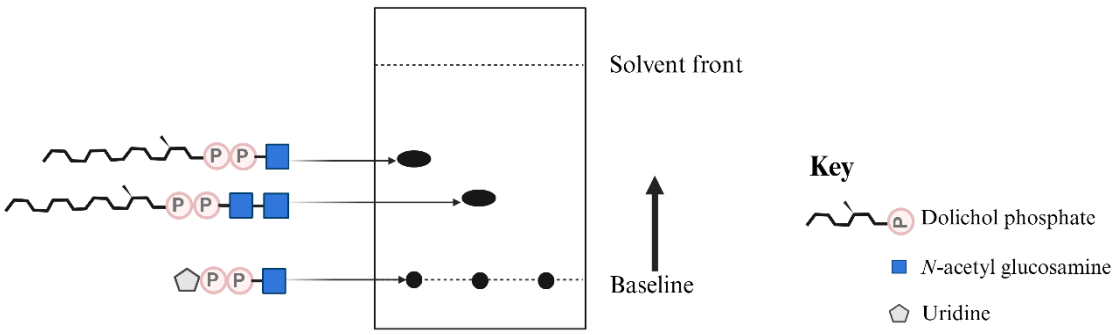


Figure 4.12 Schematic depicting movement of lipid linked oligosaccharides by TLC.

Given the issues reported above, the UDP-Glo kit was not suitable for monitoring ALG13/ALG14 kinetics and so an alternative assay was sought. As discussed, DPAGT1 was previously studied by the Davis and Carpenter labs using a ^{14}C labelled UDP-GlcNAc sugar donor. Radioactivity can be monitored moving from UDP-[^{14}C]-GlcNAc to dolichol phosphate using TLC (Figure 4.12). The polar sugar donor remains on the baseline due to its strong interaction with the silica stationary phase, whereas the lipid species are moved in the organic mobile phase up the plate. It was therefore decided to optimise this assay for monitoring of dolichol pyrophosphate $(GlcNAc)_2$ production by ALG13/ALG14. Given the increase in polarity expected when a second GlcNAc is added to the dolichol pyrophosphate GlcNAc by ALG13/ALG14, it was hoped this assay would be suitable for monitoring both steps of GlcNAc transfer when the proteins were in complex with DPAGT1.

A titration of UDP-[^{14}C]-GlcNAc was performed to assess the sensitivity of detection by phosphor screening. UDP-[^{14}C]-GlcNAc at 0.0037 MBq/µL was diluted two-fold from 80 µM to 1.25 µM before 1 µL of each concentration was spotted onto a TLC plate and run in a mobile phase of $CHCl_3$:MeOH:H_2O (65:25:4). The silica plate was allowed to dry in a fume hood before being wrapped in plastic cling film and loaded into a light-free cassette with a phosphor imaging screen. The screen was exposed to the plate for 5 days before reading for phosphorescence on a Typhoon scanner. The resulting image file was analysed using ImageJ software, to count intensity of radioactivity in each lane. Counts produced from the standard curve of UDP-[^{14}C]-GlcNAc were linear as expected (Figure 4.13), with detection even at the lowest stock concentration.

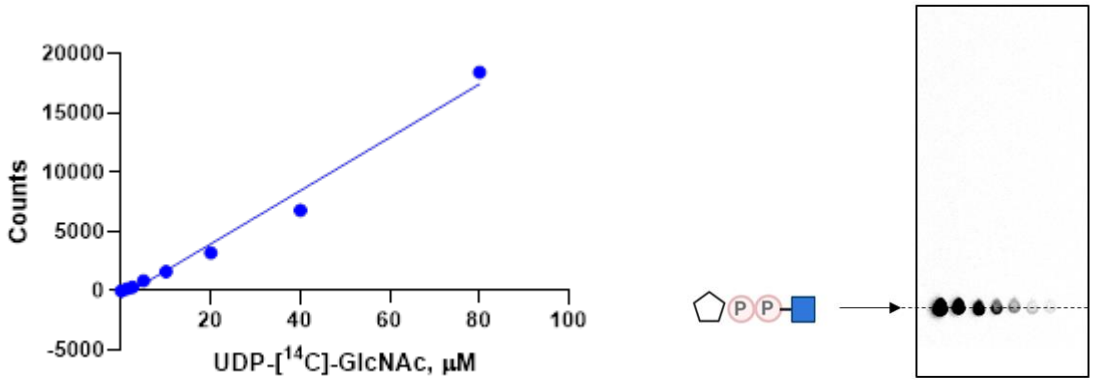


Figure 4.13 TLC assay: standard curve UDP-[^{14}C]-GlcNAc. (Left) A 7-point, 2-fold dilution series of UDP-[^{14}C]-GlcNAc from 80 µM, separated by TLC with a mobile phase of $CHCl_3$:MeOH:H_2O (65:25:4), $R^2 = 0.98$. (Right) Image of radioactivity detected on the TLC plate, with UDP-[^{14}C]-GlcNAc labelled on the baseline.

Initially, the endpoint activity of 2 µM ALG13/ALG14, DPAGT1 and DPAGT1/ALG13/ALG14 was tested to screen the best solvent system for separation of the lipid linked oligosaccharide products. The enzymes were incubated with 100 µM dolichol phosphate or dolichol pyrophosphate GlcNAc and 50 µM UDP-[^{14}C]-GlcNAc at 37 °C for 1 hour, before addition and mixing of methanol (1:1 v/v) to quench the reactions. 1 µL samples of each were spotted onto two silica plates and chromatographically separated in either chloroform: methanol: water (65:25:4) as used previously for study of DPAGT1 activity [44], or ethyl acetate: isopropanol: water (3:2:1), selected based on solvent systems used during the chemical synthesis of dolichol pyrophosphate GlcNAc by Dr Andrew Giltrap (see Chapter 5 for details). Transferase activity was observed for both ALG13/ALG14 and DPAGT1 with their respective substrates, and clear separation between the one and two GlcNAc lipid species was achieved using EtOAc:IPA:H_2O with less of the streaking which was seen for the two GlcNAc species in $CHCl_3$:MeOH:H_2O (Figure 4.14). This was therefore used for all subsequent assays monitoring ALG13/ALG14 activity. Excitingly, the complex of both

enzymes produced dolichol pyrophosphate chitobiose when supplied with dolichol phosphate, indicating a successful transfer of DPAGT1 product to ALG13/ALG14 substrate.

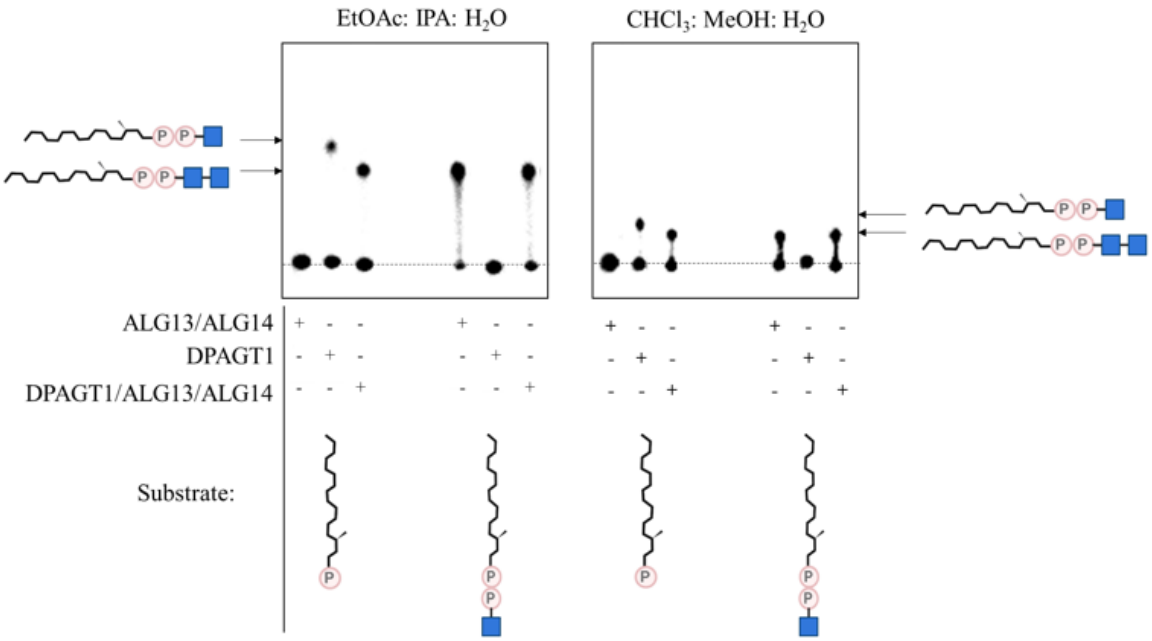


Figure 4.14 Selecting a suitable solvent: Activity of ALG13/ALG14, DPAGT1 and DPAGT1/ALG13/ALG14 with dolichol phosphate or pyrophosphate GlcNAc, separated by EtOAc: IPA: H_2O, 3: 2: 1 (left) or $CHCl_3$: MeOH: H_2O, 65: 25: 4 (right).

2 µM DPAGT1 was assessed over a 60-minute period at 37 °C as in the UMP-Glo assay. Samples from each 5-minute time point were subject to separation by TLC in EtOAc: IPA: H_2O, 3: 2: 1. For monitoring production of this mono-GlcNAc lipid species, this new solvent system gave excellent separation, and dolichol pyrophosphate GlcNAc production appeared linear across the time fame, despite some variability between time points (Figure 4.15).

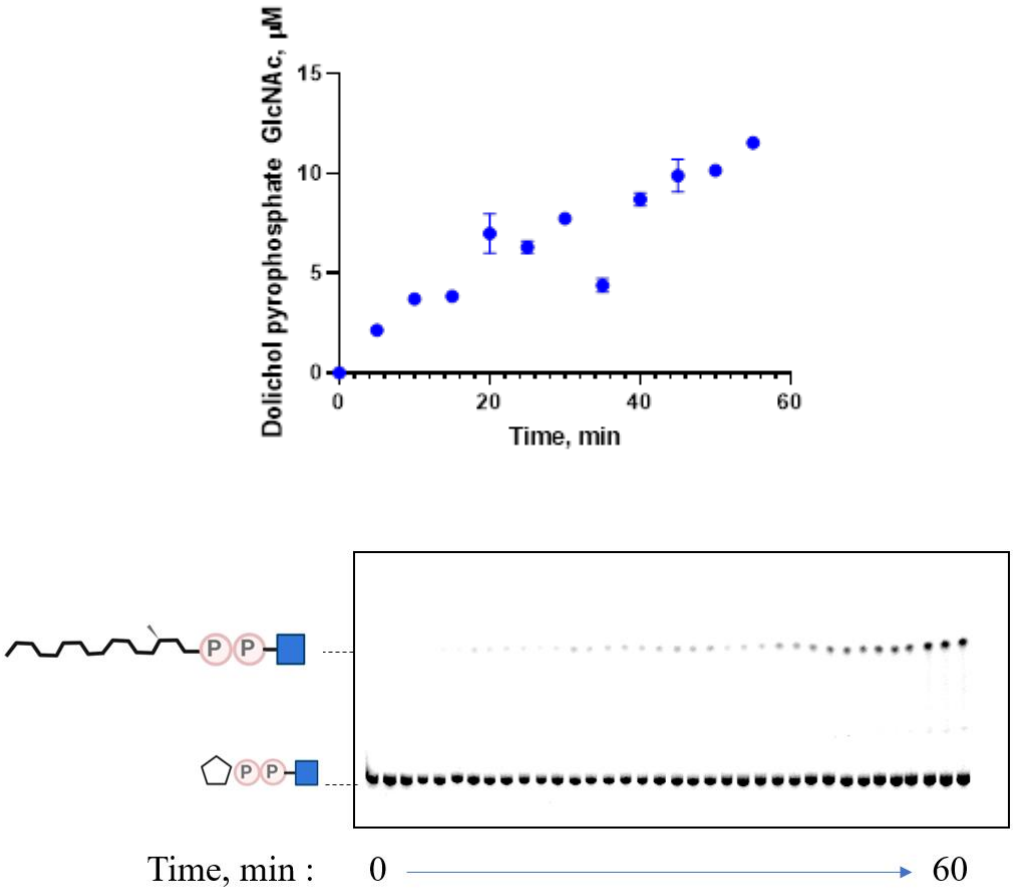


Figure 4.15 DPAGT1 time course. (Top) Dolichol pyrophosphate GlcNAc production by DPAGT1 with 100 µM of UDP-[^{14}C]-GlcNAc and dolichol phosphate, as separated by EtOAc: IPA: H_2O, 3: 2: 1 (right). (Bottom) Radioactivity detected on the TLC plate.

With this, the ligand binding kinetics of 1.5 µM DPAGT1 was assessed with saturating fixed substrate (100 µM dolichol phosphate or 50 µM UDP-[^{14}C]-GlcNAc) at 1.56 µM to 200 µM of the varied substrate. The amount of dolichol pyrophosphate GlcNAc produced after 20 minutes was used to calculate the initial velocity. Results of two biological replicates (using protein purified from two separate expressions) are summarised in Figures 4.16, with kinetic parameters reported in Table 4.3. Unfortunately, the K_M for dolichol phosphate could not be reliably determined from these data given the noise at the lowest substrate concentrations.

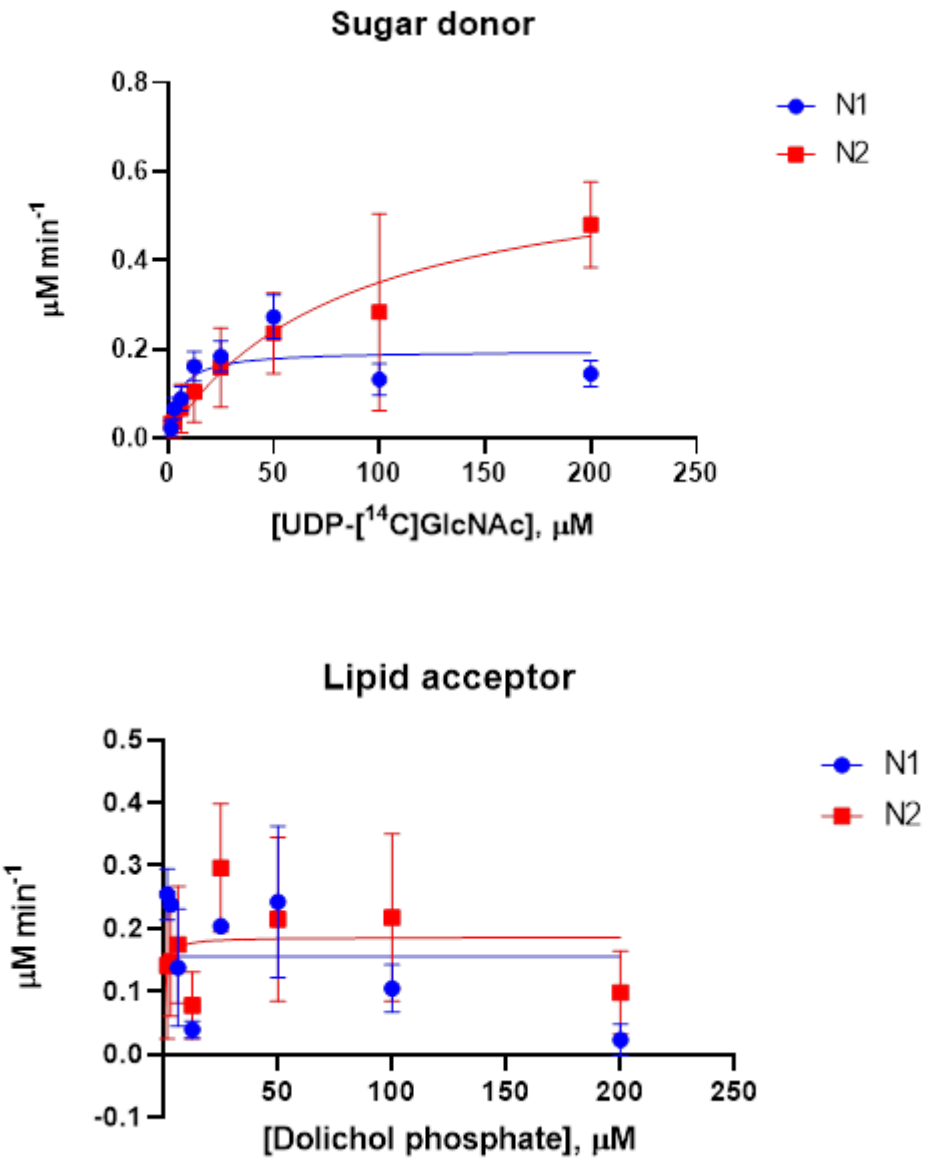


Figure 4.16 DPAGT1 ligand binding kinetics (N=2). 1.5 µM DPAGT1 was incubated with 100 µM fixed substrate and 200 µM to 1.56 µM of varied. (Top) Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters where possible, as in Table 4.2.

Table 4.3 DPAGT1 kinetics by radioactive TLC (N=2)

	K_M, μM ± SE	V_{max}, μM min^{-1} ± SE
UDP-GlcNAc	76.28 ± 40.56	0.68 ± 0.16
Dolichol phosphate	n/a	0.20 ± 0.03

4.2.2.1 ALG13/ALG14 activity

Next, methods to monitor ALG13/ALG14 activity over time were developed. First, to identify an optimum working concentration of ALG13/ALG14, the enzyme was titrated in a 6 point 1:2 dilution series of protein from 0.063-2 μM and incubated with an excess of substrates (50 and 100 μM of UDP-[^{14}C]-GlcNAc and dolichol pyrophosphate GlcNAc respectively) for 25 minutes at 37 °C (Figure 4.17). Time points were taken every 5 minutes by addition and mixing of 1:1 (v/v) of methanol. The radioactivity counts detected in the upper, product spot were normalised to the total radioactivity found in the TLC lane (i.e. baseline and upper spot combined), allowing for quantification of the production of dolichol pyrophosphate $(GlcNAc)_2$. For reactions with ALG13/ALG14 above 0.5 μM, maximum transfer was achieved after just 10 minutes. These data suggested 0.25 μM ALG13/ALG14 would be appropriate for testing reaction kinetics on this time frame.

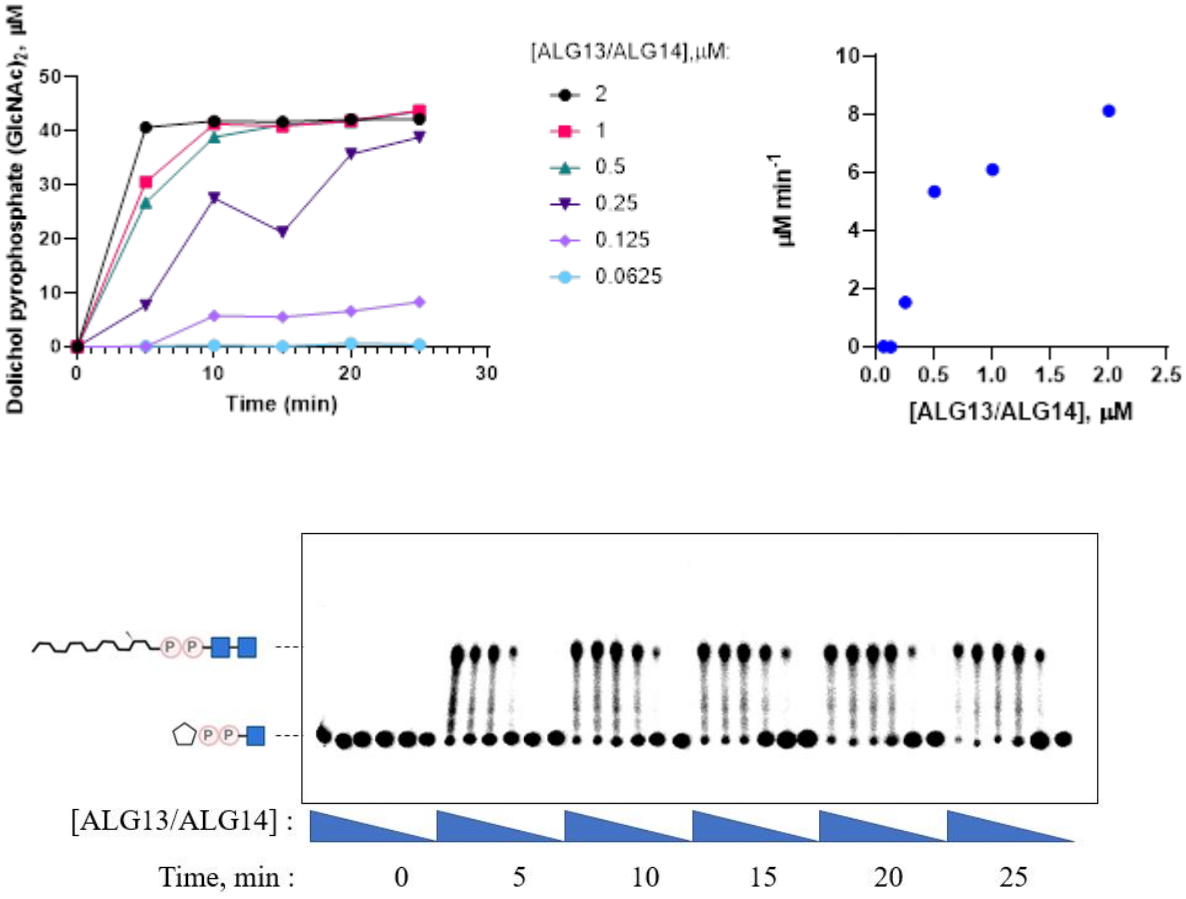


Figure 4.17 Titration of ALG13/ALG14 from 0.063-2 µM. (Top left) Dolichol pyrophosphate $(GlcNAc)_2$ production over time at a range of ALG13/ALG14 concentrations, assayed with 50 µM of UDP-[^{14}C]-GlcNAc and 100 µM dolichol pyrophosphate GlcNAc. (Top right) ALG13/ALG14 velocity against enzyme concentration. (Bottom) Radioactivity detected on the TLC plate, with schematic depicting concentration of enzyme in each lane moving from highest enzyme concentration to lowest at each time point.

However, repeated attempts to use ALG13/ALG14 at this concentration in subsequent assays resulted in reactions limited purely by substrate (Figures 4.18 and 4.19). At each dolichol pyrophosphate GlcNAc concentration, maximum product production was seen after 5 minutes, as was the case when UDP-[^{14}C]-GlcNAc was titrated at fixed lipid donor concentration. One possible explanation for this would be if the stock concentration of enzyme, [E_0], in the original titration experiment was incorrect, meaning that less enzyme was present at each concentration than expected. This would lead to the activity of ALG13/ALG14 being underestimated, and therefore result in consumption of all substrates within 5 minutes when an accurate measurement was used. This is perhaps unlikely given the experiments were performed using the same stock of purified ALG13/ALG14 meaning even an incorrect concentration should be consistent between assays. However, it is possible that given the high levels of impurities remaining after purification of the enzyme, variation existed between aliquots stored in the -80 °C freezer. Additionally, any precipitation from an unstable enzyme during the time course would lead to a lower [E_0].

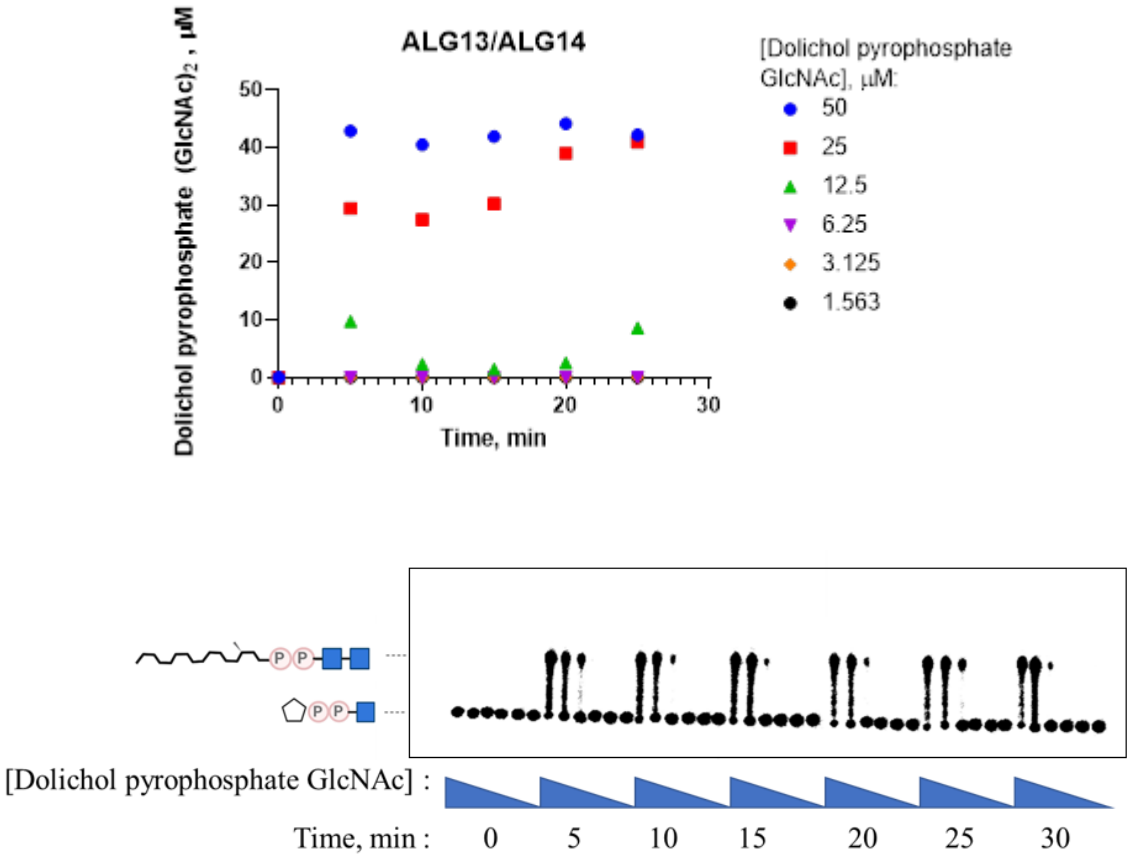


Figure 4.18 Representative results of a dolichol pyrophosphate GlcNAc titration. 0.25 µM ALG13/ALG14 was assayed over 25 minutes with 100 µM UDP-[^{14}C]-GlcNAc and 1.6- 50 µM dolichol pyrophosphate GlcNAc. Radioactivity detected on the TLC plate is shown below, with schematic representing decreasing substrate concentration at each time point.

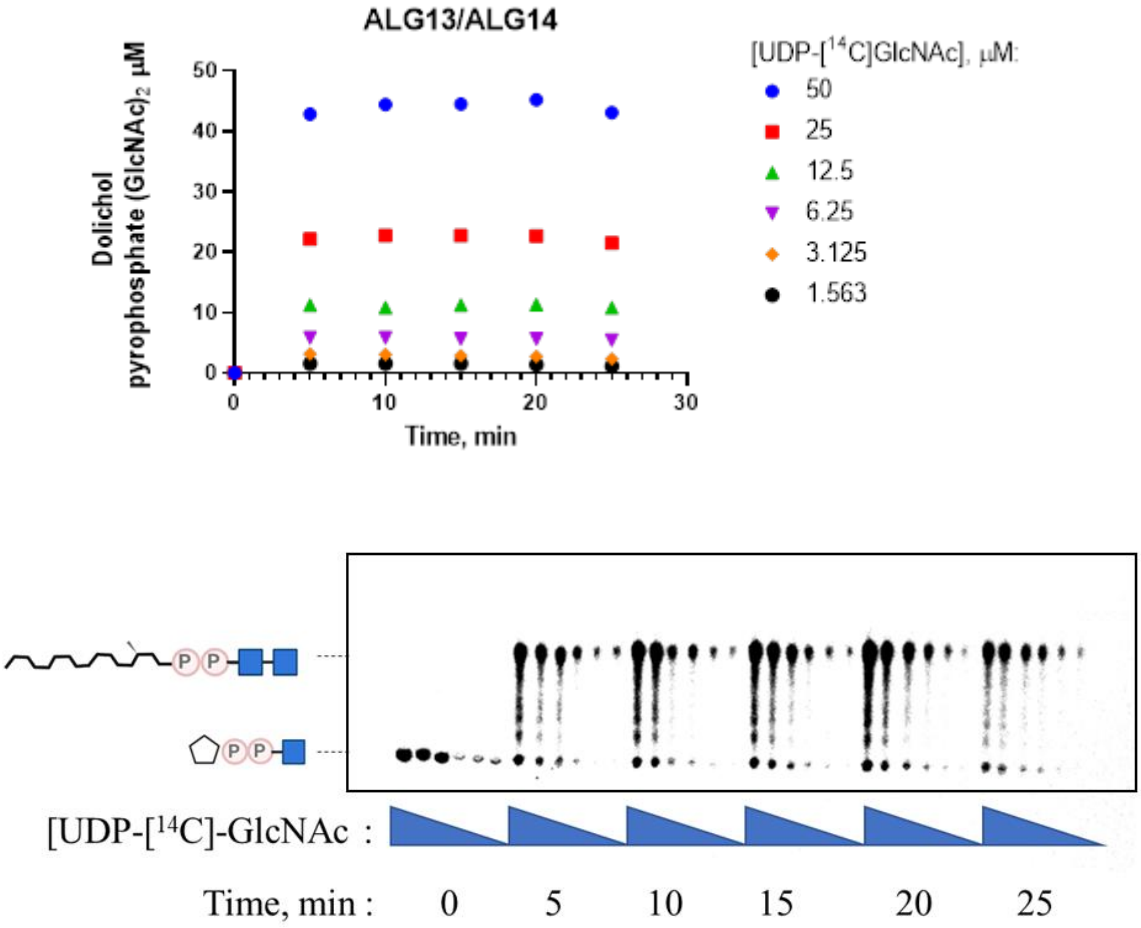


Figure 4.19 Representative results from a UDP-[^{14}C]-GlcNAc titration. 0.25 µM ALG13/ALG14 was assayed over 25 minutes with 100 µM dolichol pyrophosphate GlcNAc and 1.6 – 50 µM UDP-[^{14}C]-GlcNAc. Radioactivity detected on the TLC plate is shown below, with schematic representing decreasing substrate concentration at each time point.

To help overcome this, future preparations of ALG13/ALG14 included a high-speed centrifugation step and thorough mixing of the supernatant before aliquoting to ensure consistent ALG13/ALG14 concentrations were stored. Additionally, the stability of the enzyme during the time course was tested by monitoring the onset of protein unfolding by nano-differential scanning fluorimetry. In these experiments, ALG13/ALG14 was screened alongside DPAGT1 alone and in complex. The enzymes were monitored in assay buffer at 1

mg/mL for 1 hour at 37 °C to mimic reaction conditions, or at room temperature (22 °C). Striking increases of fluorescence were detected at 37 °C for all three stocks compared to room temperature, suggesting protein was precipitating at this temperature and therefore likely to be occurring throughout kinetic assays (Figure 4.20). One notable difference to the reaction conditions however was the lack of lipid acceptor substrate which was previously shown to increase thermostability (see Chapter 3). Despite this, it was decided to test activity of a new stock of ALG13/ALG14 at 25 °C to ensure loss of active enzyme over the time course did not occur.

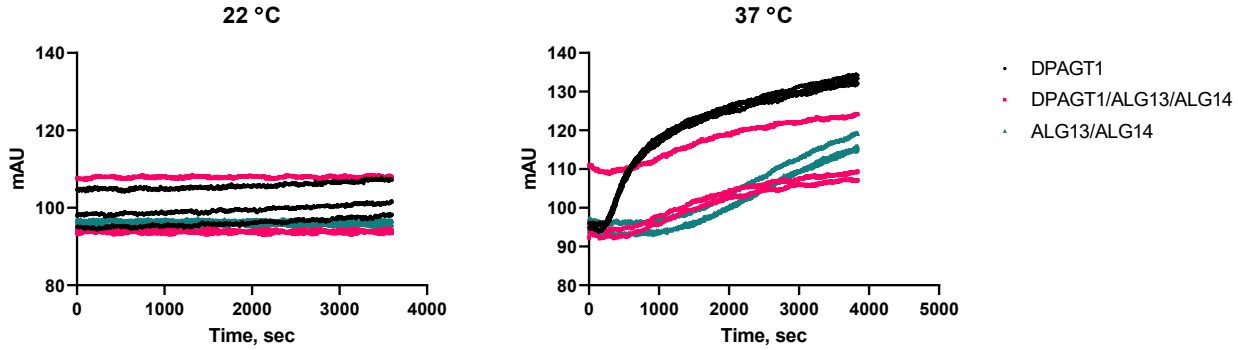


Figure 4.20 Isothermal screening by nanoDSF of ALG13/ALG14, DPAGT1 and DPAGT1/ALG13/ALG14 complex. Proteins were incubated for 1 hour at 22 °C (left) or 37 °C (right) in triplicate. mAU = arbitrary units of fluorescence, monitored at proprietary wavelengths in a Prometheus (NanoTemper) to indicate precipitation of protein in each sample.

When titrated and assayed as before with 50 µM of UDP-[^{14}C]-GlcNAc and 100 µM dolichol pyrophosphate GlcNAc, this new ALG13/ALG14 stock proved highly active (Figure 4.21). At 0.25 µM and above, the reactions were complete after 5 minutes; results that corroborate the theory that the original titration underestimated ALG13/ALG14 activity. When plotted against

velocity, it is clear that substrate is the limiting factor above an ALG13/ALG14 concentration of 0.5 µM.

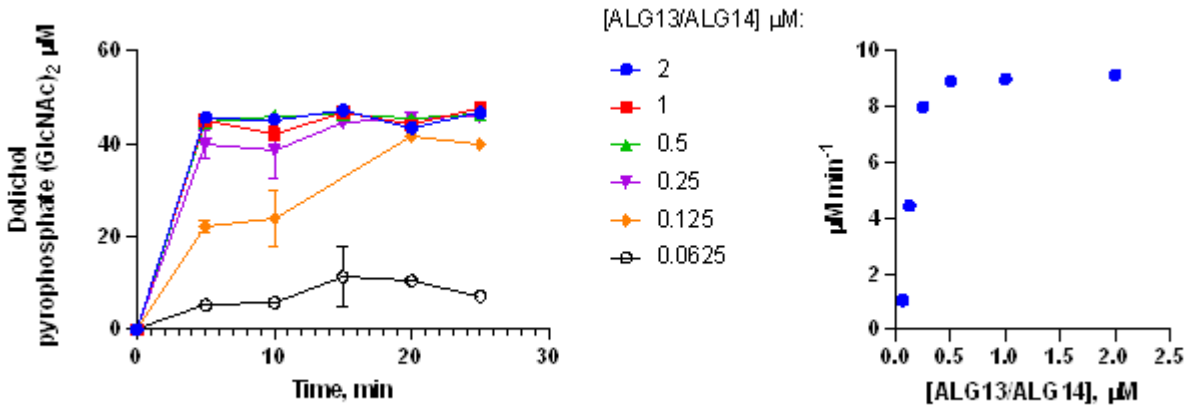


Figure 4.21 ALG13/ALG14 titration from 2-0.6 µM when assayed at 25 °C. (Left) Dolichol pyrophosphate $(GlcNAc)_2$ production over time at a range of ALG13/ALG14 concentrations, assayed with 50 µM of UDP-[^{14}C]-GlcNAc and 100 µM dolichol pyrophosphate GlcNAc. (Right) ALG13/ALG13 velocity plot against enzyme concentration.

To monitor how GlcNAc transfer by ALG13/ALG14 progressed before 5 minutes, earlier time points were taken. At 0.125 µM ALG13/ALG14 with 50 µM of UDP-[^{14}C]-GlcNAc and 100 µM dolichol pyrophosphate GlcNAc, ALG13/ALG14 had a linear phase up to just 2 minutes (Figure 4.22). This was surprising, given previous titration results with this same stock at the same substrate concentrations showed increase up to 25 minutes (Figure 4.17). Yet again there was discrepancy between activity of ALG13/ALG14 between reactions, which, given the steep increase in activity seen with just a 2-fold increase in enzyme concentration, could be due to stock concentration inaccuracies. Perhaps non-uniform mixing of the lipid substrate and enzyme stocks which are both solubilised in detergent may limit activity in some experiments.

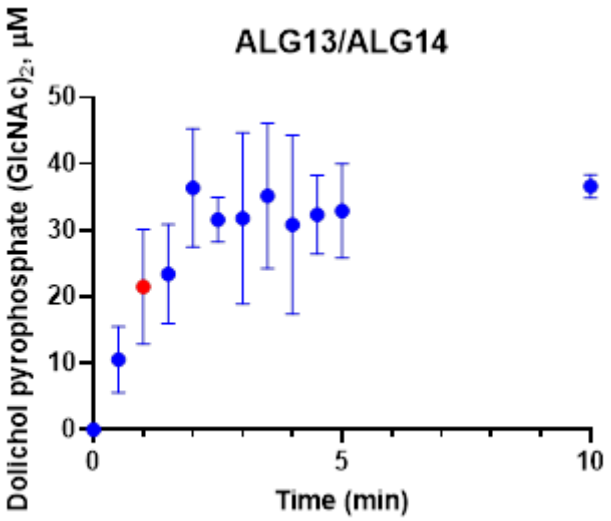


Figure 4.22 Early time points to define the linear phase of ALG13/ALG14. Dolichol pyrophosphate $(GlcNAc)_2$ production by 0.125 µM ALG13/ALG14 incubated with 50 µM of UDP-[^{14}C]-GlcNAc and 100 µM dolichol pyrophosphate GlcNAc at 25 °C. 1 minute time point is highlighted in red, as used in downstream assays.

Given linearity up to 2 minutes, it was decided to go ahead and assess ligand binding by stopping the reaction after just 1 minute. 0.125 µM ALG13/ALG14 was assessed with fixed, 100 µM dolichol pyrophosphate GlcNAc at UDP-[^{14}C]-GlcNAc from 1.56 to 200 µM, with product at 1 minute used to estimate the initial velocity. Similarly, ALG13/ALG14 was incubated with 50 µM UDP-[^{14}C]-GlcNAc and a range of dolichol pyrophosphate GlcNAc concentrations from 1.56 to 200 µM. Results are summarised in Figures 4.23 and 4.24. Michaelis-Menten curves were well defined for both substrates and a K_M' of 44.96 ± 16.42 and 13.23 ± 2.78 µM were obtained for the sugar donor and lipid acceptor substrate respectively. Due to concerns with defining a precise [E_0], k_{cat} values are not reported, instead maximum velocities of 36.00 ± 4.94 and 18.47 ± 1.07 µM min^{-1} were found (Table 4.4).

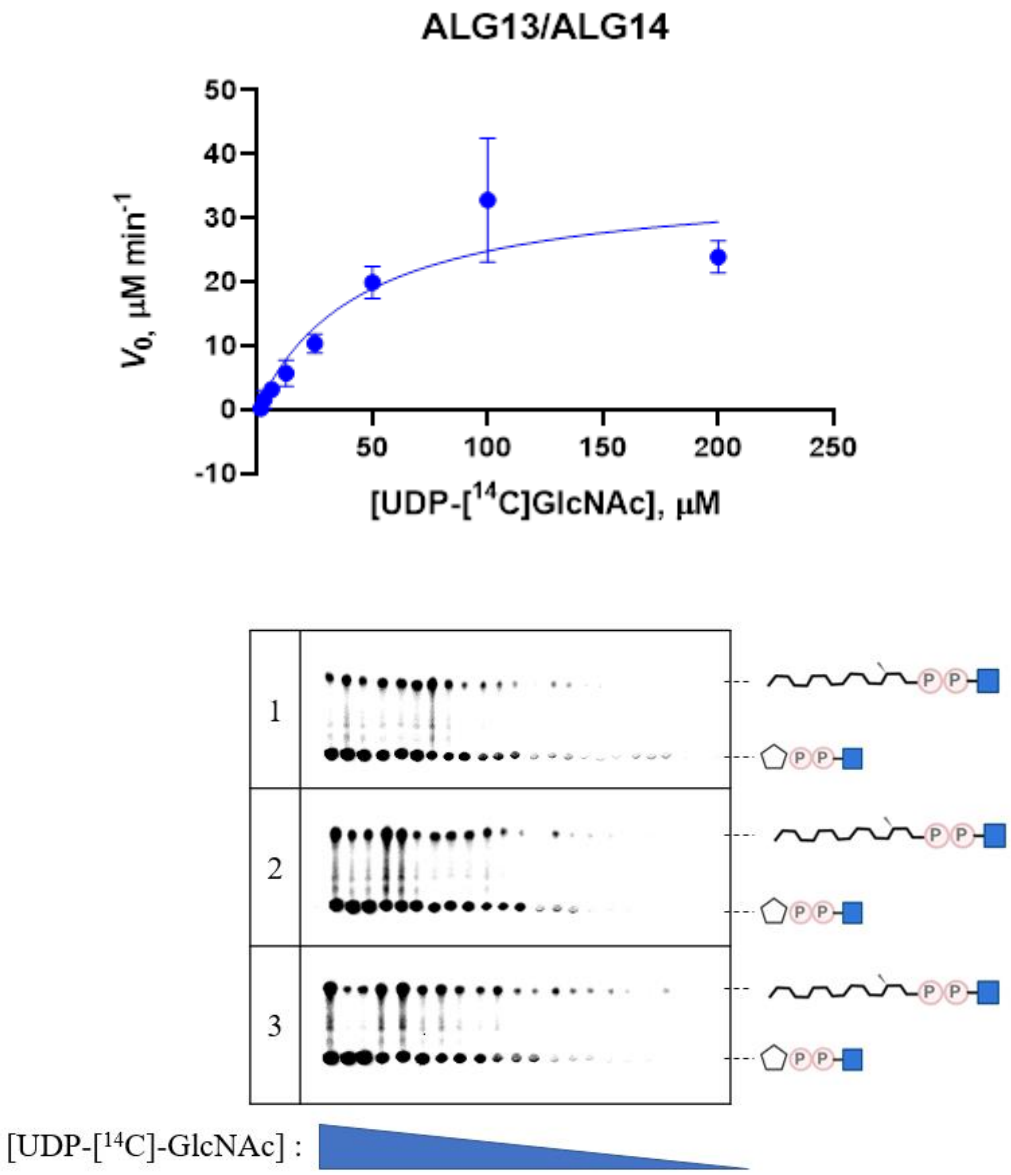


Figure 4.23 Sugar donor kinetics with ALG13/ALG14. 0.125 µM ALG13/ALG14 was incubated with 100 µM dolichol pyrophosphate GlcNAc at UDP-[^{14}C]-GlcNAc from 1.56 µM to 200 µM. (Top) Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters, as in Table 4.4. (Bottom) Radioactivity detected on TLC plates of three technical replicates, with schematic representing the titration of the sugar donor from high to low.

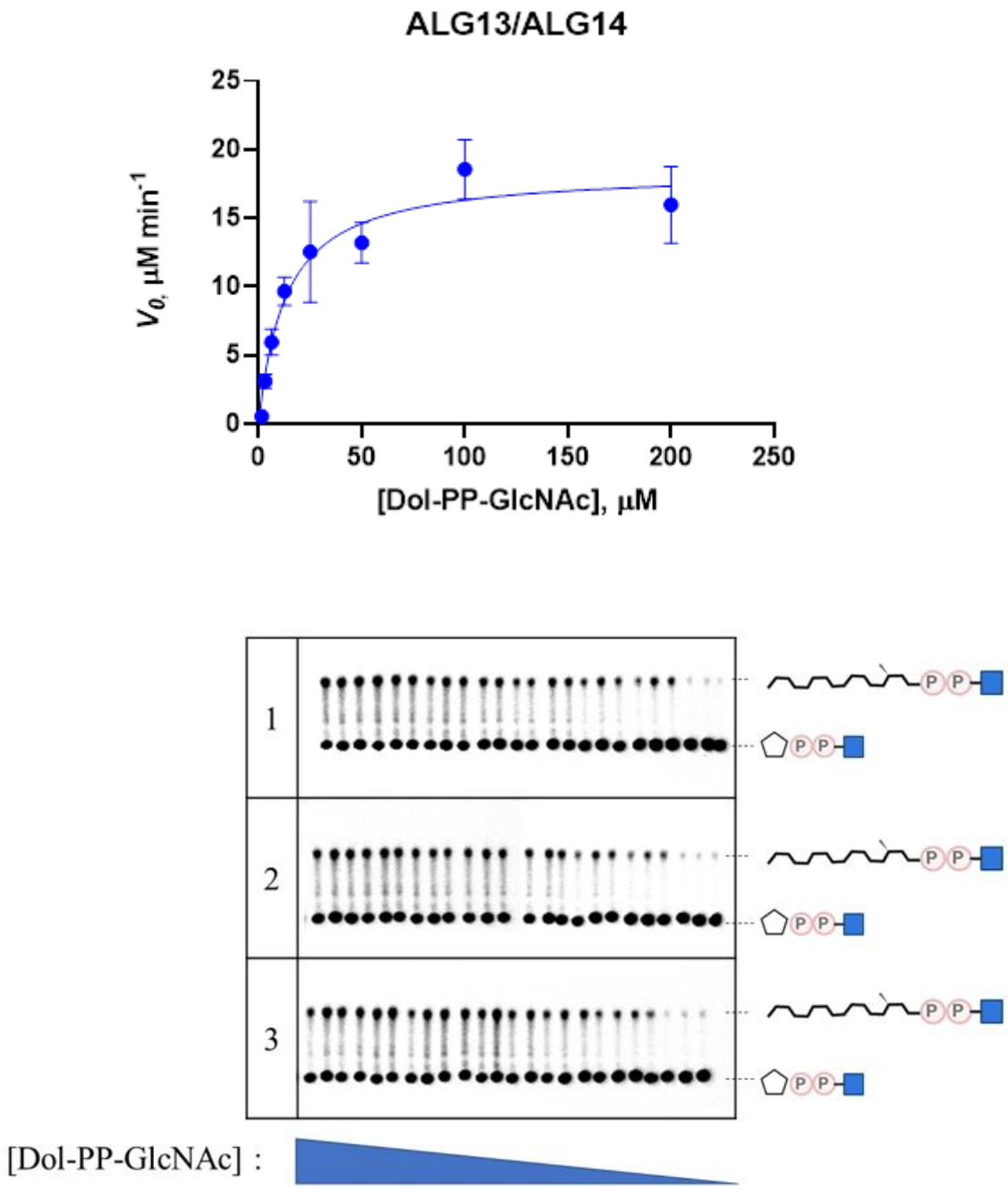


Figure 4.24 Lipid acceptor kinetics with ALG13/ALG14. 0.125 µM ALG13/ALG14 was incubated with 50 µM UDP-[^{14}C]-GlcNAc and a range of dolichol pyrophosphate GlcNAc (Dol-PP-GlcNAc) concentrations from 1.56 µM to 200 µM. (Top) Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters, as in Table 4.4.

(Bottom) Radioactivity detected on TLC plates of three technical replicates, with schematic representing the titration of the lipid acceptor from high to low.

Table 4.4 Substrate kinetics with ALG13/ALG14

	K_M, µM ± SE	V_{max}, µM min^{-1} ± SE
UDP-GlcNAc	44.96 ± 16.42	36.00 ± 4.94
Dolichol pyrophosphate GlcNAc	13.23 ± 2.78	18.47 ± 1.07

4.2.2.2 DPAGT1/ALG13/ALG14 activity

This radioactive-TLC assay was then applied to study DPAGT1 and ALG13/ALG14 when expressed and purified as a complex. As before, when the DPAGT1/ALG13/ALG14 complex was supplied with dolichol phosphate, production of both one and two GlcNAc species could be observed, with differential migration rates of the products seen by TLC (Figure 4.25). Reactions containing DPAGT1 alone where run alongside for comparison, although low levels of second GlcNAc addition were observed in these reactions too, perhaps suggesting native ALG13/ALG14 contamination with this stock.

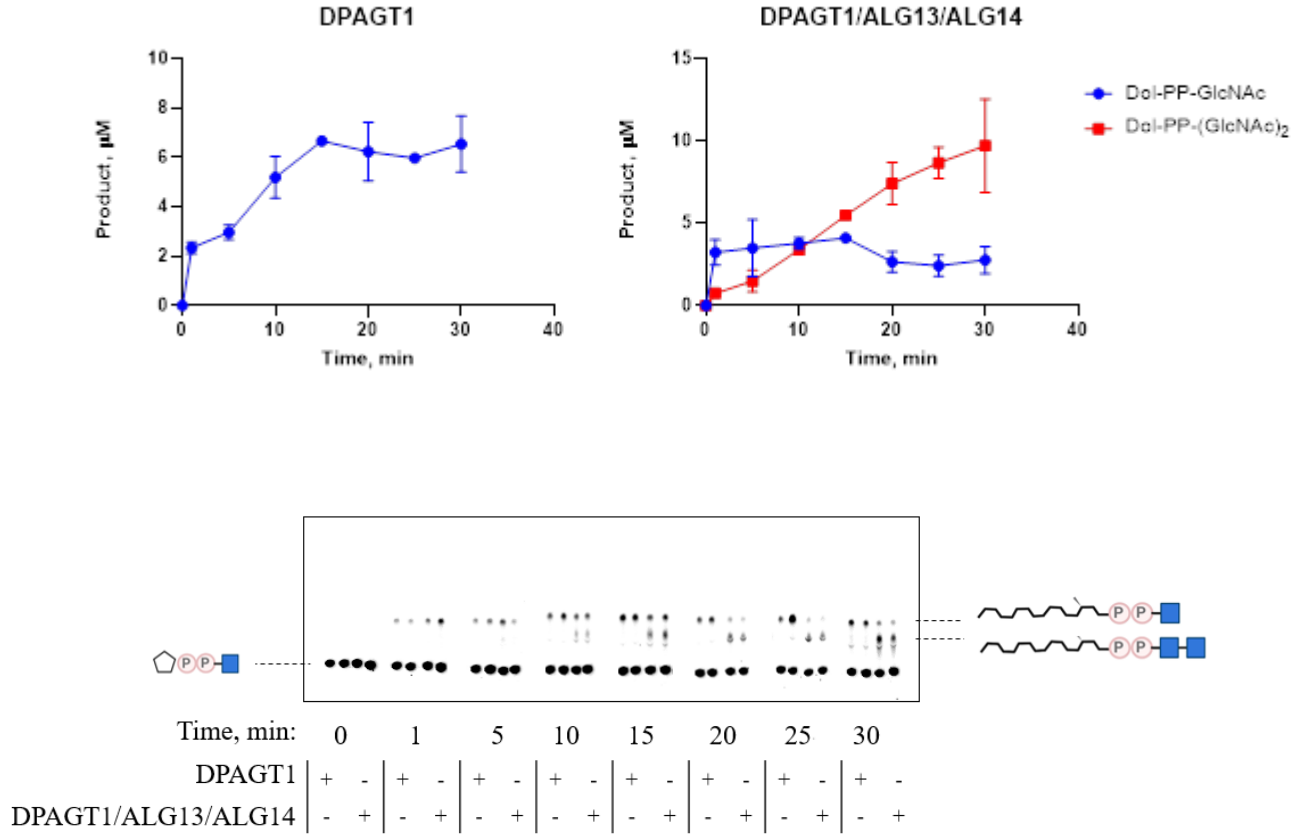


Figure 4.25 Activity of DPAGT1 and the DPAGT1/ALG13/ALG14 complex, where Dol-PP-GlcNAc = dolichol pyrophosphate GlcNAc and Dol-PP-$(GlcNAc)_2$ = dolichol pyrophosphate $(GlcNAc)_2$. 1.5 µM enzymes were supplied with 50 µM UDP-[^{14}C]-GlcNAc and 100 µM dolichol phosphate, and 2 x samples of each time point were separated by TLC in EtOAc: IPA: H_2O (3: 2: 1). A representative image of the radioactivity detected on the TLC plate is shown below.

The activity of 2 μM DPAGT/ALG13/ALG14 was analysed with 50 μM UDP-[^{14}C]-GlcNAc and 1.56- 200 μM dolichol phosphate. To aid the analysis of species on the TLC plates, labelled dolichol pyrophosphate chitobiose standards were run alongside reaction samples. These were generated by incubating ALG13/ALG14 with dolichol pyrophosphate GlcNAc and labelled UDP-[^{14}C]-GlcNAc for 30 minutes. Again, the production of both dolichol pyrophosphate GlcNAc and dolichol pyrophosphate $(GlcNAc)_2$ by the DPAGT1/ALG13/ALG14 complex was clearly observed (Figure 4.26-A). During this experiment, the zeroth minute time point was not successfully quenched and so recordings were excluded from analysis. Additionally, the phosphor screens were imaged using a Fuji FLA-5100 by kind assistance of the Ludwig Cancer Institute whilst our Typhoon reader was awaiting maintenance.

Figure 4.26-B shows quantification of the remaining time points at each dolichol phosphate concentration. Dolichol pyrophosphate GlcNAc production was seen to decline as the two GlcNAc species increases as expected. There was also a drop off in product production after 16 minutes, which may point to product instability or that the product was co-precipitating with protein.

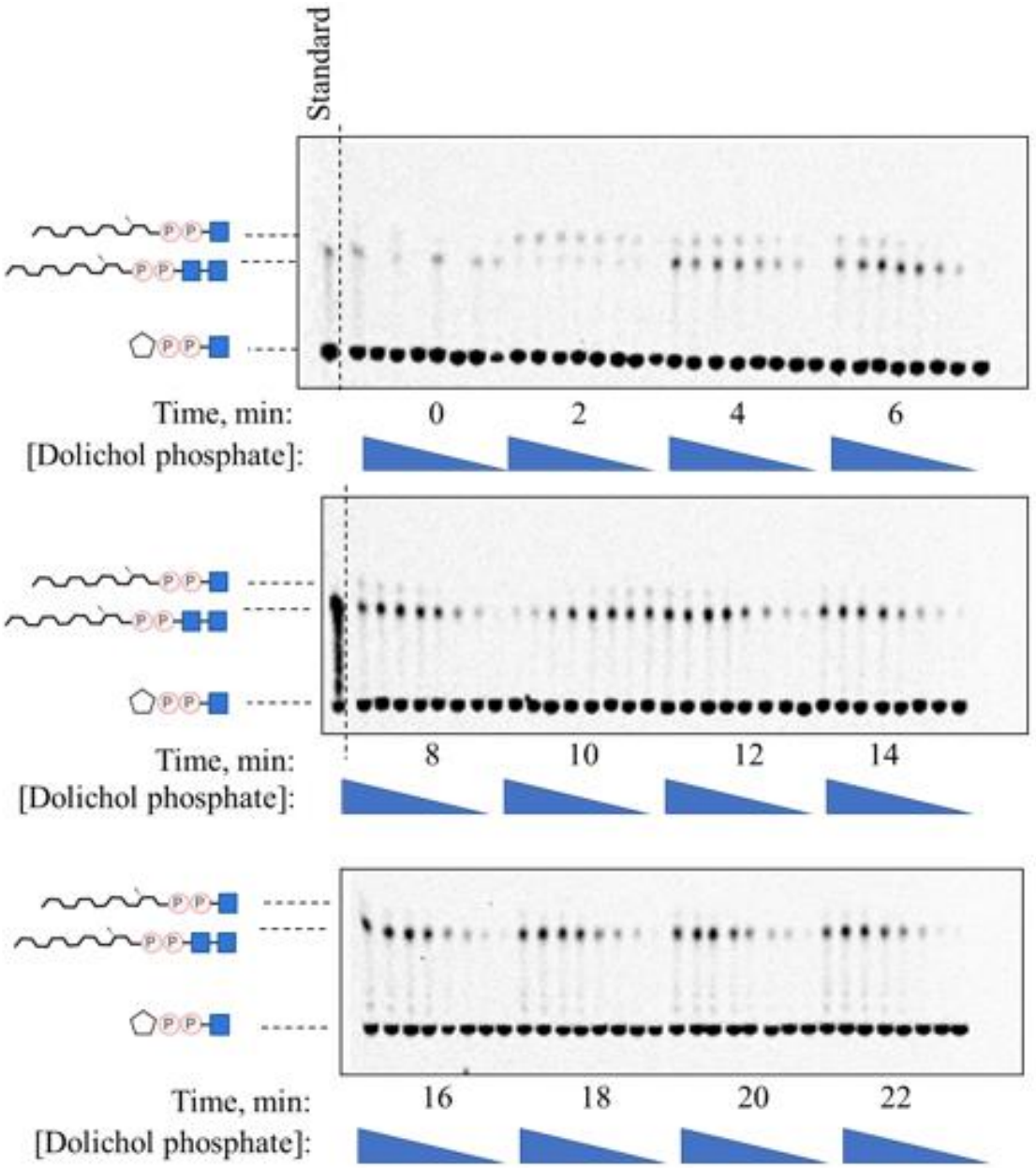


Figure 4.26(A) Time course of DPAGT1/ALG13/ALG14. Radioactivity detected on TLC plates when 2 µM DPAGT/ALG13/ALG14 was incubated with 50 µM UDP-[^{14}C]-GlcNAc and 1.56 - 200 µM dolichol phosphate. Schematics representing the titration of lipid acceptor substrate are shown below each time point. Readings from the time point at zeroth minute were excluded from analysis.

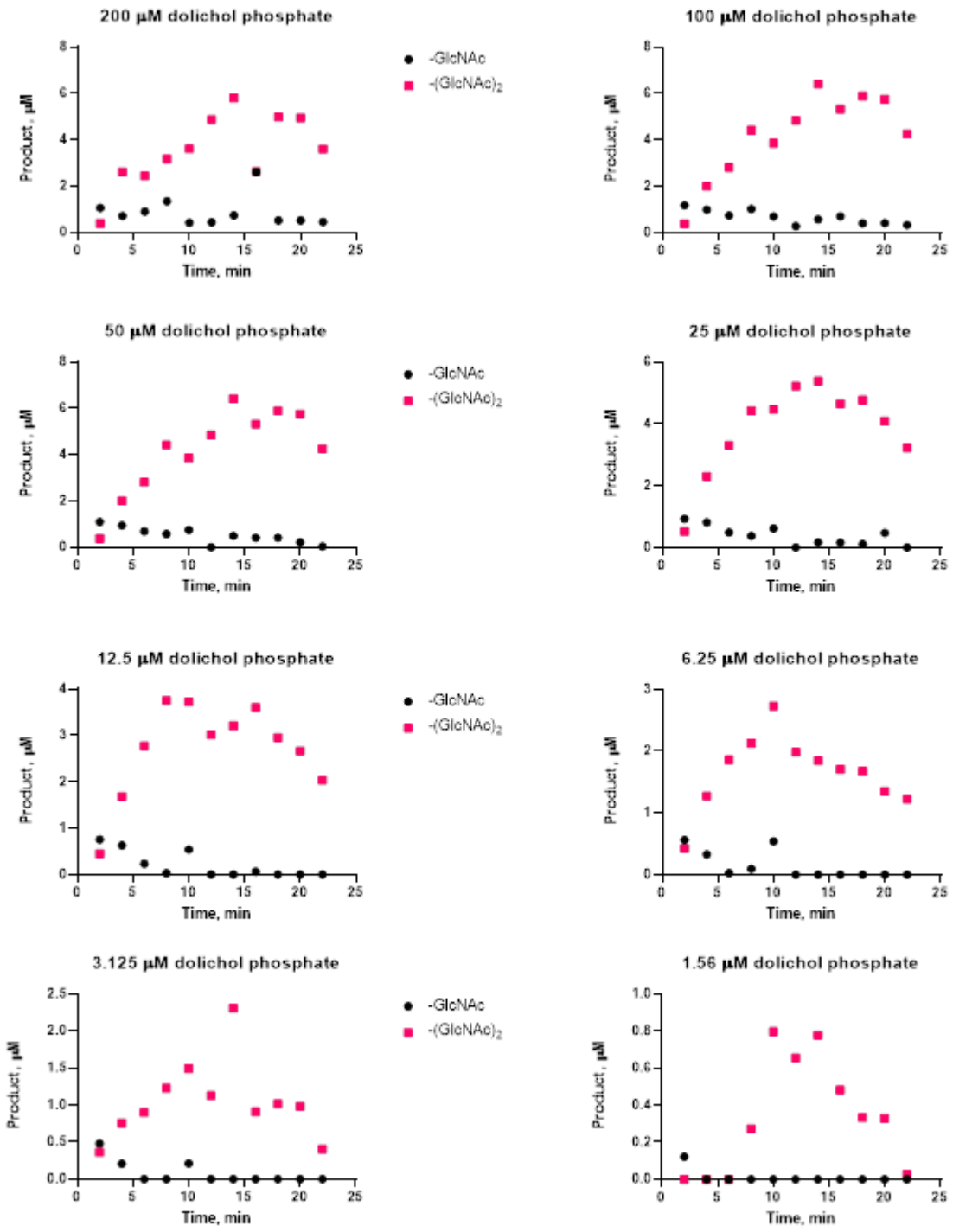


Figure 4.26(B) Time course of DPAGT1/ALG13/ALG14: GlcNAc transfer by 2 µM DPAGT/ALG13/ALG14 with 50 µM UDP-[^{14}C]-GlcNAc and 1.56 - 200 µM dolichol phosphate, where -GlcNAc = dolichol pyrophosphate GlcNAc, and –$(GlcNAc)_2$ = dolichol pyrophosphate $(GlcNAc)_2$.

In this way, the kinetics of the DPAGT1/ALG14/ALG14 complex were monitored. Given two enzymatic steps are involved, the reaction rates were compared when the enzyme complex was supplied with the first substrate, dolichol phosphate, and also with the intermediate substrate, dolichol pyrophosphate GlcNAc. The production rate of the final chitobiose product only was then compared, to give an appreciation of the comparative rates of each step. As there was a large difference in activity between DPAGT1 and ALG13/ALG14, 20- and 1-minute time points were taken when supplied with the initial or intermediate lipid substrate respectively. Furthermore, the complex was assayed at either 0.125 or 1.5 μM with each of the lipid acceptors, to allow direct comparison of parameters with those recorded when the enzymes were studied alone. Results in biological triplicate using enzyme complex purified from three separate expressions are shown in Figures 4.27 and 4.28, with resulting Michaelis-Menten parameters in Table 4.5.

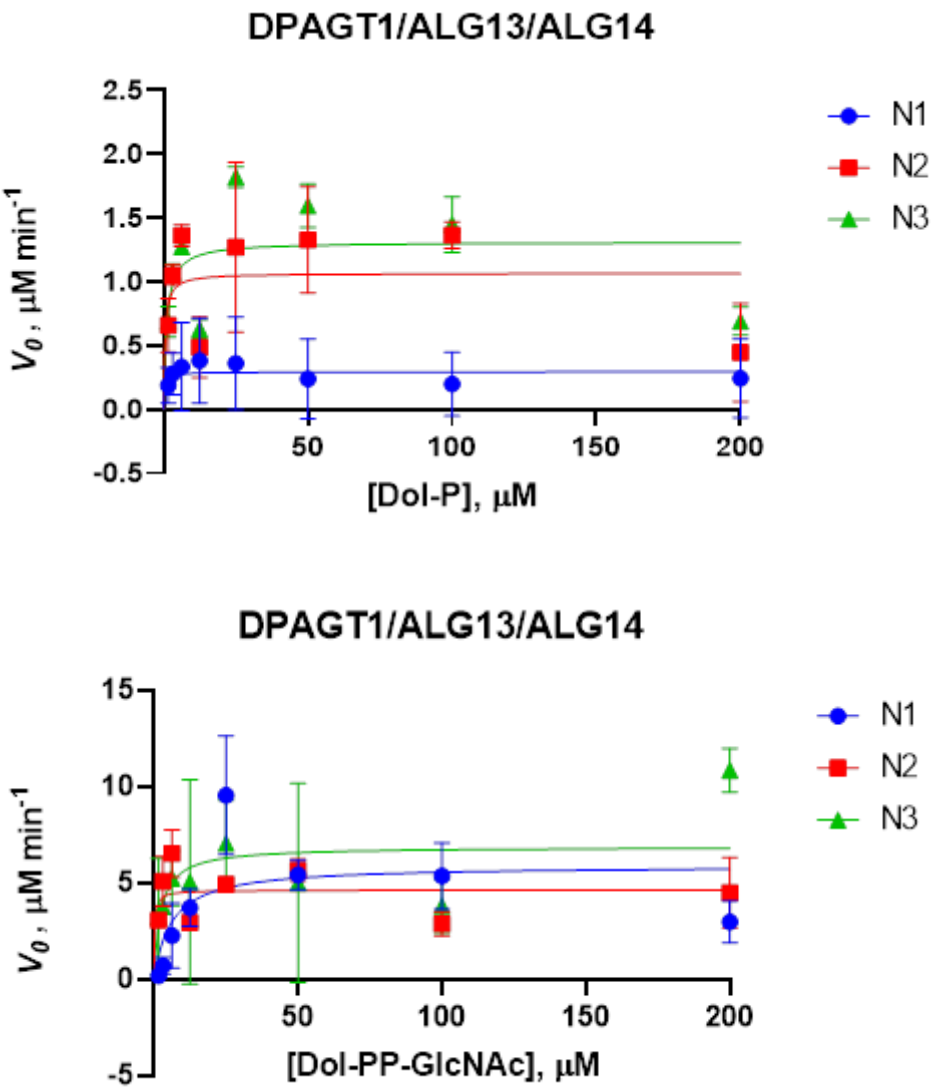


Figure 4.27 Lipid acceptor kinetics with DPAGT1/ALG13/ALG14, N=3. 1.5 µM or 0.125 µM DPAGT1/ALG13/ALG14 was incubated with 50 µM UDP-[^{14}C]-GlcNAc and a range of dolichol phosphate (Dol-P, top) or -pyrophosphate GlcNAc (Dol-PP-GlcNAc, bottom) concentrations from 200 µM to 1.56 µM. Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters, as in Table 4.5.

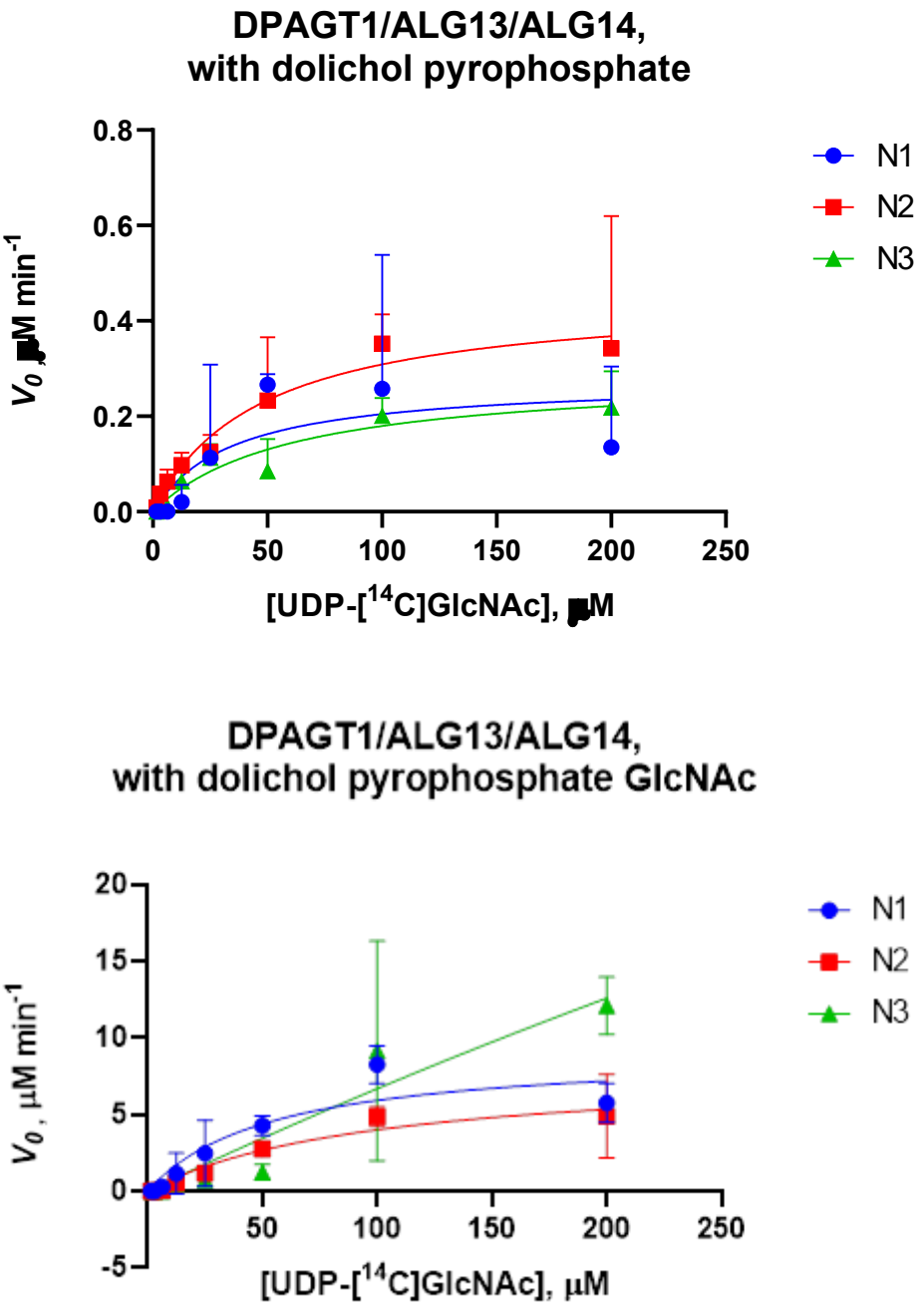


Figure 4.28 Sugar donor kinetics with DPAGT1/ALG13/ALG14. 1.5 µM or 0.125 µM DPAGT1/ALG13/ALG14 was incubated with 100 µM dolichol phosphate (Dol-P, top) or -pyrophosphate GlcNAc (Dol-PP-GlcNAc, bottom) and UDP-[^{14}C]-GlcNAc concentrations from 200 µM to 1.56 µM. Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters, as in Table 4.5.

Table 4.5 DPAG1/ALG13/ALG14 kinetics N=3 biological repeats

	K_M D, µM ± SE	V_{max} D, µM min^{-1} ± SE	K_M A, µM ± SE	V_{max} A, µM min^{-1} ± SE
Dolichol pyrophosphate GlcNAc	159.70 ± 7.31	14.79 ± 5.58	2.22 ± 1.47	5.69 ± 0.68
Dolichol phosphate	78.67 ± 58.07	0.25 ± 0.08	0.87 ± 0.60	0.84 ± 0.07

D = donor, UDP-GlcNAc, A = acceptor, dolichol phosphate or dolichol pyrophosphate GlcNAc

4.3 Discussion

In this chapter, two methods for studying the human PGT and GT enzymes, DPAGT1 and ALG13/ALG14 were explored. The use of a commercially available luminescence-based enzyme coupled assay from Promega was tested using DPAGT1 purified previously by Dr Yin Dong, and by using lipid acceptor substrate purchased from Laradon. The remaining experiments and those involving radioactive thin layer chromatography were performed using protein expressed and purified as in Chapter 3, and with full length, native dolichol substrates prepared through collaboration with Dr Andrew Giltrap. This was particularly important given the price of commercial dolichol phosphate, and also following quality issues with two batches of the dolichol phosphate received; problems that are common limitations in progressing the study of membrane associated glycosyltransferases with dolichol substrates.

4.3.1 UMP and UDP-Glo assay

To enable characterisation of DPAGT1 with the UMP-Glo kit, a number of conditions were optimised from the protocol suggested by the kit's manufacturers [181]. Most notably the need for heat inactivation as a quenching method rather than addition and mixing of the UMP detection reagent which did not stop the enzyme reaction. Promega suggest the PGT reaction is quenched through a combination of the dilution effect of addition of reagents and also from the addition of detergent. Given this assay is of a transmembrane protein and also utilises a lipid substrate, detergent was already present in the reaction buffer and this is perhaps why the reaction was not quenched. It is surprising that this was not reported when used for study of the bacterial PGTs, PglC and WecA[121, 165].

With these changes, a robust assay for DPAGT1 is presented with a reaction volume of just 10 µL in a 384 well plate. Apart from the obvious safety benefits, this assay greatly aids throughput compared to those carried out using radioactive sugar donors, allowing for multiple time points to be taken for each varied substrate concentration. This enabled initial rates to be obtained with much greater accuracy compared to the single time points used in the alternative assay which do not allow for detection of outliers.

Pleasingly, when compared to those reported previously [44], kinetic parameters obtained using the UMP-Glo assay match closely despite the difference in the detergent used for protein solubilisation (DM was used rather than OGNG). A K_M' of 11.88 ± 4.92 µM and k_{cat} of 0.53 ± 0.08 min^{-1} towards dolichol phosphate was found compared to 36.3 ± 7.2 µM and 0.20 ± 0.012 min^{-1} previously. With respect to UDP-GlcNAc, K_M' of 10.03 ± 5.12 µM and k_{cat} of 0.46 ± 0.07 min^{-1} was found compared to 4.5 ± 0.8 µM and 0.21 ± 0.007 min^{-1}. Similarly, these compare well with those found when DPAGT1 was later assayed by radioactive TLC, where a K_M' of 0.75 ± 1.00 µM and k_{cat} of 0.20 ± 0.03 min^{-1} towards dolichol phosphate and K_M' of

76.28 ± 40.56 μM and k_{cat} of 0.68 ± 0.16 min^{-1} towards UDP-GlcNAc were recorded. This gives confidence in the reliability of these measurements. This assay will therefore be of interest for those studying other PGTs with lipid acceptor substrates. Furthermore, high-throughput assays like this are required for screening of DPAGT1 inhibition by tunicamycin derivatives since selectivity for bacterial PGTs over DPAGT1 will be vital for their success as antibiotics [182].

This assay also enabled the catalytic mechanism of this bi-bi enzyme to be elucidated through the use of substrate matrix experiments. DPAGT1 activity preferentially fit to the ternary bi-bi model when compared to the ping-pong equation (Figure 4.9), confirming pervious theories that the reaction progresses in a one-step manner without a covalently modified enzyme intermediate [44]. These data do not reveal whether there is a required sequential order of binding of the sugar donor or lipid acceptor; studying the inhibitory patterns of substrate analogues or dead-end product inhibitors could help explain availability of the different enzyme forms in the normal reaction sequence, and hence enable predictions about the reaction scheme [157].

It was found that the UDP-Glo kit was unsuitable for monitoring GlcNAc transfer by ALG13/ALG14. This was because UDP-GlcNAc degraded to UDP during heat inactivation, which meant enzyme activity could not be seen above that in negative controls. Given the high level of selectivity of the kits for their respective nucleotide phosphate (Figure 4.10), if an alternative quenching method was identified that prevented this high background of UDP, the simultaneous monitoring of UDP and UMP release could be achieved.

4.3.2 ^{14}C labelled UDP-GlcNAc assays

A radioactive assay was established that utilised the power of TLC to monitor the movement of labelled sugar from the soluble donor to the lipid acceptor. This enabled the characterisation of both DPAGT1 and ALG13/ALG14. A mobile phase of ethyl acetate, isopropanol and water allowed for distinct separation of all lipid linked oligosaccharide species on the silica plates (Figures 4.14-15).

The main problem faced when assaying ALG13/ALG14 was large variability between experiments even when using the same stock of protein. This is likely due to the high levels of impurities remaining following purification, which made it difficult to get a reliable quantification of the total free enzyme concentration, [E_0]. Instead, differences in maximal rate are likely due to differences in enzyme concentration. Typically, enzyme concentration was determined using absorbance at 280 nm with a Nanodrop spectrophotometer adjusted to the expected extinction coefficient and molecular weight of the complex (ε = 44350, Mw = 44 KDa). These ALG13/ALG14 concentration measurements were found to be highly comparable to those identified using a bicinchoninic acid (BCA) assay that had been adjusted for the ratio of target protein: impurities determined using densitometry of Coomassie stained SDS-PAGE (data not shown). This meant it was not possible to properly define the turnover number of this enzyme, and V_{max} are reported rather than k_{cat}.

If a tight-binding inhibitor was available for an enzyme, it could be used to help quantify the number of active sites available and hence allow for an accurate [E_0] [183]. This can be done by titrating both enzyme concentration and concentration of the inhibitor, such that the data can be modelled to the Morrison equation with point of intersection on the x axis representing the concentration of available active sites in the stock [184].

Kinetic studies of DPAGT1 and ALG13/ALG14 alone and in complex were carried out at 25 °C following nanoDSF screening of the proteins at 37 °C that revealed precipitation within the time frame of the assay (Figure 4.20). However, this experiment was carried out in the absence of lipid acceptor substrate which was found to increase the melting temperature dramatically (Chapter 5). Future work to optimise this assay may involve screening buffer conditions that better stabilise the protein: trialling that containing a higher glycerol content than the 5% used; adjusting the ionic strength of the buffer by using a higher concentration of NaCl; or by screening for an alternative detergent and lipid system that was more protective than DM / cholesterol hemisuccinate.

GlcNAc transfer by ALG13/ALG14 was found to progress very rapidly even at reduced temperatures, with the reaction at completion after just 2-5 minutes (Figures 4.17-22). Attempts to lower enzyme and substrate concentrations to slow down the reaction progress were not successful given low amount of transfer was difficult to detect on the TLC plate (data not shown). Therefore, 1 minute was selected as a time point to obtain the initial rate of ALG13/ALG14. Here, the lipid acceptor was found to bind with only moderately higher affinity than the sugar donor, with apparent K_M' values of 13.23 ± 2.78 and 44.96 ± 16.42 μM respectively. The average V_{max} across three technical replicates was 36.00 ± 4.94 and 18.47 ± 1.07 μM min^{-1} when assayed at fixed, saturating concentration of dolichol pyrophosphate GlcNAc and UDP-GlcNAc respectively. This makes ALG13/ALG14 activity ~100-fold faster than the rate of DPAGT1. This work corroborates results from Wang *et al.* where dolichol pyrophosphate GlcNAc was reported to have a K_M' of 94.1 ± 9.4 μM towards ALG13/ALG14, with a k_{cat} at 919.5 min^{-1} ± 54.3 min^{-1}, however the authors do not report those for the sugar donor [68]. This assay remains a low-throughput option, but with more time it would be possible to assess the catalytic mechanism of ALG13/ALG14 using substrate matrix experiments.

When monitoring the complex of DPAGT1/ALG13/ALG14, it was clear that the first catalytic step by DPAGT1 is rate limiting in the synthesis of dolichol pyrophosphate chitobiose. This is because the overall rate of production of the two GlcNAc species matched that of DPAGT1 catalysing the one GlcNAc product when alone (Table 4.6). It is of course not possible to draw conclusions on ligand binding given there are two enzymatic steps governing the reaction (and hence results are italicised in Table 4.6). When provided with the intermediate lipid acceptor dolichol pyrophosphate GlcNAc, the activity of ALG13/ALG14 was found to be the same in complex as in isolation.

Finally, as the DPAGT1/ALG13/ALG14 complex was purified using 6xHis affinity tags present on both DPAGT1 and ALG13/ALG14 before SEC, it cannot be guaranteed that the total free enzyme concentration, $[E]_0$, represents only enzymes in complex and rather might contain a significant contribution of free DPAGT1 or ALG13/ALG14 enzymes. This could be addressed by using two affinity steps using unique purification tags on each component to ensure only the complex is isolated. As shown in Chapter 3, testing is required to optimise the position and type of tag used, as a C-terminal FLAG tag on DPAGT1 was found to be destabilising.

Overall, these results provide advancements in our understanding of the early N-glycosylation in humans, especially noting that the initial priming step to form the pyrophosphate linker to the dolichol anchor is rate determining. This established assay has allowed substrate preferences of the enzymes to be unpicked (Chapter 5) as well as the monitoring of pathogenic mutations of ALG13/ALG14 (Chapter 6).

Table 4.6 Summary of binding parameters measured using a TLC assay for DPAGT1 and ALG13/ALG14 alone and in complex.

		K_M ^{D}, µM ± SE	V_{max} ^{D}, µM min^{-1} ± SE	K_M ^{A}, µM ± SE	V_{max} ^{A}, µM min^{-1} ± SE
DPAGT1/ ALG13/ ALG14 complex	***Dolichol phosphate***	*78.67 ± 58.07*	*0.25 ± 0.08*	*0.87 ± 0.60*	*0.84 ± 0.07*
	Dolichol PP GlcNAc	159.70 ± 107.31	14.79 ± 5.58	2.22 ± 1.47	5.69 ± 0.68
ALG13/ ALG14		44.96 ± 16.42	36.00 ± 4.94	13.23 ± 2.78	18.47 ± 1.07
DPAGT1		76.28 ± 40.56	0.68 ± 0.16	0.75 ± 1.00	0.20 ± 0.03

where D = donor, UDP-GlcNAc, and A = acceptor lipid substrate

Chapter 5

Substrate specificity: testing a panel of native and non-native polyisoprenol and dolichol phosphates

5.1 Introduction

Polyisoprenols are present inside the cell membranes of all domains of life but vary in their saturation state, length and configuration (Figure 5.1). Differences in these lipids reflects the synthesis enzymes present, and these alternative biogenesis pathways can be mapped to evolutionary splitting events. Concomitantly, glycosyl- and oligosaccharide-transferases are known to preferentially utilise polyisoprenols of specific structure, however these preferences appear to differ even between enzymes in the same species. It was therefore of evolutionary interest to probe the substrate tolerance of human enzymes in the early N-glycosylation pathway - DPAGT1 and ALG13/ALG14 - using a range of polyisoprenol substrates. This has allowed key insights into their specificity for the chain length and saturation state of lipid acceptor substrate, and in turn poses questions about the architecture of the enzymes.

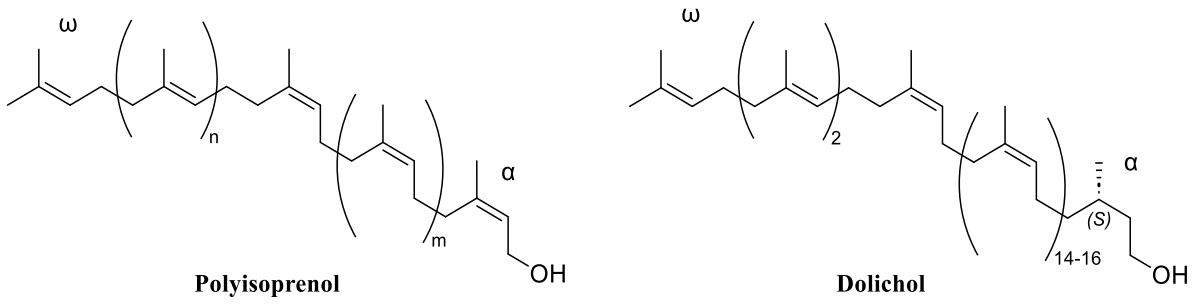


Figure 5.1 Polyisoprenols utilised in glycan biosynthesis. (Left) General structure of polyisoprenols where n represents the number of *trans* configured isoprenoids and m the

cis-configured units. (Right) Dolichols are eukaryotic polyisoprenols with a characteristic saturated α-terminal isoprenoid unit.

5.1.1 Substrate specificity of polyisoprenol glycosyltransferases

The preference of glycosyltransferases on the structure of their polyisoprenol lipid acceptor has been the subject of study for many years [159]. Rush *et al.* reported in 1998 that *Escherichia coli* phospho-GlcNAc transferase was able to utilise native C55 undecaprenol phosphates but not their C55 dolichol equivalents, suggesting saturation of the α-unit was crucial for substrate recognition [185]. Patel *et al.* compared the activity of *Salmonella enterica* WbaP (a galactose-1-phosphate transferase) with its native undecaprenol to that of lipid phosphates of various length; C10, C15, C20, C35, C50 and C95 dolichol [169]. They report very limited activity with shorter lipids and with eukaryotic C95 dolichol phosphate just 4% the activity of the native substrate. This again suggests high substrate selectivity.

During elucidation of the archaea *Methanococcus voltae* N-glycosylation pathway, Larkin *et al.* report AglK - a GlcNAc phosphate transferase – had much higher activity to native length (C55-60) dolichol phosphates to that of eukaryotic C85-105 dolichols [186]. *Campylobacter jejuni* N-glycosylation pathway enzymes PglJ, C and B were tested with a suite of polyisoprenols of various lengths, geometry and saturation [89]. It was found their activity was influenced by structure of the lipid acceptor (or donor in the case of PglB), with transfer efficiency to peptide by PglB significantly decreased with dolichol, as well as with all-*trans* orientated lipids. This again points to specific binding to an unsaturated *cis* prenol in the active site by prokaryotic enzymes.

Liu *et al.* tested the lipid carrier specificity of PglB with a panel including C10, C20, C30 and C40 polyisoprenol sugar pyrophosphates of poly-*cis* or single *trans* orientation [63]. Interestingly, they found the enzyme could utilise a C20 lipid and an all-*cis* variant, with length effecting K_M

only rather on k_{cat}. This work suggests the ω terminal *trans* configured units are not important for conferring PglB glycan transfer activity. Furthermore, MraY has been reported to have broad substrate tolerance, utilising lipids down to just 2 prenyl units (C10) and with high activity with C35, 60 and 75 polyisoprenol phosphates [161]. Both MraY and MurG were found to use lipids of both polyunsaturated and dolichol forms [187]. *Thermotoga maritima* WecA was studied by Al-Dabbagh *et al.,* and shown to have a longer minimal prenol requirement, displaying just 3% of the activity of its native lipid with C30 lipids and with no activity below that length, but was also able to use the dolichol version of the native lipid [163]. This requirement for native length lipid was corroborated by Mitachi *et al.* with *M. smegmatis* WecA [188]. The tolerance for non-native lipid acceptors by prokaryotic enzymes seems to be highly species dependent.

When it comes to the eukaryotic enzymes, preference for dolichol form lipids has been clearer cut, although much of the reported work has been performed on crude microsomal extractions rather than on isolated proteins. In 1980, Palamarczyk *et al.* solubilised glycosyl transferases from yeast membranes, and showed the requirement of GlcNAc transfer on polyisoprenols with saturated α-isoprene [189]. Longer chain lengths were found to increase affinity, with chains of below seven prenol units inactive (summarised in Figure 5.2) [189]. They observed GlcNAc transfer to C100, 80 and 55 dolichols whereas for their polyunsaturated equivalents - as well as an all *trans* orientated C45 polyunsaturated prenol and dolichol pair– they report very low transferase activity.

Low *et al*. synthesised a variety of *S*-form dolichol phosphates and tested activity of mannosyl, glycosyl and glucosaminyl transferases in microsomes prepared from rat liver [190]. By accessing polyprenols of 7 isoprene (from wood of *Betula verrucosa),* 10 (from Bay leaves*, Laurus nobilis*) 12 (from *Rhus typhina*), 17 and 18 (*Gingko biloba* leaves) and 22 (from *Sorbus suecica*), the corresponding dolichols with one extra isoprene were generated by addition of an

S-form unit using Grignard coupling. Phosphorylated substrates were used to determine kinetic parameters for GlcNAc transfer, where there was a four-fold increase in k_{cat} from dolichol phosphate with 8 isoprenes to 13. Above this chain length, there was no additional increase in maximal activity, and interestingly, the affinity for all lengths was largely unchanged (Figure 5.2).

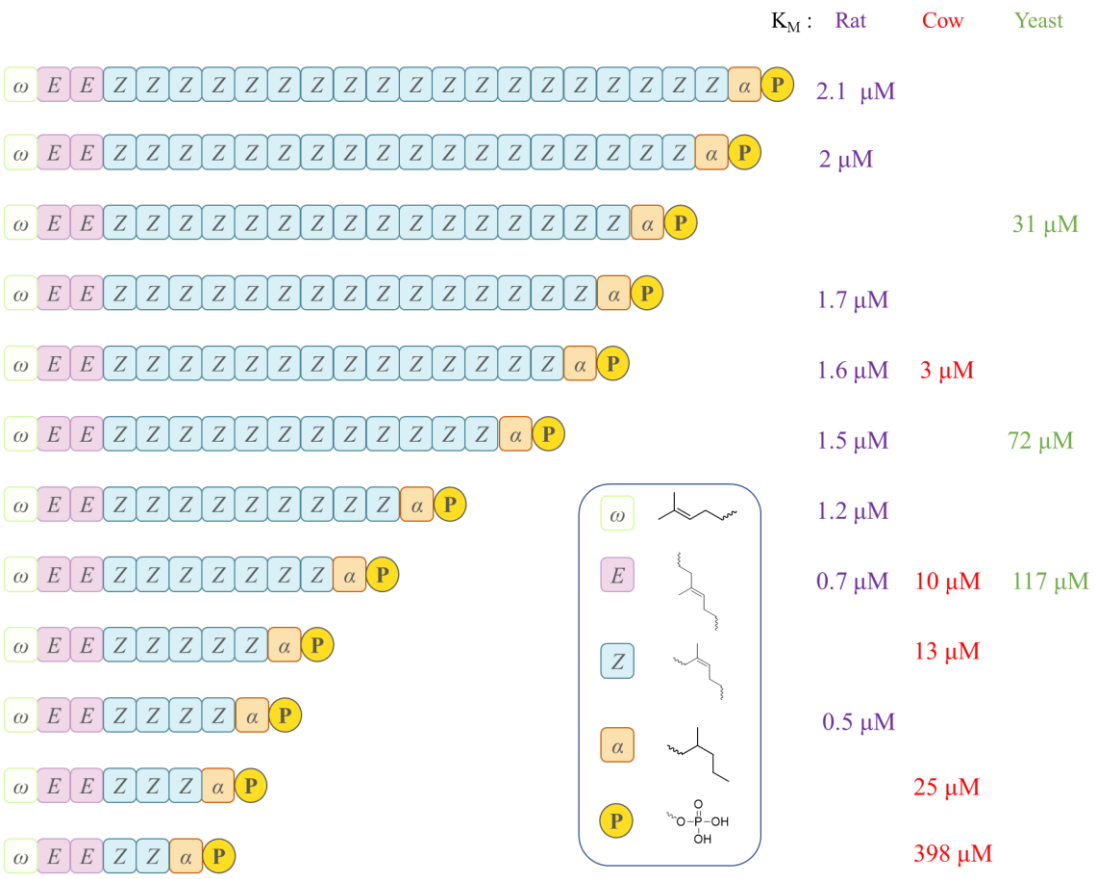


Figure 5.2 Effect of increasing number of *cis* isoprenoid units on affinity to GlcNAc transferases originating in three eukaryotic organisms: yeast [189], cow[191], rat [190]. K_M's reported are approximate values extracted from figures.

The specificity of dolichol kinase and dolichol phosphate manosyl transferase solubilised from rat liver microsomes was reported in 1992 by Chojnacki and colleagues [192], where dolichols of 11 to 32 isoprene units where compared to their fully unsaturated polyprenol equivalents.

Again, it was found that the polyunsaturated forms were poor substrates for both the kinase and mannosyl transfer compared to the dolichol substrates. Similarly, *N*-acetylglucosamine phosphate transferase activity was monitored in Chinese hamster ovary (CHO) cells [193], where the enzyme was found to have strong preference for native C90 dolichol - displaying 10-fold lower K_M and higher V_{max} than the polyunsaturated equivalent.

In recent years, short (C10, 15, 20 and 25) oligoprenols were used for study of purified yeast ALG1, 2 11 [194]. It was found that ALG2 requires at least four prenols (C20) whereas ALG1 could transfer a mannose to all chain lengths. Following the final mannose transfer by ALG11 onto C20 and C25 lipids, the resulting short chain substrates were successfully used by a protozoan OST STT3A. Recently, the Gao group reported inactivity of ALG13/ALG14 for short unnatural lipid substrates, and instead they studied activity of the enzyme using C90 dolichols semi-synthetically accessed using extraction from *Gingko biloba* [195]. This implicates a difference in chain length preferences and saturation state even between enzymes of the same pathway.

5.1.2 Synthetic and semi-synthetic approaches to access polyisoprenols

To study glycosyltransferases that utilise polyisoprenyl lipid carriers, access to pure and properly characterised lipid substrates is required. Apart from extraction from pig, rat and human livers, access to long chain lipids (C50 and above) has been achieved in various semi-synthetic ways, with an excellent review by Jeremy Boilevin and Jean-Louis Reymond summarising published work [196]. Furthermore, Cochrane *et al.* provide useful methods for accessing undecaprenol lipids from bay leaves (*Laurus nobilis)* for study of bacterial lipopolysaccharide synthesis [197].

To yield mammalian type dolichols, the terminal α-isoprenoid unit needs to be reduced to give a saturated, *S*-configured methyl group, as well as having two ω-*trans* orientated prenols. A library of dolichols of C35-C110 from plant extractions was demonstrated by Suzuki *et al.* [198],

involving coupling of an *S*-configured saturated prenoid as a chiral Grignard reagent, as used in the above mentioned work by Low *et al.* [190]. Polyisoprenols from the leaves of the *Gingko biloba* tree have therefore been long known to provide useful starting points for semi-synthetic access to human-like dolichols, with C90 lipids of the correct E_2Z_{14} geometry [120, 198, 199]. In 1986, Noyori and colleagues reported enantioselective hydrogenation of allylic alcohols using ruthenium catalysts [200] which was immediately applied for use in polyisoprenol reduction by Imperiali and Zimmerman to *Gingko*-extracted polyisoprenols [201] . This involved using high pressure for hydrogenation; a process that was made much safer by catalysts described by Wu *et al.* [202]. These were again readily applied by the Imperiali group on 12 prenoid lipids (C55-60) from the leaves of *Rhus typhina* [186] , and by Elharar *et al.* from the same starting material using a 2 hour reflux at 83 °C [203].

Previously the Davis lab has developed an array of short chain polyisoprenols for study of PglB glycosyltransferase activity by coupling appropriate sulfones and allyl chloride building blocks to yield C10, 20, 30 and 40 lipids, with poly-*trans*, one or two ω-*cis* prenols geometry [63]. Synthesis was in part based on chemistry published by Jaenicke and Siegmund in 1989, who demonstrated dolichols C20 to C55 with different geometry could be made through elongation/coupling of protected isoprenol units [204]. More recently dolichols C60 and C100 were synthesised in a microflow system using Z-selective Wittig reactions by Hirao *et al.* [205]. Together, these studies provided the perfect opportunity to probe the substrate specificity of recombinant human glycosyltransferases, DPAGT1 and ALG13/ALG14 to native and non-native lipids. The tolerance of both chain length and saturation state of the lipid acceptor molecule were probed using thoroughly characterised substrates.

5. 2 Results

5.2.1 Accessing a panel of polyisoprenols and dolichols

To assess the substrate tolerance of DPAGT1 and ALG13/ALG14 glycosyltransferases, a panel of short (C55) and long (C90) lipids of both dolichol and polyisoprenol form was designed to mimic dominant structures found in bacterial and eukaryotic membranes. Lipids bearing either a monophosphate or pyrophosphate GlcNAc were required to act as substrates for DPAGT1 and ALG13/ALG14 respectively. This lipid panel is summarised in the schematic in Figure 5.3.

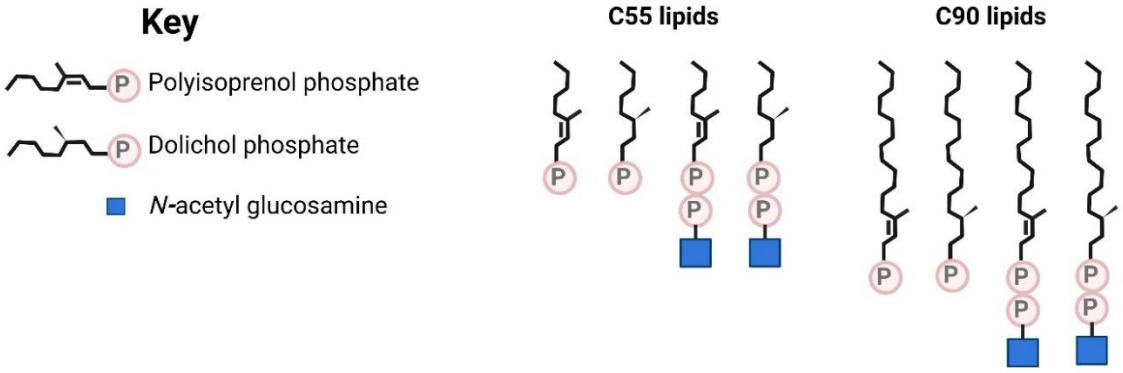


Figure 5.3 Schematic representing the structure of short (C55) and long (C90) polyisoprenoid lipids of dolichol and polyunsaturated nature. These were used to probe the tolerance of DPAGT1 and ALG13/ALG14 for substrate length and saturation state.

To start, polyisoprenoid acetates with human-like ratios of *cis* and *trans* isoprenoid units were purified from the leaves of the *Gingko biloba* tree, and deacetylated to yield long, C90 polyisoprenoid alcohols (Figure 5.4). ^{1}H NMR between 1.7- 1.5 ppm was used to confirm the length of polyisoprenol alcohols extracted as well as the ratio of *cis* and *trans* isoprene units, as described previously [63]. Figure 5.4C shows representative analysis of lipids which were 18 isoprene units long (3 *trans* +*trans* (ω), 14 *cis*,*cis*, + 1 *cis* (α)), consistent with that previously published [199].

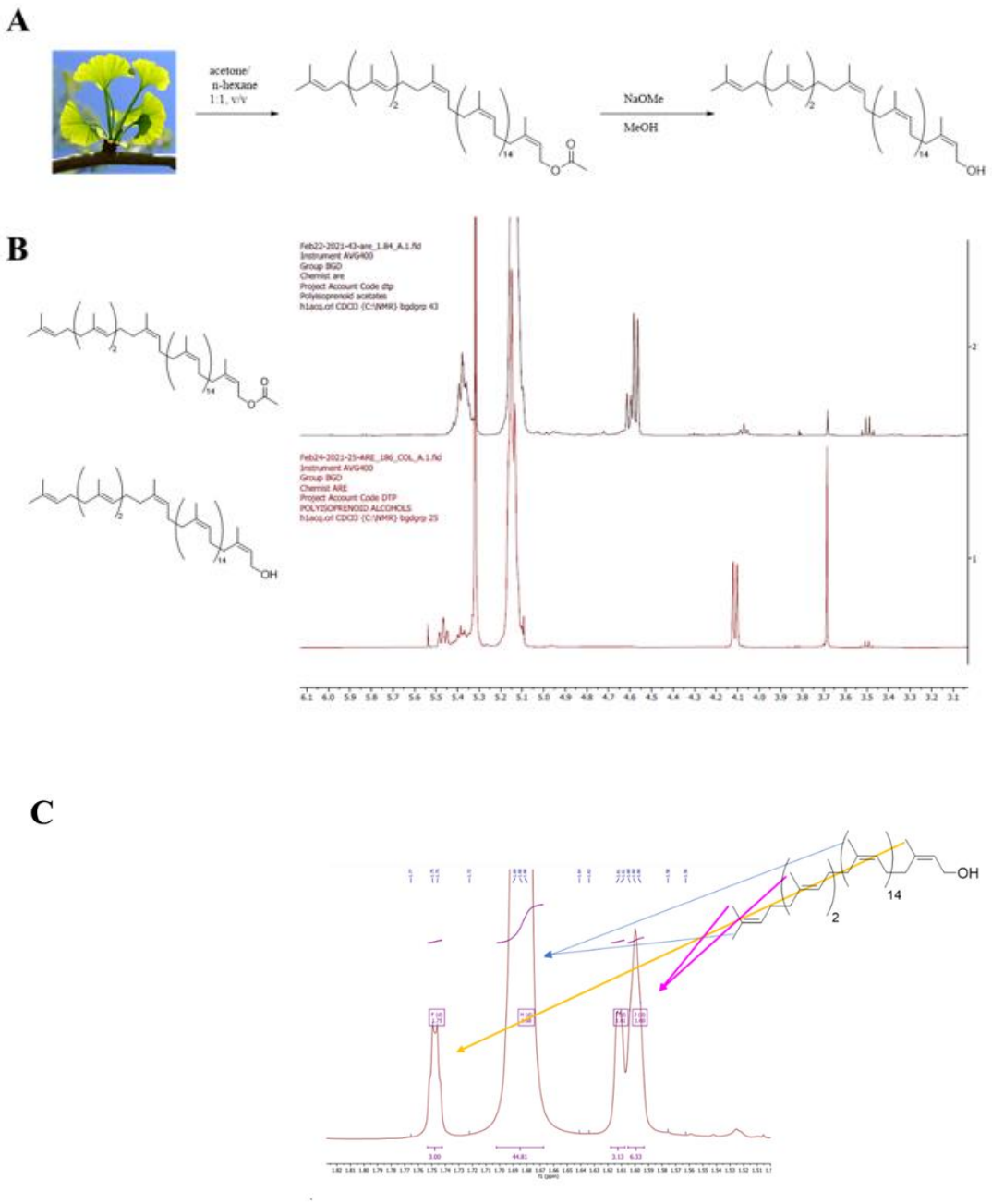


Figure 5.4 A Polyisopreonoid alcohols of human-like length were accessed from the leaves of the *Gingko biloba*. B ^{1}H NMR (400 MHz) NMR spectra from 6.1-3.1 ppm, comparing polyisoprenoid acetates (top) and alcohols (bottom) following Zemplen deacetylation. A characteristic shift of protons from 4.6 ppm to a doublet at 4.1 ppm is seen. **C** ^{1}H NMR (400 MHz) shift from 1.7 – 1.5 ppm of polyisoporenols indicating a chain with two *trans* and 14 *cis* prenoid units.

To access the dolichol version of these lipids, the terminal isoprenoid unit of the extracted polyisoprenols was reduced to give a saturated, *S*-configured methyl group. The [(*S*)-tol-binap]$RuCl_2$ catalyst was trialled for this reaction – having previously only been reported for use with shorter C55 dolichols [186, 202, 203]. The reaction was monitored by ^{1}H NMR (Figure 5.5). Despite initial success, this reaction proved difficult to reproduce, and so alternative enzymatic methods were explored in parallel.

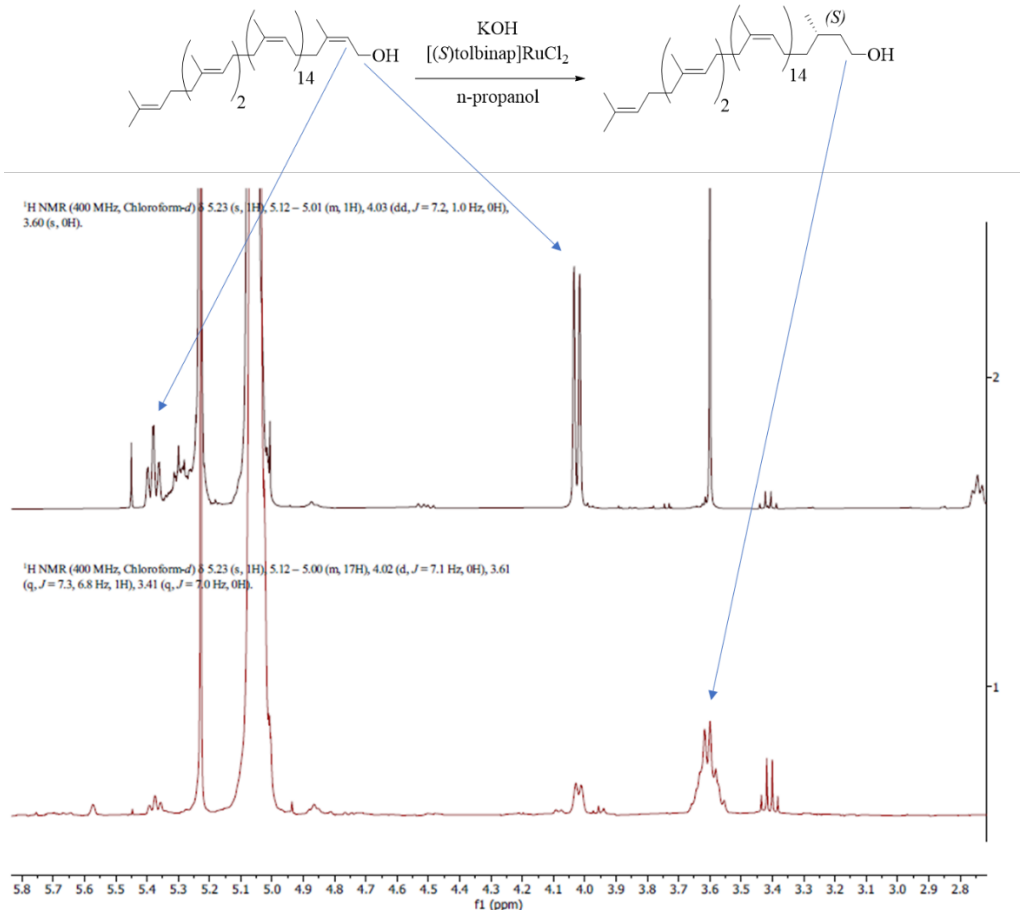


Figure 5.5 Hydrogenation of polyisoprenoid alcohols to dolichols. ^{1}H NMR (400 MHz) between 5.8- 2.8 ppm, showing the conversion of polyisoprenol (top) to dolichol (bottom). Polyisoprenoid alcohols (164 mg, 0.1319 mmol) were converted into dolichols using [(S)-tol-binap] $RuCl_2$ (14.8 mg, 0.0132 mmol) catalyst with anhydrous KOH (3.7 mg, 0.066 mmol) in dry n-propanol.

5.2.1.1 Trial use of SRD5α3 for accessing dolichols

SRD5α3 is an integral endoplasmic reticulum protein that reduces polyisoprenoid alcohols to dolichols ready for use as lipid donor molecules. It was decided to trial SRD5α3 for converting the polyisoprenols accessed from *Gingko* and bay leaves to dolichols, offering an enzymatic alternative to the [(*S*)-tol-binap]$RuCl_2$ catalyst that was proving unreliable. The cDNA sequence encoding full-length Met1 to Phe318 and an N-terminal 6xHis purification tag with a tobacco etch virus (TEV) protease cleavage site was cloned into a pFB-LIC Bse expression vector (CMD) by Jing Wang. Baculovirus were produced by transformation of DH10EMBac cells, as in Chapter 2, and used to induce expression in Sf9 insect cells. An initial expression and detergent screen was performed by Dr Andrew Quigley at the MPL. This revealed a preference of DM/CHS and DDM/CHS detergents (data not shown), and DM/CHS was selected for use in subsequent experiments.

To test the activity of the SRD5α3 expressed, an assay using crude lysate was performed. Cells from 1 L infected insect culture were lysed and homogenised in 50 ml of lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM imidazole, 5% glycerol, Roche EDTA-free protease inhibitor), and protein solubilised using DM/CHS containing buffer. 50 mL of this lysate was incubated at 37 °C with 5 mg of polyisoprenols (0.004 mmol, 1 eq) and supplemented with NADPH tetrasodium (0.0074 mmol, 1.8 eq) to avoid NADPH availability limiting the reaction. Time points at 5 minutes 30 minutes and 1, 3 and 17 hours were taken and lipids from each were extracted using DCM. Samples were compared to that from DCM extraction of crude cell lysate where low levels of natural polyisoprenols were detected.

Polyisoprenols were successfully recovered using this DCM extraction, but unfortunately ^1H NMR analysis revealed only starting polyisoprenols, with no evidence of reduction to dolichols (Figure 5.6). Attempts were made to purify SRD5α3 to allow for more concentrated incubation

with the enzyme, however SEC results indicated that most of the protein extracted and purified in DM/CHS containing buffer aggregated as there was a large void peak (Figure 5.6B). As discussed below, this work was not progressed further.

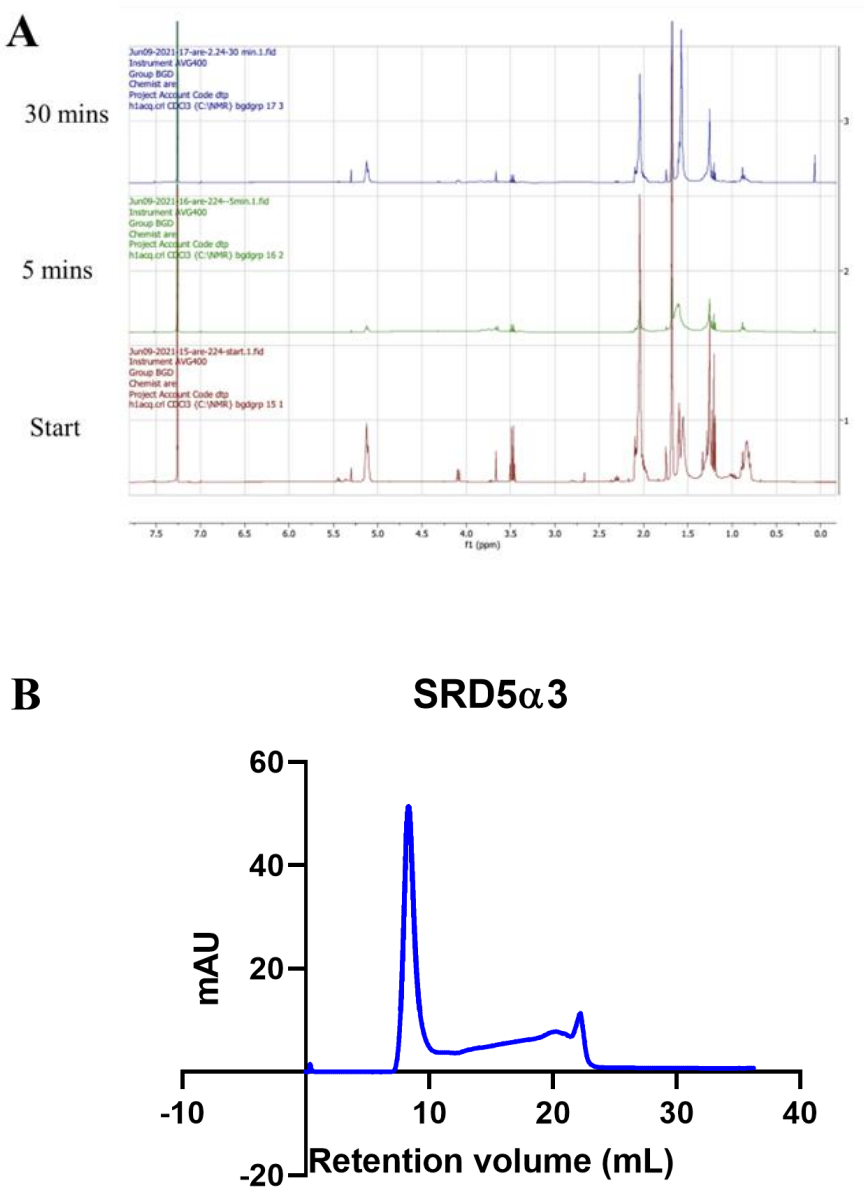


Figure 5.6. Initial use of SRD5α3 to yield dolichol lipids. A ^{1}H NMR (400 MHz) spectra from a time course with crude SRD5α3 expressing membranes, where only starting polyisoprenols were detected. **B** SEC trace showing aggregated protein during attempts to access pure enzyme.

5.2.1.3 Characterisation of the lipid panel

Given the difficulty in reproducing the selective hydrogenation of polyisoprenols to dolichols, coupled with time restraints imposed by the Covid-19 pandemic, it was decided to focus efforts elsewhere and instead access of the lipid panel was achieved by Dr Andrew Giltrap. Andrew was able to successfully synthesise C90 polyisoprenol and dolichols, initially using the same methodology reported above. These lipids were phosphorylated for use with DPAGT1, or taken forward to afford the pyrophosphate GlcNAc linked species to act as substrates for ALG13/ALG14 (as in Figure 5.7). Similarly, undecaprenol was isolated from the leaves of *Laurus nobilis*, using previously described methodology [197], and the dolichol versions of this C55 lipid were also accessed. Mass spectrometry and ^{31}P, ^{1}H and ^{13}C NMR were used to confirm the identity of all lipids of the panel, with detailed characterisation shown in the Appendix. The stocks were found to contain a range of other length of lipids from extraction from leaf matter and as would be the case *in vivo*, with C55 and C90 lengths dominating. Together, this has allowed the relationship between chain length and saturation state to be unpicked.

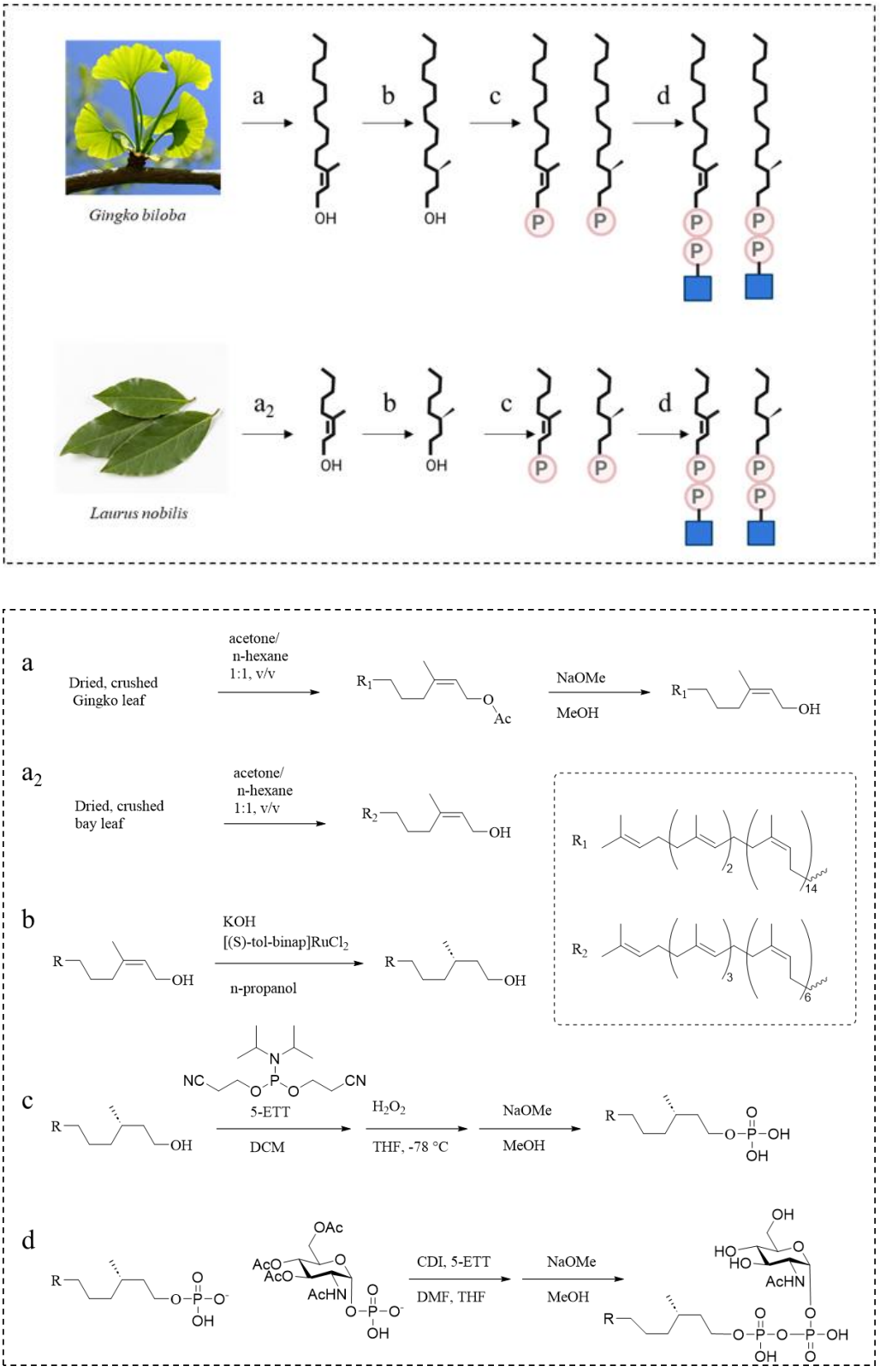


Figure 5.7 Schematic depicting methodology to access polyisoprenols of various length and saturation state from leaves. 5-ETT = 5-ethylthiotetrazole, and CDI = 1,1'-carbonyldiimidazole.

To aid characterisation of this lipid panel, a thermostability screen with DPAGT1/ALG13/ALG14 protein was performed using nanoDSF to test for successful binding of each member. DPAGT1/ALG13/ALG14 was incubated with 250 μM of lipid substrate and compared to that with buffer in DM/CHS detergent only. Samples were taken in triplicate and analysed over a temperature increase of 1 °C/min from 20 °C to 90 °C. The first derivative of fluorescence at 350 nm/ 330 nm is shown in Figure 5.8, with corresponding inflection point T_m summarised in Table 5.1. There was a very striking increase of 17 °C in T_m from protein in buffer to that with C90 dolichol pyrophosphate GlcNAc, and a 10 °C from buffer to that with C90 dolichol phosphate. This increase confers a stability increase imposed when the lipid is bound to the protein, and did not occur in control wells. Thermostability was comparable between lipids of different lengths and saturation state, suggesting they were all able to bind and stabilise the protein. This gave confidence in the synthetic substrates for functional study, as well as provided useful tools for structural biology (as in Chapter 3).

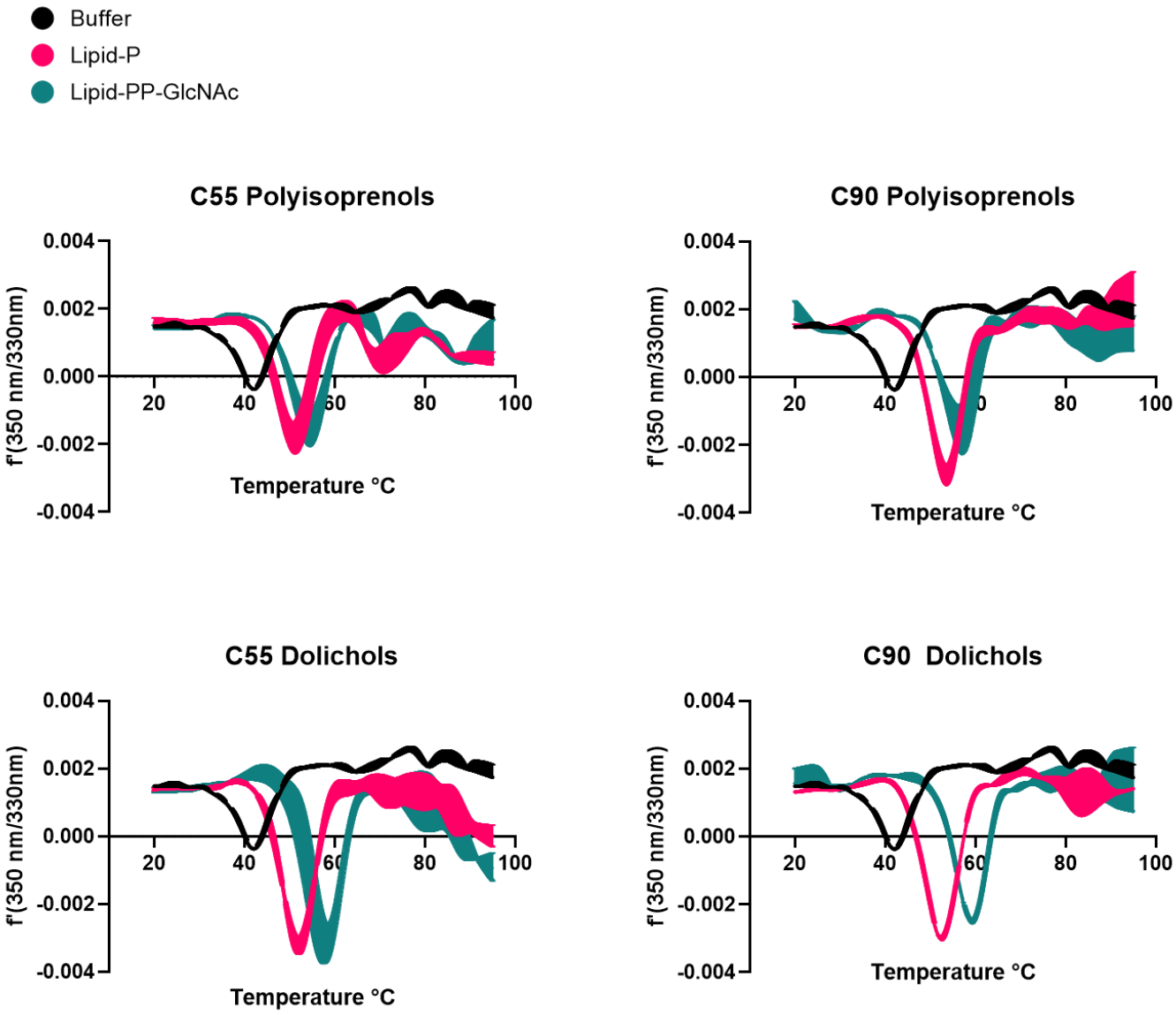


Figure 5.8 Thermostability screen of DPAGT1/ALG13/ALG14 when bound to members of the native and non-native lipid panel. Short (C55) and long (C90) lipids both in dolichol ('Dol') and polyisoprenol ('Poly') form, bearing a monophosphate or pyrophosphate GlcNAc as the substrates of DPAGT1 and ALG13/ALG14 respectively. Results shown are n=3 technical replicates.

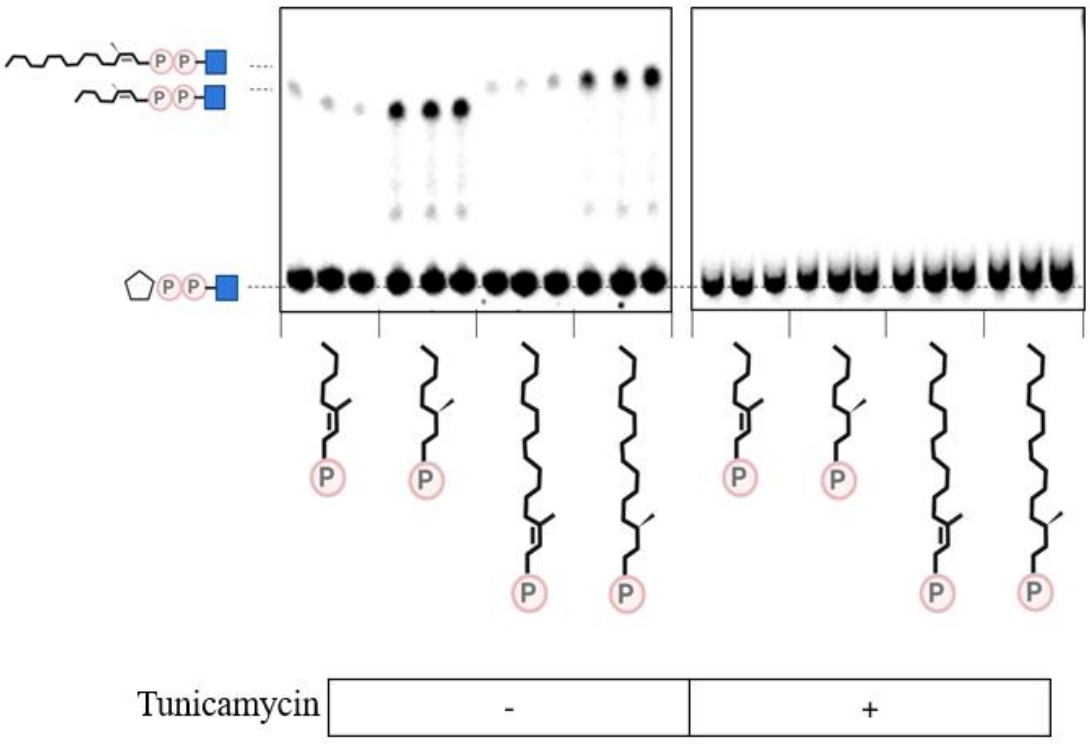
Table 5.1 Thermostability screen of DPAGT1/ALG13/ALG14 when bound to members of the native and non-native lipid panel. C55 = number of carbons, Poly = polyisoprenol, Dol = dolichol, P = phosphate, PP = pyrophosphate. Results shown are the average of n=3 technical replicates.

		T_m, °C
	Buffer	42.1
Lipid-P	C55-Poly-P	51.2
	C55-Dol-P	52.0
	C90-Poly-P	53.6
	C90-Dol-P	52.9
Lipid-PP-GlcNAc	C55-Poly-PP-GlcNAc	54.5
	C55-Dol-PP-GlcNAc	58.1
	C90-Poly-PP-GlcNAc	57.3
	C90-Dol-PP-GlcNAc	59.3

5.2.2 Activity testing

5.2.2.1 Endpoint activity assay of DPAGT1 and ALG13/ALG14

Endpoint activity of DPAGT1 and ALG13/ALG14 alone and in complex were tested with the full lipid panel of short (C55) and long (C90) polyisoprenol and dolichol phosphates and pyrophosphate GlcNAc species (depicted in Figures 5.9-11). 2 μM of each enzyme stock was incubated for 1 hour at 37 °C with an excess of labelled UDP-[^{14}C]-GlcNAc and each of the native and non-native lipid acceptor species. Transfer of the labelled GlcNAc was monitored by TLC (as in Chapter 4). A striking pattern was noted where DPAGT1 had significantly higher activity with dolichol versions of the lipids of both C55 and C90 lengths (Figure 5.9). This indicated α-saturation was limiting but length of the lipid above C55 was not. Interestingly, ALG13/ALG14 did not seem to have a preference of either length or saturation state as was able to utilise all equally (Figure 5.10). These results were recapitulated when the enzymes were expressed as a complex (Figure 5.11). Furthermore, tunicamycin was able to inhibit DPAGT1 (but not ALG13/ALG14) reactions as expected, with no product spot detected on the TLC plate (Figures 5.10 and 5.11).



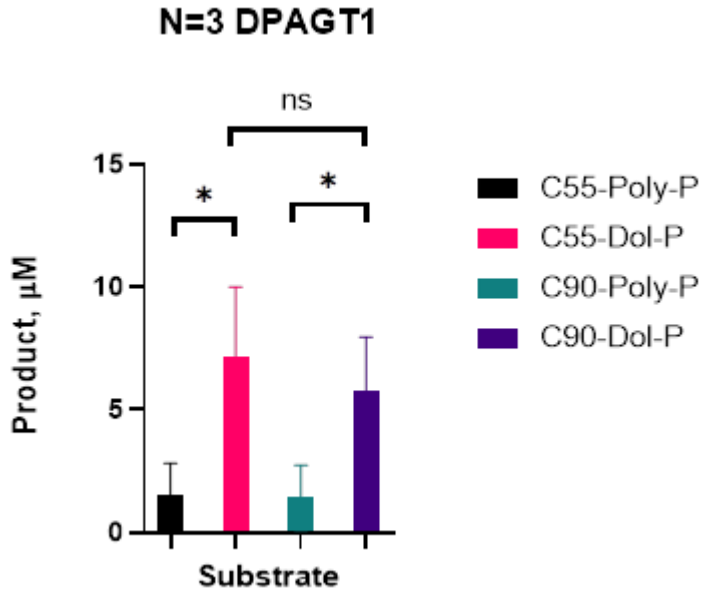
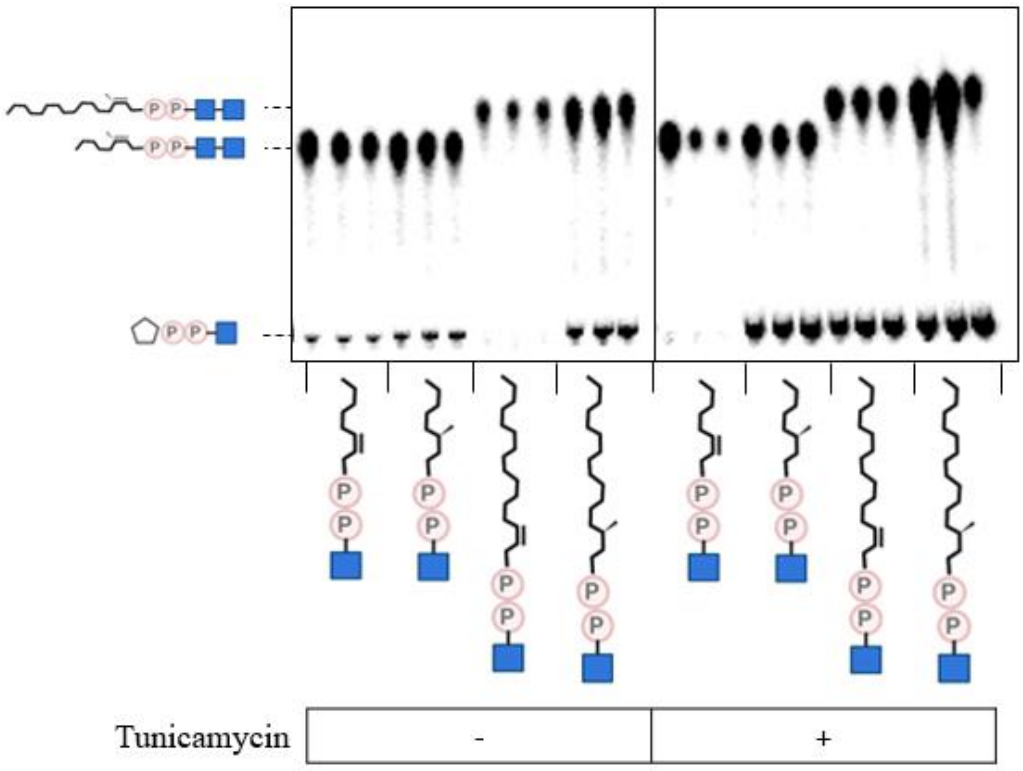


Figure 5.9 DPAGT1 endpoint activity with short (C55) and long (C90) polyisoprenol and dolichol lipids. There was no significant difference (P = >0.05) between lipids of different length whereas there was significantly higher activity with dolichol versions (C90 lipids P= 0.016, C55 lipids P= 0.015) when analysed using multiple unpaired t tests in GraphPad Prism.



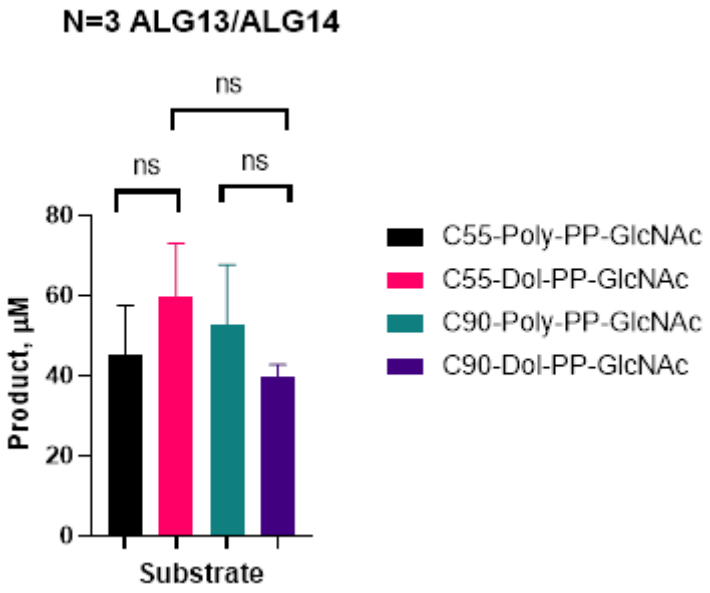


Figure 5.10 Summary of endpoint activity of ALG13/ALG14 with short (C55) and long (C90) polyisoprenol and dolichol lipids. There was no significant difference (P = >0.05) between lipids of different length or between polyisoprenol and dolichol versions, when analysed using multiple unpaired t tests in GraphPad Prism.

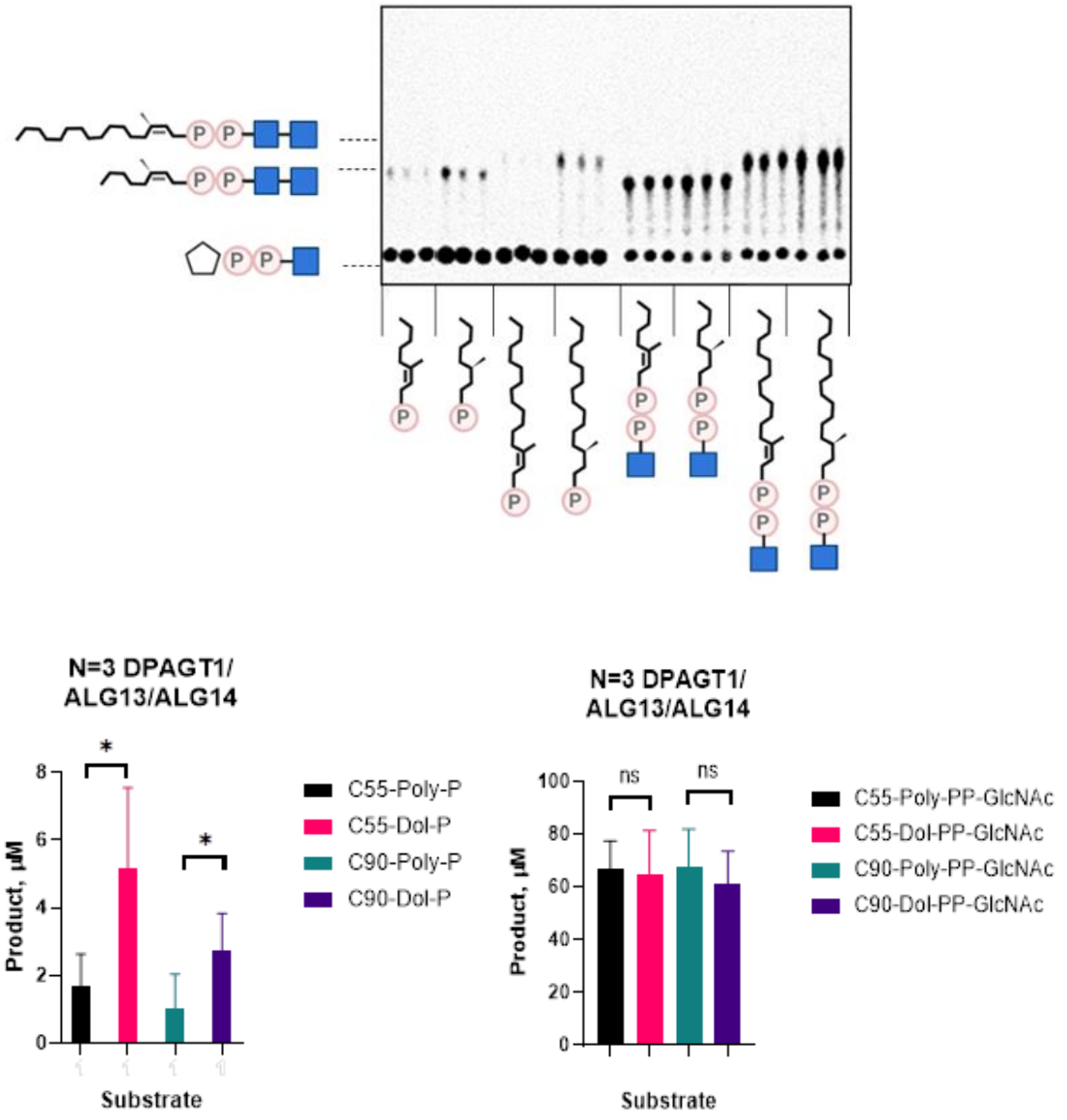


Figure 5.11 N=3 Endpoint activity screen of DPAGT1/ALG13/ALG14 with short (C55) and long (C90) polyisoprenols and dolichols. Lipids were separated from soluble reaction material using thin layer chromatography with a mobile phase of ethyl acetate: isopropanol: water (3:2:1). ImageJ was used to quantify counts and data were analysed using GraphPad Prism (D) for statistical significance using multiple t tests ($P<0.05$). N.B Phosphorescence was read using an alternative Fujifilm FLA-5100 by kind access of the Ludwig Institute rather than the Typhoon plate reader used in other experiments, hence differences in background signal.

Attempts were made to identify products using mass spectrometry, through collaboration with Dr Dan McGill. Reactions were set up as before but with unlabelled UDP-GlcNAc as the sugar donor to allow for use outside of a radiation-controlled environment. The reactions were separated by TLC and sections of the plate corresponding to the position of those detected by radioactivity were analysed using DESI-MS. Lipid signal was not able to be identified, with only what is suspected to be DM/CHS detergent signal seen across the plate (Figure 5.12).

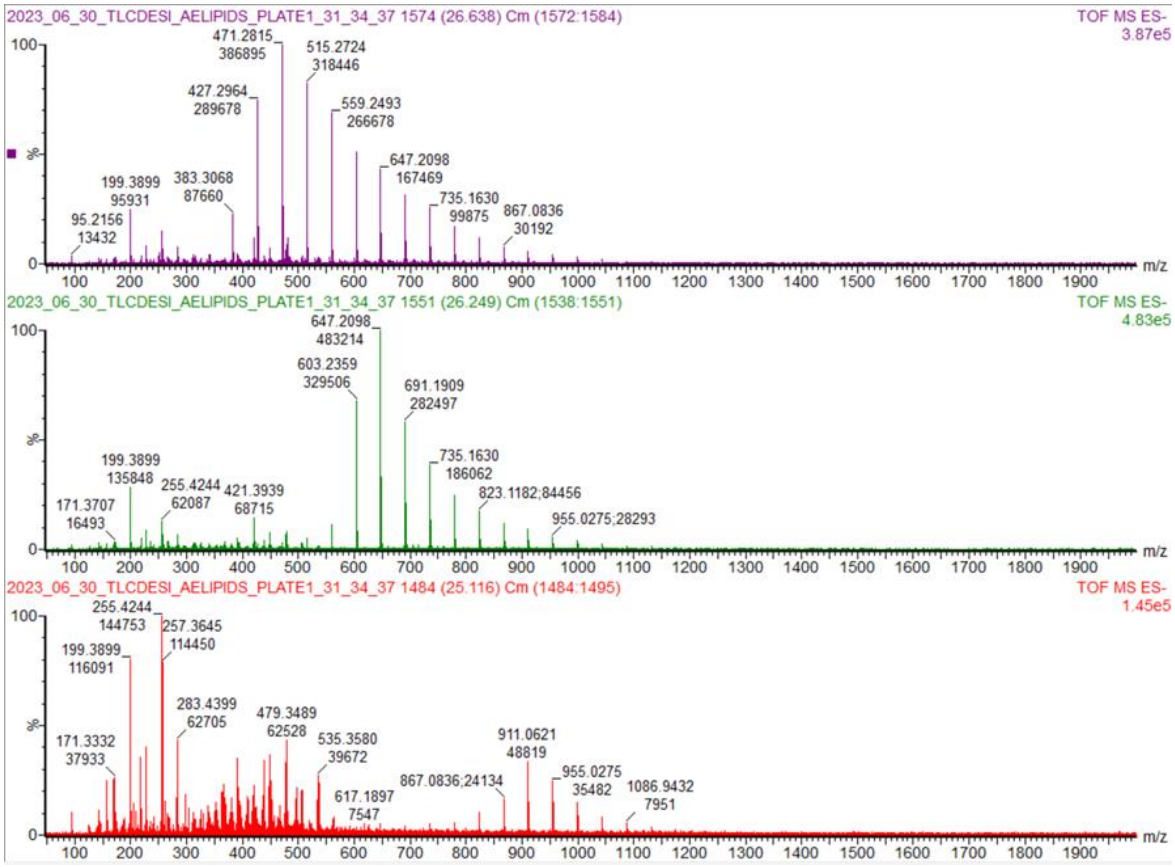


Figure 5.12 DESI-MS of DPAGT1/ALG13/ALG14 assay plates. Three spectra are shown of samples taken from spot positions identified as likely to contain C90-dolihol pyrophosphate $(GlcNAc)_2$ by comparing to positions from ^{14}C-GlcNAc reactions. Only detergent fragments are identified.

5.2.2.2 Kinetic analysis with native and non-native lipids

To further explore substrate preferences of the enzymes for the native, long (C90) dolichols over short (C55) and polyunsaturated (polyisoprenol) versions, kinetic parameters at steady state were ascertained for the DPAGT1/ALG13/ALG14 complex. 1.5 μM DPAGT1/ALG13/ALG14 was assayed at fixed, 50 μM UDP-[^{14}C]-GlcNAc with 1.6- 200 μM lipid. Experiments were carried out alongside 1.5 μM DPAGT1 or 0.25 μM ALG13/ALG14 standards with C90 and C55 lipids so that migration of the lipids on the TLC plates could be compared between experiments. Summaries of the experiments performed in biological triplicate with protein expressed on three occasions is shown in Figures 5.13 and 5.15 for the monophosphate and pyrophosphate GlcNAc species respectively. Exemplar TLC plates are shown in Figures 5.14 and 5.16, where production of both one and two-GlcNAc linked species were detected.

It was clear that when provided with polyunsaturated lipid phosphates, DPAGT1/ALG13/ALG14 was not active. For their dolichol equivalents, there was comparable k_{cat} between lipids of either C55 and C90 in length at 0.553 ± 0.15 and 0.62 ± 0.054 min^{-1}. The affinities were also comparable at 4.27 ± 4.6 and 1.27 ± 0.69 μM for C55 and C90 dolichol respectively- albeit with large errors between assays with different protein batches as discussed below.

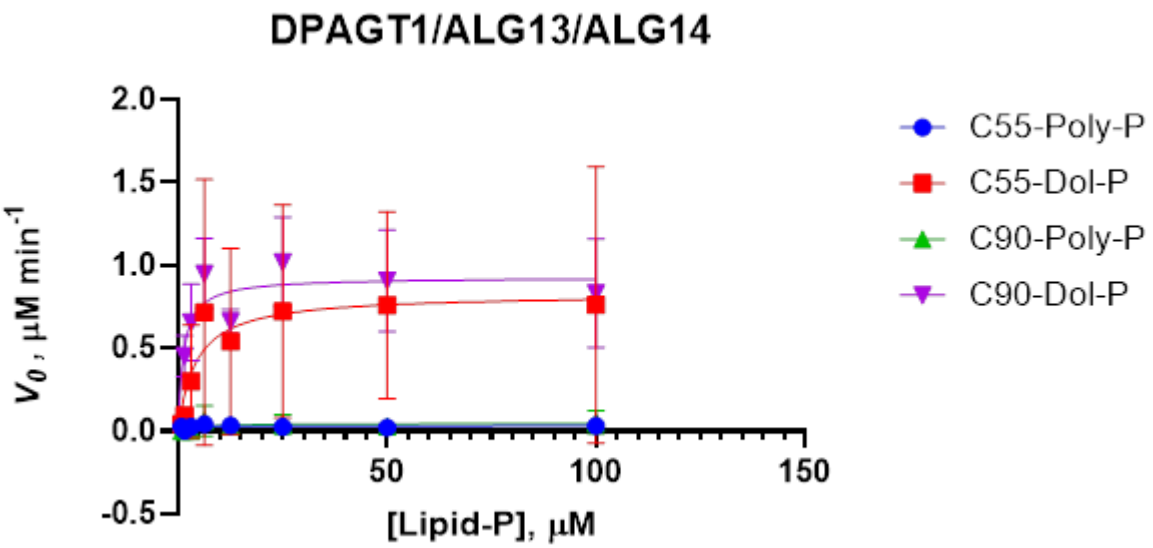


Figure 5.13 Kinetic of DPAGT1/ALG13 with native and non-native lipid substrates (N=3 biological triplicates). 1.5 µM DPAGT1/ALG13/ALG14 was incubated with 100 µM UDP-[^{14}C]-GlcNAc and between from 1.56 µM to 100 µM of C55 or C90 polyisoprenol (Poly) and dolichol (Dol) phosphates. Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters, as in Table 5.2.

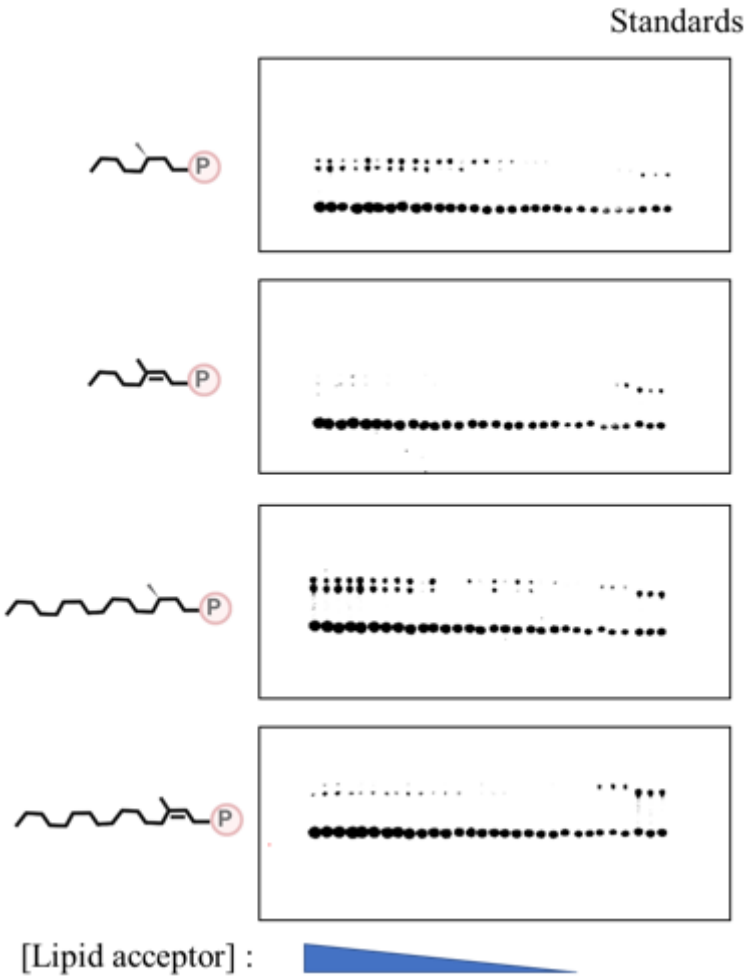


Figure 5.14 Example TLC plates from titration of each of the native and non-native lipid phosphates with DPAGT1/ALG13/ALG14. From the top TLC plate: C55 dolichol phosphate, C55 polyisoprenol phosphate, C90 dolichol phosphate and C90 polyisoprenol phosphate.

Table 5.2 DPAG1/ALG13/ALG14 N=3 biological repeats

	K_M, μM ± SE	k_{cat}, min^{-1} ± SE
C55-Poly-P	n/a	0.02 ± 0.004
C55-Dol-P	4.27 ± 4.6	0.55 ± 0.151
C90-Poly-P	n/a	0.031 ± 0.014
C90-Dol-P	1.27 ± 0.69	0.62 ± 0.054

Using pyrophosphate GlcNAc linked native and non-native lipids, the kinetics of ALG13/ALG14 within the complex were explored. As seen at end-point, all lipids could be utilised even when polyunsaturated in nature, and they tended towards the same maximal velocity (Figure 5.15). It seemed however, that the native C90-dolichol pyrophosphate GlcNAc displayed a higher affinity for the enzyme with a K_M of 2.22 ± 1.28 µM – suggesting the combination of length and saturation state drives down the K_M.

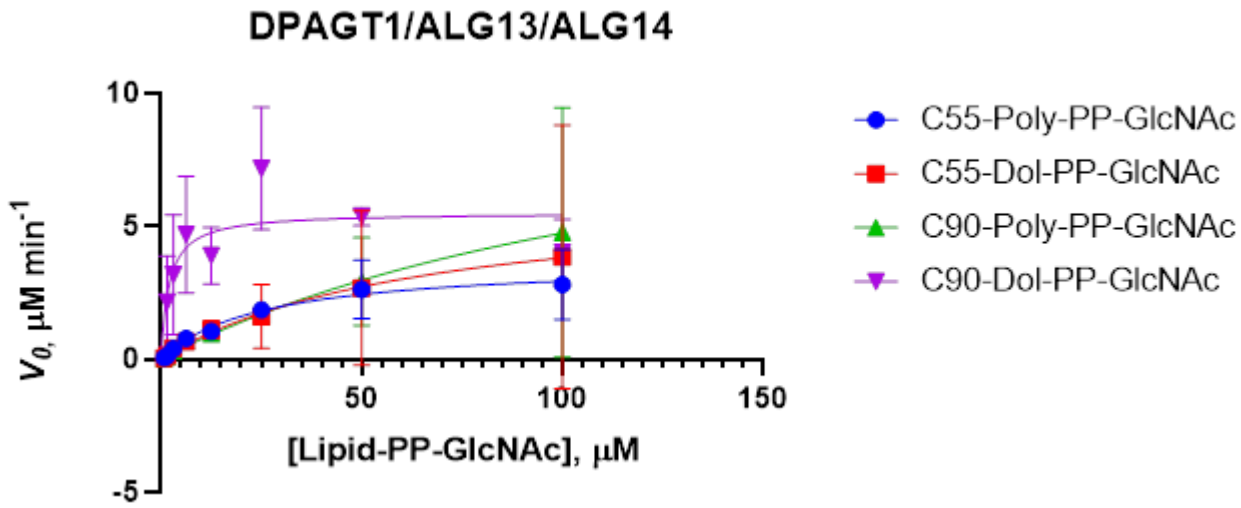


Figure 5.15 Kinetic of DPAGT1/ALG13 with native and non-native lipid substrates (N=3 biological triplicates). 0.25 µM DPAGT1/ALG13/ALG14 was incubated with 100 µM UDP-[14C]-GlcNAc and between from 1.56 µM to 100 µM of C55 or C90 polyisoprenol (Poly) and dolichol (Dol) pyrophosphate GlcNAc. Initial velocity against substrate concentration was modelled with the Michaelis-Menten equation (Equation 1, section 4.1.1) to estimate K_M and V_{max} parameters, as in Table 5.3.

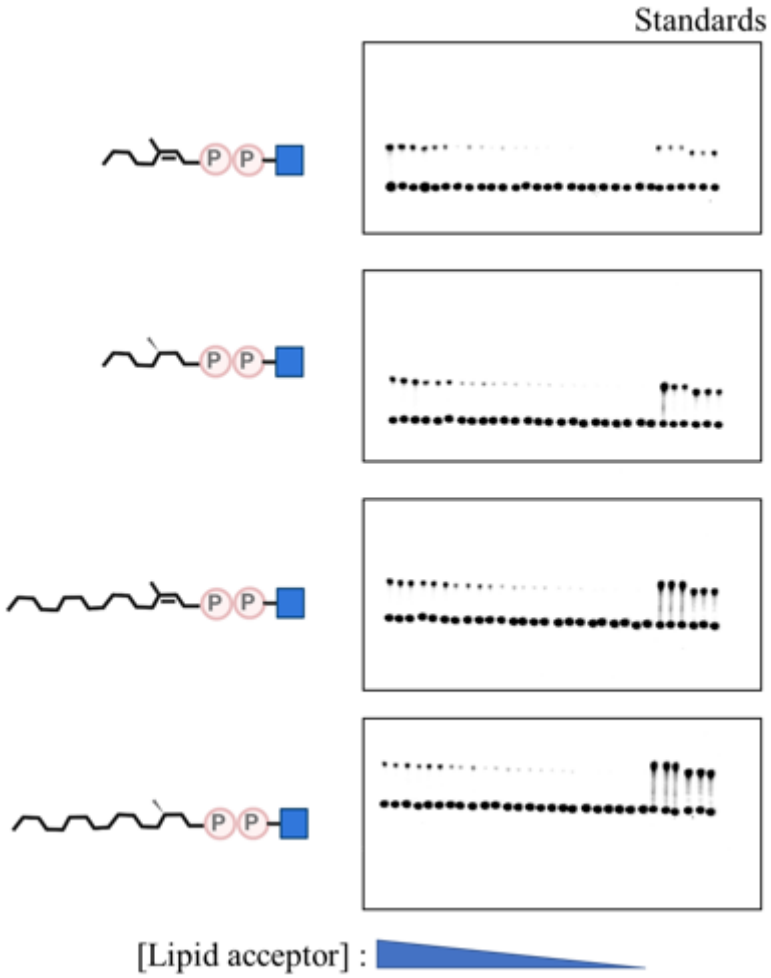


Figure 5.16 Example TLC plates from titration of each of the native and non-native lipid pyrophosphate GlcNAc. From the top TLC plate: C55 polyisoprenol pyrophosphate GlcNAc, C55 dolichol pyrophosphate GlcNAc, C90 polyisoprenol pyrophosphate GlcNAc and C90 dolichol pyrophosphate GlcNAc.

Table 5.3 DPAG1/ALG13/ALG14 N=3 biological repeats

	K_M µM ± SE	k_{cat}, min^{-1} ± SE
C55-Poly-PP-GlcNAc	25.65 ± 11.14	14.72 ± 3.72
C55-Dol-PP-GlcNAc	66.02 ± 98.9	25.28 ± 19.8
C90-Poly-PP-GlcNAc	149.10 ± 230.78	47.2 ± 49.36
C90-Dol-PP-GlcNAc	2.22 ± 1.28	22.72 ± 2.72

5.3 Discussion

The implication of eukaryotes utilising polyisoprenyl carrier lipids containing saturated α-isoprene units over the polyunsaturated equivalents used in bacteria is still not clearly understood in terms of both evolution and function [206]. By accessing C55 (11 prenol) and C90 (18 prenol) lipids in polyunsaturated and dolichol forms, the specificity of human DPAGT1 and ALG13/ALG14 glycosyltransferases were assessed. This could offer insight into the role of dolichol in human membranes.

Human length lipids were successfully accessed using plant extraction. C90 polyisoprenols containing 3 *trans*,*trans*, 14 *cis*,*cis* and 1 *cis* α were obtained from the leaves of the *Gingko biloba* tree, and ^{1}H NMR analysis matched that reported in the literature [199] as shown in Figure 5.4. Furthermore, [(*S*)-tol-binap]$RuCl_2$ was applied to convert these long polyisoprenols into dolichols for the first time, and was successful in yielding C90 dolichol on a small-scale. However, this proved to be difficult to reproduce on large scale preparations, perhaps due to poisoning of the catalysis from propanol that was not sufficiently dry or if the catalyst became exposed to air during the reaction. Fortunately, this synthesis was optimised by Dr Andrew Giltrap, who provided C90 dolichol and polyisoprenols as well as C55 lipids from extractions from bay leaves.

The phosphate and pyrophosphate GlcNAc versions of these lipids were found to dramatically stabilise the melting temperature of DPAGT1/ALG13/ALG14 indicating successful binding to the proteins (Figure 5.8). The pyrophosphate GlcNAc species of all lipid lengths and saturation states was the most stabilising, conferring a staggering 12.4, 16.0, 15.2 and 17.2 °C increase in T_m with C55 polyisoprenol, C55 dolichol, C90 polyisoprenol and C90 dolichol versions, respectively. These lipids are native and non-native versions of the DPAGT1 product and the ALG13/ALG14 substrate meaning they should bind in the active sites of both enzymes of the

complex. Given the polyunsaturated versions were later found to have low utilisation, these are of particular interest in efforts towards structural elucidation by cryoEM given they would not readily be turned over but lead to this increase in stability. It has long been known that ligand binding to proteins can confer stability to proteins, inducing less flexible and therefore more stable conformations and in turn increasing the enthalpy of unfolding [207]. Membrane proteins are typically flexible [208], and therefore it is likely DPAGT1/ALG13/ALG14 becomes substantially more ordered when bound to the lipid acceptor substrates. Interestingly, monophasic melting temperatures were observed with the complex. It is therefore likely that the signal is dominated by fluorescence from DPAGT1 given this larger protein contains 15 tryptophan resides and ALG13/ALG14 have a combined 7.

Endpoint activity experiments with each of the lipid panel were performed with DPAGT1 and ALG13/ALG14 to see whether in addition to binding, they could be accepted as substrates for the enzymes. When compared to that of the native C90 dolichol phosphate, DPAGT1 displayed 24.7% activity with the C90 polyunsaturated version. C55 dolichol phosphate had 125.0% activity, whereas C55 polyisoprenol had 26.9%. These data clearly show how the saturation state is the most significant driver of activity for DPAGT1, and that lipids of C55 length were just as acceptable as the native length lipid. Because DPAGT1 displays low activity with polyunsaturated lipids, downstream ALG13/ALG14 would not receive polyunsaturated substrates *in vivo*, and hence, the synthesis pipeline has a gatekeeper. This finding corroborates those in equivalent studies on phospho-GlcNAc transferases from yeast microsomes [189] and in CHO cells [193].

ALG13/ALG14 on the other hand were not discriminating in their activity with polyunsaturated lipids of either length, displaying in fact higher activity (132.5, 150.1, and 114.3%) with C90 polyisoprenyl, C55 polyisoprenyl or C55 dolichyl forms than native C90 dolichol pyrophosphate GlcNAc, although these were not found to be significantly higher.

McLachlan and Kragg found the same trend in microsomal extractions from CHO cells, where the homologous chitobiose forming enzyme displayed no preference to dolichol or polyisoprenols [209]. When studied as a time course, differences in the affinity of each of the lipids could be teased apart – with longer length lipids displaying a K_M lowering effect. However, these results remain putative as large errors were reported between replicates. Potential sources of error could be due to differences in the total active enzyme concentration of stocks – as discussed in Chapter 4 – although this should not impact K_M.

For DPAGT1, there was a clear V_{max} affect for dolichols over polyisoprenols (Figure 5.13). This specificity likely ensures only dolichol phosphates in the ER membrane are destined for use in N-glycan synthesis and not their polyunsaturated equivalents. It is conceivable that dolichol forms of the substrates are necessary for downstream applications – perhaps for conferring increased stability to concentrated areas of the membrane involved in sugar synthesis, or by aiding in the flipping of the two GlcNAc, five mannose linked oligosaccharide to the luminal leaflet. Furthermore, the pyrophosphate-sugar link to an alkyl pyrophosphate in dolichol is likely far more stable than to an unsaturated allylic version. Mutations in the enzyme responsible for reducing the α isoprene *in vivo*, SRD5α3, have been found in patients with congenital disorders of glycosylation (CDGs), again showing the importance of dolichol form lipids for proper N-linked glycosylation in humans. Furthermore, Kragg *et al.* found that in mutant CHO cells that accumulated polyunsaturated lipids there was four-fold less oligosaccharide linked polyisoprenols – although that is perhaps less disrupted than would be expected [210, 211].

The length of polyisoprenol lipids was not found to be a significant determining factor in utilisation by DPAGT1 or ALG13/ALG14, and therefore begs the question as to why such long, C90 lipids dominate in human cells. It is possible this extra length helps with anchoring the membrane proteins in complex or in reaching downstream interacting partners. It will

therefore be of interest to access successively shorter lipids to find the minimal requirement, and also to gain structural insight into the interplay of lipid acceptors when the enzymes are in complex.

Chapter 6

Studying pathogenic variants of ALG13/ALG14

6.1 Introduction

Congenital Disorders of Glycosylation (CDG) represent a group of over 170 monogenic disorders with extremely diverse clinical features owing to the systemic requirement of functional glycosylation across the body and during development [66, 99]. These disorders have been classified in a variety of ways since the first reported case of PMM2-CDG by Jaak Jaeken in 1980 [100]. They are now typically grouped by the pathway that their biochemical defect lies in, for example in N-glycosylation, O-glycosylation, glycosphingolipid or GPI-anchor biosynthesis [212]. Pathogenic variants of ALG13 and ALG14 have been described in patients with CDG and hence are referred to as N-glycosylation type-CDGs.

ALG13-CDG was identified in 2012 as the first CDG to be caused by a *de novo* mutation [71, 213, 214], where lysine at position 94 was replaced with glutamate leading to multi-system pathologies [214]. ALG13-CDG also represents one of only seven X-linked CDGs - the others being C1GALT1C1-, ATP6AP1-, PIGA-, SSR4-, SLC35A2-, and MAGT1-CDG [66, 215]. Interestingly, ALG13-CDG presents predominately in females although male patients have been identified. Patients typically suffer from severe infantile epilepsy, muscle hypotonia, developmental delay and, in female patients, intellectual disability [67, 213, 216]. Treatments for ALG13-CDG are currently constrained to control of symptoms, with a particular focus on management of epileptic seizures [216].

A variant where asparagine residue 107 is replaced with serine is by far the most prevalent form of ALG13-CDG [71, 216-223]. This mutation is predicted to be in a loop region (see Figure 6.1) and has been suggested to disrupt UDP-GlcNAc binding based on homology modelling with the structure of bacterial MurG [71]. Furthermore, this asparagine residue is highly conserved across species [220]. Molecular dynamics (MD) simulations were used by Mitusinska *et al.* to compare an ALG13 N107S/ ALG14 homology model to that of the wild type proteins [142]. When ALG13 was studied in isolation, the introduction of the serine was predicted to have a slight destabilising effect on protein folding (at 1.4 kcal mol^{-1} difference between the wild type and mutant) and once folded, there was increased flexibility of the pathogenic variant [142]. Recently, mutations encoding I17N, E69del and A81T variants of ALG13 have also been reported in CDG patients [71]. Interestingly, all ALG13-CDG patients have been found to display a normal transferrin glycosylation pattern which is typically the biomarker for impaired glycosylation [71, 220], although subtle abnormalities in small amounts of total plasma glycans (<0.3- 0.5 %) have been identified using mass spectrometry assays [213, 218]. This could be due to a mosaic distribution of the pathogenic variant in affected females by X chromosome inactivation, although this would not explain that seen for male patients, and in one X-inactivation study carried out for a female ALG13-CDG patient, only random inactivation was found with no skewness [224] . It has therefore been hypothesised that pathogenic variants lead to the gain of an abhorrent function elsewhere in the cell leading to disease.

In 2012, Belaya *et al.* reported *DPAGT1* as a gene in which mutations cause congenital myasthenic syndrome (CMS) [104], and the following year they also identified CMS caused by mutations in *ALG2* and *ALG14* [103]. Owing to their role in N-glycosylation, these defects in DPAGT1, ALG2 and ALG14 are classified as CDGs. CMS is a fatigable muscle disorder in which impaired signal transmission at the neuromuscular junction leads to decrement of the muscle response to nerve stimulation [103, 225]. These findings highlight the importance of

functional glycosylation on the action of receptors, with the neuromuscular synapse likely to be particularly sensitive due to the requirement of N-glycosylation for proper folding and activity of the acetylcholine receptor [103]. To date, eight missense variants of ALG14 have been reported: A11T [226], P65L[103], D74N [227], R109Q [103, 227, 228], V141G [227] , G145N [229] , I175N [228], and V197G [229] (as in Figure 6.1).

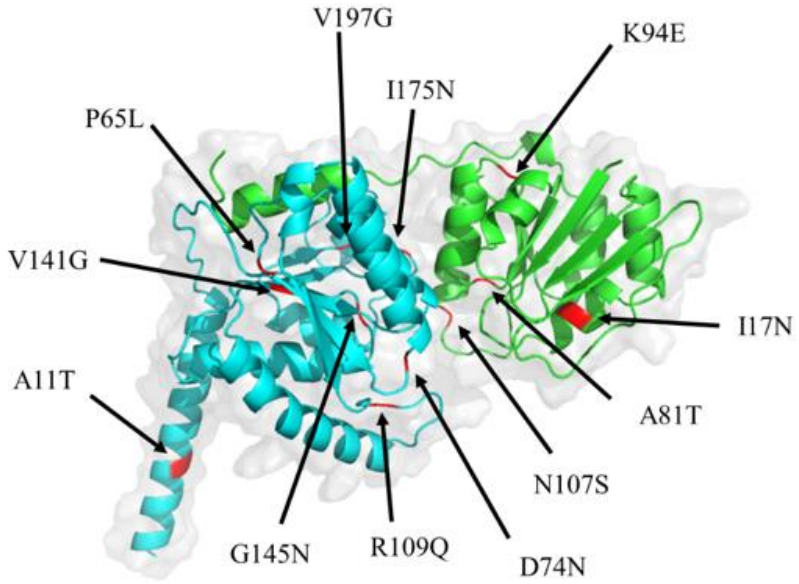


Figure 6.1 Location of missense variants of ALG13/ALG14 identified in patients with CDG. Mutations are labelled in red onto an AlphaFold Multimer [122] predicted structure of the human heterodimer (see Chapter 3 for details), where ALG13 is in green and ALG14 in blue.

The severity of disease experienced by patients with ALG13- and ALG14-CDG is extremely variable, even amongst those with the same mutation. It has been speculated that some CDG mutants may effect enzyme activity and lead to severe disease, and others may affect multi-enzyme complex formation which could result in milder forms or CMS [212]. It was therefore of interest to monitor the functional effect of select ALG13- and ALG14- CDG mutants which could in turn help inform the development of targeted treatments for patients. During the course

of this study, a detailed assessment of ALG13 and ALG14 mutants was reported by Gao *et al.* [68], providing a useful comparator to the findings, as discussed below.

6.2 Results

6.2.1 AlphaFold and AlphaMissense predictions of ALG13/ALG14 variants

The nine missense mutations in ALG13 isoform 2 or ALG14 that had been identified in patients with CDG at the start of this study were selected for study [71, 103, 214, 227, 228]. The predicted effect of these on the 3D structure of the heterodimer was made by modelling each variant using AlphaFold Multimer on Google Colab notebook [123]. Figure 6.1 and 6.2 shows those containing ALG13 and ALG14 mutations respectively, where each was aligned with the prediction for the wild-type protein using PyMOL [125]. There was predicted to be very little structural change conferred by of any of the variants with a global root mean square deviation (RMSD) of 0.6 Å for each of the ALG13 variants, and 0.2-0.3 Å for those in ALG14 (Table 6.1). This is perhaps unsurprising given AlphaFold is a based model trained on existing structures and hence when using multi-sequence alignments it is likely to predict similar overall structure for the single-change variants.

Table 6.1 Global root mean square deviation (RMSD) of alignments of the predicted structures of pathogenic ALG13/ALG14 proteins to the wildtype.

	Variant	Global RMSD, Å
ALG13	**I17N**	0.644
	A81T	0.602
	K94E	0.627
	N107S	0.619
ALG14	**D74N**	0.260
	I175N	0.274
	P65L	0.303
	R109Q	0.257
	V141G	0.259

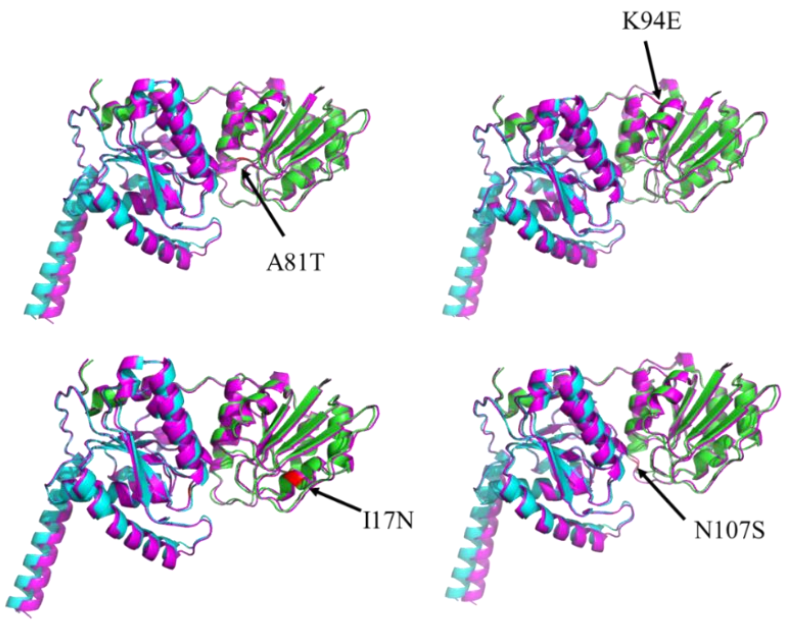


Figure 6.2 Alignment of the predicted structure of pathogenic variants of ALG13 with that of wild type ALG13/ALG14. Wild-type protein is coloured in purple, with the ALG13 containing CDG variant in green, predicted in complex with wild-type ALG14 in blue.

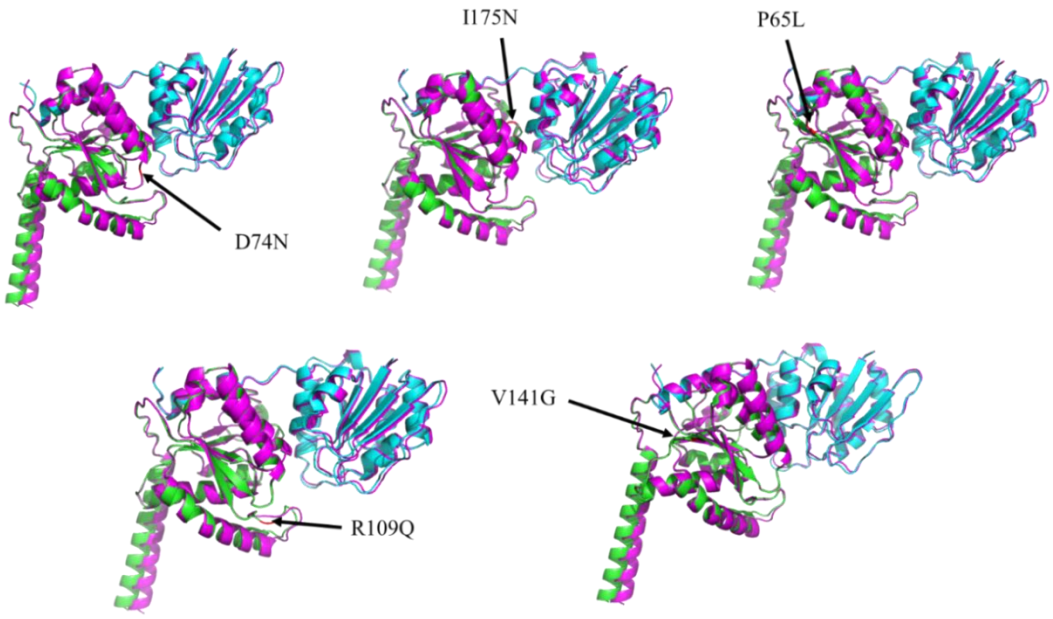


Figure 6.3 Alignment of the predicted structure of pathogenic variants of ALG14 with that of wild type ALG13/ALG14. Wild-type protein is coloured in purple, with the ALG14 containing CDG variant in green, predicted in complex with wild-type ALG13 in blue.

Recently, Cheng *et al.* utilised the advancements in protein structure prediction and population frequency data to predict the likelihood of all possible amino acid substitutions of causing disease [230]. Their published AlphaMissense database provides predictions for all human proteins whose predicted structure is available in AlphaFold2 [230]. Using the search tool developed by the Hegedűs lab [231] , the effect of a substitution at the position of each of the nine ALG13- and ALG14-CDG missense variants were evaluated. Remarkably, each disease variant is correctly predicted to be likely pathogenic (Table 6.2 and 6.3). The corresponding 'hot spot' analysis predicts missense changes in the globular C-terminus of ALG14 to be most likely to be pathogenic (Figure 6.4). As discussed in Chapter 3, only the structural prediction information for ALG13 isoform 1 is populated in the AlphaFold2 database [123], whereas it is isoform 2 that encodes the glycosylation domain that acts with ALG14 during lipid linked oligosaccharide synthesis. The predictions for ALG13 (Figure 6.4B and Table 6.2) are

therefore made assuming the effect for the 1,137 amino acid protein would be comparable to that of the 165 amino acid glycosyltransferase.

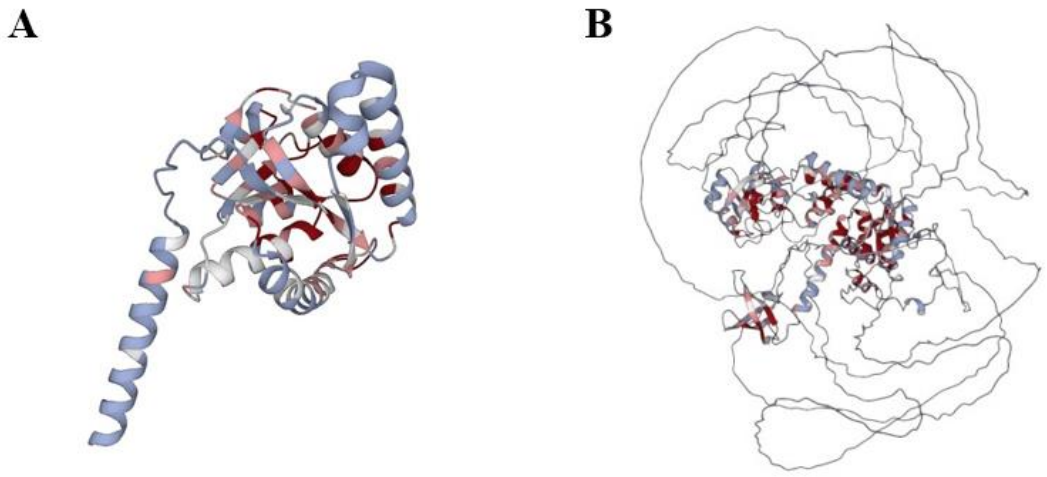


Figure 6.4 AlphaMissense HotSpot analysis of ALG14 (A) and ALG13 isoform 1 (B) where proteins are coloured by mean missense score per residue, with red = most likely pathogenic and blue = most likely benign.

Table 6.2 Summary of the predicted pathogenicity of ALG13 variants by AlphaMissense [230, 231]. Data are from that of ALG13 isoform 1, searched by 'ALG13_HUMAN, Q9N73'. Positions in bold are of those found mutated in CDG patients.

Position	Protein variant	a.a.1	a.a.2	Pathogenicity score	Pathogenicity class
17	p.Ile17Leu	I	L	0.4467	Ambiguous
	p.Ile17Phe	I	F	0.8028	Likely pathogenic
	p.Ile17Val	I	V	0.0892	Likely benign
	p.Ile17Asn	**I**	**N**	**0.9685**	**Likely pathogenic**
	p.Ile17Ser	I	S	0.8955	Likely pathogenic
	p.Ile17Thr	I	T	0.5765	Likely pathogenic
	p.Ile17Met	I	M	0.4575	Ambiguous
81	p.Ala81Pro	A	P	0.9839	Likely pathogenic
	p.Ala81Ser	A	S	0.5528	Ambiguous
	p.Ala81Thr	**A**	**T**	**0.9211**	**Likely pathogenic**
	p.Ala81Glu	A	E	0.9928	Likely pathogenic
	p.Ala81Gly	A	G	0.4847	Ambiguous
	p.Ala81Val	A	V	0.9537	Likely pathogenic
94	p.Lys94Gln	K	Q	0.6734	Likely pathogenic
	p.Lys94Glu	**K**	**E**	**0.9007**	**Likely pathogenic**
	p.Lys94Arg	K	R	0.1867	Likely benign
	p.Lys94Met	K	M	0.6062	Likely pathogenic
	p.Lys94Thr	K	T	0.7209	Likely pathogenic
	p.Lys94Asn	K	N	0.927	Likely pathogenic
107	p.Asn107Asp	N	D	0.8445	Likely pathogenic
	p.Asn107His	N	H	0.7926	Likely pathogenic
	p.Asn107Tyr	N	Y	0.9027	Likely pathogenic
	p.Asn107Ile	N	I	0.9557	Likely pathogenic
	p.Asn107Ser	**N**	**S**	**0.6215**	**Likely pathogenic**
	p.Asn107Thr	N	T	0.8184	Likely pathogenic
	p.Asn107Lys	N	K	0.9804	Likely pathogenic

Table 6.3 Summary of the predicted pathogenicity of ALG14 variants by AlphaMissense [230, 231]. Data are from that of ALG14, searched by 'ALG14_HUMAN, Q96F25'. Positions in bold are of those found mutated in CDG patients.

Position	Protein variant	a.a.1	a.a.2	Pathogenicity score	Pathogenicity class
74	p.Asp74Glu	D	E	0.9789	Likely pathogenic
	p.Asp74Ala	D	A	0.9871	Likely pathogenic
	p.Asp74Gly	**D**	**G**	**0.9788**	**Likely pathogenic**
	p.Asp74Val	D	V	0.9898	Likely pathogenic
	p.Asp74Asn	D	N	0.9547	Likely pathogenic
	p.Asp74His	D	H	0.9948	Likely pathogenic
	p.Asp74Tyr	D	Y	0.9839	Likely pathogenic
175	p.Ile175Met	I	M	0.2672	Likely benign
	p.Ile175Asn	**I**	**N**	**0.9285**	**Likely pathogenic**
	p.Ile175Ser	I	S	0.9039	Likely pathogenic
	p.Ile175Thr	I	T	0.8911	Likely pathogenic
	p.Ile175Leu	I	L	0.1328	Likely benign
	p.Ile175Phe	I	F	0.3694	Ambiguous
	p.Ile175Val	I	V	0.1149	Likely benign
65	p.Pro65Arg	P	R	0.5538	Ambiguous
	p.Pro65His	P	H	0.6741	Likely pathogenic
	p.Pro65Leu	**P**	**L**	**0.6403**	**Likely pathogenic**
	p.Pro65Ala	P	A	0.6102	Likely pathogenic
	p.Pro65Ser	P	S	0.7063	Likely pathogenic
	p.Pro65Thr	P	T	0.5412	Ambiguous
109	**p.Arg109Gln**	**R**	**Q**	**0.8284**	**Likely pathogenic**
	p.Arg109Leu	R	L	0.9744	Likely pathogenic
	p.Arg109Pro	R	P	0.9947	Likely pathogenic
	p.Arg109Gly	R	G	0.9737	Likely pathogenic
	p.Arg109Trp	R	W	0.8222	Likely pathogenic
141	p.Val141Ala	V	A	0.8683	Likely pathogenic
	p.Val141Glu	V	E	0.9878	Likely pathogenic
	p.Val141Gly	**V**	**G**	**0.8909**	**Likely pathogenic**
	p.Val141Leu	V	L	0.3013	Likely benign
	p.Val141Met	V	M	0.706	Likely pathogenic

6.2.2 Cloning of baculoviral expression vectors

To functionally assess these ALG13- and ALG14-CDG variants, attempts were made to purify pathogenic versions of the ALG13/ALG14 complex. As for the wild-type proteins, Sf9 insect cells were used for expression of the heterodimers. Baculovirus vectors containing each of the disease variants were therefore generated using site-directed mutagenesis, as described in Chapter 2. For this, a MultiBac vector containing wildtype ALG13 with N-terminally 6xHis-tagged ALG14 was used as the template in each case, with primers as in Chapter 2, Tables 2.5 and 2.6. Following mutagenesis, a screen was performed to check for successful nucleotide substitution. Both the ALG13 and ALG14 genes were amplified from the vector using PCR with the high-fidelity Herculase II Fusion DNA polymerase (Figure 6.5). The PCR product was subject to Sanger sequencing using the WIMM in-house service, and clones were selected that contained only the required mutation on the target gene (as in Figures 6.6 and 6.7). This required multiple rounds of mutagenesis, with assistance by Jing Wang.

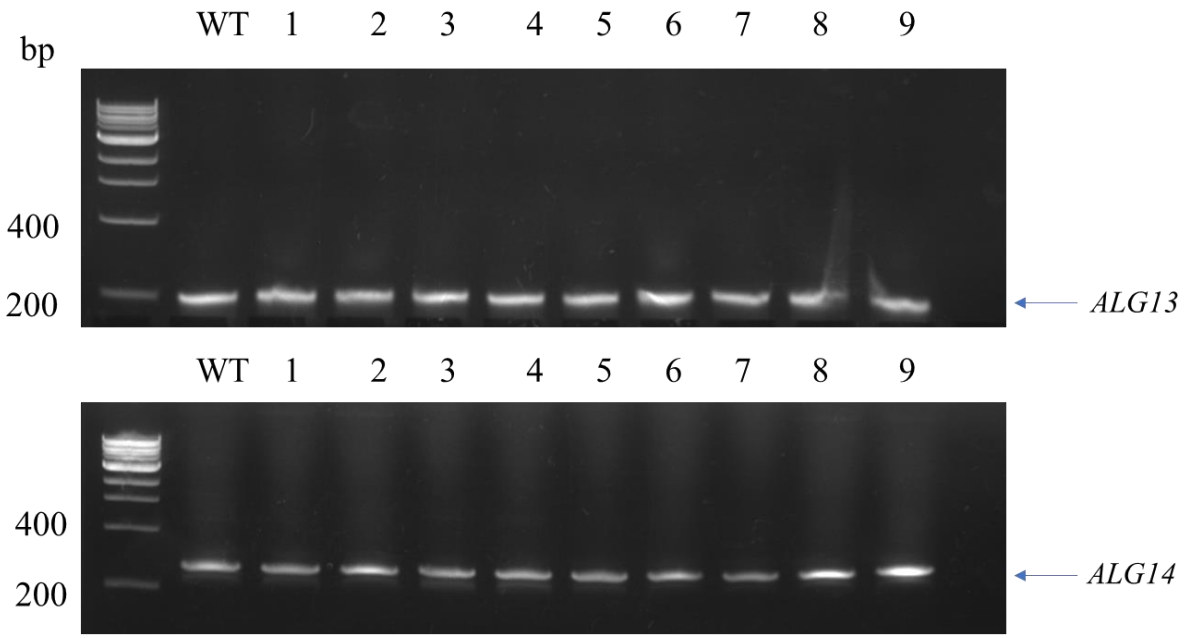


Figure 6.5 PCR screen showing successful mutagenesis of ALG13/ALG14 MultiBac plasmids, where WT = wild type ALG13/ALG14, 1= ALG13/ALG14 R109Q, 2=

ALG13/ALG14 I175N, 3= ALG13/ALG14 D74N, 4= ALG13/ALG14 V141G, 5= ALG13/ALG14 P65L, 6= ALG13 N107S/ALG14, 7= ALG13 I17N/ALG14, 8= ALG13 A81T/ALG14, 9= ALG13 K94E/ALG14.

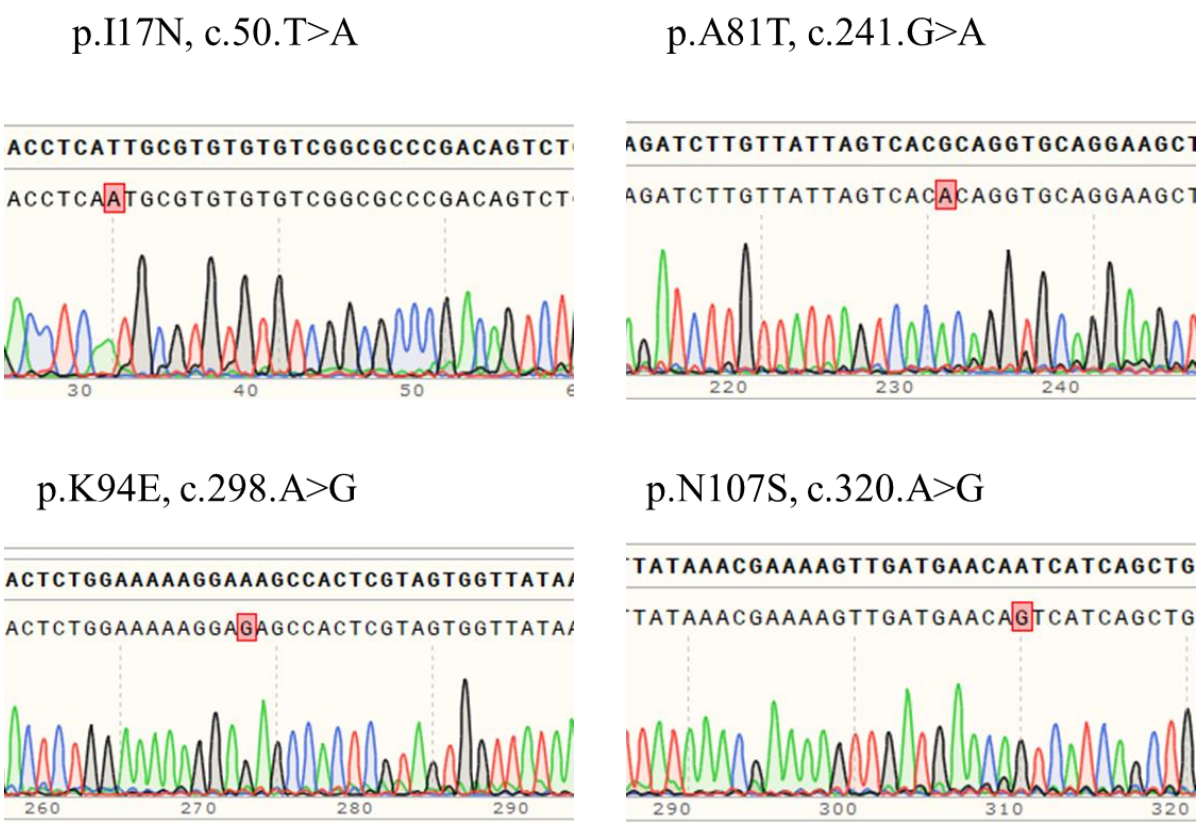
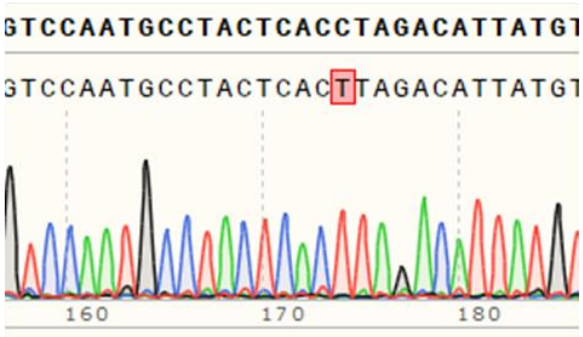
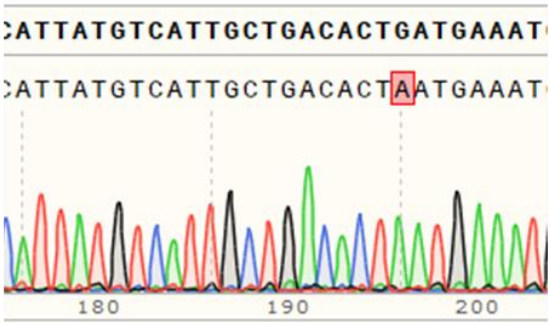


Figure 6.6 Sanger-sequencing showing successful introduction of pathogenic variants of ALG13. The disease variant and corresponding DNA mutation are labelled above each sequencing result.

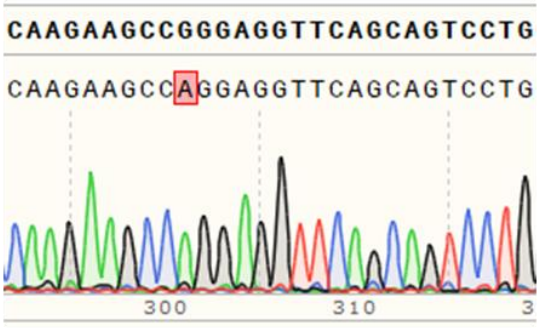
p.P65L, c.194.C>T



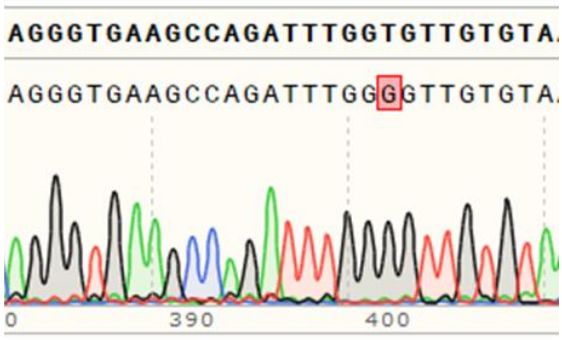
p.D74N, c.220.G>A



p.R109Q, c.326.G>A



p.V141G, c.422.T>G



p.I175N, c524.T>A

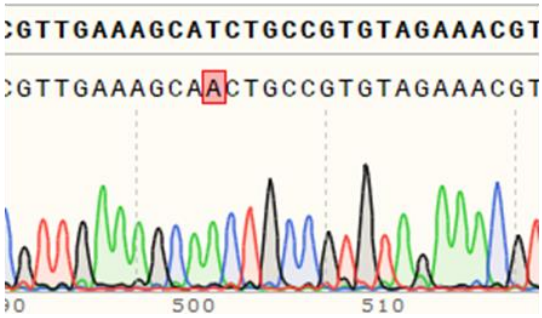


Figure 6.7 Sanger-sequencing showing successful introduction of pathogenic variants of ALG14. The disease variant and corresponding DNA mutation are labelled above each sequencing result.

6.2.3 Purification of pathogenic variants of ALG13/ALG14

MultiBac vectors encoding nine targeted ALG13/ALG14 variants were successfully recombined with bacmid DNA to access baculovirus, as described previously for the wild-type proteins. The second (P2) generation baculovirus for each variant was tested for its ability to induce expression of YFP in Sf9 cells as a proxy for infectivity, and in each case was found to infect >90% cells using fluorescent microscopy (data not shown). These successful stocks were used to infect 1 L cultures of Sf9 insect cell culture, and the expression and purification of ALG13/ALG14 proteins were trialled.

The methodology employed for the purification of ALG13/ALG14 wild-type proteins (as in Chapter 3) was applied to purify mutant ALG13/ALG14 from the membrane fraction of these cultures. The peak fractions from IMAC with Talon resin for each ALG13/ALG14 mutant were taken for high throughput purification using a Superdex 200 Increase 5/150 GL column on a fSEC fitted with an auto-sampler capability (accessed at the Membrane Protein Laboratory, Diamond Light Source). For five mutant complexes (containing ALG13 A81T, ALG14 D74N, P65L, I175N and R109Q) the elution profiles were similar to the wild-type protein with a void followed by a peak containing the target protein complex (Figures 6.8-10). For the remaining four (containing ALG14 V141G, ALG13 I17N, ALG13 K94E and ALG13 N107S), there was no detectable expression (Figures 6.8 and 6.9). Given these are known disease causing variants it was unclear whether the expression was affected by the mutations themselves or an experimental artefact. This experiment was therefore repeated in biological triplicate with Sf9 insect cell infection on three separate occasions which confirmed the same expression pattern.

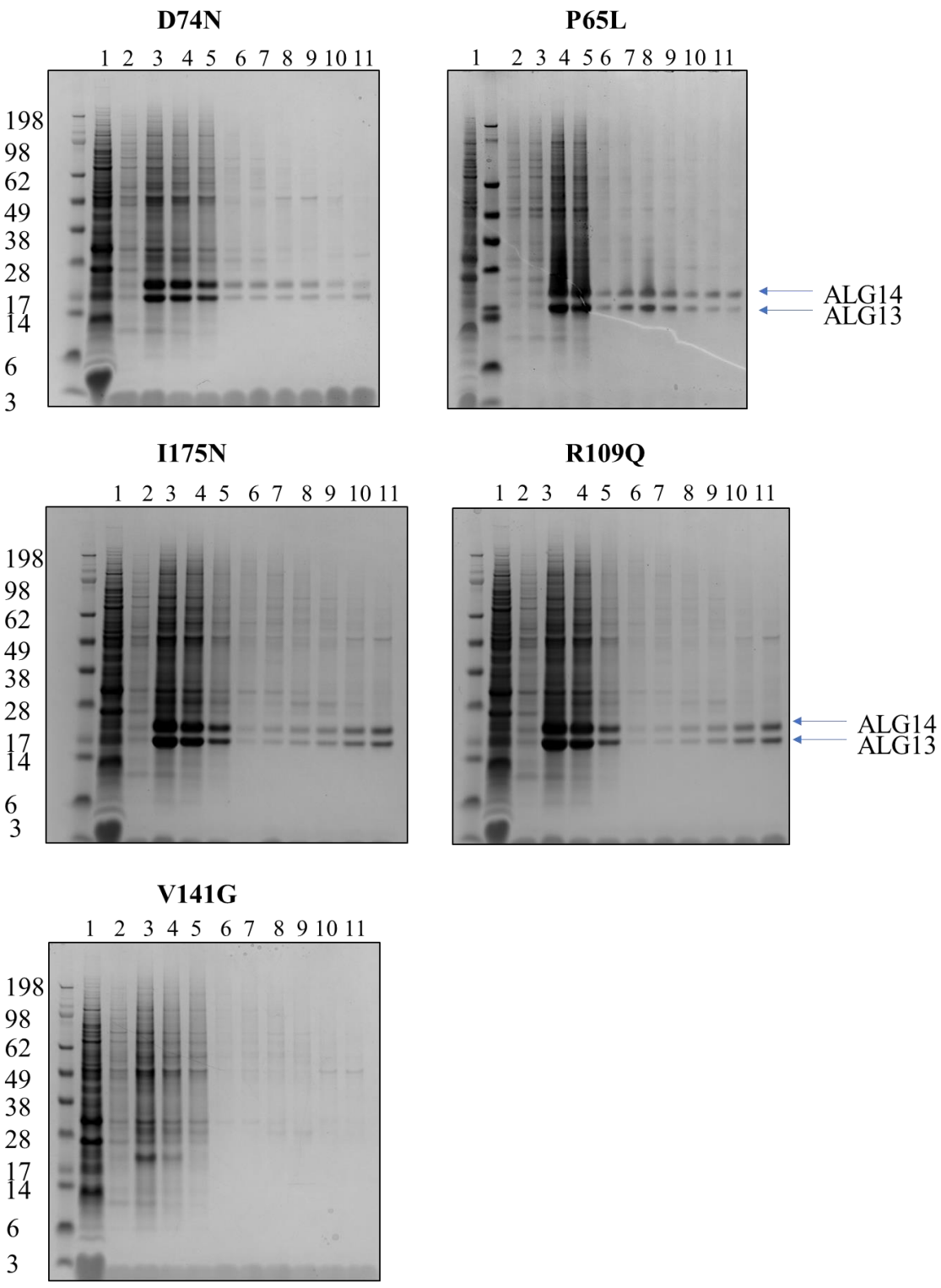


Figure 6.8 Representative SDS-PAGE gels showing progress of N=3 ALG13/AL14 mutant protein purifications, where 1= flow through, 2= wash fraction, 3-5 = IMAC elations 1, 2 and 3, and 6-12 = peak fractions from size exclusion chromatography (as in Figure 6.10)

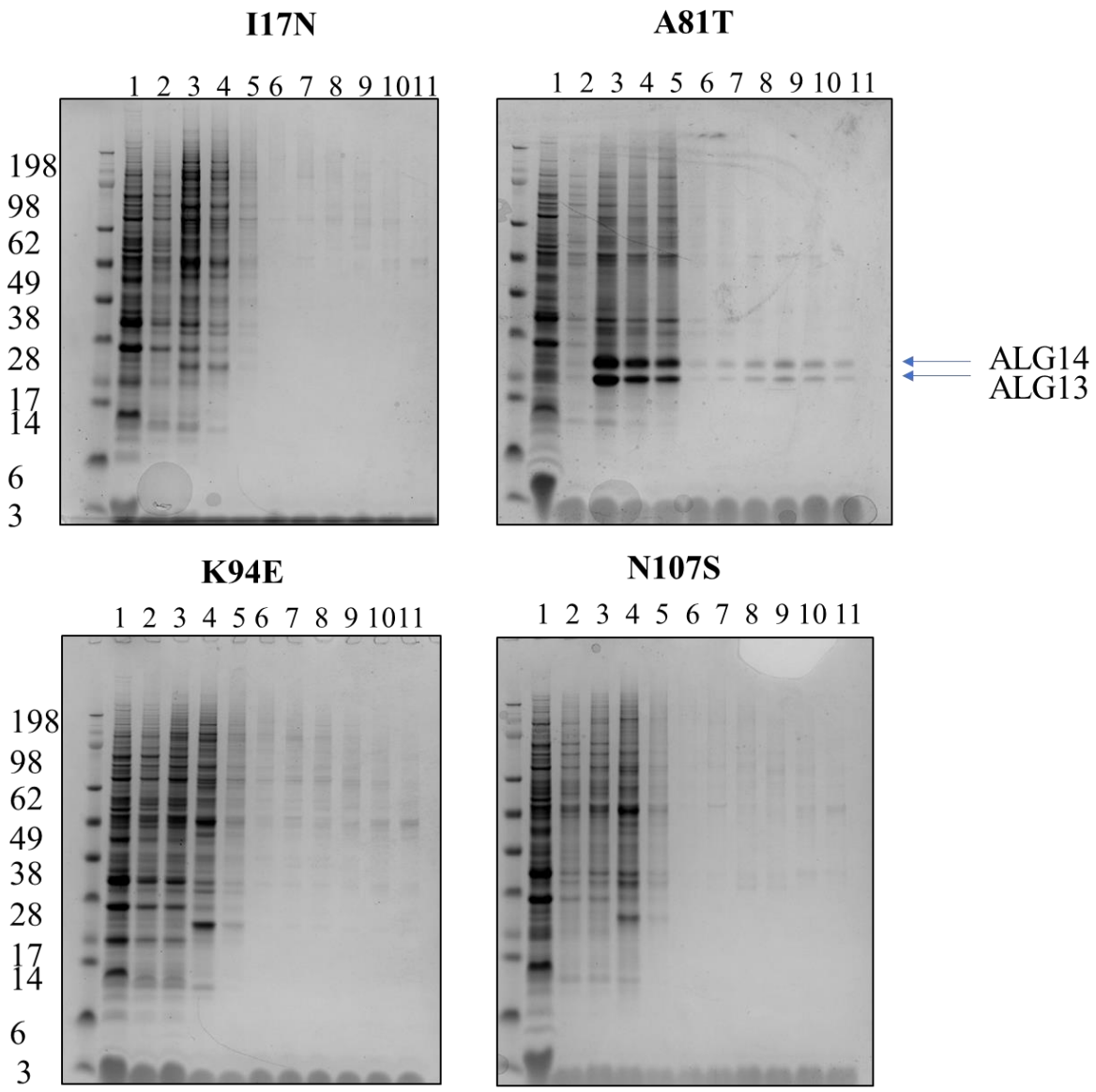


Figure 6.9 Representative SDS-PAGE gels showing progress of N=3 ALG13/AL14 mutant protein purifications, where 1= flow through, 2= wash fraction, 3-5 = IMAC elations 1, 2 and 3, and 6-12 = peak fractions from size exclusion chromatography (as in Figure 6.10)

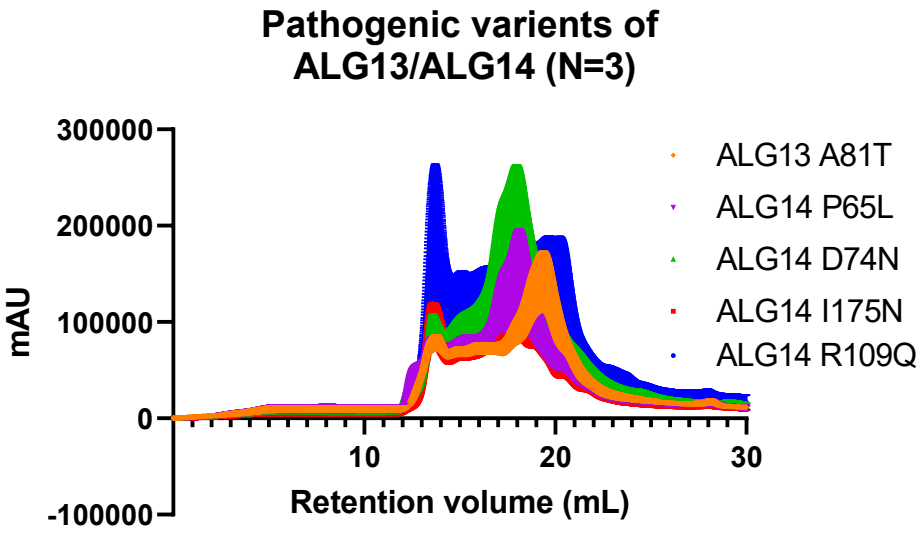


Figure 6.10 SEC profiles of 5 successfully expressed ALG13/ALG14 mutants (N=3), identified by the mutant protein within the ALG13/ALG14 heterodimeric complex. Data shown is absorbance at 280 nm (milli-absorbance units, mAU), measured using a high-throughput fSEC with Superdex 200 Increase 5/150 column (Cytiva).

To discern why there was no expression in 4 out of the 9 targeted ALG13/ALG14 variants, screens were performed at both the protein and RNA level. First, Western blots were used to probe for ALG13 and ALG14 in the crude cell lysate of infected Sf9 cells to ensure even low levels of proteins would be detected. n-dodecyl-phosphocholine (Fos-choline-12) was selected for use in protein extraction for its ability to solubilise almost all proteins so that any ALG13/ALG14 present would be detected even if it was not extracted by the DM/CHS in previous purification attempts.

Briefly, 50 mL of Sf9 insect cells at mid-log were infected with 250 µL of P2 generation baculovirus encoding ALG13/ALG14 wild type or pathogenic mutant version of the protein. Cells were harvested after 72 hours and disrupted using sonication. The membrane proteins were then solubilised from the cell lysates by addition of 1% fos-choline-12. The lysates were

cleared of cell debris by centrifugation before equal samples of the supernatant were dissolved in SDS sample buffer. The samples were loaded onto 3x SDS-PAGE gels to allow for both Coomassie staining for comparison of loading (Figure 6.11A) and Western blots for probing of ALG13 and ALG14 proteins. ALG14 was identified by its 6xHis tag and ALG13 using a protein specific antibody (Figure 6.11B).

ALG13 and ALG14 proteins were identified in the cell lysate in the same expression pattern as previous purifications apart from for ALG13 I17N/ ALG14 where both proteins were detected despite not having been successfully purified before. However, when ALG13 N107S and ALG13 K94E were expressed in the complex, no ALG14 was detected. Similarly, when ALG14 V141G was expressed there was no detectable ALG13. In each case these were the 'wild type' component of the expression vector that was not produced, and hence ruling out the theory that the pathogenic mutation itself was implicating expression.

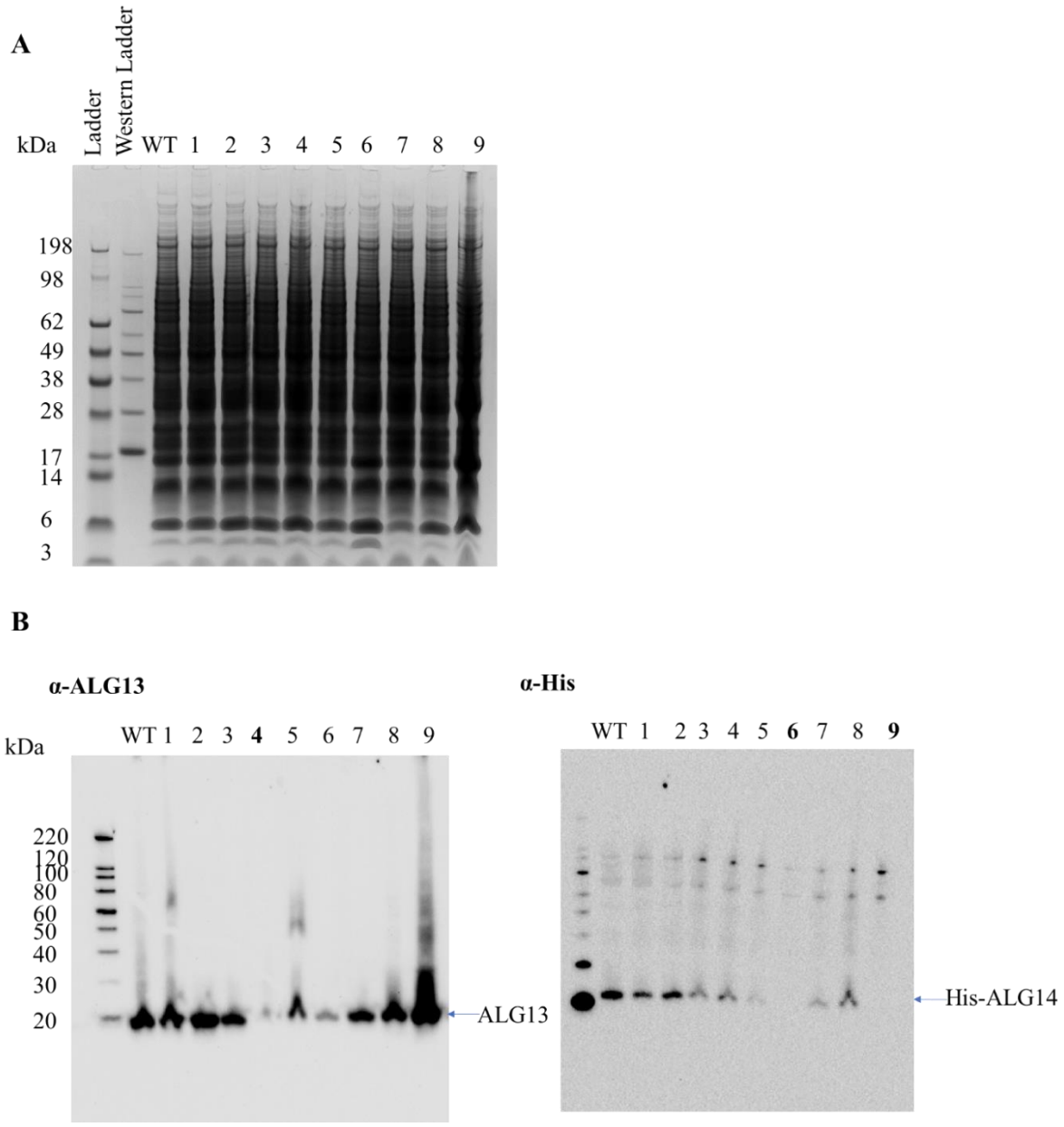


Figure 6.11 Protein detected in the fos-choline extracted cell lysate of ALG13/His-ALG14 expressing insect cells, where WT = wild type ALG13/ALG14, 1= ALG13/ALG14 R109Q, 2= ALG13/ALG14 I175N, 3= ALG13/ALG14 D74N, 4= ALG13/ALG14 V141G, 5= ALG13/ALG14 P65L, 6= ALG13 N107S/ALG14, 7= ALG13 I17N/ALG14, 8= ALG13 A81T/ALG14, 9 = ALG13 K94E/ALG14. **A** Coomassie stained SDS-PAGE gel showing total protein loaded, with samples identical to those loaded in **B**. **B** Anti-ALG13 Western blot for detection of ALG13 protein and anti-His Western blot for detection of the 6xHis tagged ALG14 protein.

These results were corroborated when screening for RNA. Total RNA was extracted from 3 mL of Sf9 insect cells expressing ALG13/ALG14 wild type or mutant proteins. Reverse transcriptase primed with ALG13 or ALG14 specific oligonucleotides was used to generate cDNA from the purified RNA. This was amplified by PCR and the resulting product was analysed by DNA electrophoresis (Figure 6.12). ALG13 encoding RNA was not detected in ALG14 V141G expressing cells, and ALG14 RNA was not detected in ALG13 N107S or ALG13 K94E constructs. As at the protein level, this is the wild-type protein out of the ALG13/ALG14 pair that is not being transcribed, whereas the RNA encoding the mutant version of the complex was detected in each case. These data point to a mutation or truncation elsewhere in the vector that has led to the lack of expression in these samples, for example in the promotor region. These are likely to have been introduced during the amplification of the MultiBac vector following site-directed mutagenesis given this was only in isolated cases. For ALG13 I17N, RNA of both mutant and wildtype protein was detected despite purifications failing in three attempted scale ups.

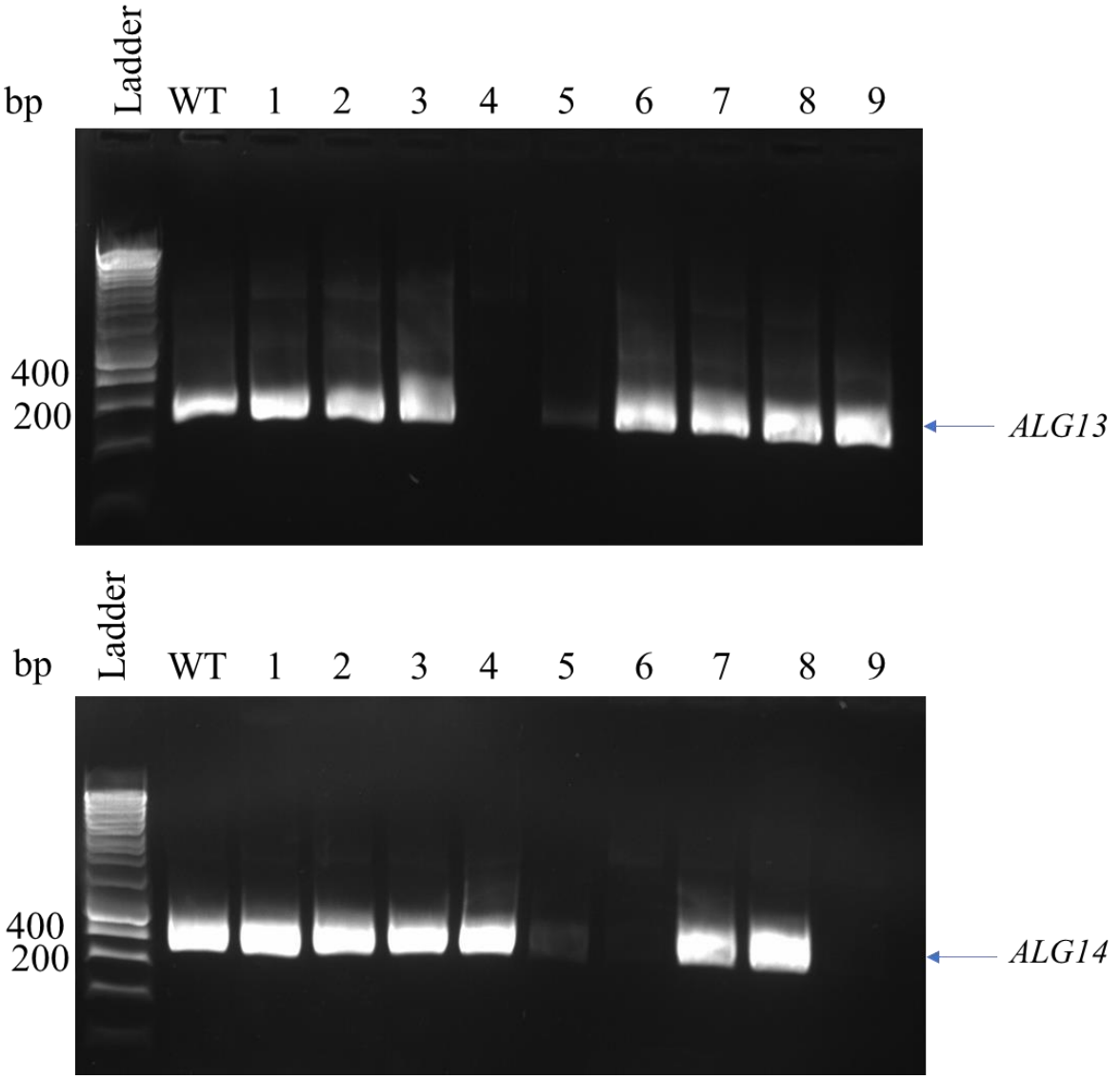


Figure 6.12 ALG13 (top) and ALG14 (bottom) gene amplified from RNA of Sf9 insect cells infected with MultiBac vectors encoding wild type or mutant versions of the proteins. WT = wild type ALG13/ALG14, 1= ALG13/ALG14 R109Q, 2= ALG13/ALG14 I175N, 3= ALG13/ALG14 D74N, 4= ALG13/ALG14 V141G, 5= ALG13/ALG14 P65L, 6= ALG13 N107S/ALG14, 7= ALG13 I17N/ALG14, 8= ALG13 A81T/ALG14, 9= ALG13 K94E/ALG14. Nb. Lane 5 shows faint levels of both ALG13 and ALG14 due to poorly amplified cDNA product.

6.2.4 Characterisation of pathogenic variants of ALG13/ALG14

Whilst cloning of new ALG13/ALG14 MultiBac expression vectors takes place, four of the successfully purified ALG13/ALG14 pathogenic variants were taken forward for characterisation. In each case, the activity of GlcNAc transfer was compared to that of the wild type using radioactive labelled UDP-[^{14}C]-GlcNAc, as in Chapters 4 and 5. To start, the end-point transfer after 1 hour by 0.5 μM protein with 50 μM UDP-[^{14}C]-GlcNAc and 100 μM dolichol-pyrophosphate GlcNAc was assessed. Interestingly, it was found that the ALG13/ALG14 P65L mutant had in fact higher activity to that of the wild type protein. ALG13/ALG14 D74N showed 84% activity whereas ALG13/ALG14 I175N displayed no glycosyltransferase activity. ALG13 A81T /ALG14 showed 24% of the activity of the wild type after 1 hour. These results are mostly consistent with that reported in the literature [195], where ALG13 A81T variants were found to have 27% activity of the wild-type protein, ALG14 D75N at 22% and ALG14 P65L at 80%. The activity of ALG14 I175N has not previously been studied.

Multiple attempts were made to perform activity time-courses with the mutant proteins to monitor whether there were changes in substrate binding or to the rate of the enzyme catalysis. Unfortunately, on repeated use of the stocks of the mutant enzyme, total activity seemed to be lost in each case (data not shown). The stocks from protein expressed on three occasions were screened using nano-DSF, where a striking decrease in melting temperature was seen for protein complex containing ALG14 mutants P65L, D74N and R109Q (Figure 6.14, data shown is from one biological repeat for clarity). Although ALG13 mutant A81T did have comparable stability to the wild type protein, these data suggested the degradation of the other protein stocks during storage at -80 °C, which may implicate stability changes conferred by the point mutations given this did not occur for the wild-type.

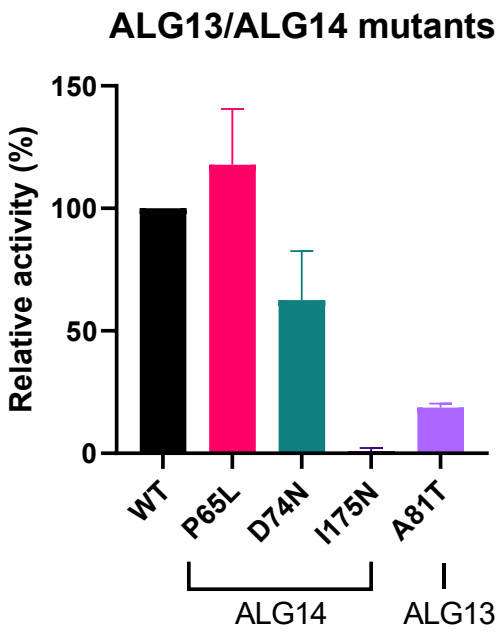


Figure 6.13 Relative activity of ALG13/ALG14 proteins bearing pathogenic mutations to wild-type (WT), where proteins are labelled by the mutant protein within the ALG13/ALG14 heterodimeric complex. Data shown is mean and standard deviation from 3 technical replicates.

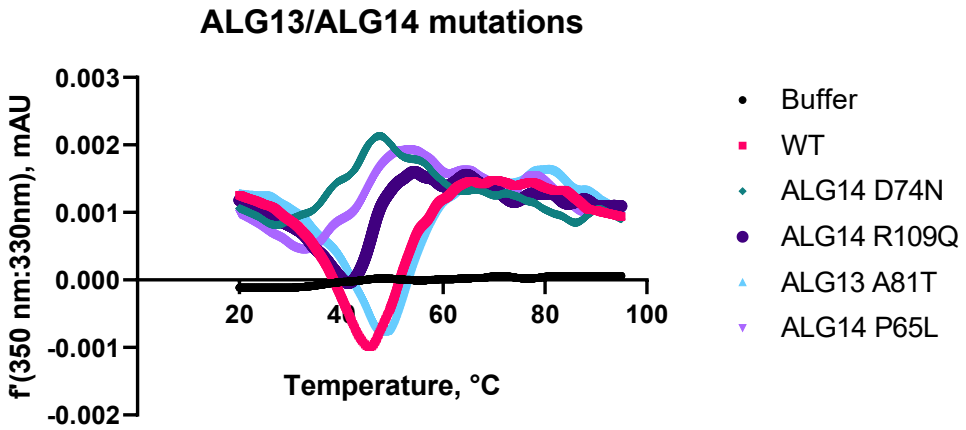


Figure 6.14 Thermostability screen of ALG13/AL14 mutant protein stocks following storage at -80 °C, where the rate of change in ratio of fluorescence at 350 nm: 330 nm was monitored as proteins were heated from 2 to 90 °C. Results are compared to WT protein and a buffer-only negative control.

6.3 Discussion

The ability for clinicians to quickly discern whether a gene variant is pathogenic allows for an accurate diagnosis to be made and ultimately leads to the correct treatment for a patient. However, it has been estimated that only 2% of the 4 million observed missense variants of human proteins have been annotated as to whether they are disease causing or benign in nature [230], making diagnosis especially challenging for rare diseases such as ALG13- and ALG14-CDG for which there are no obvious biomarkers.

It was promising that the AlphaMissense database [231] correctly predicted nine, single nucleotide polymorphisms in *ALG13* and *ALG14* to be likely pathogenic, and hence this tool offers a starting point to quickly assess the likelihood of disease and inform diagnosis [232]. This database has been found to associate pathogenicity with how evolutionarily constrained the gene is, and also on the likely impact on overall protein structure (e.g. loss or gain of aromatic amino acids or cysteines) [230, 232]. For rare diseases, relying on computational predictions alone for clinical diagnosis may be necessary, but for those in proteins where an experimentally determined 3D structure is not available or of low confidence, results need to be interpreted in context with clinical findings. Furthermore, despite 60,000 alternative isoforms being reported in the database [232], it is not complete as ALG13-CDG pathogenicity predictions could only be made for isotope 1 and not the alternative splicing variant that forms the active glycosyltransferase with ALG14 [68].

Attempts to experimentally determine the effects on protein function for nine reported missense variants of ALG13 and ALG14 proteins were hampered by a series of experimental set-backs and therefore analysis remains on-going. The first involved issues in four of the baculovirus vectors used to express the proteins in Sf9 insect cells, as summarised in Table 6.4. In each case, the cloned P2 baculovirus led to YFP production in the cells indicating infectivity and the

bacmid DNA was confirmed to contain full-length genes with no unwanted mutations (Figures 6.5 and 6.12). This meant there was no reason to doubt their use as expression vectors before use in protein production. For ALG14 V141G and ALG13 N107S and K94E the WT protein out of the ALG13/ALG14 pair was not expressed as there was no RNA transcribed, whereas the mutant variant was present. Therefore, these are the subject of future efforts to access baculovirus encoding the mutant proteins. Interestingly, for ALG13 I17N variant, both mutant and WT protein were detected in the cell lysate but this was not successfully purified, perhaps indicating proteins were misfolded or unstable.

Table 6.4 Summary of production of ALG13 and ALG14 mutant proteins, as screened on a DNA, RNA and protein level. For DNA, RNA and Protein screening, 13 = ALG13 identified, 14 = ALG14 identified. Purified = Success of large-scale purification attempts, and % Activity = to that of the wild-type protein.

	Mutation	DNA		RNA		Protein		Purified?	% Activity
		13	14	13	14	13	14		
ALG14	R109Q	✓	✓	✓	✓	✓	✓	✓	-
	I175N	✓	✓	✓	✓	✓	✓	✓	0
	D74N	✓	✓	✓	✓	✓	✓	✓	84
	V141G	✓	✓	✗	✓	✗	✓	✗	-
	P65L	✓	✓	✓	✓	✓	✓	✓	100
ALG13	N107S	✓	✓	✓	✗	✓	✗	✗	-
	I17N	✓	✓	✓	✓	✓	✓	✗	-
	A81T	✓	✓	✓	✓	✓	✓	✓	24
	K94E	✓	✓	✓	✗	✓	✗	✗	-

The second major issue was regarding protein stability and storage, which resulted in loss of activity in all purified stocks of mutant protein over time despite flash-freezing and storage at -80 °C. Future efforts to assess these variants should therefore be carried out using freshly purified stocks of the enzyme without freeze-thawing. Furthermore, assessing activity directly in crude membrane could be useful to aid stability of the target proteins, as has been used during the evaluation of bacterial sugar transferases [64, 233]. Wang *et al.* opted to express ALG13/ALG14 proteins in *E.coli* (which was justified as these human proteins have no predicted glycosylation sites) and the resultant recombinant protein was found to be active [68]. The authors were able to exploit the fact that this expression system lacks endogenous ALG13/ALG14 to enable screening of pathogenic variants of ALG13/ALG14 directly from detergent solubilised membranes without the need for purification [68]. Using this expression method would also increase the speed at which variants could be screened when identified in the clinic.

ALG14 P65L was previously found to have reduced expression compared to the wild-type protein when expressed in human embryonic kidney 293 cells [103]. When screening for end-point activity, this mutation was found to give comparable activity to wild-type (Figure 6.13), which would indicate that this variant is pathogenic due to protein deficiency rather than from a functional defect. Wang *et al.* reported that this variant had 80% the activity of wild-type ALG13/ALG14 suggesting some impairment, but this was much higher than all other disease variants they screened which showed between 8-28% activity [68]. It will now be informative to study the activity of these pathogenic variants in time course format to see how lipid acceptor or sugar donor binding is affected.

Together with findings from Wang *et al.,* these results do not explain why ALG13-CDG patients present with normal, or near-normal transferrin glycosylation pattern. As shown in Chapter 4, upstream DPAGT1 is the rate-limiting step in dolichol pyrophosphate $(GlcNAc)_2$

production as its catalytic rate is much slower than that of wild-type ALG13/ALG14. It is therefore plausible that mutations that lead to impaired GlcNAc transfer by ALG13/ALG14 go unnoticed if they do not reduce activity below that of DPAGT1, and that they may cause disease by gain of an alternative function [71]. Screening of enzyme activity with alternative nucleotide sugar substrates which would also be available in the cell would therefore be informative, as would studying their ability to disrupt complex formation with DPAGT1.

Chapter 7

Discussion and Future Perspectives

DPAGT1 and ALG13/ALG14 play crucial roles in catalysing the first steps in lipid-linked oligosaccharide synthesis during early N-glycosylation. This essential co-translational modification is required for proper protein folding and function, and pathogenic variants in these enzymes result in Congenital Disorders of Glycosylation (CDG). There remain no targeted treatments for these diseases owing in part to a lack of functional and structural understanding of the pathway.

The aim of this thesis was therefore to gain insight into the action of DPAGT1 and ALG13/ALG14 which involved expressing and purifying the membrane proteins alone and in complex, assessing their function using two experimental formats, testing the enzyme's tolerance for natural and non-natural dolichol lipid substrates, and characterising the functional effect of pathogenic variants. The key findings from this research are discussed here along with proposals for the future direction of the project.

7.1 Complex formation of human DPAGT1 and ALG13/ALG14

A baculovirus expression system in Sf9 insect cells was chosen as the method of recombinant production of human ALG13/ALG14 for its ability to express high yields of eukaryotic membrane complexes under the control of the strong polyhedrin promoter [136], as well as for its previous success for the production of upstream DPAGT1 [24]. Given there are no predicted N-glycan sites on these proteins, it was not deemed necessary to use a mammalian cell line such as human embryonic kidney (HEK) 293 cells to ensure correction glycosylation. However,

owing to the issues of stability of ALG13 in particular which was found here and by Averbeck *et al.* to be susceptible to degradation when expressed alone [69] , using HEK293 could have had the added benefit of providing human post-translational processing machinery such as chaperone proteins or proteases to remove signalling peptides [234]. This may have been especially beneficial for accessing protein containing pathogenic mutations, one of which was shown to be expressed but not successfully purified (**Chapter 6**).

Alternatively, the Gao group opted to co-express human ALG13 and ALG14 in *E.coli,* which allowed efficient production of active heterodimer from the membrane fraction [68]. Furthermore, the authors assessed its function in crude solubilised membranes without extensive purification. This would have allowed for more rapid access to relevant proteins both by quickening cloning stages with expression plasmids compared to the generation of baculovirus, and during the expansion of cell culture with rapid doubling of bacteria compared to insect cells. However, production of human integral membrane protein complexes can be challenging in *E.coli*, meaning this may prove to be unsuitable for production of the enzymes in complex with DPAGT1 which contains ten transmembrane helices [24].

The purification of human ALG13/ALG14 was achieved using an extended membrane preparation [149], IMAC and SEC (**Chapter 3**) which allowed for its detailed biochemical characterisation. Pleasingly, the human proteins were shown to exist as a functional 1:1 heterodimer in solution, as predicted from their yeast homologs [235, 236]. This finding underscores the conservation observed amongst these essential ALG enzymes. Similarly, co-expression with DAPGT1 provided the first evidence that human DPAGT1 and ALG13/ALG14 forms a multi-enzyme complex for the synthesis of dolichol pyrophosphate chitobiose, as has previously been shown for yeast ALG7 and ALG14 proteins [27, 77]. This makes the arrangement of enzymes homologous to that of bacterial MraY and MurG [132], and the predicted form of archaeal AGL24 and AGLH [74]. The remainder of the early N-glycosylation

pathway enzymes that act before the lipid linked oligosaccharide becomes flipped to the luminal leaflet appear to be less conserved across the domains of life [77] – perhaps indicating these were acquired after an evolutionary divergence.

The DPAGT1/ALG13/ALG14 complex was analysed by SEC-MALS, 2D blue native/SDS-PAGE and an anti-FLAG affinity pull down with a uniquely FLAG tagged DPAGT1. It would be beneficial to clone an alternatively tagged version of ALG13/ALG14 that would allow for a second purification step after anti-FLAG affinity to remove any unbound DPAGT1, as the 6xHis tag available on ALG14 was also present on DPAGT1. Despite utilising unique protease cleavage sites (His-ALG14 with a TEV protease site and DPAGT1-His-FLAG with a 3C protease site) both sites were not accessible for cleavage. In the same way, moving the FLAG purification tag to the N-terminus of DPAGT1 would be beneficial as the stability of the complex was found to dissociate by SEC for the C-terminal FLAG tagged construct, whilst FLAG affinity did massively improved sample purity. The C-terminus of hamster GPT was found to be crucial for its stable expression [237] and Noffz *et al.* state that decorating yeast ALG7 with an epitope tag at its C-terminus "resulted in a nonfunctional protein" being expressed whereas the protein tolerated N-terminal tagging [77].

It will now be important to show that the complex exists *in vivo* by studying its formation in a human cell line. This could be done by labelling DPAGT1 and ALG13/ALG14 with a FRET donor and acceptor fluorophore to show co-localisation in the cell, or by using a proximity labelling technique such as the use of an ascorbate peroxidase (APEX) tag to append binding partners with biotin [238]. As a control for these proximity experiments, the N-terminal domain of ALG14 - which is predicted to facilitate the formation of the complex given its role in yeast [27] - could be removed, and the functional implications on complex formation and subsequently lipid linked oligosaccharide synthesis could be monitored. However, it is possible that addition of bulky fluorophores or enzyme tags could disrupt the normal complex formation.

7.1.1 Single particle cryogenic electron microscopy (cryoEM)

Collaboration with the Membrane Protein Laboratory at Diamond Light Source aided efforts to structural elucidation of DPAGT1/ALG13/ALG14 by cryoEM – both through optimisation of sample preparation with high-throughput screens of detergents, and by preparation of EM grids, screening and data collection at the electron Bio-Imaging Centre (eBIC). Small data sets of the complex containing either FLAG and 6xHis tagged DPAGT1 were collected, where 2D class averages show density inside a detergent micelle that is likely DPAGT1 alone without association of ALG13/ALG14. This could point to the complex being weak or transient in nature, or may suggest that the native lipid environment facilitates the formation of the protein oligomerisation, for example that seen for the dimerisation of LeuT which is facilitated by cardiolipin [239]. There was a remarkable increase in stability of the complex when bound to both DPAGT1 and ALG13/ALG14 substrates and products, with C90 dolichol pyrophosphate GlcNAc increasing the melting temperature by 17 °C (**Chapter 5**). It is therefore possible that the dolichol lipid substrate itself aids complex formation in the ER.

Given biological membranes are typically 50 Å thick [240] the long chain of *cis* orientated prenol units of human C90 dolichol phosphate would be unnecessary if their function was merely to anchor the glycan to the membrane. Hence, it has been proposed to have a functional role in helping to interact multiple enzymes in the membrane. Interestingly, NMR studies of dolichol found that one C95 or C55 dolichol phosphate molecule could bind to more than one neuropeptide, suggesting it might be possible for these substrates to bind several glycosyltransferases within a multienzyme complex [241]. It was therefore hoped that incubation with dolichol phosphate species may increase the DPAGT1/ALG13/ALG14 tolerance to vitrification. Whilst processing of the latest data set of dolichol-bound complex is ongoing, it appears to be resolving DPAGT1 alone as before.

The Imperiali group have recently proposed a common methodology for the structural study of bacterial membrane bound phospho-glycosyl transferases, involving use of proteoliposomes [64]. This details using styrene maleic acid copolymer (SMA) to extract proteins directly from bacterial membranes, as well as expressing the membrane protein bearing the ubiquitin-like SUMO tag to increase expression and solubility [64]. The authors efficiently use *in vivo* cleavage of the SUMO tag by co-expressing a SUMO-protease and remove the remaining tags following affinity purification that allows for a reverse affinity step to improve purity whilst leaving full-length, untagged proteins in SMA-lipid particles (SMALPs) for cryoEM. It would be interesting to trial these methods for extracting DPAGT1/ALG13/ALG14 given direct SMA solubilisation preserves any native-lipids that could be important for stabilising the complex. As these are ER resident proteins rather than being expressed on the cell surface as in bacteria, extraction would need to be performed on prepared microsomes, as was detailed for preparing ER-vesicles for study of the protein translation machinery by tomography [75].

Use of cryogenic tomography in this way could also allow visualisation of the arrangement of a DPAGT1/ALG13/ALG14 complex in the ER membrane, and may be possible for this small complex with the rapid advancements in resolution being achieved by the field - in particular by using correlative light and electron microscopy techniques [242]. Furthermore, it could be possible to monitor whether the complex interacts with downstream ALG1, 2 and 11 enzymes, which could be imagined in order to achieve ultimate substrate channelling along the pathway. This 'super' complex would therefore be analogous to the large Mur enzyme complexes seen in some bacteria [133]. Overall, the DPAGT1/ALG13/ALG14 complex was successfully purified and initial efforts towards structural evaluation were carried out. Improvements in purification, tag placement, and use of membrane substitutes could further facilitate cryoEM.

7.2 Functional study of glycosyltransferases

Another key aspect of the project was the functional study of these enzymes. A luminescence-based enzyme coupled assay was explored to monitor the release of UMP and UDP from the reactions. This proved a robust method for studying release of UMP but not UDP and therefore a high-throughput assay for characterisation of DPAGT1 alone was carried out (**Chapter 4**). Findings from substrate matrix experiments - made possible by using this assay in 384-well format - corroborate the theory that DPAGT1 catalyses glycosyl transfer by use of a ternary bi-bi mechanism. This implicates a common use amongst PNPT enzymes given this one-step mechanism was previously found to be used by *Bacillus subtilis* MraY and *Thermotoga maritima* WecA [161]. This contrasts to the ping-pong mechanism observed for *Campylobacter concisus* PglC which belongs to the evolutionarily unrelated PHPT family [243]. It will now be important to repeat this work with independent DPAGT1 preparations to confirm these findings.

DM/CHS was selected as the detergent of choice for biochemical study of the enzymes following a screen for the most efficient methods of extraction from insect cells. It has long been noted that solubilising membrane proteins using detergents and removing the native lipid environment can result in loss of function [244, 245]. Although binding of some membrane lipids can be maintained following detergent extraction (for example lipidomic analysis revealed a bound phospholipid that was co-purified with DPAGT1 during crystallography [24]), recapitulating the ER lipid environment of the proteins using proteoliposomes could provide proteins in a more native state for assessment [64]. However, it would be difficult to quantify the amount of substrate that is encapsulated into the nanodisc for kinetic study given these ALG enzymes act on lipid substrates. For this reason, the functional assessment of bacterial phospho-galactose transferase WbaP - which was studied by cryoEM in proteoliposomes - was performed before SMALP formation [64, 246]. Studying the activity in crude membrane

extractions would also be an option, although quantification of the exact enzyme concentration would be difficult.

Fortunately, ALG13/ALG14 was found to be active in the chosen DM/CHS detergent which allowed determination of its substrate binding kinetics (**Chapter 4**). A thin layer chromatography (TLC) based assay was optimised to study glycosyl transfer, with detection of the lipid linked oligosaccharide products possible using a radioactively labelled sugar donor. This assay has enabled insight into the comparative rates of the enzymes alone and in complex with DPAGT1 as discussed below, although there remains variability between repeats which would likely be improved by increase in the quality and purity of the protein samples.

Dolichol pyrophosphate GlcNAc and UDP-GlcNAc were found to display similar affinity for ALG13/ALG14 with K_M' values of 13.23 ± 2.78 and 44.96 ± 16.42 μM respectively. These data compare well to the lipid acceptor kinetics since reported by Wang *et al.* using an LC-MS assay [68] (published during the course of this thesis), which validates these results given they were obtained from orthogonal assays in independent laboratories. Furthermore, the substrate affinities to ALG13/ALG14 were similar to that of the lipid acceptor and sugar donor with DPAGT1 (at 11.88 ± 4.92 μM and 10.03 ± 5.12 μM for dolichol phosphate and UDP-GlcNAc, respectively). However, the maximum rates of ALG13/ALG14 were 100-fold faster than that of DPAGT1. This meant when studied together, the kinetics of the formation of dolichol-pyrophosphate-$(GlcNAc)_2$ were the same as that of the first GlcNAc addition by DPAGT1, as this is much the rate limiting step. This therefore sets DPAGT1 as the key regulator of the flux of dolichol phosphate through the lipid linked oligosaccharide synthesis pathway. Interestingly, these findings also imply that any variants in ALG13/ALG14 would have to result in a significant reduction in activity for the affects to be noticed in the pathway - perhaps explaining the typically normal transferrin glycosylation pattern seen in ALG13-CDG patients [108].

However, reduction in GlcNAc transfer for many pathogenic variants was seen here and by Wang *et al.* [68] , which would suggest disrupted N-glycosylation, as discussed below.

7.2.1 Insights into ALG13/ALG14 CDG mutants

The methods developed during this project are now being applied to study the effect of pathogenic mutations in ALG13 and ALG14 identified in patients with congenital disorders of glycosylation (**Chapter 6**) which could aid in the search for targeted therapeutics.

The activity of ALG13/ ALG14 I175N, ALG13/ ALG14 D74N, ALG13 P65L/ ALG14 and ALG13 A81T/ALG14 mutant proteins were assessed before degradation of the stocks rendered them inactive. Given this did not occur for the wild-type proteins, it is possible that these mutations de-stabilise the enzymes. ALG13 bearing an asparagine at position 107 instead of serine is the most prevalent *de novo* mutation acquired in patients with ALG13-CDG [107, 142, 213, 218, 224] and has been shown to have just 8% of the GlcNAc transferase activity of the wild-type protein despite patients displaying near-normal glycosylation patterns [68]. Unfortunately, this protein was not successfully purified due to what was later found to be an issue in expression of ALG14 conferred during cloning of the MultiBac vector. Once re-cloning has concluded, it will be interesting to see if these findings are replicable in this alternative assay format, as well as to compare the action of the pathogenic mutant when in complex with DPAGT1 given this is the likely physiological conformation.

Furthermore, the activity of ALG13/ALG14 P65L was initially found to be comparable to the wild type protein, and ALG13/ALG14 D74N had 84% of wild-type activity. It would therefore be interesting to see whether these mutants have a pathogenic affect by disrupting complex formation with DPAGT1 rather than by impairing catalysis. Symptoms of these ALG14-CDGs are similar to those in patients with mutations in DPAGT1, presenting as myasthenic syndromes with impairment at the neuromuscular junction [103, 104], whereas *de novo* mutations

seen in ALG13-CDG result in severe developmental and epileptic encephalopathy [67]. It is possible these differences in symptoms reflect the differing roles of the enzymes involved, with a reduction in the tolerance of severe mutations in the more essential components, as loss of function would likely be lethal. For example, ALG14 is likely to act to anchor ALG13 to the ER membrane, facilitate complex formation with DPAGT1 as well as bind the dolichol substrate and so it may be that only mild defects become populated in this protein which ultimately result in myasthenic syndromes due to the sensitivity of the neuromuscular junction to proper N-glycosylation.

7.3 Consequences of eukaryotic dolichol use

Access to long chain polyisoprenol lipids is labour intensive and hence these substrates are expensive to buy commercially. By selecting to study bacterial and human length lipids we assessed biologically relevant natural and non-natural mimics of both DPAGT1 and ALG13/ALG14 substrates (**Chapter 5**).

It was found that human DPAGT1 is highly selective for dolichol lipids despite all substrates offering increase in thermostability. Phosphorylated polyunsaturated equivalents are present in the cytoplasmic leaflet of the ER as an intermediate of the dolichol biosynthesis pathway [85] but will not be utilised by DPAGT1. In this way the enzyme acts as the gatekeeper of the LLO synthesis pathway to ensure only the desired lipid is utilised. The conversion of polyisoprenol to dolichol by SRD5α3 will therefore be key in regulating the flux of available N-glycan substrates, which has also been shown in yeast to be the rate limiting step in dolichol synthesis [247].

Conversely, human ALG13/ALG14 was found to utilise C55 and C90 lipids with or without the saturated α-terminal unit. The shorter C55 dolichols may be useful for structural and

functional analysis of this and other ALG enzymes as they display increased solubility relative to C90 length lipids, but maintain the native α-saturation. The maximum rates of ALG13/ALG14 with all C90 and C55 lipid species were found to be within error, although a slightly higher affinity was displayed by the native, C90 dolichol phosphate GlcNAc. From the predicted structure of the active site, it was assumed that a lipid chain beyond C20 would be unlikely interact with ALG13/ALG14, however Gao *et al.* report that ALG13/ALG14 was not able to utilise phytanyl-linked lipids [68] that were previously used for study of downstream ALG1, 2 and 11[31, 174]. This may suggest there is interaction of the dolichol chain elsewhere on the protein.

Interestingly, this was observed in the crystal structural of archaeal dolichol phosphate mannose synthase in complex with GDP-mannose and C55 dolichol phosphate [248], where, as well as the coordination of the phosphate head-group and first two isoprene units by residues in the active site, the dolichol chain was found to interact with a large interface of the protein down to the second transmembrane helix where it folded into a U-shape and rested on another helix as a support. However, the authors do note that this could be a crystal lattice artefact rather than the orientation *in vivo* [248].

Previous work in this field has been limited to crude microsomes from yeast [189], cow [191] , rat [190] and CHO [193] cell extractions, and this data therefore provides molecular confirmation of the substrate tolerance of human early N-glycosylation enzymes. It would now be interesting to access dolichols that are saturated with an unnatural *R*-configured methyl group to see if these are recognised by the human proteins – substrates which could be synthesised using the commercially available Ruthenium based catalyst as for the *S*- configured methyl group used here.

7.4 Concluding remarks

In summary, interdisciplinary approaches in membrane protein purification, assay development, and enzyme kinetics have been applied to the study of DPAGT1 and ALG13/ALG14 enzymes of the early N-glycosylation pathway. The methodology presented here will be useful for application to the downstream glycosyltransferases as well as those in archaeal and bacterial membranes. Coupling the action of the DPAGT1/ALG13/ALG14 complex with ALG1, 2 and 11 could also be beneficial for the enzymatic access to lipid-linked oligosaccharide molecules, by providing native human substrates from just dolichol phosphate starting material, as has been shown for the downstream glycosyltransferases [33, 43]. This could be particularly useful for study of the elusive eukaryotic flippase proteins to complete the characterisation of the cytoplasmic leaflet N-linked glycosylation pathway, and hence provide insight for the field of CDGs.

We have shown evidence for a multi-enzyme complex governing the first steps in the human N-glycosylation pathway and developed robust methods for its expression and purification as well as provided early data for its structural evaluation. Using two different assays, we present interesting fundamental biochemistry of the substrate preference of DPAGT1 and ALG13/ALG41 which has potential evolutionary implications.

Bibliography

[1] Vigerust, D. J., and Shepherd, V. L. (2007) Virus glycosylation: role in virulence and immune interactions, *Trends in Microbiology 15*, 211-218.

[2] Raju, T. S. (2019) O-glycosylation of Proteins, In *Co- and Post-Translational Modifications of Therapeutic Antibodies and Proteins*, pp 101-117.

[3] Reily, C., Stewart, T. J., Renfrow, M. B., and Novak, J. (2019) Glycosylation in health and disease, *Nature Reviews Nephrology 15*, 346-366.

[4] Larkin, A., Imperiali, B.,. (2011) The expanding horizons of asparagine-linked glycosylation, *Biochemistry 50*, 4411-4426.

[5] Minakata, S., Manabe, S., Inai, Y., Ikezaki, M., Nishitsuji, K., Ito, Y., and Ihara, Y. (2021) Protein C-Mannosylation and C-Mannosyl Tryptophan in Chemical Biology and Medicine, *Molecules 26*, 5258.

[6] Freeze, H. H. (2006) Genetic defects in the human glycome, *Nature Reviews Genetics 7*, 537-551.

[7] Buczkowska, A., Swiezewska, E., Lefeber, D.J. (2015) Genetic defects in dolichol metabolism, *Journal of Inherited Metabolic Diseases 38*, 157-169.

[8] Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., and Clausen, H. (2020) Global view of human protein glycosylation pathways and functions, *Nature Reviews Molecular Cell Biology 21*, 729-749.

[9] Esmail, S., and Manolson, M. F. (2021) Advances in understanding N-glycosylation structure, function, and regulation in health and disease, *European Journal of Cell Biology 100*, 151186.

[10] Suzuki, T., Cummings, R.D., Aebi, M., and Parodi, A., (2022) *Essentials of Glycobiology, Chapter 39: Glycans in Glycoprotein Quality Control*, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbour (NY).

[11] Hebert, D. N., Lamriben, L., Powers, E. T., and Kelly, J. W. (2014) The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis, *Nature Chemical Biology 10*, 902-910.

[12] Appenzeller, C., Andersson, H., Kappeler, F., and Hauri, H.-P. (1999) The lectin ERGIC-53 is a cargo transport receptor for glycoproteins, *Nature Cell Biology 1*, 330-334.

[13] Watanabe, S., Kise, Y., Yonezawa, K., Inoue, M., Shimizu, N., Nureki, O., and Inaba, K. (2024) Structure of full-length ERGIC-53 in complex with MCFD2 for cargo transport, *Nature Communications 15*, 2404.

[14] Sim, G., Jeong, M., Seo, H., Kim, J., and Lee, S. (2022) The Role of N-Glycosylation in the Intracellular Trafficking and Functionality of Neuronal Growth Regulator 1, *Cells 11*.

[15] Mathys, L., François, K. O., Quandte, M., Braakman, I., and Balzarini, J. (2014) Deletion of the highly conserved N-glycan at Asn260 of HIV-1 gp120 affects folding and lysosomal degradation of gp120, and results in loss of viral infectivity, *Public Library of Science One 9*, e101181.

[16] Dahms, N., Braulke, T., and Varki, A., (2022) *Essentials of Glycobiology, Chapter 33: P-type lectins*, 4th ed., Cold Spring Harbour Laboratory Press, Cold Spring Harbor (NY).

[17] Stanley, P., Moremen, K.W., Lewis, N.E., Taniguchi, N., and Aebi, M. (2022) *Essentials of Glycobiology, Chapter 9: N-glycans*, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[18] Helenius, A., Aebi, and Markus. (2001) Intracellular Functions of N-Linked Glycans, *Science 291*, 2364-2369.

[19] Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity, *Nature Reviews Immunology 8*, 874-887.

[20] Giovannone, N., Liang, J., Antonopoulos, A., Geddes Sweeney, J., King, S. L., Pochebit, S. M., Bhattacharyya, N., Lee, G. S., Dell, A., Widlund, H. R., Haslam, S. M., and Dimitroff, C. J. (2018) Galectin-9 suppresses B cell receptor signaling and is regulated by I-branching of N-glycans, *Nature Communications 9*, 3287.

[21] Earl, L. A., Bi, S., and Baum, L. G. (2010) N- and O-glycans modulate galectin-1 binding, CD45 signaling, and T cell death, *Journal of Biological Chemisry 285*, 2232-2244.

[22] Burda, P., and Aebi, M. (1999) The dolichol pathway of N-linked glycosylation, *Biochimica et Biophysica Acta (BBA) - General Subjects 1426*, 239-257.
[23] Huffaker, T. C., and Robbins, P. W. (1982) Temperature-sensitive yeast mutants deficient in asparagine-linked glycosylation, *Journal of Biological Chemistry 257*, 3203-3210.
[24] Dong, Y. Y., Wang, H., Pike, A. C. W., Cochrane, S. A., Hamedzadeh, S., Wyszyński, F. J., Bushell, S. R., Royer, S. F., Widdick, D. A., Sajid, A., Boshoff, H. I., Park, Y., Lucas, R., Liu, W. M., Lee, S. S., Machida, T., Minall, L., Mehmood, S., Belaya, K., Liu, W. W., Chu, A., Shrestha, L., Mukhopadhyay, S. M. M., Strain-Damerell, C., Chalk, R., Burgess-Brown, N. A., Bibb, M. J., Barry Iii, C. E., Robinson, C. V., Beeson, D., Davis, B. G., and Carpenter, E. P. (2018) Structures of DPAGT1 Explain Glycosylation Disease Mechanisms and Advance TB Antibiotic Design, *Cell 175*, 1045-1058.e1016.
[25] Lehrman, M. A. (1994) Commentary: A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases, *Glycobiology 4*, 768-771.
[26] Lehrman, M. A. (1991) Biosynthesis of N-acetylglucosamine-P-P-dolichol, the committed step of asparagine-linked oligosaccharide assembly, *Glycobiology 1*, 553-562.
[27] Lu, J., Takahashi, T., Ohoka, A., Nakajima, K., Hashimoto, R., Miura, N., Tachikawa, H., and Gao, X. D. (2012) Alg14 organizes the formation of a multiglycosyltransferase complex involved in initiation of lipid-linked oligosaccharide biosynthesis, *Glycobiology 22*, 504-516.
[28] Gao, X. D., Nishikawa, A., and Dean, N. (2004) Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum, *Glycobiology 14*, 559-570.
[29] Dean, N., and Gao, X. D. (2014) Heterodimeric Alg13/Alg14 UDP-GlcNAc Transferase (ALG13,14), In *Handbook of Glycosyltransferases and Related Genes* (Taniguchi, N., Honke, K., Fukuda, M., Narimatsu, H., Yamaguchi, Y., and Angata, T., Eds.), pp 1231-1238, Springer Japan, Tokyo.
[30] Wang, X., Weldeghiorghis, T., Zhang, G., Imperiali, B., and Prestegard, J. H. (2008) Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase, *Structure 16*, 965-975.
[31] Li, S. T., Wang, N., Xu, S., Yin, J., Nakanishi, H., Dean, N., and Gao, X. D. (2017) Quantitative study of yeast Alg1 beta-1, 4 mannosyltransferase activity, a key enzyme involved in protein N-glycosylation, *Biochimica et Biophysica Acta (BBA)- General Subjects 1861*, 2934-2941.
[32] Wilson, I. B. H., Webberley, M. C., Revers, L., and Flitsch, S. L. (1995) Dolichol is not a necessary moiety for lipid-linked oligosaccharide substrates of the mannosyltransferases involved in in vitro N-linked-oligosaccharide assembly, *Biochemical Journal 310*, 909-916.
[33] Ramírez, A. S., Boilevin, J., Lin, C. W., Ha, G. B., Janser, D., Aebi, M., Darbre, T., Reymond, J. L., and Locher, K. P. (2017) Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins, *Glycobiology 27*, 726-733.
[34] Li, S. T., Wang, N., Xu, X. X., Fujita, M., Nakanishi, H., Kitajima, T., Dean, N., and Gao, X. D. (2018) Alternative routes for synthesis of N-linked glycans by Alg2 mannosyltransferase, *Federation of American Societies for Experimental Biology Journal 32*, 2492-2506.
[35] Xiang, M. H., Xu, X. X., Wang, C. D., Chen, S., Xu, S., Xu, X. Y., Dean, N., Wang, N., and Gao, X. D. (2022) Topological and enzymatic analysis of human Alg2 mannosyltransferase reveals its role in lipid-linked oligosaccharide biosynthetic pathway, *Communications Biology 5*, 117.
[36] O'Reilly, M. K., Zhang, G., and Imperiali, B. (2006) In Vitro Evidence for the Dual Function of Alg2 and Alg11: Essential Mannosyltransferases in N-Linked Glycoprotein Biosynthesis, *Biochemistry 45*, 9593-9603.
[37] Cipollo, J. F., Trimble, R. B., Chi, J. H., Yan, Q., and Dean, N. (2001) The Yeast ALG11 Gene Specifies Addition of the Terminal α1,2-Man to the Man5GlcNAc2-PP-dolicholN-Glycosylation Intermediate Formed on the Cytosolic Side of the Endoplasmic Reticulum*210, *Journal of Biological Chemistry 276*, 21828-21840.
[38] Rush, J. S. (2015) Role of Flippases in Protein Glycosylation in the Endoplasmic Reticulum, *Lipid Insights 8*, 45-53.
[39] Schenk, B., Fernandez, F., and Waechter, C. J. (2001) The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum, *Glycobiology 11*, 61r-70r.

[40] Helenius, J., Ng, D. T. W., Marolda, C. L., Walter, P., Valvano, M. A., and Aebi, M. (2002) Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein, *Nature 415*, 447-450.
[41] Sanyal, S., Frank, C. G., and Menon, A. K. (2008) Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum, *Biochemistry 47*, 7937-7946.
[42] Frank, C. G., Sanyal, S., Rush, J. S., Waechter, C. J., and Menon, A. K. (2008) Does Rft1 flip an N-glycan lipid precursor?, *Nature 454*, E3-E4.
[43] Bloch, J. S., Pesciullesi, G., Boilevin, J., Nosol, K., Irobalieva, R. N., Darbre, T., Aebi, M., Kossiakoff, A. A., and Reymond, J. L. L., K. P. (2020) Structure and mechanism of the ER-based glucosyltransferase ALG6, *Nature 579*, 443-447.
[44] Dong, Y. Y., Wang, H., Pike, A.C.W., Cochrane, S.A., Hamedzadeh, S., Wyszynski, F.J., Bushell, S.R., Royer, S.F., Widdick, D.A., Sajid, A., Boshoff, H.I., Park, Y., Lucas, R., Liu, W., Lee, S.S., Machida, T., Minall, L., Mehmood, S., Belaya, K., Liu, W., Chu, A., Shrestha, L., Mukhopadhyay, S.M.M., Strain-Damerell, C., Chalk,R., Burgess-Brown, N.A., Bibb, M.J., Barry III, C.E., Robinson, C.V., Beeson, D., Davis, B.G., Carpenter, E.P. (2018) Structures of DPAGT1 Explain Glycosylation Disease Mechanisms and Advance TB Antibiotic Design, *Cell 175*, 1045-1058.
[45] Eichler, J., Imperiali, B., . (2018) Biogenesis of Asparagine-Linked Glycoproteins Across Domains of Life- Similarities and Differences, *Americal Chemical Society Chemical Biology 13*, 833-837.
[46] Nyathi, Y., Wilkinson, B.M., Pool, M.R., . (2013) Co-translational targeting and translocation of proteins to the endoplasmic reticulum, *Biochimica et Biophysica Acta- Molecular Cell Research 1833*, 2392-2402.
[47] Bai, L., Wang, T., Zhao, G., Kovach, A., and Li, H. (2018) The atomic structure of a eukaryotic oligosaccharyltransferase complex, *Nature 555*, 328-333.
[48] Mescher, M. F., and Strominger, J. L. (1976) Purification and characterization of a prokaryotic glycoprotein from the cell envelope of Halobacterium salinarium, *Journal of Biological Chemistry 251*, 2005-2014.
[49] Vershinin, Z., Zaretsky, M., Guan, Z., and Eichler, J. (2021) Revisiting N-glycosylation in Halobacterium salinarum: Characterizing a dolichol phosphate- and glycoprotein-bound tetrasaccharide, *Glycobiology 31*, 1645-1654.
[50] Weerapana, E., and Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems, *Glycobiology 16*, 91R-101R.
[51] Alaimo, C., Catrein, I., Morf, L., Marolda, C. L., Callewaert, N., Valvano, M. A., Feldman, M. F., and Aebi, M. (2006) Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides, *The European Molecular Biology Organization (EMBO) Journal 25*, 967-976.
[52] Bugg, T. D. H., and Brandish, P. E. (1994) From peptidoglycan to glycoproteins: Common features of lipid-linked oligosaccharide biosynthesis, *Federation of European Microbiological Societies (FEMS) Microbiology Letters 119*, 255-262.
[53] Meyer, B. H., and Albers, S. V. (2013) Hot and sweet: protein glycosylation in Crenarchaeota, *Biochemical Society Transactions 41*, 384-392.
[54] Larkin, A., Chang, M. M., Whitworth, G. E., and Imperiali, B. (2013) Biochemical evidence for an alternate pathway in N-linked glycoprotein biosynthesis, *Nature Chemical Biology 9*, 367-373.
[55] Price, N. P., and Momany, F. A. (2005) Modeling bacterial UDP-HexNAc: polyprenol-P HexNAc-1-P transferases, *Glycobiology 15*, 29R-42R.
[56] Meyer, B. H., Shams-Eldin, H., and Albers, S.-V. (2017) AglH, a thermophilic UDP-N-acetylglucosamine-1-phosphate:dolichyl phosphate GlcNAc-1-phosphotransferase initiating protein N-glycosylation pathway in Sulfolobus acidocaldarius, is capable of complementing the eukaryal Alg7, *Extremophiles 21*, 121-134.
[57] Lehrer, J., Vigeant, K. A., Tatar, L. D., and Valvano, M. A. (2007) Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating

the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide, *Journal of Bacteriology 189*, 2618-2628.
[58] Chung, B. C., Zhao, J., Gillespie, R. A., Kwon, D. Y., Guan, Z., Hong, J., Zhou, P., and Lee, S. Y. (2013) Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis, *Science 341*, 1012-1016.
[59] Hug, I., and Feldman, M. F. (2010) Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria, *Glycobiology 21*, 138-151.
[60] Yoo, J., Mashalidis, E. H., Kuk, A. C. Y., Yamamoto, K., Kaeser, B., Ichikawa, S., and Lee, S.-Y. (2018) GlcNAc-1-P-transferase–tunicamycin complex structure reveals basis for inhibition of N-glycosylation, *Nature Structural & Molecular Biology 25*, 217-224.
[61] Dean, N., and Gao, X.-D. (2014) Dolichyl-Phosphate (UDP-N-Acetylglucosamine) N-Acetylglucosaminephospho transferase 1 (GlcNAc-1-P Transferase) (DPAGT1), In *Handbook of Glycosyltransferases and Related Genes* (Taniguchi, N., Honke, K., Fukuda, M., Narimatsu, H., Yamaguchi, Y., and Angata, T., Eds.), pp 1223-1230, Springer Japan, Tokyo.
[62] Saldías, M. S., Patel, K., Marolda, C. L., Bittner, M., Contreras, I., and Valvano, M. A. (2008) Distinct functional domains of the Salmonella enterica WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit, *Microbiology (Reading) 154*, 440-453.
[63] Liu, F., Vijayakrishnan, B., Faridmoayer, A., Taylor, T. A., Parsons, T. B., Bernardes, G. J. L., Kowarik, M., and Davis, B. G. (2014) Rationally Designed Short Polyisoprenol-Linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation, *Journal of the American Chemical Society 136*, 566-569.
[64] Dodge, G. J., Bernstein, H. M., and Imperiali, B. (2023) A generalizable protocol for expression and purification of membrane-bound bacterial phosphoglycosyl transferases in liponanoparticles, *Protein Expression and Purification 207*, 106273.
[65] Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022, *Nucleic Acids Res 50*, W276-W279.
[66] CDG Hub 2024,https://www.cdghub.com
[67] Frontiers in Congenital Disorders of Glycosylation Consortium ALG13-Congenital Disorder of Glycosylation (ALG13-CDG) 2024,https://fcdgc.rarediseasesnetwork.org/diseases-studied/alg13-cdg
[68] Wang, C. D., Xu, S., Chen, S., Chen, Z. H., Dean, N., Wang, N., and Gao, X. D. (2022) An in vitro assay for enzymatic studies on human ALG13/14 heterodimeric UDP-N-acetylglucosamine transferase, *Frontiers in Cell and Developmental Biology 10*.
[69] Averbeck, N., Gao, X. D., Nishimura, S., and Dean, N. (2008) Alg13p, the catalytic subunit of the endoplasmic reticulum UDP-GlcNAc glycosyltransferase, is a target for proteasomal degradation, *Molecular Biology of the Cell 19*, 2169-2178.
[70] Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics 25*, 1189-1191.
[71] Ng, B. G., Eklund, E. A., Shiryaev, S. A., Dong, Y. Y., Abbott, M.-A., Asteggiano, C., Bamshad, M. J., Barr, E., Bernstein, J. A., Chelakkadan, S., Christodoulou, J., Chung, W. K., Ciliberto, M. A., Cousin, J., Gardiner, F., Ghosh, S., Graf, W. D., Grunewald, S., Hammond, K., Hauser, N. S., Hoganson, G. E., Houck, K. M., Kohler, J. N., Morava, E., Larson, A. A., Liu, P., Madathil, S., McCormack, C., Meeks, N. J. L., Miller, R., Monaghan, K. G., Nickerson, D. A., Palculict, T. B., Papazoglu, G. M., Pletcher, B. A., Scheffer, I. E., Schenone, A. B., Schnur, R. E., Si, Y., Rowe, L. J., Serrano Russi, A. H., Russo, R. S., Thabet, F., Tuite, A., Villanueva, M. M., Wang, R. Y., Webster, R. I., Wilson, D., Zalan, A., Undiagnosed Diseases Network, U. o. W. C. f. M. G., Wolfe, L. A., Rosenfeld, J. A., Rhodes, L., and Freeze, H. H. (2020) Predominant and novel de novo variants in 29 individuals with ALG13 deficiency: Clinical description, biomarker status, biochemical analysis, and treatment suggestions, *Journal of Inherited Metabolic Disease 43*, 1333-1348.

[72] Ha, S., Walker, D., Shi, Y., Walker, S. (2000) The 1.9 A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis, *Protein Science 9*, 1045-1052.
[73] Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., Walker, S. (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases, *Proceedings of the National Academy of Sciences of the United States of America 100*, 845-849.
[74] Meyer, B. H., Adam, P. S., Wagstaff, B. A., Kolyfetis, G. E., Probst, A. J., Albers, S. V., and Dorfmueller, H. C. (2022) Agl24 is an ancient archaeal homolog of the eukaryotic N-glycan chitobiose synthesis enzymes, *eLife 11*.
[75] Gemmer, M., Chaillet, M. L., van Loenhout, J., Cuevas Arenas, R., Vismpas, D., Gröllers-Mulderij, M., Koh, F. A., Albanese, P., Scheltema, R. A., Howes, S. C., Kotecha, A., Fedry, J., and Förster, F. (2023) Visualization of translation and protein biogenesis at the ER membrane, *Nature 614*, 160-167.
[76] Pfeffer, S., Dudek, J., Gogala, M., Schorr, S., Linxweiler, J., Lang, S., Becker, T., Beckmann, R., Zimmermann, R., and Förster, F. (2014) Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon, *Nature Communications 5*, 3072.
[77] Noffz, C., Keppler-Ross, S., and Dean, N. (2009) Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle, *Glycobiology 19*, 472-478.
[78] Asturias, F. J., Chadick, J. Z., Cheung, I. K., Stark, H., Witkowski, A., Joshi, A. K., and Smith, S. (2005) Structure and molecular organization of mammalian fatty acid synthase, *Nature Structural & Molecular Biology 12*, 225-232.
[79] Günenc, A. N., Graf, B., Stark, H., and Chari, A. (2022) Fatty Acid Synthase: Structure, Function, and Regulation, In *Macromolecular Protein Complexes IV: Structure and Function* (Harris, J. R., and Marles-Wright, J., Eds.), pp 1-33, Springer International Publishing, Cham.
[80] Sweetlove, L. J., and Fernie, A. R. (2018) The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation, *Nature Communications 9*, 2136.
[81] Kean, E. L., Wei, Z., Anderson, V. E., Zhang, N., and Sayre, L. M. (1999) Regulation of the Biosynthesis of N-Acetylglucosaminylpyrophosphoryldolichol, Feedback and Product Inhibition *Journal of Biological Chemistry 274*, 34072-34082.
[82] Bonten, E. J., Annunziata, I., and d'Azzo, A. (2014) Lysosomal multienzyme complex: pros and cons of working together, *Cellular and Molecular Life Sciences 71*, 2017-2032.
[83] Valtersson, C., van Duÿn, G., Verkleij, A. J., Chojnacki, T., de Kruijff, B., and Dallner, G. (1985) The influence of dolichol, dolichol esters, and dolichyl phosphate on phospholipid polymorphism and fluidity in model membranes, *Journal of Biological Chemistry 260*, 2742-2751.
[84] Ekström, T. J., Chojnacki, T., and Dallner, G. (1987) The alpha-saturation and terminal events in dolichol biosynthesis, *Journal of Biological Chemistry 262*, 4090-4097.
[85] Cantagrel, V., and Lefeber, D. J. (2011) From glycosylation disorders to dolichol biosynthesis defects: a new class of metabolic diseases, *Journal of Inherited Metabolic Disease 34*, 859-867.
[86] Eichler, J., and Imperiali, B. (2018) Stereochemical Divergence of Polyprenol Phosphate Glycosyltransferases, *Trends in Biochemical Sciences 43*, 10-17.
[87] Jones, M. B., Rosenberg, J. N., Betenbaugh, M. J., and Krag, S. S. (2009) Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life, *Biochimica et Biophysica Acta 1790*, 485-494.
[88] Swiezewska, E., Sasak, W., Mańkowski, T., Jankowski, W., Vogtman, T., Krajewska, I., Hertel, J., Skoczylas, E., and Chojnacki, T. (1994) The search for plant polyprenols, *Acta Biochimica Polonica 41*, 221-260.
[89] Chen, M. M., Weerapana, E., Ciepichal, E., Stupak, J., Reid, C. W., Swiezewska, E., and Imperiali, B. (2007) Polyisoprenol Specificity in the Campylobacter jejuni N-Linked Glycosylation Pathway, *Biochemistry 46*, 14342-14348.
[90] Eichler, J., and Guan, Z. (2017) Lipid sugar carriers at the extremes: The phosphodolichols Archaea use in N-glycosylation, *Biochim Biophys Acta Mol Cell Biol Lipids 1862*, 589-599.

[91] Taguchi, Y., Fujinami, D., and Kohda, D. (2016) Comparative Analysis of Archaeal Lipid-linked Oligosaccharides That Serve as Oligosaccharide Donors for Asn Glycosylation *Journal of Biological Chemistry 291*, 11042-11054.
[92] Hennet, T. (2012) Diseases of glycosylation beyond classical congenital disorders of glycosylation, *Biochimica et Biophysica Acta 1820*, 1306-1317.
[93] Schenk, B., Fernandez, F., Waechter, C.J., . (2001) The ins(ide) and outs(ide) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum, *Glycobiology 11*, 61R-70R.
[94] Naparstek, S., Guan, Z., and Eichler, J. (2012) A predicted geranylgeranyl reductase reduces the ω-position isoprene of dolichol phosphate in the halophilic archaeon, Haloferax volcanii, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1821*, 923-933.
[95] Shridas, P., Waechter, C.J. (2006) Human dolichol kinase, a polytopic endoplasmic reticulum membrane protein with a cytoplasmically orientated CTP-binding site, *Journal of Biological Chemistry 28*, 31696-31704.
[96] Manat, G., Roure, S., Auger, R., Bouhss, A., Barreteau, H., Mengin-Lecreulx, D., and Touzé, T. (2014) Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier, *Microbial Drug Resistance 20*, 199-214.
[97] Hartley, M. D., and Imperiali, B. (2012) At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates, *Archives of Biochemistry and Biophysics 517*, 83-97.
[98] Yu, J., Ma, P. J., Shi, D. J., Li, S. M., and Wang, C. L. (2008) Homologous comparisons of photosynthetic system I genes among cyanobacteria and chloroplasts, *Journal of Integrative Plant Biology 50*, 929-940.
[99] Chang, I. J., He, M., and Lam, C. T. (2018) Congenital disorders of glycosylation, *Annals of Translational Medicine 6*, 477.
[100] Jaeken, J., Vanderschueren-Lodeweyckx, M., Casaer, P., Snoeck, L., Corbeel, L., Eggermont, E., and Eeckels, R. (1980) Familial psychomotor retardation with markedly fluctuating serum prolactin, FSH and GH levels, partial TBG-deficiency, increased serum arylsulphatase A and increased CSF protein: a new syndrome?: 90, *Pediatric Research 14*, 179-179.
[101] Francisco, R., Brasil, S., Poejo, J., Jaeken, J., Pascoal, C., Videira, P., A.,, and dos Reis Ferreira, V. (2023) Congenital disorders of glycosylation (CDG): state of the art in 2022, *Orphanet Journal of Rare Diseases 18*, 329.
[102] Grünewald, S., Matthijs, G., and Jaeken, J. (2002) Congenital Disorders of Glycosylation: A Review, *Pediatric Research 52*, 618-624.
[103] Cossins, J., Belaya, K., Hicks, D., Salih, M. A., Finlayson, S., Carboni, N., Liu, W. W., Maxwell, S., Zoltowska, K., Farsani, G. T., Laval, S., Seidhamed, M. Z., Donnelly, P., Bentley, D., McGowan, S. J., Müller, J., Palace, J., Lochmüller, H., and Beeson, D. (2013) Congenital myasthenic syndromes due to mutations in ALG2 and ALG14, *Brain 136*, 944-956.
[104] Belaya, K., Finlayson, S., Cossins, J., Liu, W. W., Maxwell, S., Palace, J., and Beeson, D. (2012) Identification of DPAGT1 as a new gene in which mutations cause a congenital myasthenic syndrome, *Annals of the New York Academy of Sciences 1275*, 29-35.
[105] Jaeken, J., Vleugels, W., Régal, L., Corchia, C., Goemans, N., Haeuptle, M. A., Foulquier, F., Hennet, T., Matthijs, G., and Dionisi-Vici, C. (2009) RFT1-CDG: deafness as a novel feature of congenital disorders of glycosylation, *Journal of Inherited Metabolic Disease 32 Suppl 1*, S335-338.
[106] Biffi, S., Tamaro, G., Bortot, B., Zamberlan, S., Severini, G. M., and Carrozzi, M. (2007) Carbohydrate-deficient transferrin (CDT) as a biochemical tool for the screening of congenital disorders of glycosylation (CDGs), *Clincal Biochemistry 40*, 1431-1434.
[107] Datta, A. N., Bahi-Buisson, N., Bienvenu, T., Buerki, S. E., Gardiner, F., Cross, J. H., Heron, B., Kaminska, A., Korff, C. M., Lepine, A., Lesca, G., McTague, A., Mefford, H. C., Mignot, C., Milh, M., Piton, A., Pressler, R. M., Ruf, S., Sadleir, L. G., de Saint Martin, A., Van Gassen, K., Verbeek, N. E., Ville, D., Villeneuve, N., Zacher, P., Scheffer, I. E., and Lemke, J. R. (2021) The phenotypic spectrum of X-linked, infantile onset ALG13-related developmental and epileptic encephalopathy, *Epilepsia 62*, 325-334.

[108] Berry, G. T., Freeze, H. H., and Morava, E. (2021) Is X-linked, infantile onset ALG13-related developmental and epileptic encephalopathy a congenital disorder of glycosylation?, *Epilepsia 62*, 335-336.
[109] Yuste-Checa, P., Vega, A. I., Martín-Higueras, C., Medrano, C., Gámez, A., Desviat, L. R., Ugarte, M., Pérez-Cerdá, C., and Pérez, B. (2017) DPAGT1-CDG: Functional analysis of disease-causing pathogenic mutations and role of endoplasmic reticulum stress, *Public Library of Science (PLoS) One 12*, e0179456.
[110] Altassan, R., Péanne, R., Jaeken, J., Barone, R., Bidet, M., Borgel, D., Brasil, S., Cassiman, D., Cechova, A., Coman, D., Corral, J., Correia, J., de la Morena-Barrio, M. E., de Lonlay, P., Dos Reis, V., Ferreira, C. R., Fiumara, A., Francisco, R., Freeze, H., Funke, S., Gardeitchik, T., Gert, M., Girad, M., Giros, M., Grünewald, S., Hernández-Caselles, T., Honzik, T., Hutter, M., Krasnewich, D., Lam, C., Lee, J., Lefeber, D., Marques-da-Silva, D., Martinez-Monseny, A. F., Moravej, H., Õunap, K., Pascoal, C., Pascreau, T., Patterson, M., Quelhas, D., Raymond, K., Sarkhail, P., Schiff, M., Seroczyńska, M., Serrano, M., Seta, N., Sykut-Cegielska, J., Thiel, C., Tort, F., Vals, M.-A., Videira, P., Witters, P., Zeevaert, R., and Morava, E. (2019) International clinical guidelines for the management of phosphomannomutase 2-congenital disorders of glycosylation: Diagnosis, treatment and follow up, *Journal of Inherited Metabolic Disease 42*, 5-28.
[111] Van Schaftingen, E., and Jaeken, J. (1995) Phosphomannomutase deficiency is a cause of carbohydrate-deficient glycoprotein syndrome type I, *Federation of European Biochemical Societies (FEBS) Letters 377*, 318-320.
[112] Silvaggi, N. R., Zhang, C., Lu, Z., Dai, J., Dunaway-Mariano, D., and Allen, K. N. (2006) The X-ray Crystal Structures of Human α-Phosphomannomutase 1 Reveal the Structural Basis of Congenital Disorder of Glycosylation Type 1a *, *Journal of Biological Chemistry 281*, 14918-14926.
[113] Pirard, M., Achouri, Y., Collet, J.-F., Schollen, E., Matthijs, G., and Schaftingen, E. V. (1999) Kinetic properties and tissular distribution of mammalian phosphomannomutase isozymes, *Biochemical Journal 339*, 201-207.
[114] Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR), *Nucleic Acids Res 18*, 6069-6074.
[115] Ciccarone, V. C., Polayes, D. A., and Luckow, V. A. (1998) Generation of Recombinant Baculovirus DNA in E.coli Using a Baculovirus Shuttle Vector, *Methods Mol Med 13*, 213-235.
[116] Berger, I., Fitzgerald, D.J., Richmond, T.J., . (2004) Baculovirus expression system for heterologous multiprotein complexes, *Nature Biotechnology 22*.
[117] Stetsenko, A., and Guskov, A. (2017) An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization, *Crystals 7*, 197.
[118] Structural Genomics Consortium, SATurn platform,https://www.thesgc.org/saturn
[119] Wittig, I., Braun, H. P., and Schägger, H. (2006) Blue native PAGE, *Nat Protocols 1*, 418-428.
[120] Ibata, K., Mizuno, M., Takigawa, T., and Tanaka, Y. (1983) Long-chain betulaprenol-type polyprenols from the leaves of Ginkgo biloba, *Biochemical Journal 213*, 305-311.
[121] Das, D., Walvoort, M. T. C., Lukose, V., and Imperiali, B. (2016) A Rapid and Efficient Luminescence-based Method for Assaying Phosphoglycosyltransferase Enzymes, *Scientific Reports 6*, 33412.
[122] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2021) Protein complex prediction with AlphaFold-Multimer, *bioRxiv*, 2021.2010.2004.463034.
[123] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature*.

[124] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all, *Nature Methods 19*, 679-682.
[125] Schrödinger, L. (2015) PyMOL Molecular Graphics System version 2.0.
[126] Parker, J. L., and Newstead, S. (2012) Current trends in α-helical membrane protein crystallization: an update, *Protein Science 21*, 1358-1365.
[127] Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., and Cheng, Y. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM, *Nature Methods 10*, 584-590.
[128] Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., and Scheres, S. H. W. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3, *eLife 7*, e42166.
[129] Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination, *Nature Methods 14*, 290-296.
[130] Rohou, A., and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs, *Journal of Structural Biology 192*, 216-221.
[131] Gao, X.-D., Tachikawa, H., Sato, T., Jigami, Y., Dean, N. (2005) Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation, *Journal of Biological Chemistry 280*, 36254-36262.
[132] Mohammadi, T., Karczmarek, A., Crouvoisier, M., Bouhss, A., Mengin-Lecreulx, D., and den Blaauwen, T. (2007) The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in Escherichia coli, *Molecular Microbiology 65*, 1106-1121.
[133] Laddomada, F., Miyachiro, M. M., Jessop, M., Patin, D., Job, V., Mengin-Lecreulx, D., Le Roy, A., Ebel, C., Breyton, C., Gutsche, I., and Dessen, A. (2019) The MurG glycosyltransferase provides an oligomeric scaffold for the cytoplasmic steps of peptidoglycan biosynthesis in the human pathogen Bordetella pertussis, *Science Reports 9*, 4656.
[134] Bac-to-Bac® Baculovirus Expression System 2024,https://tools.thermofisher.com/content/sfs/manuals/bactobac_man.pdf
[135] Kollewe, C., and Vilcinskas, A. (2013) Production of Recombinant Proteins in Insect Cells, *American Journal of Biochemistry and Biotechnology 9*.
[136] Berger, I., Fitzgerald, D. J., and Richmond, T. J. (2004) Baculovirus expression system for heterologous multiprotein complexes, *Nature Biotechnology 22*, 1583-1587.
[137] Arachea, B. T., Sun, Z., Potente, N., Malik, R., Isailovic, D., and Viola, R. E. (2012) Detergent selection for enhanced extraction of membrane proteins, *Protein Expression and Purification 86*, 12-20.
[138] Kühlbrandt, W. (2014) The Resolution Revolution, *Science 343*, 1443-1444.
[139] Raman, S., Lange, O. F., Rossi, P., Tyka, M., Wang, X., Aramini, J., Liu, G., Ramelot, T. A., Eletsky, A., Szyperski, T., Kennedy, M. A., Prestegard, J., Montelione, G. T., and Baker, D. (2010) NMR Structure Determination for Larger Proteins Using Backbone-Only Data, *Science 327*, 1014-1018.
[140] Bloch, J. S., Pesciullesi, G., Boilevin, J., Nosol, K., Irobalieva, R. N., Darbre, T., Aebi, M., Kossiakoff, A. A., Reymond, J.-L., and Locher, K. P. (2020) Structure and mechanism of the ER-based glucosyltransferase ALG6, *Nature 579*, 443-447.
[141] Harrison, P. J., Vecerkova, T., Clare, D. K., and Quigley, A. (2023) A review of the approaches used to solve sub-100 kDa membrane proteins by cryo-electron microscopy, *Journal of Structural Biology 215*, 107959.
[142] Mitusińska, K., Góra, A., Bogdańska, A., Rożdżyńska-Świątkowska, A., Tylki-Szymańska, A., and Jezela-Stanek, A. (2022) Structural Analysis of the Effect of Asn107Ser Mutation on Alg13 Activity and Alg13-Alg14 Complex Formation and Expanding the Phenotypic Variability of ALG13-CDG, *Biomolecules 12*, 398.
[143] Averbeck, N., Keppler-Ross, S., and Dean, N. (2007) Membrane Topology of the Alg14 Endoplasmic Reticulum UDP-GlcNAc Transferase Subunit, *Journal of Biological Chemistry 282*, 29081-29088.

[144] Brown, K., Vial, S. C., Dedi, N., Westcott, J., Scally, S., Bugg, T. D., Charlton, P. A., and Cheetham, G. M. (2013) Crystal structure of the Pseudomonas aeruginosa MurG: UDP-GlcNAc substrate complex, *Protein Pept Lett 20*, 1002-1008.
[145] Gao, X. D., Moriyama, S., Miura, N., Dean, N., Nishimura, S.I. (2008) Interaction between the C Termini of Alg13 and Alg14 Mediates Formation of the Active UDP-N-acetylglucosamine Transferase Complex, *Journal of Biological Chemistry 283*, 32534-32541.
[146] Schägger, H., and von Jagow, G. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form, *Analytical Biochemistry 199*, 223-231.
[147] Gimpl, G., and Fahrenholz, F. (2002) Cholesterol as stabilizer of the oxytocin receptor, *Biochimica et Biophysica Acta 1564*, 384-392.
[148] Gimpl, G. (2016) Interaction of G protein coupled receptors and cholesterol, *Chemistry and Physics of Lipids 199*, 61-73.
[149] Stevens, R. (2008) Large Scale Preparation of Membrane Containing Over-expressed Proteins from SF9 Cells, *Joint Centre for Integral Membrane Protein Technologies (JCIMPT) 1*.
[150] Sinz, A. (2018) Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein–Protein Interactions: Where Are We Now and Where Should We Go from Here?, *Angewandte Chemie International Edition 57*, 6390-6396.
[151] Jones, A. X., Cao, Y., Tang, Y.-L., Wang, J.-H., Ding, Y.-H., Tan, H., Chen, Z.-L., Fang, R.-Q., Yin, J., Chen, R.-C., Zhu, X., She, Y., Huang, N., Shao, F., Ye, K., Sun, R.-X., He, S.-M., Lei, X., and Dong, M.-Q. (2019) Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers, *Nature Communications 10*, 3911.
[152] Doerr, A. (2016) Single-particle cryo-electron microscopy, *Nature Methods 13*, 23-23.
[153] Kampjut, D., Steiner, J., and Sazanov, L. A. (2021) Cryo-EM grid optimization for membrane proteins, *iScience 24*, 102139.
[154] D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M., and Kühlbrandt, W. (2019) Protein denaturation at the air-water interface and how to prevent it, *eLife 8*, e42747.
[155] Bayburt, T. H., Grinkova, Y. V., and Sligar, S. G. (2002) Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins, *Nano Letters 2*, 853-856.
[156] Weissenberger, G., Henderikx, R. J. M., and Peters, P. J. (2021) Understanding the invisible hands of sample preparation for cryo-EM, *Nature Methods 18*, 463-471.
[157] Segel, I. H. (1993) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, Wiley.
[158] Cleland, W. W. (1963) The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations, *Biochimica et Biophysica Acta 67*, 104-137.
[159] Lukose, V., Walvoort, M. T. C., and Imperiali, B. (2017) Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface, *Glycobiology 27*, 820-833.
[160] Keller, R. K., Boon, D. Y., and Crum, F. C. (1979) N-Acetylglucosamine- 1 -phosphate transferase from hen oviduct: solubilization, characterization, and inhibition by tunicamycin, *Biochemistry 18*, 3946-3952.
[161] Al-Dabbagh, B., Olatunji, S., Crouvoisier, M., El Ghachi, M., Blanot, D., Mengin-Lecreulx, D., and Bouhss, A. (2016) Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily, *Biochimie 127*, 249-257.
[162] Liu, Y., 刘, 垚., Rodrigues, J. P. G. L. M., Bonvin, A. M. J. J., Zaal, E. A., Berkers, C. R., Heger, M., Gawarecka, K., Swiezewska, E., Breukink, E., and Egmond, M. R. (2016) New Insight into the Catalytic Mechanism of Bacterial MraY from Enzyme Kinetics and Docking Studies, *Journal of Biological Chemistry 291*, 15057-15068.
[163] Al-Dabbagh, B., Mengin-Lecreulx, D., and Bouhss, A. (2008) Purification and characterization of the bacterial UDP-GlcNAc: undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA, *J Bact 190*, 7141-7146.
[164] Ray, L. C., Das, D., Entova, S., Lukose, V., Lynch, A. J., Imperiali, B., and Allen, K. N. (2018) Membrane association of monotopic phosphoglycosyl transferase underpins function, *Nature Chemical Biology 14*, 538-541.

[165] Das, D., Kuzmic, P., and Imperiali, B. (2017) Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate, *Proceedings of the National Academy of Sciences 114*, 7019-7024.
[166] Mitachi, K., Yun, H. G., Gillman, C. D., Skorupinska-Tudek, K., Swiezewska, E., Clemons, W. M., Jr., and Kurosu, M. (2020) Substrate Tolerance of Bacterial Glycosyltransferase MurG: Novel Fluorescence-Based Assays, *American Chemical Society Infectious Disease 6*, 1501-1516.
[167] Behrens, N. H., and Tábora, E. (1978) Dolichol intermediates in the glycosylation of proteins, *Methods in Enzymology 50*, 402-435.
[168] Bouhss, A., Crouvoisier, M., Blanot, D., and Mengin-Lecreulx, D. (2004) Purification and Characterization of the Bacterial MraY Translocase Catalyzing the First Membrane Step of Peptidoglycan Biosynthesis, *Journal of Biological Chemistry 279*, 29974-29980.
[169] Patel, K. B., Ciepichal, E., Swiezewska, E., and Valvano, M. A. (2012) The C-terminal domain of the Salmonella enterica WbaP (UDP-galactose:Und-P galactose-1-phosphate transferase) is sufficient for catalytic activity and specificity for undecaprenyl monophosphate, *Glycobiology 22*, 116-122.
[170] Hyland, S. A., and Anderson, M. S. (2003) A high-throughput solid-phase extraction assay capable of measuring diverse polyprenyl phosphate: sugar-1-phosphate transferases as exemplified by the WecA, MraY, and MurG proteins, *Analytical Biochemistry 317*, 156-165.
[171] Sharma, C. B. C., B., Lehle, L., and Tanner, W. (1982) Solubilization and Characterization of the Initial Enzymes of the Dolichol Pathway from Yeast, *European Journal of Biochemistry 126*, 319-325.
[172] Kaushal, G. P., and Elbein, A. D. (1986) Purification and Properties of UDP-GlcNAc:Dolichyl-Pyrophosphoryl-GlcNAc GlcNAc Transferase from Mung Bean Seedling 1, *Plant Physiology 81*, 1086-1091.
[173] Tai, V. W. F., O'Reilly, M. K., and Imperiali, B. (2001) Substrate specificity of N-acetylglucosaminyl(diphosphodolichol) N-acetylglucosaminyl transferase, a key enzyme in the dolichol pathway, *Bioorganic and Medicinal Chemistry 9*, 1133-1140.
[174] Li, S. T., Lu, T. T., Xu, X. X., Ding, Y., Li, Z., Kitajima, T., Dean, N., Wang, N., and Gao, X. D. (2019) Reconstitution of the lipid-linked oligosaccharide pathway for assembly of high-mannose N-glycans, *Nature Communications 10*, 1813.
[175] Stachyra, T., Dini, C., Ferrari, P., Bouhss, A., van Heijenoort, J., Mengin-Lecreulx, D., Blanot, D., Biton, J., and Le Beller, D. (2004) Fluorescence Detection-Based Functional Assay for High-Throughput Screening for MraY, *Antimicrobial Agents and Chemotherapy 48*, 897-902.
[176] Li, J.-J., and Bugg, T. D. H. (2004) A fluorescent analogue of UDP-N-acetylglucosamine: application for FRET assay of peptidoglycan translocase II (MurG), *Chemical Communications*, 182-183.
[177] Schouten, J. A., Bagga, S., Lloyd, A. J., de Pascale, G., Dowson, C. G., Roper, D. I., and Bugg, T. D. (2006) Fluorescent reagents for in vitro studies of lipid-linked steps of bacterial peptidoglycan biosynthesis: derivatives of UDPMurNAc-pentapeptide containing d-cysteine at position 4 or 5, *Molecular Biosystems 2*, 484-491.
[178] Shapiro, A. B., Jahić, H., Gao, N., Hajec, L., and Rivin, O. (2012) A high-throughput, homogeneous, fluorescence resonance energy transfer-based assay for phospho-N-acetylmuramoyl-pentapeptide translocase (MraY), *Journal of Biomolecular Screening 17*, 662-672.
[179] Ramírez, A. S., Boilevin, J., Biswas, R., Gan, B. H., Janser, D. l., Aebi, M., Darbre, T., Reymond, J. L., and Locher, K. P. (2017) Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs, *Glycobiology 27*, 525-535.
[180] Qi, X., Ma, M., Wang, L., Zhang, Y., Jiang, R., Bai, L., and Li, Y. (2015) Biochemical characterization of a novel bifunctional glycosyl-1-phosphate transferase involved in the exopolysaccharide biosynthesis, *Biochemical and Biophysical Research Communications 465*, 113-118.
[181] Engel, L., Alves, J., Hennek, J., Goueli, S. A., and Zegzouti, H. (2021) Utility of Bioluminescent Homogeneous Nucleotide Detection Assays in Measuring Activities of

Nucleotide-Sugar Dependent Glycosyltransferases and Studying Their Inhibitors, *Molecules 26*.
[182] Astani, E. K., Malek Zadeh, S., Hsu, N.-S., Lin, K.-H., Sardari, S., and Li, T.-L. (2022) Intermolecular Interactions of Nucleoside Antibiotic Tunicamycin with On-Target MraYCB-TUN and Off-Target DPAGT1-TUN in the Active Sites Delineated by Quantum Mechanics/Molecular Mechanics Calculations, *American Chemical Society Omega 7*, 32970-32987.
[183] Copeland, R. A. (2000) Tight Binding Inhibitors, In *Enzymes* 2 ed., pp 305-317.
[184] Williams, J. W., and Morrison, J. F. (1979) The kinetics of reversible tight-binding inhibition, In *Methods Enzymol*, pp 437-467, Academic Press.
[185] Rush, J. S., Gao, N., Lehrman, M. A., and Waechter, C. J. (2008) Recycling of Dolichyl Monophosphate to the Cytoplasmic Leaflet of the Endoplasmic Reticulum after the Cleavage of Dolichyl Pyrophosphate on the Lumenal Monolayer, *Journal of Biological Chemistry 283*, 4087-4093.
[186] Larkin, A., Chang, M.M., Whitworth, G.E., Imperiali, B., . (2013) Biochemical evidence for an alternative pathway in N-linked glycoprotein biosynthesis, *Nature Chemical Biology 9*, 367-373.
[187] Breukink, E., van Heusden, H. E., Vollmerhaus, P. J., Swiezewska, E., Brunner, L., Walker, S., Heck, A. J., and de Kruijff, B. (2003) Lipid II is an intrinsic component of the pore induced by nisin in bacterial membranes, *Journal of Biological Chemistry 278*, 19898-19903.
[188] Mitachi, K., Siricilla, S., Yang, D., Kong, Y., Skorupinska-Tudek, K., Swiezewska, E., Franzblau, S. G., and Kurosu, M. (2016) Fluorescence-based assay for polyprenyl phosphate-GlcNAc-1-phosphate transferase (WecA) and identification of novel antimycobacterial WecA inhibitors, *Analtyical Biochemistry 512*, 78-90.
[189] Palamarczyk, G., Lehle, L., Mankowski, T., Chojnacki, T., and Tanner, W. (1980) Specificity of Solubilized Yeast Glycosyl Transferases for Polyprenyl Derivatives, *European Journal of Biochemistry 105*, 517-523.
[190] Löw, P., Peterson, E., Mizuno, M., Takigawa, T., Chojnacki, T., and Dallner, G. (1986) Effectivity of dolichyl phosphates with different chain lengths as acceptors of nucleotide activated sugars, *Bioscience Reports 6*, 677-683.
[191] Berendes, R., and Jaenicke, L. (1992) Short-Chain Dolichols of Defined Chain Length as Cofactors in Reactions of the Microsomal Dolichyl-Phosphate Cycle and Transglycosylations, *Biological Chemistry 373*, 35-42.
[192] Szkopińska, A., Swieżewska, E., and Chojnacki, T. (1992) On the specificity of dolichol kinase and DolPMan synthase towards isoprenoid alcohols of different chain length in rat liver microsomal membrane, *International Journal of Biochemistry 24*, 1151-1157.
[193] McLachlan, K. R., and Krag, S. S. (1992) Substrate specificity of N-acetylglucosamine 1-phosphate transferase activity in Chinese hamster ovary cells, *Glycobiology 2*, 313-319.
[194] Ramírez, A. S., Boilevin, J., Lin, C.-W., Ha Gan, B., Janser, D., Aebi, M., Darbre, T., Reymond, J.-L., and Locher, K. P. (2017) Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins, *Glycobiology 27*, 726-733.
[195] Wang, C.-D., Xu, S., Chen, S., Chen, Z.-H., Dean, N., Wang, N., and Gao, X.-D. (2022) An in vitro assay for enzymatic studies on human ALG13/14 heterodimeric UDP-N-acetylglucosamine transferase, *Frontiers in Cell and Developmental Biology 10*.
[196] Boilevin, J. M., and Reymond, J.-L. (2018) Synthesis of Lipid-Linked Oligosaccharides (LLOs) and Their Phosphonate Analogues as Probes To Study Protein Glycosylation Enzymes, *Synthesis 50*, 2631-2654.
[197] Cochrane, R. V. K., Alexander, F. M., Boland, C., Fetics, S. K., Caffrey, M., and Cochrane, S. A. (2020) From plant to probe: semi-synthesis of labelled undecaprenol analogues allows rapid access to probes for antibiotic targets, *Chemical Communications 56*, 8603-8606.
[198] Suzuki, S., Mori, F., Takigawa, T., Ibata, K., Ninagawa, Y., Nishida, T., Mizuno, M., Tanakaa, Y., . (1983) Synthesis of mammalian dolichos from plant polyprenols, *Tetrahedron Letters 24*, 5103-5106.

[199] Takigawa, T., Ibata, K., Mizuno, M., . (1989) Synthesis of mammalian dolichols from plant polyprenols, *Chemistry and Physics of Lipids 51*, 171-182.
[200] Takaya, H., Ohta, T., Sayo, N., Kumobayashi, H., Akutagawa, S., Inoue, S., Kasahara, I., and Noyori, R. (1987) Enantioselective hydrogenation of allylic and homoallylic alcohols, *Journal of the American Chemical Society 109*, 1596-1597.
[201] Imperiali, B., and Zimmerman, J. W. (1988) Synthesis of dolichols via asymmetric hydrogenation of plant polyprenols, *Tetrahedron Letters 29*, 5343-5344.
[202] Wu, R., Beauchamps, M.G., Laquidara, J.M., Sowa Jr, J.R., . (2012) Ruthenium-catalyzyed Asymmetric Transfer Hydrogenation of Allylic Alcohols by an Enantioselective Isomerisation/Transfer Hydrogenation Mechanism *Angewandte Chemie International Edition 51*, 2106-2110.
[203] Elharar, Y., Podilapu, A. R., Guan, Z., Kulkarni, S. S., and Eichler, J. (2017) Assembling Glycan-Charged Dolichol Phosphates: Chemoenzymatic Synthesis of a Haloferax volcanii N-Glycosylation Pathway Intermediate, *Bioconjugate Chemistry 28*, 2461-2470.
[204] Jaenicke, L., and Siegmund, H.-U. (1989) Synthesis and characterization of dolichols and polyprenols of designed geometry and chain length, *Chemistry and Physics of Lipids 51*, 159-170.
[205] Hirao, K., Ono, R., Manabe, Y., Masui, S., Atomi, H., and Fukase, K. (2020) Total Syntheses of C60- and C100-Dolichols, *The Journal of Organic Chemistry 85*, 11549-11559.
[206] Rush, J. S., Rick, P. D., and Waechter, C. J. (1997) Polyisoprenyl phosphate specificity of UDP-GIcNAc:undecaprenyl phosphate N-acetylglucosaminyl 1-P transferase from E.coli, *Glycobiology 7*, 315-322.
[207] Fukada, H., Sturtevant, J. M., and Quiocho, F. A. (1983) Thermodynamics of the binding of L-arabinose and of D-galactose to the L-arabinose-binding protein of Escherichia coli, *Journal of Biological Chemistry 258*, 13193-13198.
[208] Errasti-Murugarren, E., Bartoccioni, P., and Palacín, M. (2021) Membrane Protein Stabilization Strategies for Structural and Functional Studies, *Membranes (Basel) 11*.
[209] McLachlan, K. R., and Krag, S. S. (1994) Three enzymes involved in oligosaccharide-lipid assembly in Chinese hamster ovary cells differ in lipid substrate preference, *Journal of Lipid Research 35*, 1861-1868.
[210] Krag, S. S. (1998) The Importance of Being Dolichol, *Biochemical and Biophysical Research Communications 243*, 1-5.
[211] Hall, C. W., McLachlan, K. R., Krag, S. S., and Robbins, A. R. (1997) Reduced utilization of Man5GlcNAc2-P-P-lipid in a Lec9 mutant of Chinese hamster ovary cells: analysis of the steps in oligosaccharide-lipid assembly, *Journal of Cellular Biochemistry 67*, 201-215.
[212] Ng, B. G., and Freeze, H. H. (2018) Perspectives on Glycosylation and Its Congenital Disorders, *Trends in Genetics 34*, 466-476.
[213] Alsharhan, H., He, M., Edmondson, A. C., Daniel, E. J. P., Chen, J., Donald, T., Bakhtiari, S., Amor, D. J., Jones, E. A., Vassallo, G., Vincent, M., Cogné, B., Deb, W., Werners, A. H., Jin, S. C., Bilguvar, K., Christodoulou, J., Webster, R. I., Yearwood, K. R., Ng, B. G., Freeze, H. H., Kruer, M. C., Li, D., Raymond, K. M., Bhoj, E. J., and Sobering, A. K. (2021) ALG13 X-linked intellectual disability: New variants, glycosylation analysis, *Journal of Inherited Metabolic Disease*, 1-12.
[214] Timal, S., Hoischen, A., Lehle, L., Adamowicz, M., Huijben, K., Sykut-Cegielska, J., Paprocka, J., Jamroz, E., van Spronsen, F. J., Körner, C., Gilissen, C., Rodenburg, R. J., Eidhof, I., Van den Heuvel, L., Thiel, C., Wevers, R. A., Morava, E., Veltman, J., and Lefeber, D. J. (2012) Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing, *Human Molecular Genetics 21*, 4151-4161.
[215] Barone, R., Fiumara, A., and Jaeken, J. (2014) Congenital Disorders of Glycosylation with Emphasis on Cerebellar Involvement, *Seminars in Neurology 34*, 357-366.
[216] Shah, R., Johnsen, C., Pletcher, B. A., Edmondson, A. C., Kozicz, T., and Morava, E. (2023) Long-term outcomes in ALG13-Congenital Disorder of Glycosylation, *American Journal of Medical Genetics Part A 191*, 1626-1631.
[217] Alsharhan, H., He, M., Edmondson, A. C., Daniel, E. J. P., Chen, J., Donald, T., Bakhtiari, S., Amor, D. J., Jones, E. A., Vassallo, G., Vincent, M., Cogné, B., Deb, W., Werners, A. H., Jin,

S. C., Bilguvar, K., Christodoulou, J., Webster, R. I., Yearwood, K. R., Ng, B. G., Freeze, H. H., Kruer, M. C., Li, D., Raymond, K. M., Bhoj, E. J., and Sobering, A. K. (2021) ALG13 X-linked intellectual disability: New variants, glycosylation analysis, and expanded phenotypes, *J Inherit Metab Dis 44*, 1001-1012.

[218] Galama, W. H., Verhaagen-van den Akker, S. L. J., Lefeber, D. J., Feenstra, I., and Verrips, A. (2018) ALG13-CDG with Infantile Spasms in a Male Patient Due to a De Novo ALG13 Gene Mutation, *Journal of Inherited Metabolic Disease Reports 40*, 11-16.

[219] de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-van Silfhout, A. T., Koolen, D. A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B. B., Brunner, H. G., Veltman, J. A., and Vissers, L. E. (2012) Diagnostic exome sequencing in persons with severe intellectual disability, *The New England Journal of Medicine (NEJM) 367*, 1921-1929.

[220] Smith-Packard, B., Myers, S. M., and Williams, M. S. (2015) Girls with Seizures Due to the c.320A>G Variant in ALG13 Do Not Show Abnormal Glycosylation Pattern on Standard Testing, *Journal of Inherited Metabolic Diseases (JIMD) Reports 22*, 95-98.

[221] Kobayashi, Y., Tohyama, J., Kato, M., Akasaka, N., Magara, S., Kawashima, H., Ohashi, T., Shiraishi, H., Nakashima, M., Saitsu, H., and Matsumoto, N. (2016) High prevalence of genetic alterations in early-onset epileptic encephalopathies associated with infantile movement disorders, *Brain and Development 38*, 285-292.

[222] Bastaki, F., Bizzari, S., Hamici, S., Nair, P., Mohamed, M., Saif, F., Malik, E. M., Al-Ali, M. T., and Hamzeh, A. R. (2018) Single-center experience of N-linked Congenital Disorders of Glycosylation with a Summary of Molecularly Characterized Cases in Arabs, *Annals of Human Genetics 82*, 35-47.

[223] Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E. E., Epstein, M. P., Glauser, T., Goldstein, D. B., Han, Y., Heinzen, E. L., Hitomi, Y., Howell, K. B., Johnson, M. R., Kuzniecky, R., Lowenstein, D. H., Lu, Y.-F., Madou, M. R. Z., Marson, A. G., Mefford, H. C., Esmaeeli Nieh, S., O'Brien, T. J., Ottman, R., Petrovski, S., Poduri, A., Ruzzo, E. K., Scheffer, I. E., Sherr, E. H., Yuskaitis, C. J., Abou-Khalil, B., Alldredge, B. K., Bautista, J. F., Berkovic, S. F., Boro, A., Cascino, G. D., Consalvo, D., Crumrine, P., Devinsky, O., Dlugos, D., Epstein, M. P., Fiol, M., Fountain, N. B., French, J., Friedman, D., Geller, E. B., Glauser, T., Glynn, S., Haut, S. R., Hayward, J., Helmers, S. L., Joshi, S., Kanner, A., Kirsch, H. E., Knowlton, R. C., Kossoff, E. H., Kuperman, R., Kuzniecky, R., Lowenstein, D. H., McGuire, S. M., Motika, P. V., Novotny, E. J., Ottman, R., Paolicchi, J. M., Parent, J. M., Park, K., Poduri, A., Scheffer, I. E., Shellhaas, R. A., Sherr, E. H., Shih, J. J., Singh, R., Sirven, J., Smith, M. C., Sullivan, J., Lin Thio, L., Venkat, A., Vining, E. P. G., Von Allmen, G. K., Weisenberg, J. L., Widdess-Walsh, P., Winawer, M. R., Epi, K. C., and Epilepsy Phenome/Genome, P. (2013) De novo mutations in epileptic encephalopathies, *Nature 501*, 217-221.

[224] Hamici, S., Bastaki, F., and Khalifa, M. (2017) Exome sequence identified a c.320A > G ALG13 variant in a female with infantile epileptic encephalopathy with normal glycosylation and random X inactivation: Review of the literature, *European Journal of Medical Genetics 60*, 541-547.

[225] Evangelista, T., Hanna, M., and Lochmüller, H. (2015) Congenital Myasthenic Syndromes with Predominant Limb Girdle Weakness, *Journal of Neuromuscular Disorders 2*, S21-s29.

[226] Gang, Q., Bettencourt, C., Brady, S., Holton, J. L., Healy, E. G., McConville, J., Morrison, P. J., Ripolone, M., Violano, R., Sciacco, M., Moggio, M., Mora, M., Mantegazza, R., Zanotti, S., Wang, Z., Yuan, Y., Liu, W. W., Beeson, D., Hanna, M., and Houlden, H. (2022) Genetic defects are common in myopathies with tubular aggregates, *Annals of Clinical and Translational Neurology 9*, 4-15.

[227] Schorling, D. C., Rost, S., Lefeber, D. J., Brady, L., Müller, C. R., Korinthenberg, R., Tarnopolsky, M., Bönnemann, C. G., Rodenburg, R. J., Bugiani, M., Beytia, M., Krüger, M., van der Knaap, M., and Kirschner, J. (2017) Early and lethal neurodegeneration with myasthenic and myopathic features: A new ALG14-CDG, *Neurology 89*, 657-664.

[228] Palombo, F., Piccolo, B., Saccani, E., Fiorini, C., Capristo, M., Caporali, L., Pisani, F., and Carelli, V. (2021) A novel ALG14 missense variant in an alive child with myopathy,

epilepsy, and progressive cerebral atrophy, *American Journal of Medical Genetics Part A 185*, 1918-1921.
[229] Katata, Y., Uneoka, S., Saijyo, N., Aihara, Y., Miyazoe, T., Koyamaishi, S., Oikawa, Y., Ito, Y., Abe, Y., Numata-Uematsu, Y., Takayama, J., Kikuchi, A., Tamiya, G., Uematsu, M., and Kure, S. (2022) The longest reported sibling survivors of a severe form of congenital myasthenic syndrome with the ALG14 pathogenic variant, *American Journal of Medical Genetics Part A 188*, 1293-1298.
[230] Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., and Avsec, Ž. (2023) Accurate proteome-wide missense variant effect prediction with AlphaMissense, *Science 381*, eadg7492.
[231] Tordai, H., Torres, O., Csepi, M., Padányi, R., Lukács, G. L., and Hegedűs, T. (2023) Lightway access to AlphaMissense data that demonstrates a balanced performance of this missense mutation predictor, *bioRxiv*, 2023.2010.2030.564807.
[232] Minton, K. (2023) Predicting variant pathogenicity with AlphaMissense, *Nature Reviews Genetics 24*, 804-804.
[233] Patel, K. B., Furlong, S. E., and Valvano, M. A. (2010) Functional analysis of the C-terminal domain of the WbaP protein that mediates initiation of O antigen synthesis in Salmonella enterica, *Glycobiology 20*, 1389-1401.
[234] Thomas, P., and Smart, T. G. (2005) HEK293 cell line: A vehicle for the expression of recombinant proteins, *Journal of Pharmacological and Toxicological Methods 51*, 187-200.
[235] Chantret, I., Dancourt, J., Barbat, A., and Moore, S. E. H. (2005) Two Proteins Homologous to the N- and C-terminal Domains of the Bacterial Glycosyltransferase Murg Are Required for the Second Step of Dolichyl-linked Oligosaccharide Synthesis in Saccharomyces cerevisiae, *Journal of Biological Chemistry 280*, 9236-9242.
[236] Gao, X. D., Tachikawa, H., Sato, T., Jigami, Y., and Dean, N. (2005) Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation, *Journal of Biological Chemistry 280*, 36254-36262.
[237] Zara, J., and Lehrman, M. A. (1994) Role of the carboxyl terminus in stable expression of hamster UDP-GlcNAc:dolichol-P GlcNAc-1-P transferase, *Journal of Biological Chemistry 269*, 19108-19115.
[238] Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., and Ting, A. Y. (2013) Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging, *Science 339*, 1328-1331.
[239] Gupta, K., Donlan, J. A. C., Hopper, J. T. S., Uzdavinys, P., Landreh, M., Struwe, W. B., Drew, D., Baldwin, A. J., Stansfeld, P. J., and Robinson, C. V. (2017) The role of interfacial lipids in stabilizing membrane protein oligomers, *Nature 541*, 421-424.
[240] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., (2002) *Moleuclar Biology of the Cell: The Lipid Bilayer*, 4th ed., Garland Science, New York.
[241] Zhou, G. P., and Troy, F. A., 2nd. (2003) Characterization by NMR and molecular modeling of the binding of polyisoprenols and polyisoprenyl recognition sequence peptides: 3D structure of the complexes reveals sites of specific interactions, *Glycobiology 13*, 51-71.
[242] Liedtke, J., Depelteau, J. S., and Briegel, A. (2022) How advances in cryo-electron tomography have contributed to our current view of bacterial cell biology, *Journal of Structural Biology:X (JSBX) 6*, 100065.
[243] Das, D., Kuzmic, P., and Imperiali, B. (2017) Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate, *Proceedings of the National Academy of Sciences of the United States of America 114*, 7019-7024.
[244] Laganowsky, A., Reading, E., Allison, T. M., Ulmschneider, M. B., Degiacomi, M. T., Baldwin, A. J., and Robinson, C. V. (2014) Membrane proteins bind lipids selectively to modulate their structure and function, *Nature 510*, 172-175.
[245] Krishnarjuna, B., and Ramamoorthy, A. (2022) Detergent-Free Isolation of Membrane Proteins and Strategies to Study Them in a Near-Native Membrane Environment, *Biomolecules 12*, 1076.

[246] Dodge, G. J., Anderson, A. J., He, Y., Liu, W., Viner, R., and Imperiali, B. (2023) Mapping the architecture of the initiating phosphoglycosyl transferase from S. enterica O-antigen biosynthesis in a liponanoparticle, *bioRxiv*.
[247] Cantagrel, V., Lefeber, D. J., Ng, B. G., Guan, Z., Silhavy, J. L., Bielas, S. L., Lehle, L., Hombauer, H., Adamowicz, M., Swiezewska, E., De Brouwer, A. P., Blümel, P., Sykut-Cegielska, J., Houliston, S., Swistun, D., Ali, B. R., Dobyns, W. B., Babovic-Vuksanovic, D., van Bokhoven, H., Wevers, R. A., Raetz, C. R., Freeze, H. H., Morava, E., Al-Gazali, L., and Gleeson, J. G. (2010) SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder, *Cell 142*, 203-217.
[248] Gandini, R., Reichenbach, T., Tan, T.-C., and Divne, C. (2017) Structural basis for dolichylphosphate mannose biosynthesis, *Nature Communications 8*, 120.

Appendices

Table S1. DNA constructs and protein sequences of target proteins used in this study

Target	Tag	DNA Sequence	Protein Sequence
SRD5α3	N-terminal 6X His tag with TEV protease site	ATGGGCCACCATCATCATCATCATTCTTCTGGTGT AGATCTGGGTACCGAGAACCTGTACTTCCAATCCA TGGCTCCCTGGGCGGAGGCCGAGCACTCGGCGCT GAACCCGCTGCGCGCGGTGTGGCTCACGCTGACC GCCGCCTTCCTGCTGACCCTACTGCTGCAGCTCCT GCCGCCCGGCCTGCTCCCGGGCTGCGCGATCTTCC AGGACCTGATCCGCTATGGGAAAACCAAGTGTGG GGAGCCGTCGCGCCCCGCCGCCTGCCGAGCCTTTG ATGTCCCCAAGAGATATTTTTCCCACTTTTATATCA TCTCAGTGCTGTGGAATGGCTTCCTGCTTTGGTGC CTTACTCAATCTCTGTTCCTGGGAGCACCTTTTCCA AGCTGGCTTCATGGTTTGCTCAGAATTCTCGGGGC GGCACAGTTCCAGGGAGGGGAGCTGGCACTGTCT GCATTCTTAGTGCTAGTATTTCTGTGGCTGCACAG CTTACGAAGACTCTTCGAGTGCCTCTACGTCAGTG TCTTCTCCAATGTCATGATTCACGTCGTGCAGTACT GTTTTGGACTTGTCTATTATGTCCTTGTTGGCCTAA CTGTGCTGAGCCAAGTGCCAATGGATGGCAGGAA TGCCTACATAACAGGGAAAAATCTATTGATGCAA GCACGGTGGTTCCATATTCTTGGGATGATGATGTT CATCTGGTCATCTGCCCATCAGTATAAGTGCCATG TTATTCTCGGCAATCTCAGGAAAAATAAAGCAGG AGTGGTCATTCACTGTAACCACAGGATCCCATTTG GAGACTGGTTTGAATATGTTTCTTCCCCTAACTACT TAGCAGAGCTGATGATCTACGTTTCCATGGCCGTC ACCTTTGGGTTCCACAACTTAACTTGGTGGCTAGT GGTGACAAATGTCTTCTTTAATCAGGCCCTGTCTG CCTTTCTCAGCCACCAATTCTACAAAAGCAAATTT GTCTCTTACCCGAAGCATAGGAAAGCTTTCCTACC ATTTTTGTTTTGACAGTAAAGGTGGATACGGATCC GAATTCGAGCTCCGTCGACAAGCTT	MGHHHHHHSSGVDLGTENLYFQSMA PWAEAEHSALNPLRAVWLTLTAAFLL TLLLQLLPPGLLPGCAIFQDLIRYGKT KCGEPSRPAACRAFDVPKRYFSHFYII SVLWNGFLLWCLTQSLFLGAPFPSWL HGLLRILGAAQFQGGELALSAFLVLV FLWLHSLRRLFECLYVSVFSNVMIHV VQYCFGLVYYVLVGLTVLSQVPMDG RNAYITGKNLLMQARWFHILGMMMF IWSSAHQYKCHVILGNLRKNKAGVVI HCNHRIPFGDWFEYVSSPNYLAELMI YVSMAVTFGFHNLTWWLVVTNVFFN QALSAFLSHQFYKSKFVSYPKHRKAF LPFLF

His-DPAGT1	N-terminal 6xHis Tag with TEV protease site	ATGGGCCACCATCATCATCATCATTCTTCTGGTGT AGATCTGGGTACCGAGAACCTGTACTTCCAATCCA TGTGGGCCTTCTCGGAATTGCCCATGCCGCTGCTG ATCAATTTGATCGTCTCGCTGCTGGGATTTGTGGC CACAGTCACCCTCATCCCGGCCTTCCGGGGCCACT TCATTGCTGCGCGCCTCTGTGGTCAGGACCTCAAC AAAACCAGCCGACAGCAGATCCCAGAATCCCAGG GAGTGATCAGCGGTGCTGTTTTCCTTATCATCCTCT TCTGCTTCATCCCTTTCCCCTTCCTGAACTGCTTTG TGAAGGAGCAGTGTAAGGCATTCCCCCACCATGA ATTTGTGGCCCTGATAGGTGCCCTCCTTGCCATCT GCTGCATGATCTTCCTGGGCTTTGCGGATGATGTA CTGAATCTGCGCTGGCGCCATAAGCTGCTGCTACC TACAGCTGCCTCACTACCTCTCCTCATGGTCTATTT CACCAACTTTGGCAACACGACCATTGTGGTGCCCA AGCCCTTCCGCCCGATACTTGGCCTGCATCTGGAC TTGGGAATCCTGTACTATGTCTACATGGGGCTGCT GGCAGTGTTCTGTACCAATGCCATCAATATCCTAG CAGGAATTAACGGCCTAGAGGCTGGCCAGTCACT AGTCATTTCTGCTTCCATCATTGTCTTCAACCTGGT AGAGTTGGAAGGTGATTGTCGGGATGATCATGTCT TTTCCCTCTACTTCATGATACCCTTTTTTTTCACCA CTTTGGGATTGCTCTACCACAACTGGTACCCATCA CGGGTGTTTGTGGGAGATACCTTCTGTTACTTTGCT GGCATGACCTTTGCCGTGGTGGGCATCTTGGGACA CTTCAGCAAGACCATGCTACTATTCTTCATGCCCC AGGTGTTCAACTTCCTCTACTCACTGCCTCAGCTC CTGCATATCATCCCCTGCCCTCGCCACCGCATACC CAGACTCAATATCAAGACAGGCAAACTGGAGATG AGCTATTCCAAGTTCAAGACCAAGAGCCTCTCTTT CTTGGGCACCTTTATTTTAAAGGTGGCAGAGAGCC TCCAGCTGGTGACAGTACACCAGAGTGAGACTGA AGATGGTGAATTCACTGAATGTAACAACATGACCC TCATCAACTTGCTACTTAAAGTCCTTGGGCCCATA CATGAGAGAAACCTCACATTGCTCCTGCTGCTGCT GCAGATCCTGGGCAGTGCCATCACCTTCTCCATTC GATATCAGCTCGTTCGACTCTTCTATGATGTCTGA CAGTAAAGGTGGATACGGATCCGAATTCGAGCTC CGTCGACAAGCTT	MGHHHHHHSSGVDLGTENLYFQSM WAFSELPMPLLINLIVSLLGFVATVTLI PAFRGHFIAARLCGQDLNKTSRQQIPE SQGVISGAVFLIILFCFIPFPFLNCFVKE QCKAFPHHEFVALIGALLAICCMIFLG FADDVLNLRWRHKLLLPTAASLPLLM VYFTNFGNTTIVVPKPFRPILGLHLDL GILYYVYMGLLAVFCTNAINILAGING LEAGQSLVISASIIVFNLVELEGDCRD DHVFSLYFMIPFFFTTLGLLYHNWYPS RVFVGDTFCYFAGMTFAVVGILGHFS KTMLLFFMPQVFNFLYSLPQLLHIIPC PRHRIPRLNIKTGKLEMSYSKFKTKSL SFLGTFILKVAESLQLVTVHQSETEDG EFTECNNMTLINLLLKVLGPIHERNLT LLLLLLQILGSAITFSIRYQLVRLFYDV

DPAGT1-His-FLAG	C-terminal His-FLAG tag with 3C protease site	ATGGGCCACCATCATCATCATCATTCTTCTGGTGT AGATCTGGGTACCGAGAACCTGTACTTCCAATCCA TGTGGGCCTTCTCGGAATTGCCCATGCCGCTGCTG ATCAATTTGATCGTCTCGCTGCTGGGATTTGTGGC CACAGTCACCCTCATCCCGGCCTTCCGGGGCCACT TCATTGCTGCGCGCCTCTGTGGTCAGGACCTCAAC AAAACCAGCCGACAGCAGATCCCAGAATCCCAGG GAGTGATCAGCGGTGCTGTTTTCCTTATCATCCTCT TCTGCTTCATCCCTTTCCCCTTCCTGAACTGCTTTG TGAAGGAGCAGTGTAAGGCATTCCCCCACCATGA ATTTGTGGCCCTGATAGGTGCCCTCCTTGCCATCT GCTGCATGATCTTCCTGGGCTTTGCGGATGATGTA CTGAATCTGCGCTGGCGCCATAAGCTGCTGCTACC TACAGCTGCCTCACTACCTCTCCTCATGGTCTATTT CACCAACTTTGGCAACACGACCATTGTGGTGCCCA AGCCCTTCCGCCCGATACTTGGCCTGCATCTGGAC TTGGGAATCCTGTACTATGTCTACATGGGGCTGCT GGCAGTGTTCTGTACCAATGCCATCAATATCCTAG CAGGAATTAACGGCCTAGAGGCTGGCCAGTCACT AGTCATTTCTGCTTCCATCATTGTCTTCAACCTGGT AGAGTTGGAAGGTGATTGTCGGGATGATCATGTCT TTTCCCTCTACTTCATGATACCCTTTTTTTTCACCA CTTTGGGATTGCTCTACCACAACTGGTACCCATCA CGGGTGTTTGTGGGAGATACCTTCTGTTACTTTGCT GGCATGACCTTTGCCGTGGTGGGCATCTTGGGACA CTTCAGCAAGACCATGCTACTATTCTTCATGCCCC AGGTGTTCAACTTCCTCTACTCACTGCCTCAGCTC CTGCATATCATCCCCTGCCCTCGCCACCGCATACC CAGACTCAATATCAAGACAGGCAAACTGGAGATG AGCTATTCCAAGTTCAAGACCAAGAGCCTCTCTTT CTTGGGCACCTTTATTTTAAAGGTGGCAGAGAGCC TCCAGCTGGTGACAGTACACCAGAGTGAGACTGA AGATGGTGAATTCACTGAATGTAACAACATGACCC TCATCAACTTGCTACTTAAAGTCCTTGGGCCCATA CATGAGAGAAACCTCACATTGCTCCTGCTGCTGCT GCAGATCCTGGGCAGTGCCATCACCTTCTCCATTC GATATCAGCTCGTTCGACTCTTCTATGATGTCTGA CAGTAAAGGTGGATACGGATCCGAATTCGAGCTC CGTCGACAAGCTT	MGHHHHHHSSGVDLGTENLYFQSM WAFSELPMPLLINLIVSLLGFVATVTLI PAFRGHFIAARLCGQDLNKTSRQQIPE SQGVISGAVFLIILFCFIPFPFLNCFVKE QCKAFPHHEFVALIGALLAICCMIFLG FADDVLNLRWRHKLLLPTAASLPLLM VYFTNFGNTTIVVPKPFRPILGLHLDL GILYYVYMGLLAVFCTNAINILAGING LEAGQSLVISASIIVFNLVELEGDCRD DHVFSLYFMIPFFFTTLGLLYHNWYPS RVFVGDTFCYFAGMTFAVVGILGHFS KTMLLFFMPQVFNFLYSLPQLLHIIPC PRHRIPRLNIKTGKLEMSYSKFKTKSL SFLGTFILKVAESLQLVTVHQSETEDG EFTECNNMTLINLLLKVLGPIHERNLT LLLLLLQILGSAITFSIRYQLVRLFYDV
His-ALG13	N-terminal 6x His Tag with TEV cleavage site	ATGGGCCACCATCATCATCATCATTCTTCTGGTGT AGATCTGGGTACCGAGAACCTGTACTTCCAATCCA TGAAGTGCGTGTTTGTTACCGTAGGGACCACCAGC TTTGACGACCTCATTGCGTGTGTGTCGGCGCCCGA CAGTCTGCAAAAAATCGAGAGCCTTGGTTACAACC GACTTATCCTGCAAATTGGTAGAGGAACGGTGGTA CCTGAACCCTTCAGTACTGAGTCGTTTACTCTGGA TGTTTACAGGTACAAGGATTCCTTGAAAGAAGACA TTCAGAAAGCAGATCTTGTTATTAGTCACGCAGGT GCAGGAAGCTGTTTGGAGACTCTGGAAAAAGGAA AGCCACTCGTAGTGGTTATAAACGAAAAGTTGATG AACAATCATCAGCTGGAACTGGCAAAGCAGCTAC ACAAAGAGGGTCATCTCTTCTATTGTACCTGCAGC ACGCTTCCTGGGCTGTTACAGTCAATGGACTTATC AACACTGAAATGTTATCCTCCTGGCCAGCCAGAAA AATTTTCTGCATTTTTGGATAAAGTTGTTGGATTAC AAAAATGACAGTAAAGGTGGATACTCTAGAGCCT GCAGTCTCG	MGHHHHHHSSGVDLGTENLYFQSMK CVFVTVGTTSFDDLIACVSAPDSLQKI ESLGYNRLILQIGRGTVVPEPFSTESFT LDVYRYKDSLKEDIQKADLVISHAGA GSCLETLEKGKPLVVVINEKLMNNHQ LELAKQLHKEGHLFYCTCSTLPGLLQ SMDLSTLKCYPPGQPEKFSAFLDKVV GLQK

ALG13	Untagged	ATGAAGTGCGTGTTTGTTACCGTAGGGACCACCAG CTTTGACGACCTCATTGCGTGTGTGTCGGCGCCCG ACAGTCTGCAAAAAATCGAGAGCCTTGGTTACAA CCGACTTATCCTGCAAATTGGTAGAGGAACGGTGG TACCTGAACCCTTCAGTACTGAGTCGTTTACTCTG GATGTTTACAGGTACAAGGATTCCTTGAAAGAAG ACATTCAGAAAGCAGATCTTGTTATTAGTCACGCA GGTGCAGGAAGCTGTTTGGAGACTCTGGAAAAAG GAAAGCCACTCGTAGTGGTTATAAACGAAAAGTT GATGAACAATCATCAGCTGGAACTGGCAAAGCAG CTACACAAAGAGGGTCATCTCTTCTATTGTACCTG CAGCACGCTTCCTGGGCTGTTACAGTCAATGGACT TATCAACACTGAAATGTTATCCTCCTGGCCAGCCA GAAAAATTTTCTGCATTTTTGGATAAAGTTGTTGG ATTACAAAAATGAGCAGAGAACCTCTACTTCCAAT CGGATTACAAGGATGACGACGATAAGTGATCTAG A	MKCVFVTVGTTSFDDLIACVSAPDSL QKIESLGYNRLILQIGRGTVVPEPFSTE SFTLDVYRYKDSLKEDIQKADLVISH AGAGSCLETLEKGKPLVVVINEKLMN NHQLELAKQLHKEGHLFYCTCSTLPG LLQSMDLSTLKCYPPGQPEKFSAFLD KVVGLQK
His-ALG14	N-terminal 6x His Tag with TEV cleavage site	ATGGGCCACCATCATCATCATCATTCTTCTGGTGT AGATCTGGGTACCGAGAACCTGTACTTCCAATCCA TGGTGTGCGTTCTCGTTCTAGCTGCGGCCGCAGGA GCTGTGGCGGTTTTCCTAATCCTGCGAATATGGGT AGTGCTTCGTTCCATGGACGTTACGCCCCGGGAGT CTCTCAGTATCTTGGTAGTGGCTGGGTCCGGTGGG CATACCACTGAGATCCTGAGGCTGCTTGGGAGCTT GTCCAATGCCTACTCACCTAGACATTATGTCATTG CTGACACTGATGAAATGAGTGCCAATAAAATAAA TTCTTTTGAACTAGATCGAGCTGATAGAGACCCTA GTAACATGTATACCAAATACTACATTCACCGAATT CCAAGAAGCCGGGAGGTTCAGCAGTCCTGGCCCT CCACCGTTTTCACCACCTTGCACTCCATGTGGCTCT CCTTTCCCCTAATTCACAGGGTGAAGCCAGATTTG GTGTTGTGTAACGGACCAGGAACATGTGTTCCTAT CTGTGTATCTGCCCTTCTCCTTGGGATACTAGGAA TAAAGAAAGTGATCATTGTCTACGTTGAAAGCATC TGCCGTGTAGAAACGTTATCCATGTCCGGAAAGAT TCTGTTTCATCTCTCAGATTACTTCATTGTTCAGTG GCCGGCTCTGAAAGAAAAGTATCCCAAATCGGTG TACCTTGGGCGAATTGTTTGACAGTAAAGGTGGAT ACTCTAGAGCCTGCAGTCTCG	MGHHHHHHSSGVDLGTENLYFQSMV CVLVLAAAAGAVAVFLILRIWVVLRS MDVTPRESLSILVVAGSGGHTTEILRL LGSLSNAYSPRHYVIADTDEMSANKI NSFELDRADRDPSNMYTKYYIHRIPRS REVQQSWPSTVFTTLHSMWLSFPLIH RVKPDLVLCNGPGTCVPICVSALLLGI LGIKKVIIVYVESICRVETLSMSGKILF HLSDYFIVQWPALKEKYPKSVYLGRI V

Experimental and Characterisation of the polyisoprenol panel (used in Chapter 3, 4 and 5), as written and performed by Dr Andrew Giltrap:

Solvents and chemicals

Chemicals and solvents were purchased from commercial suppliers in the United Kingdom (Acros, Alfa Aesar, Carbosynth, Fisher Scientific, Fluorochem, Sigma Aldrich, VWR) and used as received unless stated otherwise. Unless available as in-house supply, anhydrous solvents were purchased from Sigma Aldrich and stored under argon in a septum-capped bottle. All solvents were either analytical or HPLC grade. Reactions requiring anhydrous conditions were carried out with flame-dried reaction vessels under an inert atmosphere of nitrogen or argon.

Chromatography

Thin layer chromatography (TLC) was performed on Merck Millipore aluminium TLC plate, silica gel coated with fluorescent indicator F254 and visualised under UV irradiation at 254 nm, or by staining with aqueous potassium permanganate or p-anisaldehyde dip followed by subsequent drying with a heat gun. Flash column chromatography was performed using Geduran® Silica Gel 60 (0.040-0.063 mm) or The Teledyne ISCO CombiFlash NextGen 100 Flash Chromatography System.

Nuclear Magnetic Resonance (NMR) spectrometry

All NMR spectra were recorded in commercially available deuterated solvents on a Bruker AVIIIHD 400 nanobay (^{1}H at 400, ^{13}C at 101, ^{31}P 162) or a Bruker Avance Neo 600 with BBO cryoprobe (^{1}H at 601, ^{13}C at 151, ^{31}P 243). All NMR data were processed using Mestrenova 193 v.14.1.0. Chemical shifts are quoted in parts per million (ppm) relative to the centre of the solvent peak for ^{1}H and ^{13}C spectral data. All coupling constants are reported in Hz to the nearest 0.1 Hz. ^{13}C spectra are ^{1}H decoupled and ^{31}P-^{13}C heteronuclear coupling on ^{13}C spectra are reported.

LRMS

Low resolution mass spectra were recorded on an Agilent 6120 Quadrupole spectrometer equipped with an electrospray ion source. Methanol or acetonitrile were used as the carrier solvents.

HRMS

High resolution mass spectra were recorded on a Waters Xevo G2-S QTof in negative mode equipped with an electrospray ion source. Substrates were dissolved in a mixture of DCM/MeOH (~3:1) and were infused directly into the source. Lockspray consisted of leuenkephalin in water/H_2O (1:1).

A note on lipid lengths

The lipids extracted from Ginkgo leaves and Bay leaves are distributions of lengths. In the case of Gingko the major length is C90, with C80, C85, C95 all present to a degree. In the case of Bay leaves, C55 is the major length with C50 and C60 also present. As such, molecular weights have been calculated on the most abundant lipid length C90 for Gingko derived lipids and C55 for Bay leaf derived lipids.

NMR considerations

NMRs of all lipid phosphates and lipid pyrophosphates were carried out in a mixture of MeOD and $CDCl_3$ (1:9 to 1:3) depending on the sample. All samples were locked and shimmed to MeOD. The amphiphilic nature of these compounds means they exhibit very broad NMR spectra, presumably due to formation of micelles and higher order structures, which makes interpretation very difficult. As such integrals are approximate. Importantly, all ^{31}P peaks were observed.

General Procedures

Lipid Phosphate Formation (General Procedure A)

To a vigorously stirred solution of 5-ethylthiotetrazole (2.7 equiv.) and lipid alcohol (1 equiv.) in THF (0.024 M wrt to lipid) was added bis(2-cyanoethyl) diisopropylphosphoramidite (3 equiv.). The solution was allowed to stir for 2 h and monitored by TLC. If SM remained extra bis(2-cyanoethyl) diisopropylphosphoramidite (0.8 equiv.) was added and the reaction continued to be monitored by TLC. Once complete, the mixture was cooled to -30 °C and hydrogen peroxide (2.5 mL per mmol of lipid alcohol, 30% aqueous) was added dropwise. The mixture was then allowed to warm to room temperature and stirred for 2 h. The solution was then cooled to 0 °C and quenched slowly with saturated aqueous Na_2SO_3. The mixture was then diluted with EtOAc and $NaHCO_3$, then the aqueous phase was rextracted with EtOAc (2x). The combined organics were then washed with brine, then dried over Na_2SO_4, filtered and concentrated *in vacuo*.

To a solution of the crude protected phosphate in MeOH + THF (5:2, 0.034 M wrt to lipid) was added a solution of NaOMe in MeOH (0.83 mL per mmol substrate, 25 wt%). The mixture was allowed to stir for 16 h at which point TLC indicated a small amount of SM present and so extra NaOMe (0.1 mL) was added and the mixture was stirred for a further 3 h at which point TLC showed consumption of SM. The solution was then neutralised with excess Dowex(NH_4^+), filtered and concentrated *in vacuo.* The crude product was then purified by flash chromatography (9:1:0:0.1 → 80:20:2:0.1 → 65:25:5:0.1 $CHCl_3$:MeOH:H_2O:NH_4OH solution) to yield the lipid phosphate as a sticky solid.

GlcNAc(OAc)$_3$-PP-lipid Pyrophosphate Formation Conditions (General Procedure B)

To a solution of GlcNAc(OAc)$_3$-P (1.4 equiv) in DMF (0.12 M) was added CDI (7 equiv, 5 equiv cf to GlcNAc-P). The mixture was allowed to stir for 2 h, before MeOH (2 eq. per excess of CDI i.e. 11.2 eq.) was added and stirred for 30 min. A solution of Lipid-P (1 equiv) in DCM (0.08 M) and DMF (minimal) was added. The solution then went cloudy and so extra DCM may be added. To the solution was then added 5-ethyl thioterazole (1 equiv) and a further portion of DCM (minimal) to ensure solubility. The mixture was allowed to stir at RT for 5-14 days before being concentrated with a stream of N_2 overnight. The crude product was then purified by flash chromatography (80:20:2:0.1, $CHCl_3$:MeOH:H_2O:NH_4OH (25% sol'n)) and concentrated *in vacuo* to yield the panel of lipid-PP-GlcNAc(OAc)$_3$ as sticky solids.

GlcNAc(OAc)$_3$-PP-lipid OAc Deprotection Conditions (General Procedure C)

To a solution of protected lipid-PP-GlcNAc(OAc)$_3$ (1 equiv) in a mixture of MeOH:THF (0.01 M, 1:1) was added NaOMe (0.5 equiv, 1 wt% in MeOH solution). The mixture was then stirred at RT for 40 min before being treated with Dowex resin (NH_4^+) and stirred for about 20 min before being filtered and concentrated *in vacuo*. TLC analysis shows that there is some decomposition to the lipid-P as well as formation of the product. The crude product was then purified by flash chromatography (4:2:1, EtOAC:iPrOH:H_2O) and concentrated *in vacuo* to yield the panel of lipid-PP-GlcNAc as sticky solids.

Synthesis of α-GlcNAc-1-PO_3^-

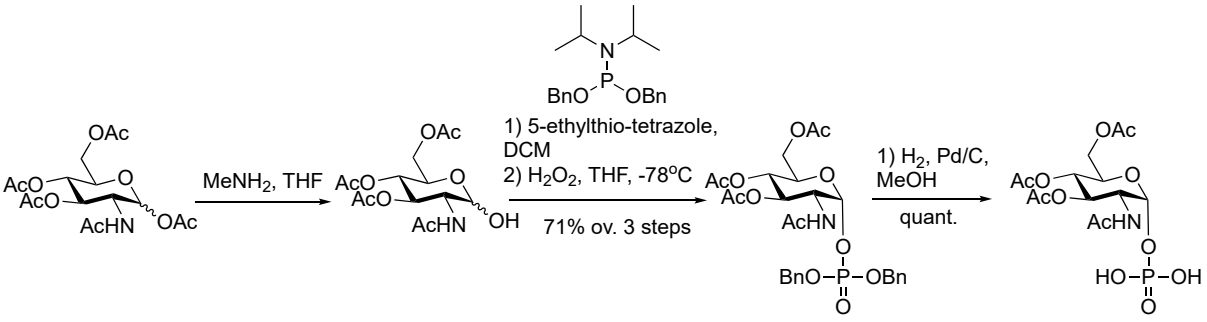
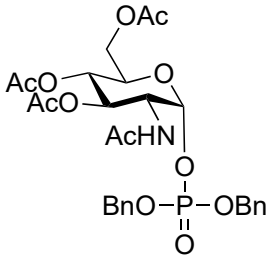


Figure 1. Synthesis of α-GlcNAc-1-PO_3^-

α-$PO(OBn)_2$-GlcNAc$(OAc)_3$



To a solution of perOAc-GlcNAc (2.0 g, 5.14 mmol) in THF (25 mL) was added 8M $MeNH_2$ in EtOH (1.28 mL, 10.3 mmol). The solution was stirred at room temperature for 2 hours and then concentrated *in vacuo* to yield the lactol which was used without purification.

To a vigorously stirred solution of 5-ethylthiotetrazole (3.14 g, 24.2 mmol) and bis(benzyl) diisopropylphosphoramidite (5.17 mL, 15.42 mmol) in dichloromethane (25 mL) was added crude lactol (5.14 mmol) in dichloromethane (25 mL). The solution was allowed to stir for 2 h before being diluted with dichloromethane (150 mL) and washed with saturated aqueous $NaHCO_3$ (1 x 100 mL), water (1 x 100 mL) and brine (1 x 100 mL). The organic layer was then dried over Na_2SO_4, filtered and concentrated *in vacuo*.

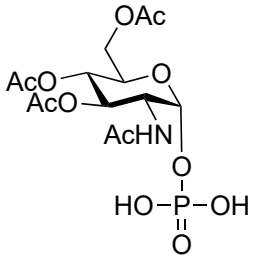
To a solution of crude phosphite in THF (100 mL) at -78 °C was added hydrogen peroxide (8.5 mL, 30% aqueous). The mixture was then allowed to warm to room temperature and stirred for 3 h. The solution was then cooled to 0 °C and quenched slowly with saturated aqueous Na_2SO_3 (10 mL). The mixture was then diluted with EtOAc (200 mL), washed with saturated aqueous $NaHCO_3$ (1 x 1000 mL), water (1 x 100 mL) and brine (1 x 100 mL). The organic layer was then dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude product was then purified by flash chromatography (1:49 →1:24 MeOH:DCM) to yield the benzyl phosphate as a white solid (2.22 g, 71%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.41 – 7.29 (m, 8H), 7.30 – 7.22 (m, 2H), 5.87 (d, *J* = 9.2 Hz, 1H), 5.66 (dd, *J* = 6.1 (C-P), 3.3 Hz, 1H), 5.21 – 5.00 (m, 6H), 4.43 – 4.32 (m, 1H), 4.12 (dd, *J* = 12.5, 4.0 Hz, 1H), 3.99 (ddd, *J* = 9.9, 4.0, 2.2 Hz, 1H), 3.92 (dd, *J* = 12.5, 2.3 Hz, 1H), 2.01 (s, 3H), 1.99 (s, 6H) 1.71 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 171.2, 170.6, 170.4, 169.2, 135.4 (d, *J* = 6.3 Hz), 135.3 (d, *J* = 6.3 Hz), 129.1, 128.90 (d, *J* = 4.1 Hz), 128.4, 128.2 (d, *J* = 4.1 Hz), 127.8, 96.4 (d, *J* = 6.7 Hz), 70.2, 70.1, 70.1 (d, *J* = 4.7 Hz) 69.8, 67.5, 61.4, 51.9 (d, *J* = 7.8 Hz), 22.8, 20.7, 20.7, 20.6.

In agreement with ref[1]

α-GlcNAc-1-PO_3^-



To a solution of benzyl phosphate (2.20 g, 3.6 mmol) in MeOH (40 mL) was added Pd/C (10 wt%, 396 mg, 0.36 mmol). The mixture was then evacuated and then immediately placed under nitrogen. This was repeated 3 times, before the solution was evacuated and then placed under an atmosphere of hydrogen, before being stirred for 16 h. The mixture was the evacuated and placed under nitrogen, and this was repeated, before the mixture was filtered over celite and concentrated *in vacuo* to give the desired product as a white foam (1.54 g, quant.).

^{1}H NMR (400 MHz, MeOD) δ 5.58 (dd, *J* = 6.6 (C-P), 3.4 Hz, 1H), 5.29 (dd, *J* = 10.9, 9.3 Hz, 1H), 5.10 (t, *J* = 9.8 Hz, 1H), 4.33 – 4.27 (m, 2H), 4.24 (ddd, *J* = 10.3, 3.9, 2.2 Hz, 1H), 4.13 (dd, *J* = 12.2, 2.2 Hz, 1H), 2.06 (s, 3H), 2.02 (s, 3H), 1.99 (s, 3H), 1.95 (s, 3H).

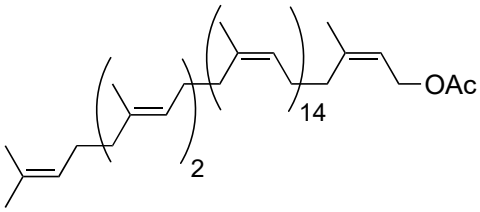
^{13}C NMR (101 MHz, MeOD) δ 173.7, 172.4, 171.9, 171.2, 95.9 (d, *J* = 6.2 Hz), 71.6, 70.3, 69.7, 62.8, 53.1 (*J* = 8.7), 22.4, 20.6, 20.6, 20.6.

^{31}P NMR (162 MHz, MeOD) δ -2.2.

In agreement with ref[1]

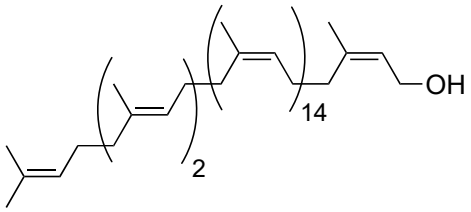
Lipid Synthesis

Gingko extract



Dried *Gingko biloba* leaves (250 g) purchased from Amazon were suspended in a mixture of acetone:petroleum ether (1 L, 1:1) and stirred for 15 days at room temperature. Three independent batches were extracted side-by-side, and all three were then combined for subsequent work up and purification. The crude solutions were filtered over celite and concentrated *in vacuo*. The crude residue was the initially purified via a vacuum silica column (petroleum ether → 95:5 Petroleum ether:Et_2O) and 6 g of crude product was obtained. Following this the partially purified product was further purified using the Teledyne combilfash system, with a gradient of petroleum ether → 95:5 Petroleum ether:Et_2O using an 80 g column at 60 $mLmin^{-1}$. This was repeated a second time to yield the desired C90 polyisoprenol acetates (4.3 g).

C90-polyunsaturated

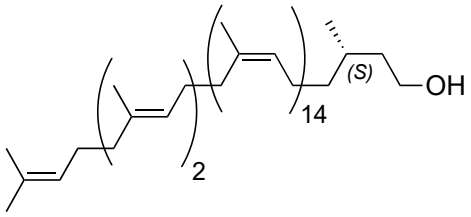


To a solution of the above acetates (4.28 g, 3 mmol) (assuming average MW of lipids = 1201) in THF (20 mL) and MeOH (10 mL) was added K_2CO_3 (2.07 g, 15 mmol). The mixture was left to stir for 16 h before being diluted with petroleum ether (200 mL) and the organic phase washed with H_2O (100 mL) and brine (100 mL) before being dried over Na_2SO_4, filtered and concentrated *in vacuo* and then purified by flash chromatography (0 → 20% Et_2O:petroleum ether on a 80 g column over 15 mins) to yield the pure C90 polyisoprenol as a yellow oil which solidifies on freezing (2.50 g, 72 %).

^{1}H NMR (601 MHz, $CDCl_3$) δ 5.45 (td, *J* = 7.2, 1.6 Hz, 1H), 5.16 – 5.06 (m, 18H), 4.09 (d, *J* = 7.2 Hz, 2H), 2.12 – 1.95 (m, 74H), 1.75 (s, 3H), 1.68 (m, 45H), 1.61 (s, 3H), 1.61 – 1.59 (m, 6H).

^{13}C NMR (151 MHz, $CDCl_3$) δ 140.0, 136.2, 135.5, 135.5, 135.4, 135.4, 135.3, 135.3, 135.1, 125.2, 125.1, 125.1, 125.0, 124.7, 124.6, 124.5, 124.4, 124.3, 59.1, 39.9, 39.9, 32.4, 32.3, 32.3, 32.1, 26.9, 26.8, 26.6, 26.5, 26.5, 26.4, 25.8, 23.6, 23.6, 23.5, 17.8, 16.1, 16.1.

Dolichols



Dry *n*-propanol (from Sigma Aldrich under suba-seal) was dried overnight with 4 Å molecular sieves (pre-activated in furnace at 350 °C). Then undecaprenol (300 mg, 0.25 mmol) was dissolved in toluene in a 2-neck flask and rotavaped to dryness and then dried on high-vac for 2 h. Then a dry stir bar followed by (*S*)-Tol-binap $RuCl_2$ (*p*-cymene) (28 mg, 0.025 mmol) and KOH (2.8 mg, 0.05 mmol) were added to the flask and it was immediately vac → N_2 → vac and left for 2 h. The dry *n*-propanol weas then sparged for 30 min using a balloon of Argon and then 25 mL was added to the 2 neck flask containing the other components under N_2 and then this was again dried by (vac → N_2) x 3 before being left to stir at rt for 16 h. The solution was then concentrated *in vacuo* and the crude product was then purified by flash chromatography (0 → 10% EtOAc:petroleum ether on a 12 g column over 10 mins) to yield the to yield the C90 dolichols as a colourless liquid (103 mg, 33%) + impure aldehyde (176 mg, 59%)

To a solution of aldehyde (176 mg, 0.146 mmol) in THF (2 mL) and MeOH (1 mL) was added $NaBH_4$ (12.5 mg, 0.293 mmol) and stirred at room temperature for 30 mins. The mixture was the concentrated *in vacuo* before being diluted with HCl (10 mL, 0.1 M) and washed with DCM (3 x 10 mL) before being dried over Na_2SO_4, filtered and concentrated *in vacuo.* The crude material was then purified by flash chromatography (0 → 10% EtOAc:petroleum ether on a 12 g column over 10 mins) to yield the to yield the C90 dolichols as a colourless liquid (85 mg, 48%). This was combined with the product to yield the desired C90 dolichols overall in 65% yield.

C90 polyisoprenol (300 mg, 0.25 mmol) was dissolved in toluene in a 2-neck flask and then rotavaped to dryness, and left on high vac for 3 h. (*S*)-Tol-binap $RuCl_2$ (*p*-cymene) (28 mg, 0.025 mmol) and KOH (2.8 mg, 0.05 mmol) were added to the flask and it was immediately vac → N_2 → vac and left for 3 h. Dried and purged *n*-propanol (5 mL) was then added and the solution was degassed by vac → N_2 (x 4). The solution was then heated to 80 °C and stirred for 2.5 h. NMR demonstrated 1:1 ratio of SM:product. An extra portion of catalyst (28 mg, 0.025 mmol) and KOH (2.8 mg, 0.05 mmol) were added followed

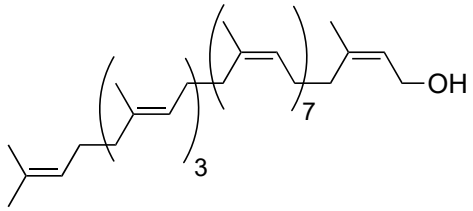
by degassing the solvent and heating at 80 °C for 3 h before analysis by NMR. A third portion of catalyst and KOH were then added and the solution stirred at RT overnight which led to complete conversion to the product. The solution was then concentrated *in vacuo* and the crude product was then purified by flash chromatography (0 → 10% EtOAc:petroleum ether) to yield the C90 dolichols as a colourless liquid (166 mg, 55%).

^{1}H NMR (601 MHz, $CDCl_3$) δ 5.12 (q, *J* = 7.9 Hz, 20H), 3.75 – 3.60 (m, 2H), 1.61 (s, 4H), 1.61 – 1.59 (m, 9H), 1.42 – 1.31 (m, 2H), 1.22 – 1.13 (m, 2H), 0.90 (d, *J* = 6.5 Hz, 3H).

^{13}C NMR (151 MHz, $CDCl_3$) δ 135.5, 135.4, 135.4, 135.3, 135.1, 135.1, 131.4, 125.6, 125.2, 125.2, 125.1, 124.5, 124.4, 124.3, 61.4, 40.1, 39.9, 39.9, 37.6, 32.4, 32.3, 32.1, 29.4, 26.9, 26.8, 26.6, 26.5, 25.8, 25.4, 23.6, 23.6, 19.7, 17.8, 16.1, 16.1.

$[\alpha]_D^{25}$ = -0.41 (c = 2.95, $CHCl_3$)

Undecaprenol



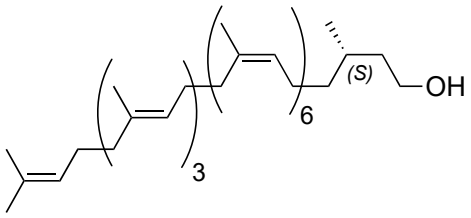
Dried bay leaves (100 g, crushed by hand) were suspended in a large round bottom flask containing acetone (800 mL) and *n*-hexane (200 mL). The suspension was stirred at round temperature for 9 days before being filtered over celite and concentrated *in vacuo*. The residue was then suspended in a mixture of *n*-hexane (100 mL), EtOH (500 mL) and aqueous KOH (70 mL, 15%). The mixture was refluxed at 90 °C for 1 h and then cooled back to room temperature. At this point water (330 mL) and Et_2O (330 mL) were added and shaken vigorously in a large separating funnel and allowed to partition. The Et_2O layer was isolated, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude product was then purified by flash chromatography (0:20→1:19→1:9→3:17 Et_2O:Pet. ether) to yield undecaprenol as a yellow oil which solidified upon freezing (544 mg).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.44 (td, *J* = 7.2, 1.5 Hz, 1H), 5.18 – 5.05 (m, 12H), 4.09 (dt, *J* = 7.2, 0.9 Hz, 2H), 2.15 – 1.95 (m, 49H), 1.74 (q, *J* = 1.1 Hz, 3H), 1.68 (p, *J* = 1.1 Hz, 25H), 1.60 (dt, *J* = 3.1, 1.4 Hz, 14H).

LRMS (ESI) ($M+H^+$(-OH)) C_{55} 749.6, C_{50} 681.5, C_{60} 817.6

In agreement with ref[2]

C55-Dol



Dry *n*-propanol (from Sigma Aldrich under suba-seal) was dried overnight with 4 Å molecular sieves (pre-activated in furnace at 350 °C). Then undecaprenol (355 mg, 0.5 mmol) was dissolved in toluene in a 2-neck flask and rotavaped to dryness and then dried on high-vac for 2 h. Then a dry stir bar followed by (*S*)-Tol-binap $RuCl_2$ (*p*-cymene) (56 mg, 0.05 mmol) and KOH (5.6 mg, 0.1 mmol) were added to the flask and it was immediately vac → N_2 → vac and left for 2 h. The dry *n*-propanol weas then sparged for 30 min using a balloon of Argon and then 25 mL was added to the 2 neck flask containing the other components under N_2 and then this was again dried by (vac → N_2) x 3 before being left to stir at rt for 16 h. The solution was then concentrated *in vacuo* and the crude product was then purified by flash chromatography (0 → 10% EtOAc:petroleum ether on a 12 g column over 10 mins) to yield the to yield the C55 dolichols as a colourless liquid (142 mg, 40%) + impure aldehyde (184 mg, 52%)

To a solution of aldehyde (184 mg, 0.257 mmol) in THF (4 mL) and MeOH (2 mL) was added $NaBH_4$ (25 mg, 0.518 mmol) and stirred at room temperature for 30 mins. The mixture was the concentrated *in vacuo* before being diluted with HCl (20 mL, 0.1 M) and washed with DCM (3 x 20 mL) before being dried over Na_2SO_4, filtered and concentrated *in vacuo.* The crude material was then purified by flash chromatography (0 → 10% EtOAc:petroleum ether on a 12 g column over 10 mins) to yield the to yield the C55 dolichols as a colourless liquid (80 mg, 43%). This was combined with the product to yield the desired C55 dolichols overall in 62% yield.

^{1}H NMR (601 MHz, $CDCl_3$) δ 5.17 – 5.07 (m, 10H), 3.74 – 3.62 (m, 2H), 2.11 – 2.01 (m, 31H), 2.01 – 1.95 (m, 7H), 1.69 – 1.67 (m, 20H), 1.61 (d, *J* = 1.3 Hz, 4H), 1.60 (d, *J* = 2.8 Hz, 10H), 1.43 – 1.30 (m, 2H), 1.17 (dddd, *J* = 13.3, 9.7, 7.7, 5.7 Hz, 1H), 0.90 (d, *J* = 6.5 Hz, 3H).

^{13}C NMR (151 MHz, $CDCl_3$) δ 135.5, 135.4, 135.4, 135.4, 135.3, 135.1, 135.1, 135.0, 131.4, 125.6, 125.2, 125.2, 125.2, 125.1, 124.5, 124.4, 124.4, 124.3, 61.4, 40.1, 39.9, 39.9, 37.6, 32.4, 32.3, 32.1, 29.4, 26.9, 26.8, 26.8, 26.6, 26.5, 25.8, 25.4, 23.6, 23.6, 19.7, 17.8, 16.2, 16.1.

$[\alpha]_D^{25}$ = -0.39 (c = 1.53, $CHCl_3$)

Synthesis of lipid phosphates

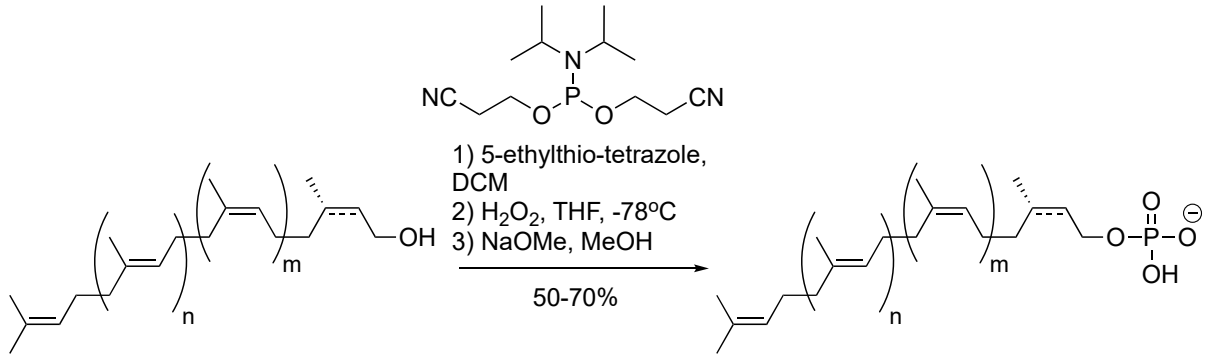
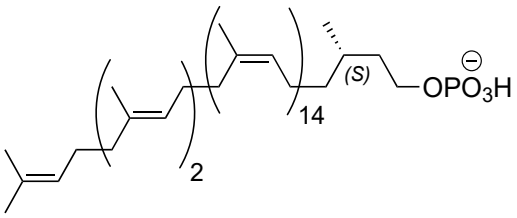


Figure 2. General synthesis of lipid phosphate

Dolichol Phosphate



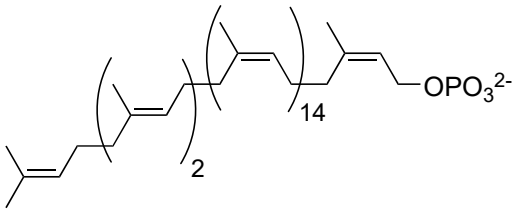
According to *General Procedure A* dolichols (167 mg, 0.139 mmol) was phosphorylated to yield dolichol phosphate as a sticky solid (96 mg, 52%).

^{1}H NMR (601 MHz, 9:1 $CDCl_3$:MeOD, lock MeOD) δ 5.10 – 4.99 (m, 16H), 2.07 – 1.85 (m, 66H), 1.61 (s, 42H), 1.54 (d, *J* = 1.2 Hz, 3H), 1.53 (t, *J* = 1.6 Hz, 6H).

^{31}P NMR (243 MHz, 9:1 $CDCl_3$:MeOD, lock MeOD) δ 2.83.

HRMS (ESI-); (*m/z*) calc. for $C_{90}H_{148}O_4P$: 1324.1120 $[M-H]^-$; found: 1324.1166 (for C90)

C90-poly-Phosphate



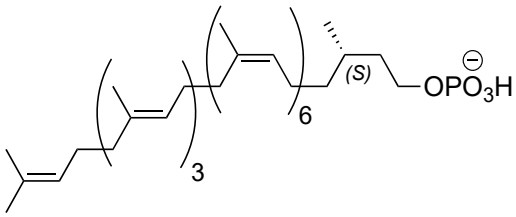
According to *General Procedure A* C90-polyisoprenol (300 mg, 0.25 mmol) was phosphorylated to yield the C90 polyisoprenol phosphate as a sticky solid (236 mg, 71 %).

^{1}H NMR (400 MHz, 1:3 MeOD+CDCl$_3$) δ 5.32 (s, 1H), 5.15 – 5.01 (m, 16H), 2.09 – 1.90 (m, 62H), 1.64 (d, *J* = 1.6 Hz, 41H), 1.57 (d, *J* = 1.3 Hz, 3H), 1.56 (s, 6H).

^{31}P NMR (162 MHz, 1:3 MeOD+CDCl$_3$) δ 0.5.

HRMS (ESI-); (*m/z*) calc. for $C_{90}H_{146}O_4P$: 1322.0964 [M-H]$^-$; found: 1322.1028 (for C90)

C55-dol-Phosphate



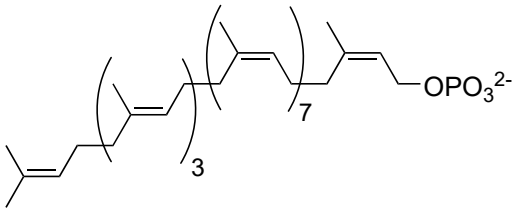
According to *General Procedure A* C55 dolichols (170 mg, 0.24 mmol) was phosphorylated to yield C55 dolichol phosphate as a sticky solid (128 mg, 62%).

^{1}H NMR (400 MHz, 1:3 MeOD+CDCl$_3$) δ 5.17 – 4.98 (m, 12H), 3.82 (d, *J* = 39.5 Hz, 2H), 2.13 – 1.89 (m, 42H), 1.64 (d, *J* = 3.0 Hz, 22H), 1.57 (s, 2H), 1.56 (s, 12H).

^{31}P NMR (162 MHz, 1:3 MeOD+CDCl$_3$) δ 1.1.

HRMS (ESI-); (*m/z*) calc. for $C_{55}H_{92}O_4P$: 847.6738 [M-H]$^-$; found: 847.6774 (for C55)

Undecaprenol Phosphate



To a vigorously stirred solution of 5-ethylthiotetrazole (439.5 mg, 3.38 mmol) and bis(2-cyanoethyl) diisopropylphosphoramidite (570 μL, 2.18 mmol) in dichloromethane (7.65 mL) was added undecaprenol (544 mg, 0.765 mmol) in dichloromethane (7.65 mL). The solution was allowed to stir for 2 h before being diluted with dichloromethane (30 mL) and washed with saturated aqueous $NaHCO_3$ (1 x 20 mL), water (1 x 20 mL) and brine (1 x 20 mL). The organic layer was then dried over Na_2SO_4, filtered and concentrated *in vacuo*.

To a solution of crude phosphite in THF (15.3 mL) at -78 °C was added hydrogen peroxide (1.71 mL, 30% aqueous). The mixture was then allowed to warm to room temperature and stirred for 2 h. The solution was then cooled to 0 °C and quenched slowly with saturated aqueous Na_2SO_3 (5 mL). The mixture was then diluted with EtOAc (50 mL), washed with saturated aqueous $NaHCO_3$ (1 x 20 mL), water (1 x 20 mL) and brine (1 x 20 mL). The organic layer was then dried over Na_2SO_4, filtered and concentrated *in vacuo*.

To a solution of the crude protected phosphate in MeOH (13.8 mL) was added a solution of NaOMe in MeOH (0.568 mL, 25 wt%). The mixture was allowed to stir for 16 h at which point TLC indicated a small amount of SM present. A small amount of toluene (2 mL) and NaOMe (0.1 mL) was added and the mixture was stirred for a 3 h at which point TLC showed consumption of SM. The solution was then neutralised with excess Dowex(NH_4^+), filtered and concentrated *in vacuo.* The crude product was then purified by flash chromatography (9:1:0:0.1 → 65:25:5:0.1 $CHCl_3$:MeOH:H_2O:NH_4OH solution) to yield undecaprenol phosphate as a yellow oil (327 mg, 50%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.38 (t, *J* = 6.4 Hz, 1H), 5.17 – 5.04 (m, 12H), 4.43 (dd, *J* = 6.8 (P-H), 6.8 Hz, 2H), 2.14 – 1.92 (m, 46H), 1.72 – 1.57 (m, 39H).

^{31}P NMR (162 MHz, $CDCl_3$) δ 0.50.

HRMS (ESI-); (*m/z*) calc. for $C_{55}H_{90}O_4P$: 845.6582 $[M-H]^-$; found: 845.6601 (for C55)

In agreement with ref[2]

Synthesis of lipid-PP-GlcNAc

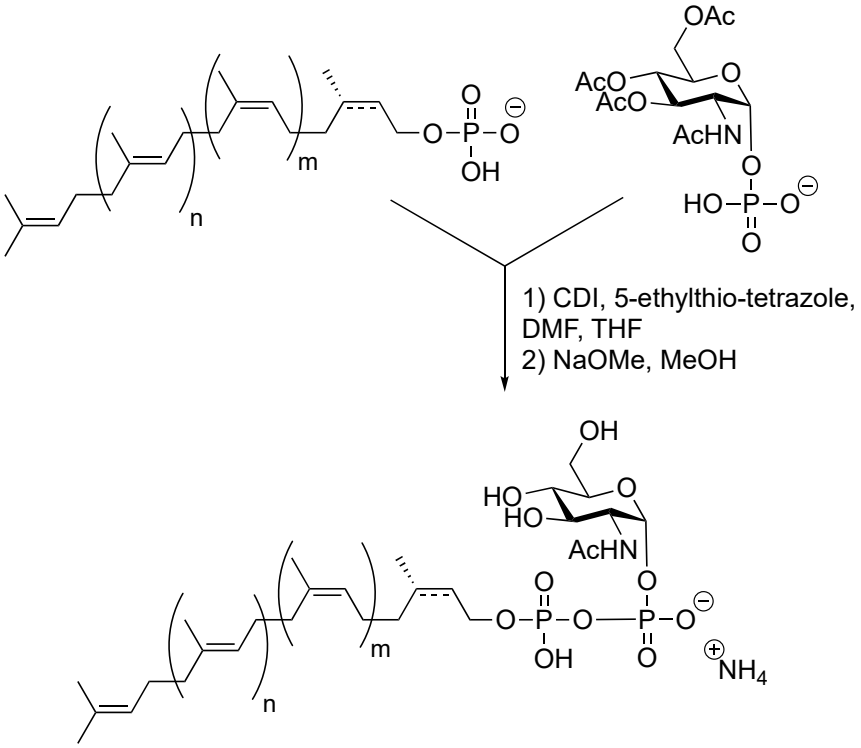
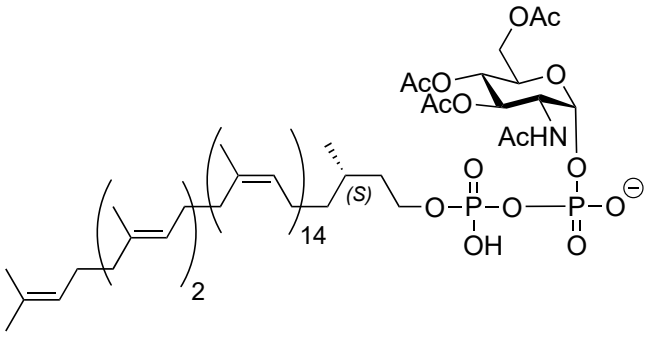


Figure 3. General Synthesis of Lipid-PP-GlcNAc

Dol-PP-GlcNAc(OAc)$_3$



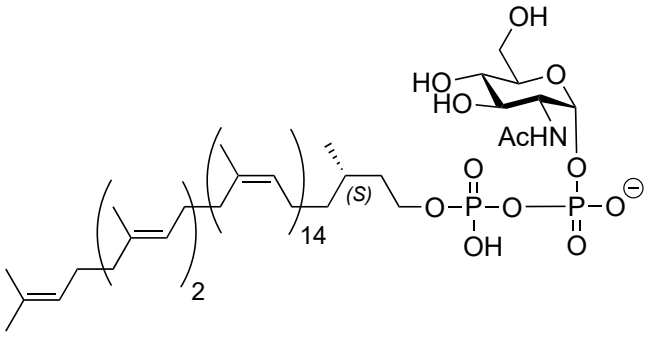
According to *General Procedure B* dolichols (35 mg, 0.026 mmol) was coupled to yield dolichol pyrophosphate as a sticky solid (37.2 mg, 83%).

^{1}H NMR (601 MHz, 1:3 MeOD+CDCl$_3$) δ 5.60 (s, 1H), 5.24 (dd, *J* = 7.6, 4.7 Hz, 1H), 5.15 – 4.93 (m, 17H), 4.31 – 4.14 (m, 3H), 4.12 – 4.03 (m, 1H), 3.96 – 3.85 (m, 2H), 2.10 – 1.87 (m, 78H), 1.65 – 1.58 (m, 46H), 1.54 (d, *J* = 1.3 Hz, 3H), 1.54 – 1.52 (m, 6H), 1.52 – 1.45 (m, 1H), 1.39 – 1.31 (m, 1H), 1.30 – 1.24 (m, 1H), 1.08 (tdd, *J* = 15.5, 7.0, 4.0 Hz, 1H), 0.82 (d, *J* = 6.6 Hz, 3H).

^{31}P NMR (243 MHz, MeOD) δ -11.4, -13.3.

HRMS (ESI-); (*m/z*) calc. for $C_{104}H_{167}NO_{15}P_2$: 866.5928 $[M-2H]^{2-}$; found: 866.5955 (for C90)

Dol-PP-GlcNAc



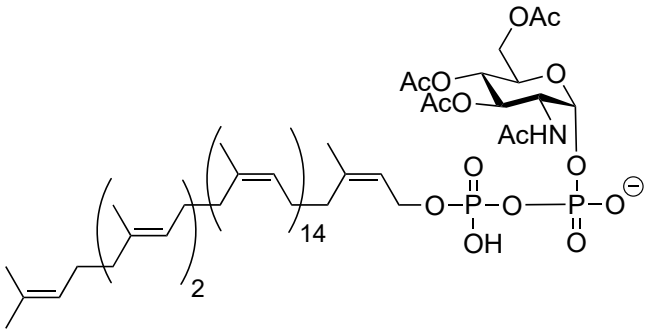
According to *General Procedure C* dolichol-PP-GlcNAc(OAc)$_3$ (35 mg, 0.022 mmol) was deprotected to yield pyrophosphate as a sticky solid (18 mg, 56%).

^{1}H NMR (400 MHz, 1:3 MeOD+CDCl$_3$) δ 5.55 (s, 1H), 5.50 (s, 1H), 5.07 (m, 16H), 3.95 (s, 4H), 3.71 – 3.60 (m, 1H), 3.54 (d, *J* = 10.7 Hz, 1H), 3.27 – 3.15 (m, 1H), 2.71 (s, 2H), 2.09 – 1.90 (m, 63H), 1.65 – 1.62 (m, 40H), 1.57 (d, *J* = 1.3 Hz, 2H), 1.55 (s, 6H), 0.86 (d, *J* = 6.7 Hz, 3H).

^{31}P NMR (162 MHz, 1:3 MeOD+CDCl$_3$) δ -10.1, -12.4.

HRMS (ESI-); (*m/z*) calc. for $C_{98}H_{161}NO_{12}P_2$: 803.5769 $[M-2H]^{2-}$; found: 803.5750 (for C90)

C90-polyisoprenol-PP-GlcNAc(OAc)$_3$



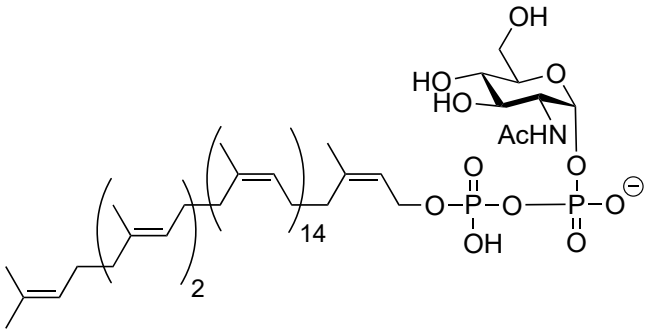
According to *General Procedure B* C90 polyisoprenol (50 mg, 0.046 mmol) was coupled to yield C90 polyisoprenol pyrophosphate as a sticky solid (39.5 mg, 50%).

^{1}H NMR (601 MHz, 1:3 MeOD+CDCl$_3$) δ 5.60 (s, 1H), 5.30 (t, *J* = 6.9 Hz, 1H), 5.24 (d, *J* = 6.8 Hz, 0H), 5.12 – 4.96 (m, 20H), 4.41 (t, *J* = 7.4 Hz, 2H), 4.30 – 4.16 (m, 3H), 4.11 – 4.02 (m, 1H), 2.07 – 1.88 (m, 81H), 1.66 (s, 3H), 1.61 (q, *J* = 2.1 Hz, 48H), 1.60 (s, 3H), 1.54 (d, *J* = 1.3 Hz, 3H), 1.53 (t, *J* = 1.7 Hz, 6H).

^{31}P NMR (243 MHz, 1:3 MeOD+CDCl$_3$) δ -11.4, -13.3.

HRMS (ESI-); (*m/z*) calc. for $C_{104}H_{165}NO_{15}P_2$: 865.5849 $[M-2H]^{2-}$; found: 865.5869 (for C90)

C90-polyisoprenol-PP-GlcNAc



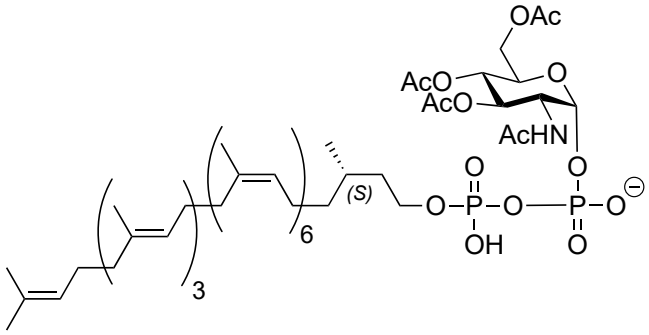
According to *General Procedure C* C90 polyisoprenol-PP-GlcNAc(OAc)$_3$ (37 mg, 0.021 mmol) was deprotected to yield deprotected pyrophosphate as a sticky solid (17.2 mg, 51%).

^{1}H NMR (400 MHz, 1:3 MeOD+CDCl$_3$) δ 5.55 (s, 1H), 5.35 (s, 1H), 5.18 – 4.95 (m, 18H), 4.45 (s, 1H), 3.95 (s, 2H), 3.68 (d, *J* = 13.0 Hz, 1H), 3.60 – 3.49 (m, 1H), 3.23 (s, 1H), 2.70 (s, 1H), 2.09 – 1.94 (m, 69H), 1.70 (s, 2H), 1.66 – 1.62 (m, 48H), 1.57 (d, *J* = 1.3 Hz, 3H), 1.55 (s, 6H).

^{31}P NMR (162 MHz, 1:3 MeOD+CDCl$_3$) δ -10.3, -12.2.

HRMS (ESI-); (*m/z*) calc. for $C_{98}H_{159}NO_{12}P_2$: 802.0674 $[M-2H]^{2-}$; found: 802.0743 (for C90)

C55dol-PP-GlcNAc(OAc)$_3$



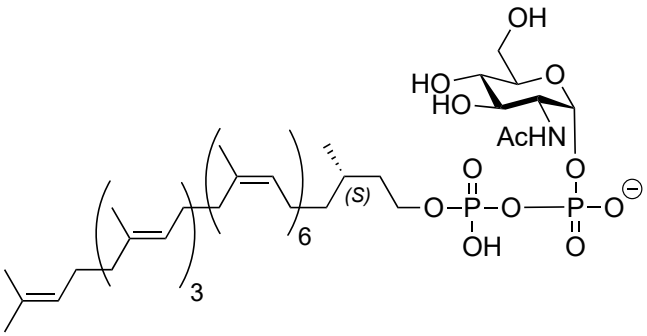
According to *General Procedure B* C55 dolichols (40 mg, 0.046 mmol) was coupled to yield C55 dolichol pyrophosphate as a sticky solid (40 mg, 69%).

^{1}H NMR (601 MHz, 1:3 MeOD+CDCl$_3$) δ 5.59 (s, 1H), 5.24 (dd, *J* = 9.4 Hz, 1H), 5.05 (dt, *J* = 14.9, 10.1 Hz, 11H), 4.32 – 4.14 (m, 3H), 4.12 – 4.03 (m, 1H), 3.91 (s, 2H), 2.07 – 1.87 (m, 47H), 1.63 – 1.60 (m, 20H), 1.60 (s, 3H), 1.54 (s, 3H), 1.53 (d, *J* = 2.8 Hz, 9H), 1.50 – 1.45 (m, 1H), 1.39 – 1.32 (m, 1H), 1.30 – 1.25 (m, 1H), 1.11 – 1.04 (m, 1H), 0.82 (d, *J* = 6.6 Hz, 3H).

^{31}P NMR (243 MHz, 1:3 MeOD+CDCl$_3$) δ -11.4, -13.3.

HRMS (ESI-); (*m/z*) calc. for $C_{69}H_{111}NO_{15}P_2$: 627.8720 $[M-2H]^{2-}$; found: 627.8767 (for C55)

C55dol-PP-GlcNAc



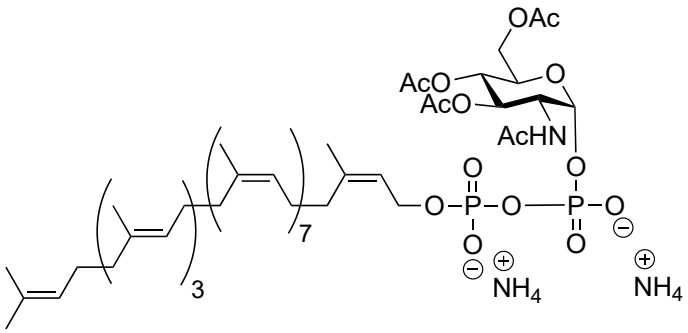
According to *General Procedure C* C55-dolichol-PP-GlcNAc(OAc)$_3$ (38 mg, 0.03 mmol) was deprotected to yield C55 dolichol pyrophosphate as a sticky solid (16.6 mg, 49%).

^{1}H NMR (400 MHz, 1:3 MeOD+CDCl$_3$) δ 5.53 (s, 2H), 5.14 – 5.00 (m, 13H), 3.68 (d, *J* = 12.7 Hz, 1H), 3.55 (s, 1H), 3.22 (s, 1H), 2.71 (s, 1H), 2.00 (qd, *J* = 14.1, 8.7 Hz, 53H), 1.66 – 1.61 (m, 28H), 1.58 – 1.57 (m, 5H), 1.56 (s, 13H), 0.87 (d, *J* = 6.8 Hz, 3H).

^{31}P NMR (162 MHz, 1:3 MeOD+CDCl$_3$) δ -9.9, -11.9.

HRMS (ESI-); (*m/z*) calc. for $C_{63}H_{106}NO_{12}P_2$: 1130.7195 $[M-H]^-$; found: 1130.7224 (for C55)

Und-PP-GlcNAc(OAc)$_3$



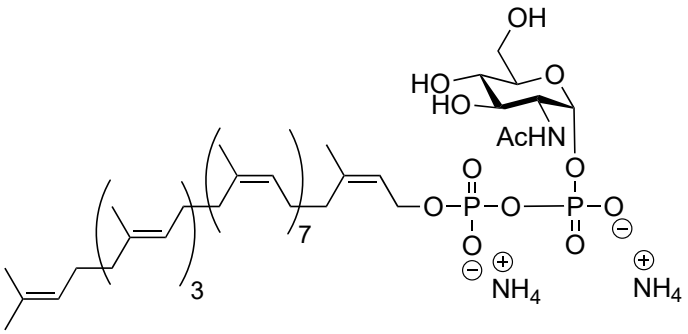
According to *General Procedure B* Und-P (60 mg, 0.070 mmol) was coupled to yield the pyrophosphate as a white solid (74 mg, 84%).

^{1}H NMR (400 MHz, CDCl$_3$+MeOD) δ 5.66 (dd, *J* = 7.8, 3.4 Hz, 1H), 5.37 (dt, *J* = 7.5, 4.1 Hz, 1H), 5.34 – 5.24 (m, 1H), 5.15 – 5.02 (m, 12H), 4.47 (t, *J* = 7.0 Hz, 2H), 4.38 – 4.24 (m, 3H), 4.10 (d, *J* = 11.8 Hz, 1H), 2.12 – 1.92 (m, 58H), 1.71 (s, 3H), 1.68 – 1.62 (m, 14H), 1.58 (s, 3H), 1.57 (s, 9H).

^{31}P NMR (162 MHz, CDCl$_3$+MeOD) δ -11.4, -13.6.

In agreement with ref[2]

Und-PP-GlcNAc



According to *General Procedure C* Und-PP-GlcNAc(OAc)$_3$ (70 mg, 0.056 mmol) was deprotected to yield the deprotected pyrophosphate as a sticky white solid (30 mg, 47%).

^{1}H NMR (400 MHz, CDCl$_3$+MeOD) δ 5.56 (s, 1H), 5.40 – 5.33 (m, 1H), 5.15 – 5.02 (m, 12H), 3.96 (s, 2H), 3.74 – 3.64 (m, 1H), 3.58 (s, 1H), 3.24 (d, *J* = 12.8 Hz, 2H), 2.11 – 1.91 (m, 55H), 1.71 (s, 3H), 1.65 (m, 25H), 1.58 – 1.58 (m, 3H), 1.57 (s, 12H).

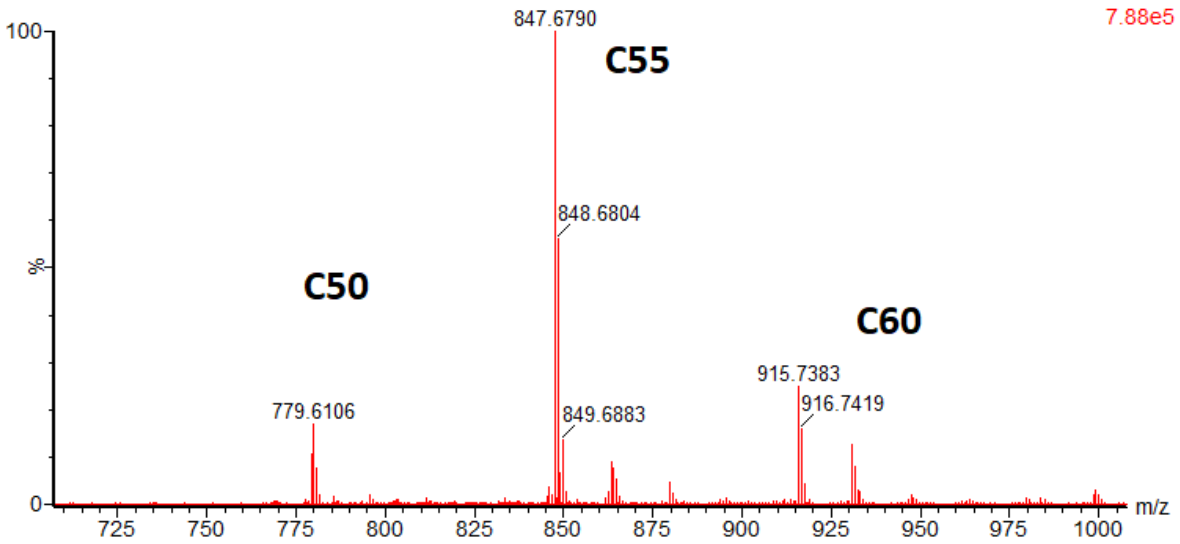
^{31}P NMR (162 MHz, CDCl$_3$+MeOD) δ -10.2, -12.3.

HRMS (ESI-); (*m/z*) calc. for $C_{63}H_{104}NO_{12}P_2$: 1128.7039 $[M-H]^-$; found: 1128.7051 (for C55)

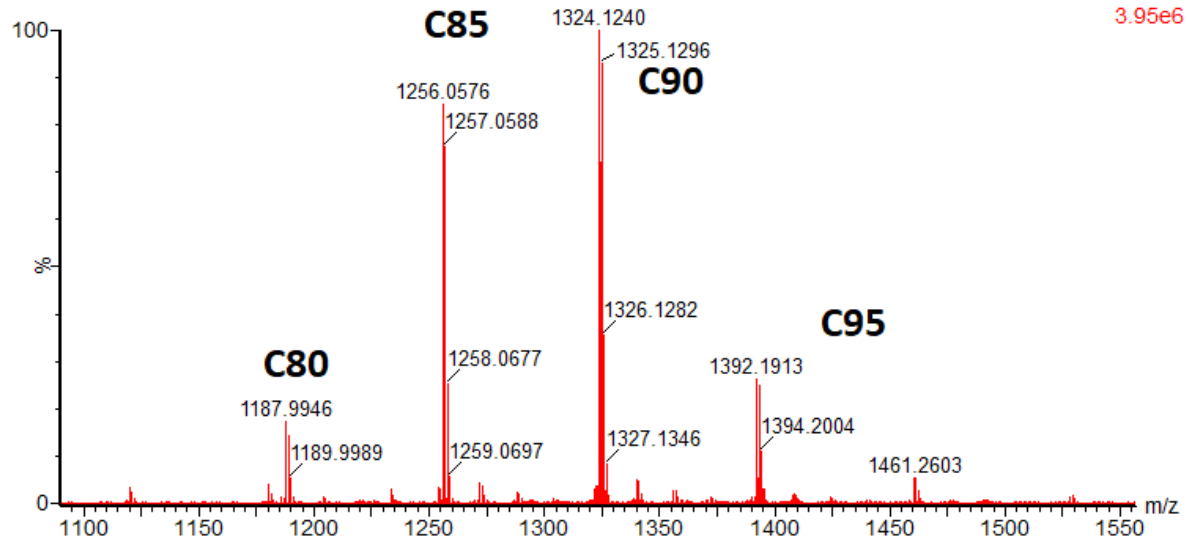
In agreement with ref[2]

Exemplar Spectra

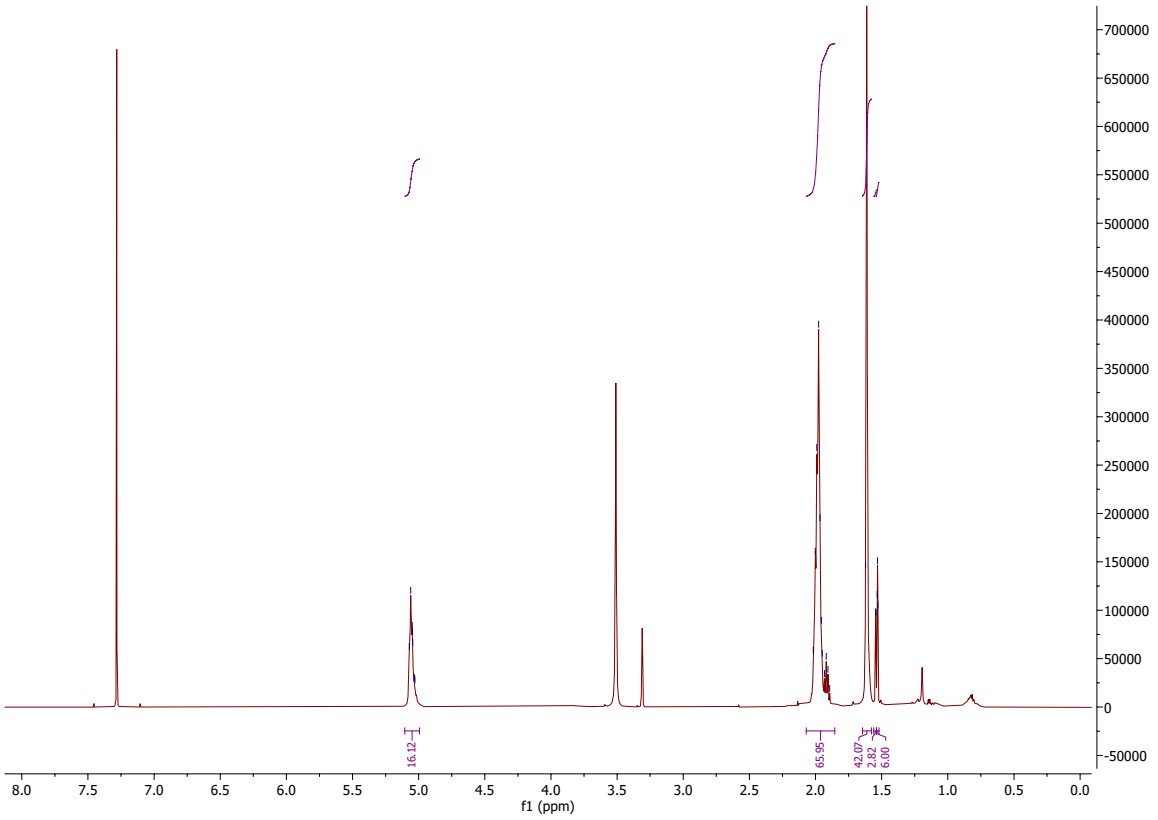
C55-Dol-P

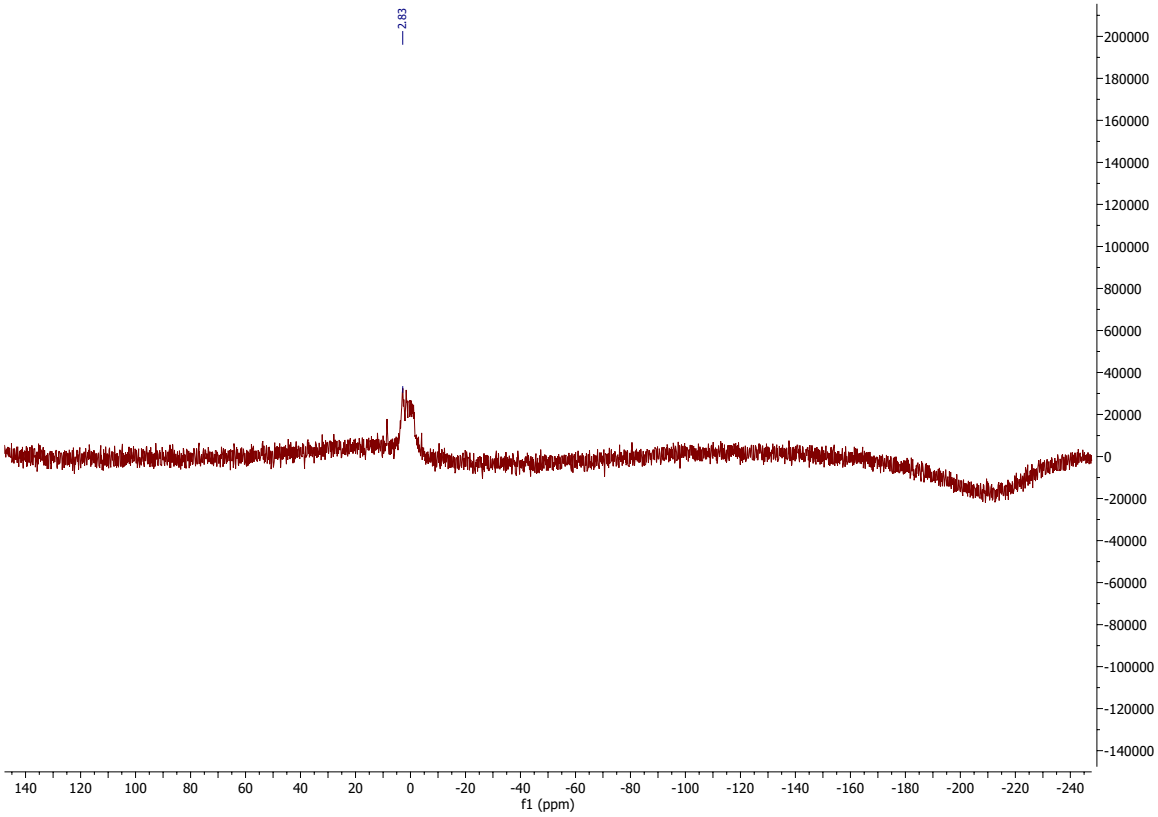


C90-Dol-P

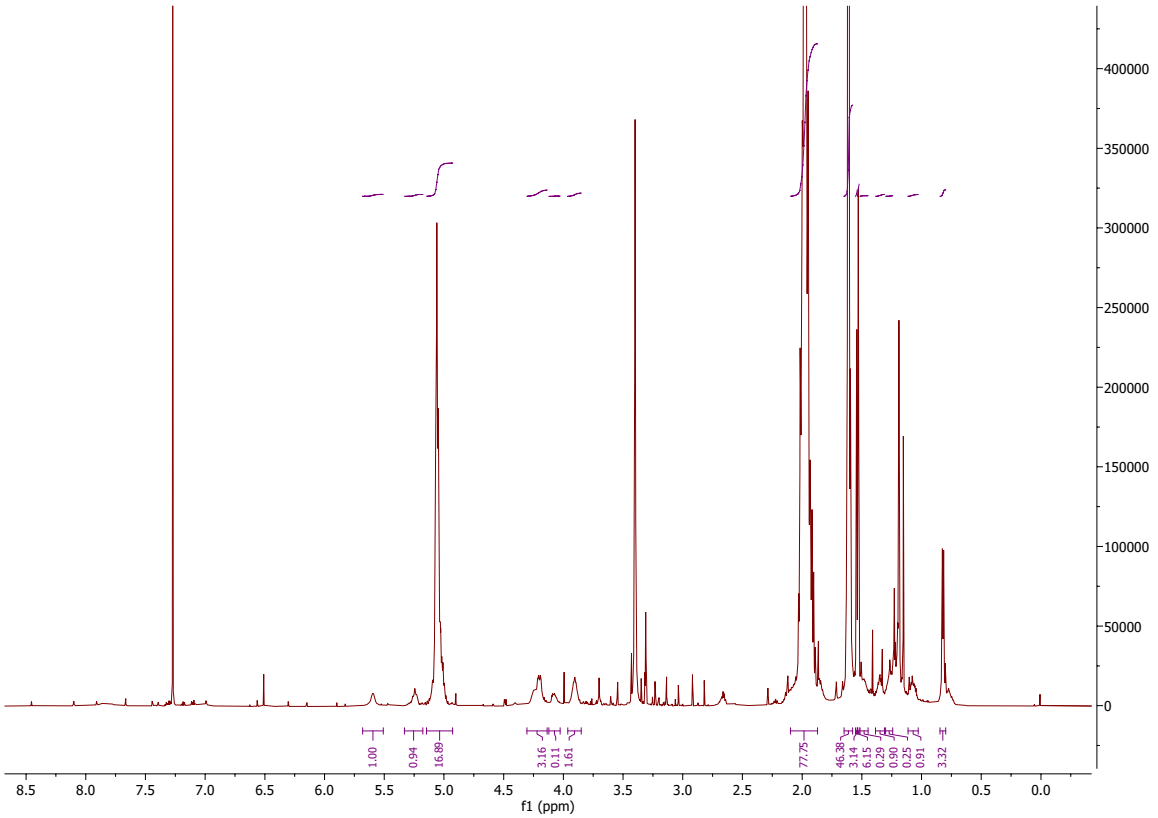


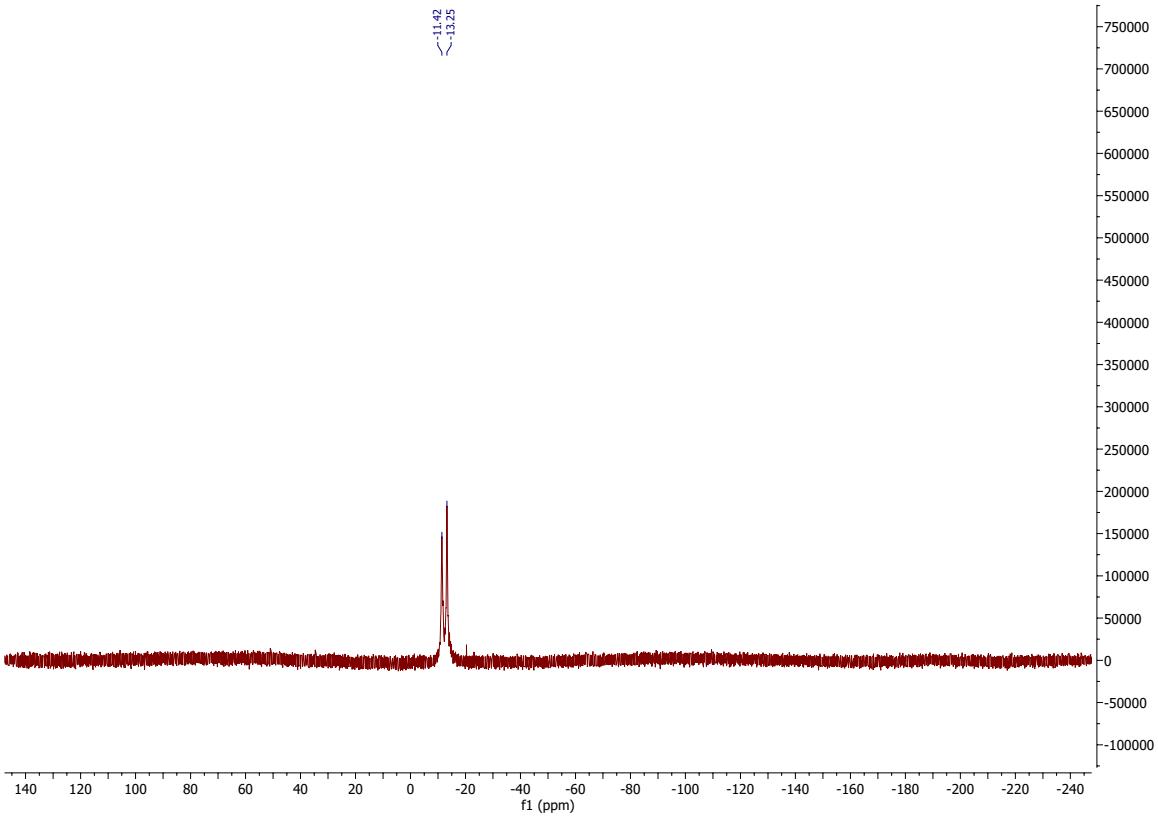
C90-Dol-P



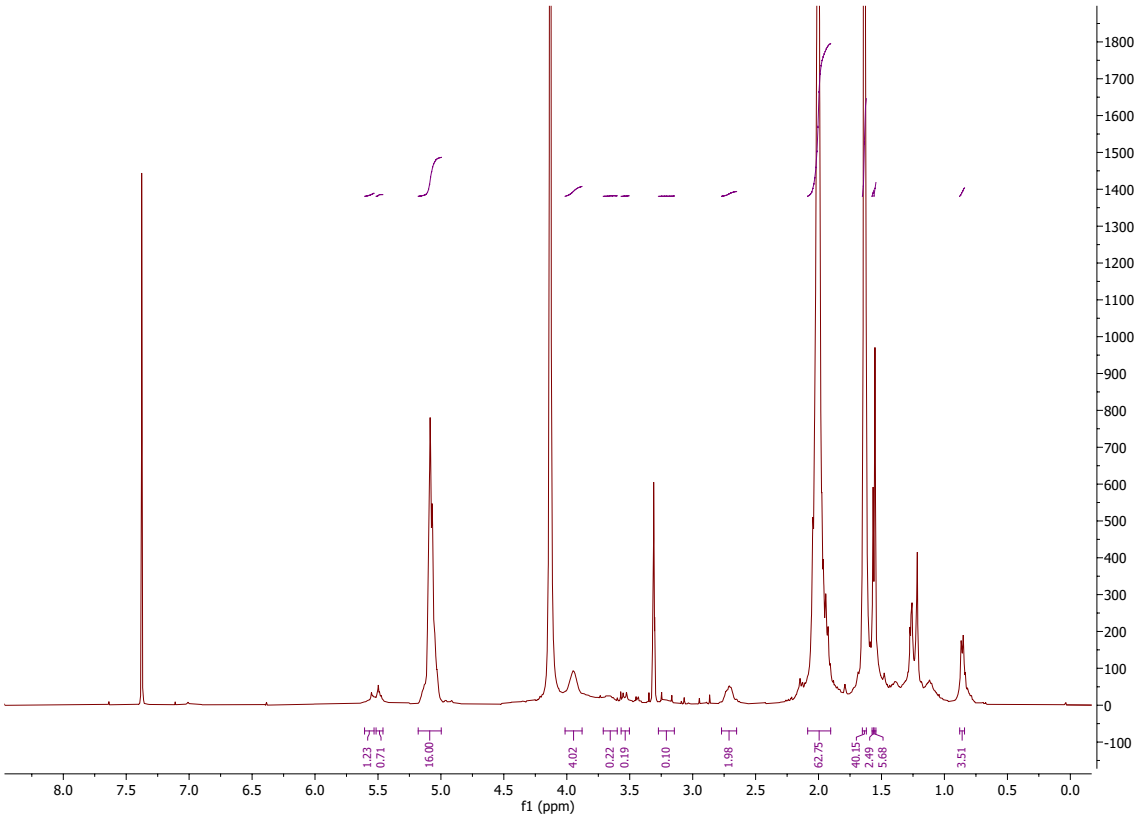


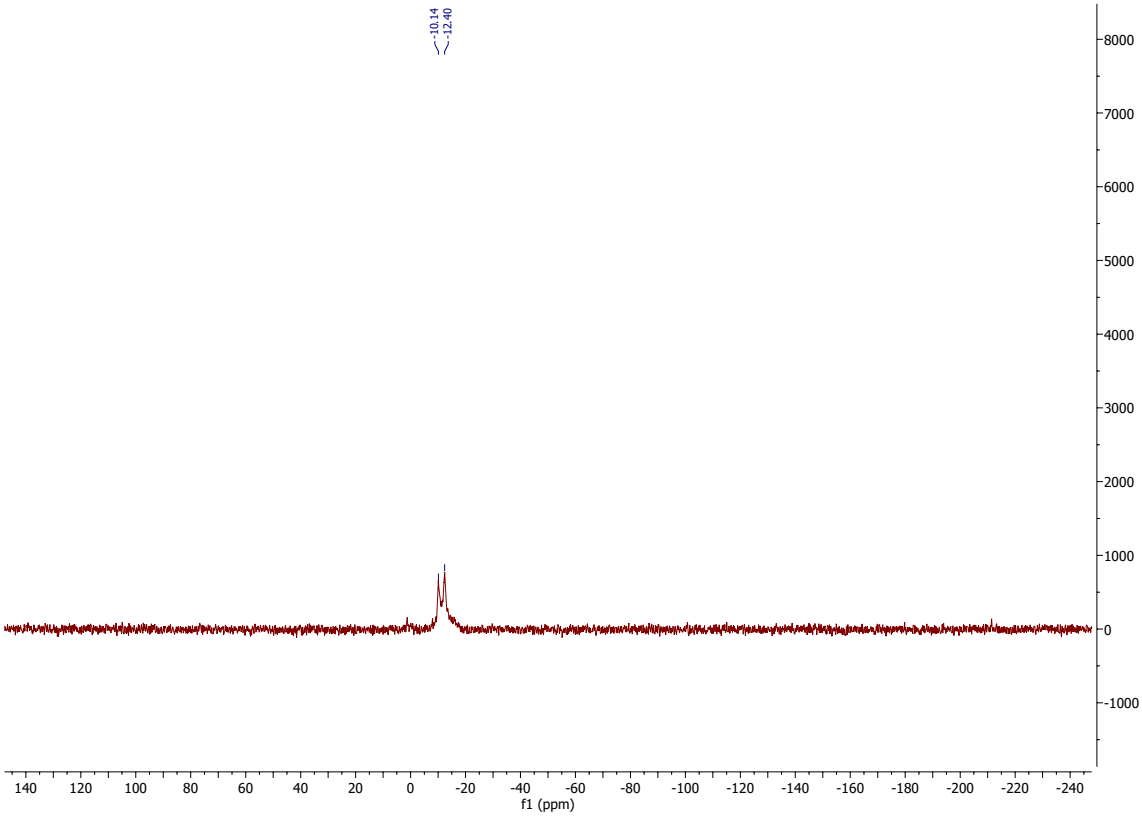
Dol-PP-GlcNAc(OAc)$_3$





Dol-PP-GlcNAc





References

(1) Liu, F.; Vijayakrishnan, B.; Faridmoayer, A.; Taylor, T. A.; Parsons, T. B.; Bernardes, G. J. L.; Kowarik, M.; Davis, B. G. *Journal of the American Chemical Society* **2014**, *136* (2), 566-569.

(2) Dong, Y. Y.; Wang, H.; Pike, A. C. W.; Cochrane, S. A.; Hamedzadeh, S.; Wyszyński, F. J.; Bushell, S. R.; Royer, S. F.; Widdick, D. A.; Sajid, A.; et al. *Cell* **2018**, *175* (4), 1045-1058.e1016.