DISS. ETH NO. 26652

STRUCTURAL AND MECHANISTIC INSIGHTS INTO EUKARYOTIC N-GLYCAN BIOSYNTHESIS AND HUMAN VITAMIN B12 TRANSPORT

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

Presented by

JOËL SHARON BLOCH

MSc ETH in Biology, Major in Structural Biology and Biophysics, ETH Zurich

Born on 02.03.1989

Citizen of
Zürich ZH and of Germany

Accepted on the recommendation of

Prof. Dr. K. Locher, examiner
Prof. Dr. R. Glockshuber, co-examiner
Prof. Dr. R Dutzler, co-examiner

2020

II

"The best way to have a good idea is to have lots of ideas."

Linus Pauling

Acknowledgements

I would like to thank Prof. Kaspar Locher for giving me the opportunity to work on multiple challenging and really exciting projects in his lab and for his tremendous support throughout my studies. I especially enjoyed the countless scientific discussions from which many good ideas sparked and which, especially during hard times of the projects, where always very motivational.

I would like to thank the members my thesis committee Prof. Rudi Glockshuber and Prof. Raymund Dutzler, as well as the chair of my examination Prof. Manfred Kopf.

Concerning the studies on the substrate specificity of the beta domain of transcobalamin, I would like to thank Dr. Markus Ruetz and Prof. Markus Kräutler for their collaboration in providing me with Anti-B12.

Concerning the studies on nanobody binders for transcobalamin and CD320, I would like to thank Prof. Edward Quadros and Jeffrey Sequeira for their, collaboration in visualizing the endocytosis of TCNB4 by human cells. I would like to thank Dr. Saša Štefanić for help with alpaca immunizations, blood collection, and testing of the specific immune response by ELISA. I thank Prof. Raymund Dutzler and Yvonne Neldner for technical support and the primers and vectors used for the generation of the VHH DNA library. I thank Dr. Ana Ramirez for help with the generation of the VHH library.

I would like to thank the Beamline Staff at the Swiss Light Source at the Paul Scherrer Institute for their assistance in X-ray data collection.

I would like to thank Dr. Amer Alam for introducing me to the transcobalamin project, to the lab and for many fruitful discussions and Bernadette Prinz for introducing me to the lab and cell culture.

Concerning the work on ER-luminal N-glycosylation, I would like to thank Giorgio Pesciullesi, Dr. Jérémy Boilevin, Dr. Tamis Darbre and Prof. Jean-Louis Reymond for

Parts of this thesis have been published

Structure of the human transcobalamin beta domain in four distinct states

Bloch, J. S., Ruetz, M., Kräutler, B., Locher, K. P. (2017) PLOS ONE 12(9)

Structure and mechanism of the ER-based glucosyltransferase ALG6

Bloch, J.S., Pesciullesi, G., Boillevin, J., Nosol, K., Irobalieva, R.N., Darbre, T., Aebi, M., Kossiakoff, A.A., Reymond, J.L., Locher, K.P. (2020) Nature, published online 26.02.2020

Related Publication

Structural basis of transcobalamin recognition by human CD320 receptor

Alam, A., Woo, J. S., Schmitz, J., Prinz, B., Root, K., Chen, F., **Bloch, J. S.**, Zenobi, R., Locher, K. P. (2016) Nature Communications, 7: 12100

Table of Contents

3.   A VERSATILE SET OF NANOBODIES TARGETING THE HUMAN, TRANSCOBALAMIN MEDIATED VITAMIN B12 UPTAKE ROUTE

XI

Summary

Structural biology allows us to shed light onto the smallest processes of life. The genetic information that is stored in the DNA of every living cell holds the building plans for proteins: These tiny molecules, which are often only at the size of a few nanometers, are responsible most of the biochemical functions in life. By means of X-ray crystallography and cryo-EM we are able to determine the structures of these proteins at atomic resolution, which makes it possible to rationalize the biochemical functions that are carried out by these proteins.

The first part of this thesis describes structural and mechanistic studies on transcobalamin mediated vitamin B12 uptake in human cells. In humans, vitamin B12 is essential for DNA synthesis and for certain metabolic processes. Humans are dependent on oral uptake of vitamin B12 from animal sources or from supplements. After ingestion, the vitamin is transported into the bloodstream, via a multistep pathway involving different carrier- and transport proteins. Transcobalamin is a soluble two-domain protein that is secreted into the bloodstream and specifically recognizes and binds vitamin B12 by clamping it between its two binding domains. The transcobalamin B12 complex is captured by the cell surface receptor CD320 and the complex is taken up into the cell by receptor mediated endocytosis.

As transcobalamin binds to vitamin B12 at a virtually irreversibly high affinity, studying its binding to different ligands without denaturing of the protein poses a great difficulty. We have heterologously expressed the beta-domain of transcobalamin and co-crystallized it in the apo form, as well as in complex with vitamin B12, the B12 related molecule cobinamide and synthetic antivitamin B12, a compound that cannot be further processed by the cellular enzymes and is designated for clinical studies in which it is used to artificially induce vitamin B12 deficiency in humans or animals. Using X-ray crystallography, we have obtained structures at high-resolution which helped to rationalize the binding modes of the respective substrates. In addition, we could prove biochemically, that the two domains of transcobalamin only assemble upon binding of vitamin B12.

The second part of the thesis describes the structural and mechanistic studies on the ER-luminal glucosyltransferase ALG6: N-glycosylation is one of the most common post translational modifications and is found in all domains of life. In eukaryotes a lipid-linked oligosaccharide is generated by the stepwise enzymatic addition of hexoses

onto a membrane embedded dolichylphosphate precursor molecule. In the lumen of the endoplasmic reticulum, the oligosaccharide is then transferred onto an acceptor peptide of a secreted- or membrane protein by oligosaccharyltransferase. The glycan helps in protein folding, quality control and after diversification in the Golgi as an identification tag of the respective cell or protein. The lipid-linked oligosaccharide biosynthesis starts in the cytoplasm and after the addition of one phosphate, two N-acetylglucosamine molecules and five mannoses, the lipid linked oligosaccharide is flipped into the ER-lumen. There the six membrane integral glycosyltransferases ALG3, ALG9, ALG12, ALG6, ALG8 and ALG10 add four additional mannoses and three glucoses in a step wise fashion, leading to a final $Dol-PP-GlcNAc_2Man_9Glc_3$ molecule. I have been studying the structure and mechanism of ALG6, which adds the first glucose to the LLO.

Building on the previously established chemoenzymatic synthesis of $Dol-PP-GlcNAc_2Man_5$, we used purified enzymes ALG3, ALG9 and ALG12 and synthetic substrate molecules to reconstitute a part the pathway of ER-luminal enzymatic LLO synthesis. We obtained pure $Dol-PP-GlcNAc_2Man_9$, which is the acceptor substrate of ALG6. In addition, we were able to generate every substrate intermediate of the enzymatic pathway. With the help of the $Dol-PP-GlcNAc_2Man_9$ and synthetic Dol-P-Glc donor substrate we could demonstrate the function of ALG6.

We reconstituted ALG6 into lipid nanodisc and with the help of a Fab binder from a synthetic library we were able to obtain a structure of ALG6 at 3.0 Å resolution by cryo-EM. An additional structure at 3.9 Å was determined with the protein reconstituted in detergent and with the synthetic donor-substrate analog $Dol_{25}-P-Glc$ bound to it. The structure of the substrate-bound enzyme in combination with site directed mutagenesis helped to identify the catalytic residue Asp69, which is likely to act as a general base in ALG6 and which hints to a possible catalytic mechanism. ALG6 is a member of the GT-C superfamily. The GT-C superfamily categorizes membrane integral glycosyltransferases which require isoprenoid-inked sugar donor substrates. Structurally, the GT-C superfamily is only poorly defined. By comparing our structure of ALG6 to other published structures of GT-C enzymes, we were able to identify common motifs in GT-C enzymes and we observed an evolutionarily conserved modular design of GT-C enzymes.

Our described method for the chemo-enzymatic synthesis of eukaryotic LLOs and LLO intermediates provides and unprecedent tool to study enzymes involved in LLO synthesis as well as oligosaccharyltransferases.

Our structural findings not only explain the function of ALG6 and give therefore insight into ER-based LLO biosynthesis, but also, they help to rationalize the architecture and mechanism of GT-C enzymes in general.

With ALG6 being 64 kDa in size and the complexed Fab 48 kDa, the structure of ALG6:6AG9-Fab complex at 3.0 Å resolution, is at this time reportedly the highest resolved structure of a membrane-bound protein complex resolved by cryo-EM. We therefore also demonstrate the usefulness of Fab technology for the study of small membrane proteins and describe a useful tool for further studies of e.g. small membrane integral glycosyltransferases.

Zusammenfassung

Strukturbiologie hilft uns die kleinsten Prozesse des Lebens zu erklären. Die genetische Information, die in der DNS, jeder lebenden Zelle aufbewahrt wird enthält die Baupläne für Proteine. Diese Makromoleküle, sind für die meisten biochemischen Funktionen in der belebten Welt verantwortlich. Röntgenkristallographie und Kryo-Elektronenmikroskopie ermöglichen es uns, die dreidimensionalen Strukturen dieser Proteine mit atomarer Auflösung aufzuklären. Dies ermöglicht es, die biochemischen Funktionen dieser Proteine zu erklären.

Der erste Teil dieser Dissertation beschreibt strukturelle und mechanistische Studien zur Transcobalamin-abhängigen Aufnahme von Vitamin B12 in menschlichen Zellen. Vitamin B12 ist essenziell für die DNS Synthese und für gewisse metabolische Prozesse im Menschen. Menschen, sind auf die Aufnahme von Vitamin B12 aus tierischen Quellen oder aus Nahrungsergänzungsmitteln angewiesen. Nach der Nahrungsaufnahme gelangt das Vitamin B12 über mehrere Schritte, durch verschiedene Transportproteine in die Blutbahn. Transcobalamin ist ein lösliches Zweidomänenprotein, welches in die Blutbahn sekretiert wird. Es bindet spezifisch an Vitamin B12 indem es das Vitamin, zwischen den beiden Proteindomänen einklemmt. Der Komplex von Transcobalamin und Vitamin B12 wird dann von dem Zelloberflächenrezeptorprotein CD320 gebunden und mit Hilfe von Clathrin-vermittelter Endozytose in die Zelle aufgenommen.

Da Transcobalamin mit einer nahezu unendlich hohen Affinität an Vitamin B12 bindet, stellt das Studieren der Bindungsinteraktionen von Transcobalamin mit anderen Liganden, ohne dafür Transcobalamin denaturieren zu müssen, eine grosse Hürde dar. Um diese Hürde zu Überschreiten haben wir die beta-Domäne von Transcobalamin heterolog exprimiert und in der Abwesenheit eines Liganden, sowie als Komplex mit Vitamin B12, dem verwandten Molekül Cobinamid, sowie synthetischem Antivitamin B12 kristallisiert. Antivitamin B12 ist eine synthetisch veränderte Form von Vitamin B12, welche von der Zellmaschinerie nicht weiterverarbeitet werden kann. Es dient dazu für klinische Studien in Patienten oder Tieren kurzzeitig einen künstlichen Vitamin B12 Mangel zu erzeugen. Mit Hilfe von hochauflösender Röntgenkristallographie konnten wir die Strukturen der jeweiligen Komplexe ermitteln, welche uns Aufschluss über die jeweiligen Substratbindungsmodi von Transcobalamin gaben. Zusätzlich gelang es uns biochemisch aufzuzeigen, dass

sich die beiden Domänen von Transcobalamin nur durch die Zugabe von Vitamin B12 zusammenfügen.

Der Zweite Teil dieser Dissertation beschreibt strukturelle und mechanistische Studien an der ER-luminalen Glukosyltransferase ALG6: N-Glykosilierung ist eine der am häufigsten vorkommenden posttranslationalen Proteinmodifikationen und kann in allen Domänen des Lebens gefunden werden. In Eukaryonten wird ein Lipid-gebundenes Oligosaccharid durch das enzymkatalysierte, schrittweise Hinzufügen von Hexosen an einen Membrangebundenen Dolichylphosphat-Vorläufer generiert. Im Lumen des Endoplasmatischen Retikulum (ER) wird das Oligosaccharid dann durch die Oligosaccharyltransferase an eine Asparagin-Seitenkette eines sekretorischen Proteins oder eines Membranproteins gehängt. Das N-Glykan hilft dem Protein bei der Proteinfaltung und ist an der Qualitätskontrolle der Faltung vor dem Weitertransport zum Golgi-Aparatus beteiligt. Nach einer ausgiebigen Diversifizierung durch eine Vielzahl von Glukosidasen und Glykosyltransferasen im Golgi erreicht das Glykan seine endgültige Zelltyp- und Glykoprotein-spezifische kovalente Struktur. Die Biosynthese des lipidgebundenen Oligosaccharids beginnt im Zytoplasma. Nach dem Anfügen einer Phosphatgruppe, zwei N-Acetylglucosamin Molekülen und fünf Mannosen an Dolicholphosphat wird das lipidgebundene Oligosaccharid ins Lumen des ER übergeführt. Von dort aus fügen die sechs membranintegralen Glykosyltransferasen ALG3, ALG9, ALG12, ALG6, ALG8 und ALG10 schrittweise, gesamthaft vier Mannosen und drei Glukosen an, was zum finalen Dol-PP-$GlcNAc_2Man_9Glc_3$ Molekül führt. In dieser Doktorarbeit habe ich mich mit der Struktur und dem enzymatischen Mechanismus von ALG6 befasst, welches die erste Glukose an das Oligosaccharid anfügt.

Aufbauend auf der bereits etablierten, chemoenzymatischen Synthese von Dol-PP-$GlcNAc_2Man_5$ haben wir mit Hilfe der aufgereinigten Enzyme ALG3, ALG9 und ALG12, sowie mit synthetisch hergestellen Substratmolekülen einen Teil des enzymatischen, ER-luminalen LLO-Synthese-Stoffwechselwegs nachgebaut. Es gelang uns reines Dol-PP-$GlcNAc_2Man_9$ aufzureinigen, welches das Substrat von ALG6 darstellt. Zusätzlich konnten wir alle Zwischenprodukte des Stoffwechselwegs herstellen. Mit Hilfe des Dol-PP-$GlcNAc_2Man_9$ sowie synthetisch hergestelltem Dol-P-Glc gelang es uns die Aktivität von ALG6 zu demonstrieren.

Wir betteten ALG6 in eine Nanodisc ein und mit Hilfe von eines Antikörper Fab-Fragments, das aus einer Synthetischen Bibliothek stammte, gelang es uns mit Kryo-

EM eine Struktur von ALG6 mit einer Auflösung von 3.0 Å zu ermitteln. Eine zusätzliche Struktur mit einer Auflösung von 3.9 Å wurde ermittelt. Hierbei wurde in Detergenz gelöstes, mit Fab komplexiertes ALG6 welches ebenfalls mit dem synthetisch hergestellten Donor-Substrat Dol-P-Glc komplexiert war verwendet. Die Struktur des Substrat-gebundenen Enzyms, sowie gezielte Mutagenese des *ALG6* Gens ermöglichten es, die katalytische Seitenkette Asp69, welche vermutlich als allgemeine Base fungiert, zu identifizieren. Dies gab uns ebenfalls Hinweise auf einen möglichen Reaktionsmechanismus von ALG6. ALG6 gehört zur GT-C Superfamilie Membran-integraler Glykosyltransferasen, welche Isoprenoid-gebundene Zucker als Donor-Substrate verwenden. Mit unseren Strukturen konnten wir gemeinsame Motive in GT-C Enzymen aufzeigen und haben eine evolutionär konservierte, modulare Architektur der GT-C Enzyme entdeckt.

Die von uns beschriebene Methode für die chemo-enzymatische Synthese von eukaryotischen LLOs und LLO-Zwischenprodukten ist ein noch einzigartiges Werkzeug zur Charakterisierung der in der LLO-Synthese involvierten Enzyme.

Die Strukturellen Einblicke in dieser Studie erklären nicht nur die Funktion von ALG6 und vermitteln dadurch Einsicht in die ER-basierte Biosynthese von LLOs, sondern tragen auch zum generellen Verständnis der Architektur und der Mechanismen von GT-C Enzymen bei.

Mit einer Masse von lediglich 64 kDa für ALG6 und 48 kDa für das Fab-Antikörper Fragment, stellt Struktur des ALG6:6AG9-Fab Komplexes mit einer Auflösung von 3.0 Å, die momentan am höchsten aufgelöste Kryo-EM Struktur eines membrangebundenen Komplexes dieser Grösse dar. Wir konnten somit die Nützlichkeit der Fab-Fragmente für die Strukturaufklärung von kleinen Membranproteinen demonstrieren. Solche Fab-Fragmente könnten sich für die Strukturaufklärungen von weiteren Membranproteinen, wie zum Beispiel kleinen, Membran-integralen Glykosyltransferasen als nützlich erweisen.

1. Introduction

In this thesis I have been working on two different topics, which are both related to structural biology. Initially, I had been studying mechanistic aspects of human TC:CD320 mediated vitamin B12 transport using X-ray crystallography and nanobody technology. After obtaining promising results on the structural basis of substrate recognition of transcobalamin and the successful development of conformational nanobody binders, which turned out to be useful as crystallization chaperones, as well as for the targeting and destruction of cancer cells, I wanted to tackle a new challenge. I aimed to work on a eukaryotic membrane protein and, in view of the recent "resolution revolution[1]" in cryo-electron microscopy, I decided to embrace this newly emerging technology and therefore to shift my focus from X-ray crystallography to cryo-EM. For the second part of my PhD, I took on a new project, where I would study the small, membrane integral, eukaryotic glycosyltransferase ALG6 using cryo-EM.

1.1 Vitamin B12

1.1.1 The structure of vitamin B12

Vitamin B12, also rereferred to as nature's most beautiful cofactor[2] is an essential cofactor in humans[3]. The water-soluble molecule is composed of a corrin-ring, in which the four central nitrogen atoms are coordinating a cobalt atom[3-6] (Fig. 1). The cobalt atom, which poses the catalytic group of B12 is further coordinated from one face of the corrin ring by a nitrogen atom of a 5,6-dimethylbenzimidazole ribonucleotide moiety that is tethered to the corrin ring by a flexible linker[3,4]. The sixth coordination site of the cobalt atom, which is located on the opposite face of the corrin ring can be occupied by a variety of different ligands[3]. These include a hydroxyl group (hydroxocobalamin), a cyano group (cyanocobalamin, vitamin B12), a methyl group or a 5'-deoxyadenosyl group[3]. Due to its chemical complexity and large number of chiral centers, the formal synthesis of vitamin B12, as a proof of principle has been a long-standing quest, which was finally completed in a collaboration of the groups of R. Woodward and A. Eschenmoser[7]. To this day, the synthesis marks as a milestone of total synthesis in organic chemistry and as a result of this R. Woodward was awarded with the Nobel prize in chemistry.
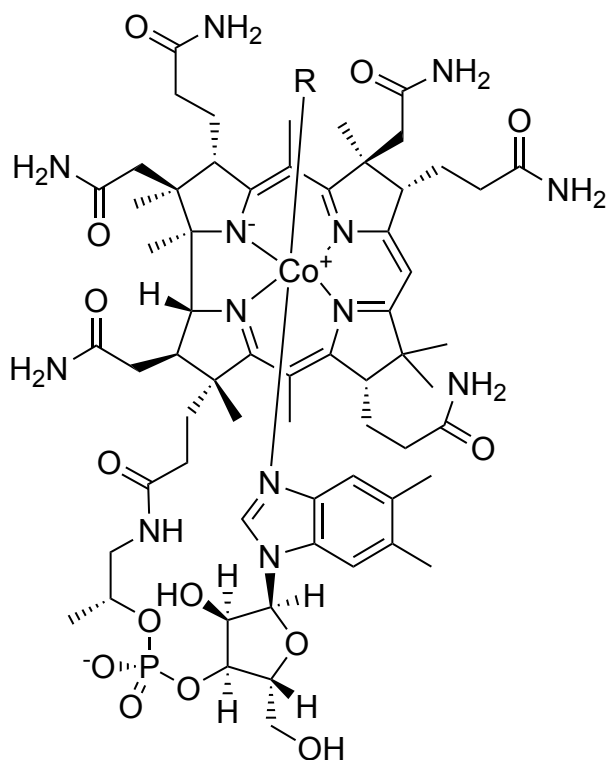
Figure 1 Structure of cobalamin. The structure of vitamin B12 consists of a corrin ring with the four central nitrogen atoms coordinating a central cobalt atom[3-6]. The corrin ring has various substituent groups including a flexibly linked 5,6-dimethylbenzimidazole ribonucleotide moiety that is coordinating the central Co. The sixth Co-coordinating group, denoted as "R" can be replaced by various groups including a hydroxyl group, a methyl group or cyano group. Cyano substituted cobalamin is referred to as cyanocobalamin or vitamin B12.

1.1.2 The physiological role of vitamin B12

Vitamin B12 is essential for at least two enzymes in humans[3]. The molecule is required as a cofactor for the mitochondrial conversion of methylmalonyl-CoA to succinyl-CoA via methylmalonyl-CoA mutase[3]. A process, that is required for metabolizing a variety of organic compounds including odd-chain fatty acids and branched chain amino acids.

The other known function of vitamin B12 in the human body is its role as a cofactor for methionine synthase which catalyzes the methylation of cytoplasmic homocysteine, yielding methionine[3]. This amino acid, which is essential for ribosomal protein synthesis, is also required for DNA methylation, which indirectly makes vitamin B12 essential for DNA synthesis.

1.1.3 The eukaryotic vitamin B12 uptake route

B12 is produced only by certain bacteria and archaea[8]. Some of these bacteria reside in the gut of certain herbivores, where they are part of the gut flora. The produced B12

can then be absorbed by the animal[9]. For this reason, omnivores like humans depend on the dietary uptake of B12 from animal products such as meat, fish, milk and eggs[10]. Another option for individuals following a vegan diet is the uptake of vitamin B12 from supplements. There, the vitamin is produced in Bacteria and after the purification of cobalamin with activated carbon, cyanocobalamin respectively vitamin B12 is yielded[8].
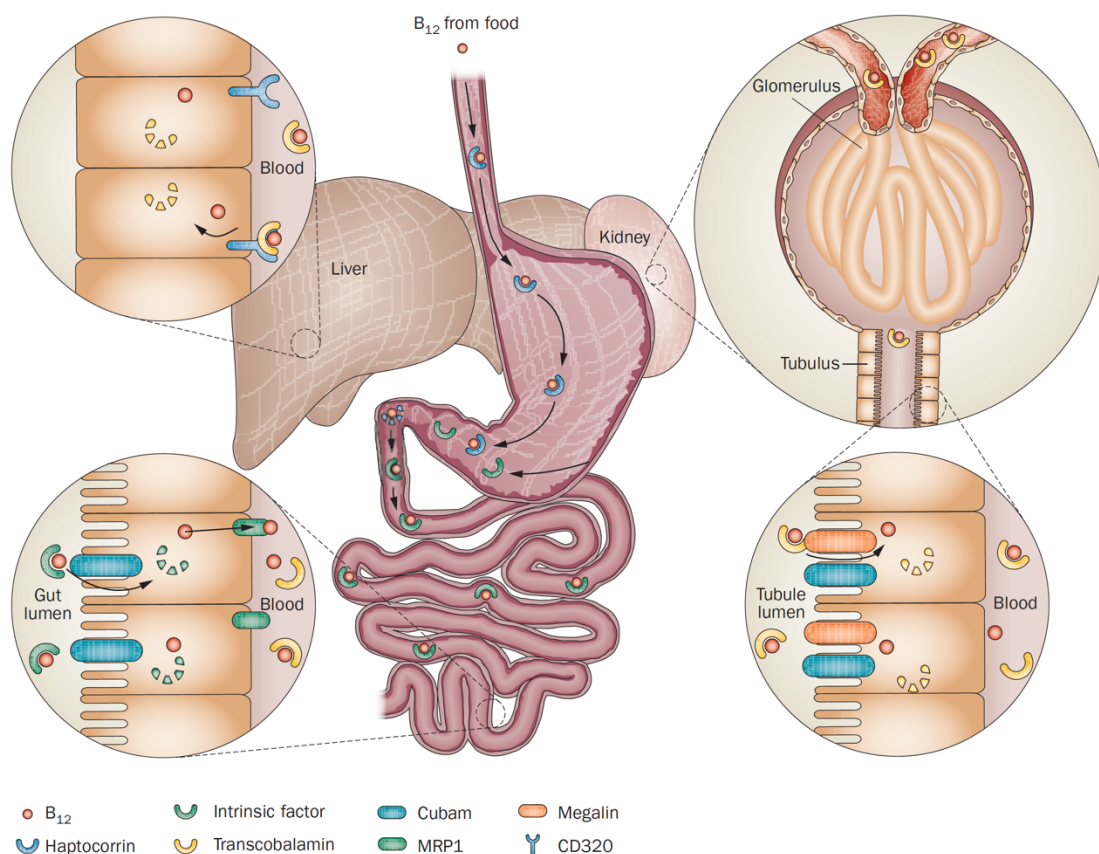
Figure 2 The human vitamin B12 uptake route. Upon oral ingestion of B12 containing foods, B12 is sequestered by Haptocorrin (HC), which escorts it safely through the stomach. After degradation of HC in the duodenum, B12 is released and re-bound by intrinsic factor (IF). The IF:B12 complex is taken up by enterocytes via binding to the cell surface receptor cubam and subsequent, receptor mediated endocytosis. From there B12 is extruded into the bloodstream by the ABC transporter MRP1. In the blood B12 is bound by transcobalamin (TC) which delivers the vitamin to cells via binding to the receptor CD320 and subsequent receptor mediated endocytosis. Megalin in the glomerulus prevents from renal loss of TC. Figure and content of legend from[3].

During the oral consumption of B12 containing foods (Fig. 2), the vitamin is captured in the mouth by the salivary secreted, soluble B12 carrier protein haptocorrin (HC)[3,11]. HC safely escorts the precious vitamin trough the stomach, protecting it from proteolytic degradation[11]. In the duodenum, HC is degraded by pancreatic proteases and B12 gets released and is subsequently recaptured by the B12 carrier protein intrinsic factor (IF)[11], which in turn is resistant to proteolysis by these enzymes[3,12].

Holo-IF is specifically recognized by cubam, a cell, surface receptor protein that is displayed on the enterocytes of the ileum[13,14]. The complex of Cubam:IF:B12 is then being endocytosed and after lysosomal degradation of IF, B12 is released into the cytoplasm. The vitamin is then extruded into the bloodstream of the portal vein by the ABC transporter ABCC1 also known as MRP1[15]. In the blood plasma, B12 is bound by either one of the soluble and structurally related carrier proteins transcobalamin (TC) or HC[16,17]. HC has a broad substrate specificity and can also bind other cobalamins, or inactive forms of B12[3]. The role of HC in the blood plasma is not fully understood[18]. One hypothesis is than HC is thought to act as a scavenger protein by sequestering corrins that cannot be metabolized or that are potentially harmful[18]. Another hypothesis is that the role of HC could also be in helping to control the cellular uptake of vitamin B12 by rendering the sequestered B12 unavailable to TC[18]. TC on the other hand is highly selective for B12 and is responsible for the cellular uptake of the vitamin[3,17]. Holo-TC is recognized by the LDL-type cell surface receptor CD320 and enters the cells via receptor mediated endocytosis[19]. When the endocytosed complex reaches the endosome, the low pH causes the receptor and TC to dissociate[20]. CD320 has been proposed to be recycled to the cell surface, whereas TC is being trafficked to the lysosome[3]. There TC is proteolytically degraded and B12 gets released and is subsequently extruded into the cytoplasm. The transmembrane proteins LMBD1 and ABCD1 have been proposed to be responsible for the lysosomal extrusion of B12[21,22].

1.1.4 Targeting of vitamin B12 transport for cancer therapy

Since B12 is linked to DNA synthesis, for many years the cellular uptake-route of B12 via TC:CD320 has raised an interest in cancer therapy[23]. CD320 which has been shown to be downregulated in quiescent cells, is specifically upregulated in highly dividing cells such as cancer cells[24]. Attempts in utilizing the cellular uptake route of B12 for cancer therapy and imaging have included the derivatization of B12 with radioactive markers, as well as the use of antibody drug conjugates, targeting CD320[25-27]. Although fulfilling their purposes in the lab, none of the approaches has led do a drug on the marketplace[23].
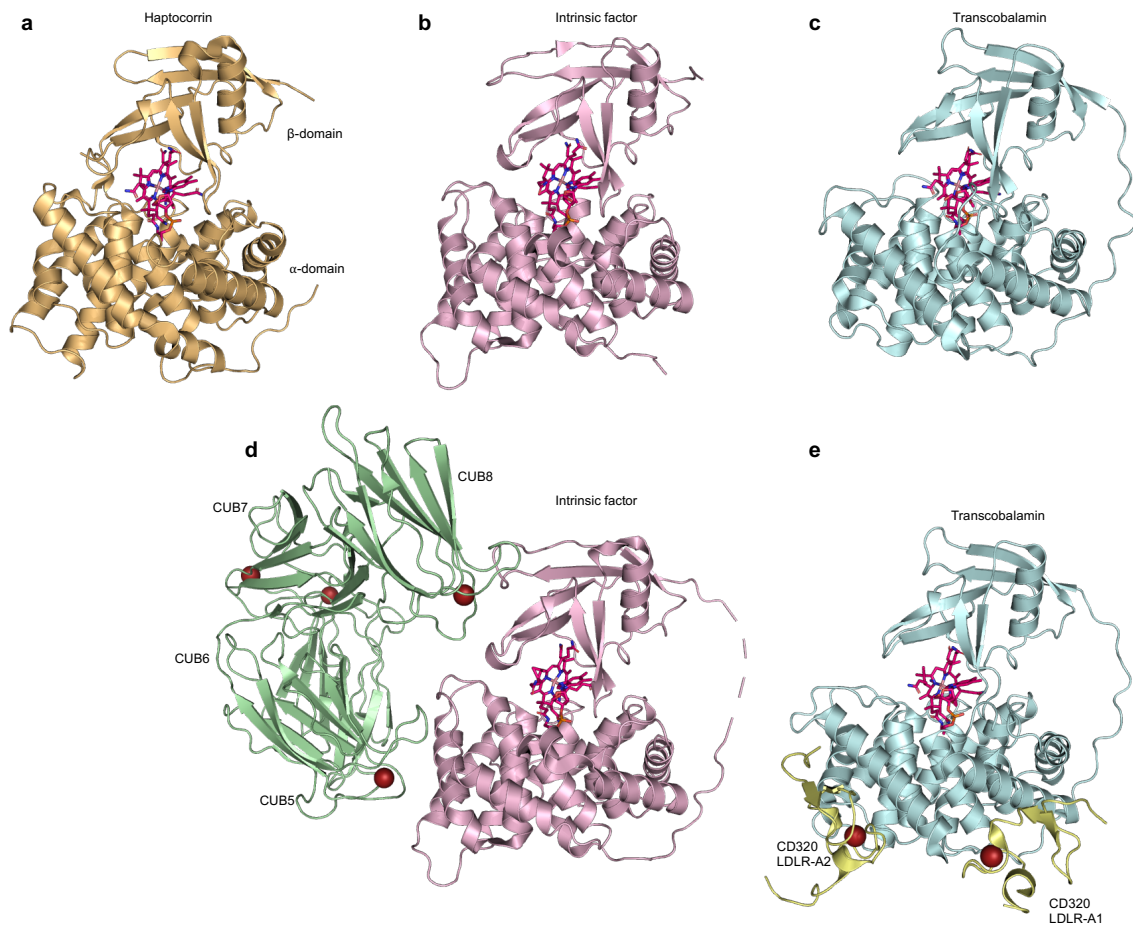
Figure 3 Structural basis of human vitamin B12 transport. Human B12 carrier proteins of known structure are shown in cartoon representation with bound B12 molecules shown as pink sticks and bound calcium ions as red spheres. **a**, Haptocorrin (orange, PDB ID 4KKI[28]), **b**, intrinsic factor (pink, PDB ID 2PMV[29]) **c**, transcobalamin (cyan, PDB ID 3BB5[30]), **d**, Intrinsic factor (pink) in complex with cubilin (green, PDB ID 3KQ4[31]), **e**, transcobalamin (cyan) in complex with CD320 (yellow, PDB ID 4ZRP[32]). The structures share a conserved two-domain fold (labeled in a), with an N-terminal α-domain (bottom) and a C-terminalβ-domain (top) that are tethered by a flexible linker. The Proteins also share a common B12 binding mode, where the alpha and the beta-domain dimerize upon B12 binding, as it has been experimentally demonstrated for IF[33] and TC[34].

1.1.5 Structural basis of mammalian vitamin B12 transport

The first structures of human and bovine TC (Fig. 3) were solved in 2006 by L. Randaccio and coworkers[30]. The structures revealed a two-domain fold composed of an alpha helix bundle (α-domain) and a beta sheet rich domain (β-domain). The two domains are tethered by a flexible linker and bind both to B12 from opposite sides. Subsequently, the structures of IF[29], IF in complex with binding domains of cubam[31] and of HC were solved by X-ray crystallography[28]. A decade after the structure determination of TC, K. Locher and coworkers solved the structure of TC in complex with the soluble part of CD320[32]. The structure revealed the binding interface of TC:CD320 and helped in rationalizing the pH-dependent endosomal release of TC

from CD320. Nonetheless, open questions about the TC:CD320 receptor ligand interactions remained: The EGF-like domain of the receptor CD320, as well as a loop in TC, which is located close to the binding interface with LDLR-A1 of CD320, could not be resolved by the structure, but would be needed to be able to fully rationalize the TC:CD320 receptor-ligand binding interactions.

1.2 The resolution revolution in cryo-electron microscopy

For decades, the only possible ways for high resolution structure determination were X-ray and neutron diffraction crystallography, as well as solution NMR. Solution NMR was very limited to relatively small proteins and therefore never became a broadly used structural technique. X-ray crystallography on the other hand was extremely useful for the structure determination of globular and stable proteins and to this day is the most used method for protein structure determination. The need for well-ordered protein crystals however, brought various limitations for this technique. X-ray crystallography became particularly challenging for the work with proteins that were unfavorable for forming rigid lattice contacts. Among these were unstable proteins, such as membrane proteins or large and therefore inherently flexible complexes, such as ribosomes or viruses. Another limitation of X-ray crystallography was the need for large amounts of highly concentrated protein, for the often extensive screening of crystallization conditions. High yields of protein are particularly challenging to obtain with certain membrane proteins as well as with large protein complexes that cannot be overexpressed.

In single particle cryo-EM one could use a comparably small sample of the solubilized protein of interest[35] and most importantly one wouldn't need crystals. The sample would be dispersed over a pinhead sized grid with very small holes, leading to a thin film of solution, which then could be vitrified in liquid ethane and subsequently illuminated and recorded under the electron microscope[35]. Single particle cryo-electron microscopy (cryo-EM) had been used for a long time for the low-resolution structure determination of large macromolecules. The technique was frequently used to determine the shape of macromolecular molecules, which could then be used to dock in individual high-resolution crystal structures[36-40]. However, due to the nature of the indirect CCD detectors, these microscopes were restricted to resolving blobs at low resolution[41].

Figure 4 X-ray crystallography vs single particle cryo-EM. Schematic comparison of the protein structure determination techniques X-ray crystallography and single particle cryo-EM. Left: In X-ray crystallography a single crystal composed of millions of proteins is illuminated by an X-ray beam. The recorded diffraction pattern is then used to calculate the three-dimensional structure of the unit cell, respectively of the proteins in the crystal. Right: In Single particle cryo-EM a protein sample in solution is vitrified in a thin layer of ice. The ice is then illuminated with an electron beam and thousands of images containing the projections of hundreds of hundreds of proteins are recorded. The individual protein particles are then combined and averaged to obtain a three-dimensional structure of the average particle respectively protein. Figure from[1]

In 2014 when I started my doctoral studies, the so-called resolution revolution[41] had just begun. After a technological improvement of the detectors for electron microscopes, new direct electron detectors made it possible to record data at atomic resolution[41]. With this improvement a paradigm shift in structural biology had started. By averaging the projections of thousands of measured single particles, it became possible to determine structures of proteins at side-chain resolution[42-45] (Fig. 4). In the beginning, the technique was only used for large complex macromolecules, since the size of the target molecule plays an important role in the 3-dimensional reconstruction[46]. Nonetheless, the resolved targets became smaller and smaller until in 2013 Y. Cheng and co-workers determined the first structure of a membrane protein at side-chain resolution by cryo-EM[47].

Our group followed the new technological development early on and started shifting the focus of the entire lab into cryo-EM. Given the technological revolution, I also decided to tackle a cryo-EM project, despite by background in X-ray crystallography. The protein, that I decided to work on however, was still considered too small for conventional methods in cryo-EM and would therefore pose a new challenge in cutting the edge of the scope of cryo-EM.

1.3 N-linked glycosylation

N-linked glycosylation of proteins is one of the most common post translational modification in eukaryotes and has been observed in all kingdoms of life[48,49]. More than half of all proteins in nature have been estimated to be glycosylated[50], as well as virtually all proteins found in the human serum[51] and virtually all proteins in the cell membranes of animals[52].
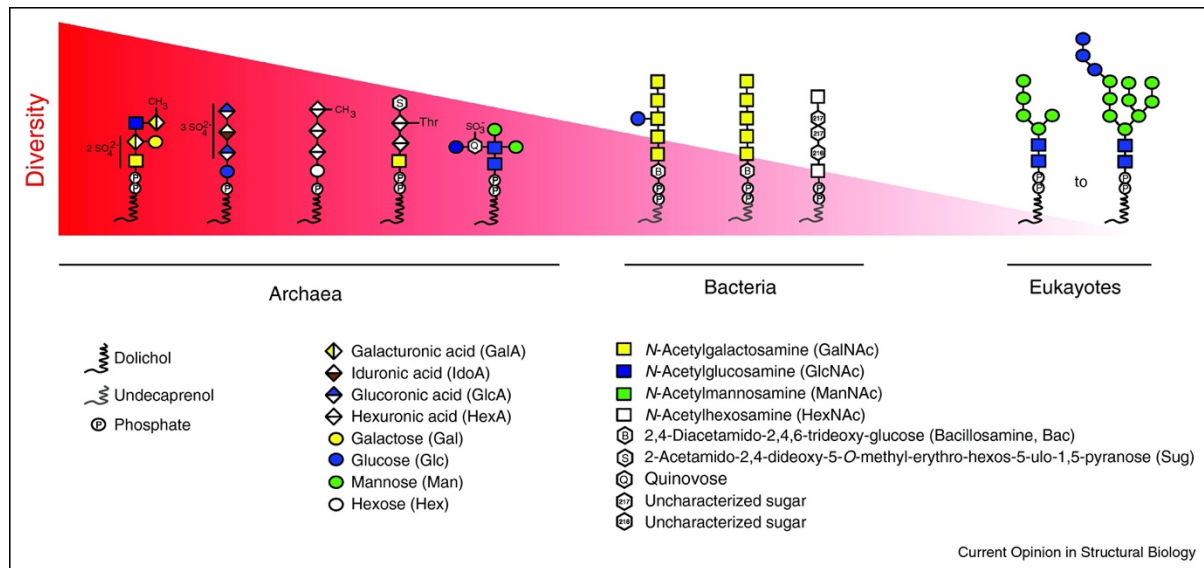
Figure 5 Lipid linked oligosaccharide diversity. Overview of representative structures of lipid-linked oligosaccharide (LLO) donor substrates used in N-glycosylation in prokaryotes and eukaryotes. In archaea the diversity of the used LLOs is the highest. Bacteria have an intermediate diversity of used LLOs, and eukaryotes are almost exclusively restricted to a Dol-PP-GlcNAc$_2$Man$_9$Glc$_3$ as N-glycan donor LLO. Figure from[48].

With 90% of all glycoproteins having been predicted to be N-glycosylated, N-glycosylation is the most common form of protein glycosylation, followed by O-glycosylation[50]. In the process of N-linked glycosylation oligosaccharyltransferase (OST) transfers an initially lipid-linked oligosaccharide from the lipid onto an asparagine residue in an Asn-X-Ser/Thr sequon, where X denotes any amino acid except proline, of a secretory- or membrane protein[53,54]. In eukaryotes this process takes place in the lumen of the endoplasmic reticulum (ER) and in prokaryotes in the periplasm. In the eukaryotic cell, the freshly N-glycosylated protein undergoes steps in glycan trimming and quality control in the ER, before it is being trafficked into the Golgi[55]. There numerous possible steps in glycan modification lead to a large variety of possible final glycans[55]. The combined glycans respectively the glycome displayed on the surface on a cell serve as a marker which makes it possible to differentiate

between different species, different tissues and even between different developmental stages of tissues[55]. For example, the blood type of an individual is determined by the glycosylated lipids and glycoproteins that are displayed by their erythrocytes[55].
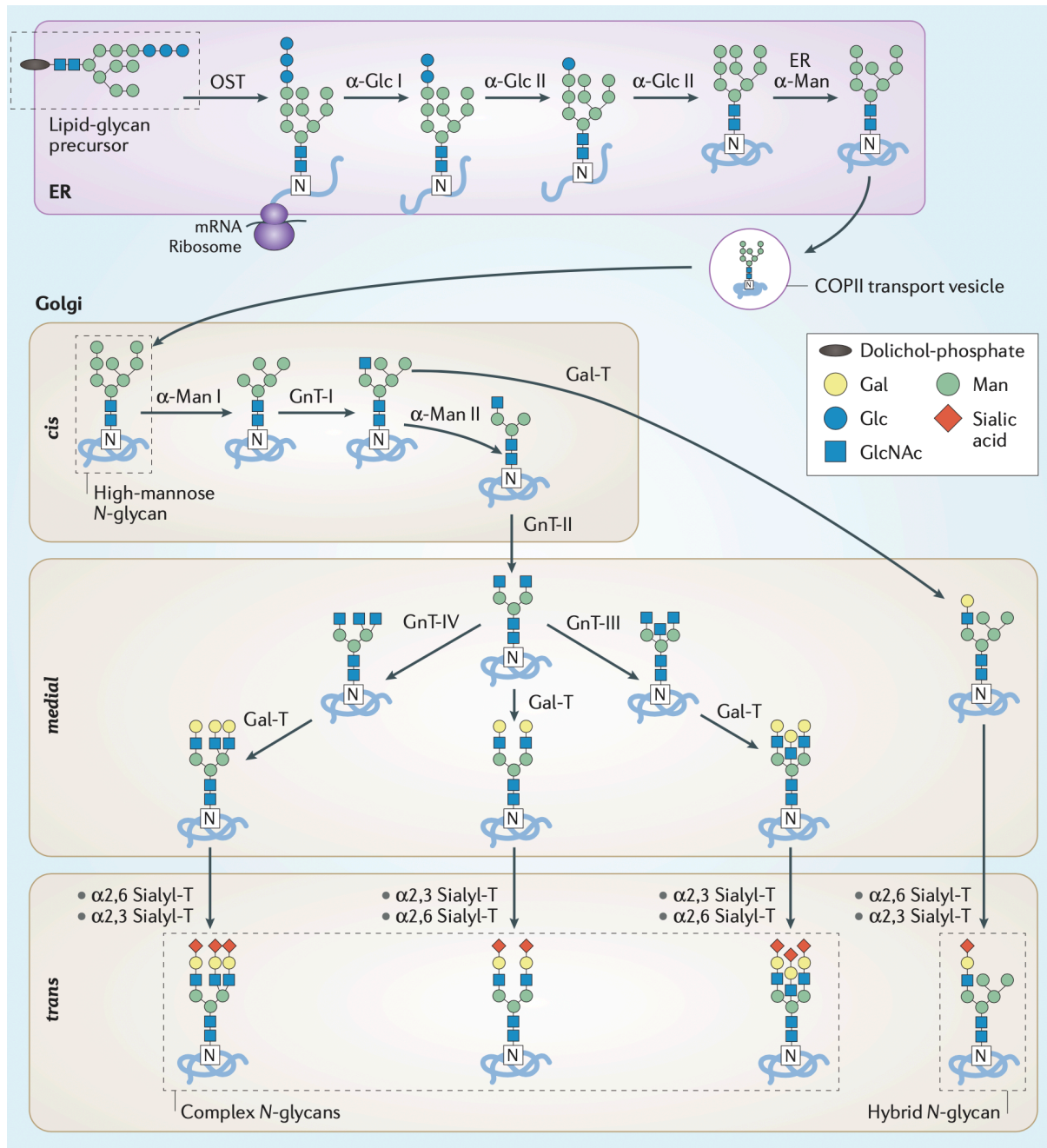
Figure 6 Human N-glycan processing and diversification. In higher eukaryotes such as mammals the N-glycan undergoes a number of steps of diversifications after the oligosaccharyltransfer onto the protein. In the ER, all glucoses and one mannose are removed from the glycan before the glycoprotein moves into the Golgi. There, glycosylhydrolases remove more hexoses and a variety of glycosyltransferases add different sugars, leading to a wide variety of possible final complex and hybrid N-glycans. The figure shows only a small fraction of final possible outcomes. Figure from[55].

The spectrum of different glycans that can be generated by the cellular machinery and displayed on a cell surface is so large and complex that one also refers to glycosylation as the third alphabet or language of life[56].

1.3.1 Structure of the LLO

Among all kingdoms of life N-glycans are generated in a step wise synthesis, leading to a lipid linked oligosaccharide (LLO), which is transferred *en bloc* onto an acceptor peptide by OST[57]. The lipidic, membrane embedded part of the LLO is comprised of a dolichol moiety in archaea and eukaryotes or an undecaprenol moiety in bacteria[57]. The oligosaccharide is linked to the lipid moiety via a pyro- or in certain archaea a monophosphate group[48]. Whereas in archaea the LLO structures are highly diverse, the diversity is smaller in bacteria and highly conserved in eukaryotes[48] (Fig. 5). With the exception of a few eukaryotic species that have diversified early in evolution, the eukaryotic LLO is represented exclusively by a Dol-PP-$GlcNAc_2Man_9Glc_3$ molecule[48,58]. Here, however the diversification of the N-glycan follows a after the oligosaccharyl-transfer. As mentioned earlier, in the ER and later in the Golgi a number of glycosidases and glycosyl transferases trim down and rebuild the N-glycan, resulting in a plethora of possible final complex glycans[55] (Fig. 6).

1.3.2 Assembly of the eukaryotic LLO

Beginning with an ER membrane resident dolichylphosphate molecule that is facing the cytoplasm, ALG7 and ALG13/ALG14 add one phosphate and the first two GlcNAc molecules to the emerging LLO using cytosolic UDP-GlcNAc as donor substrate[59] (Fig. 7). Subsequently, the cytosolic, glycosyltransferases ALG1, ALG2 and ALG11 add a total of five mannoses leading to $Man_5GlcNAc_2$-PP-Dol[59]. Then, the glycan is flipped into the ER lumen[59]. The protein RFT1 has been proposed to be the responsible flippase[60], which is however under debate[61-64]. Subsequently, six evolutionarily related transmembrane proteins of the glycosyltransferase superfamily class C (GT-C) complete the buildup of the emerging LLO by adding additional hexoses to it[59,65-67]. ALG3, ALG9 and ALG12 transfer a total of four mannoses and ALG6, ALG8 and ALG10 transfer each one glucose molecule to the LLO[59]. The donor-substrates of these ALG enzymes are either dolichylphosphate-linked mannose (Dol-P-Man) or dolichylphosphate-linked glucose (Dol-P-Glc)[59]. The reason for these ER-

resident sugar donor-substrates to be lipid-linked may be for evolutionary reasons: In pre-eukaryotic times, respectively for bacteria, what is now the ER lumen would have been the periplasm. Therefore, soluble, secreted sugars such as GDP-Man would have been lost quickly by diffusion, which is prevented by the linkage to the membrane[68,69]. The sugar donor molecules in the Golgi are again soluble nucleotide-linked sugars which are brought into the organelle by transport proteins[70]. Also the Golgi resident glycosyl transferases are all soluble, since GT-C enzymes are only thought to be found in the ER or in the plasma membrane[71]. The reason why the LLO synthesis is completed in the ER lumen and not in the cytosol, may be, because it might not be possible to flip larger hydrophilic groups than a $Man_5GlcNAc_2$ oligosaccharide across the membrane.
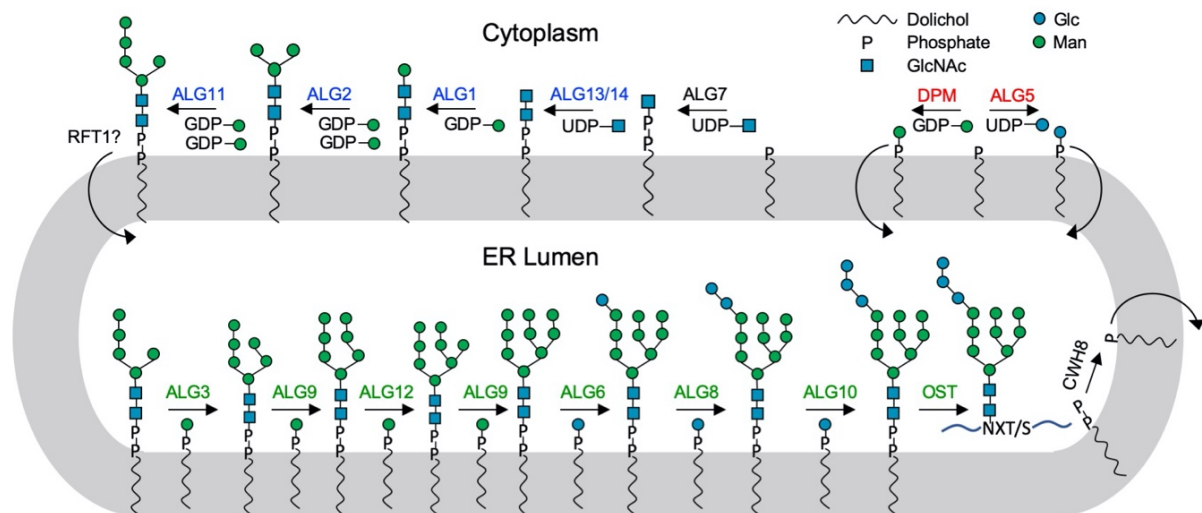
Figure 7 Biosynthetic pathway of N-glycan biosynthesis. Schematic of the biosynthetic reactions and reaction intermediates involved in lipid-linked oligosaccharide synthesis and transfer of the LLO onto an acceptor peptide, by OST[59]. The responsible enzymes are indicated above the arrows. Donor substrates are indicated below the arrows. Glycosyltransferases of the superfamily GT-A are colored in red, GT-B in blue and GT-C in green[65]. Note that ALG7 does not fall into any of the canonical families of glycoylstransferases. Figure adapted from[72].

Whereas genetic studies have helped to assign the enzymes responsible for the glycosyltransfer reactions[57,73-80], structurally and mechanistically only very little is known about the enzymatic pathway of LLO biosynthesis. In 2011 the first high-resolution structure of the bacterial OST PglB was reported[81], which was followed by high-resolution structures of archaeal-[82], yeast-[83,84] and recently human OST[85]. Also known are the structures of the N-acetylglucosaminephosphotransferase ALG7[86] and of the Dolichol phosphate mannose synthase DPMS[87], which is responsible for the regeneration of Dol-P to Dol-P-Man. The cytosolic assembly of the LLO, catalyzed by

ALG1, ALG2 and ALG11 has been studied mechanistically and a robust protocol for the chemoenzymatic production of pure Dol-PP-GlcNAc$_2$Man$_5$ has been established[88]. Also, the function of different OSTs have been studied extensively[85,89-91]. However, very little is known about the ER-luminal ALG pathway. Apart from the genomic assignment of the respective enzymes[57,92-94], the one pot synthesis of phytanyl-linked PP-GlcNAc$_2$Man$_9$ using enzymes ALG1 to ALG9 has been demonstrated[95]. Structural and detailed mechanistic understanding of ER-luminal ALG enzymes however remained absent.

1.3.3 Glycosyltransferase folds

The large number of known glycosyltransferases have been categorized into the three structural superfamilies GT-A, GT-B and GT-C[71]. Members of the superfamilies GT-A and GT-B are soluble proteins composed of either one (GT-A) or two (GT-B) Rossman-fold like domains[96] (Fig. 8). Variable loops on the common scaffold enable these structurally similar GTs to accommodate a large variety of substrates[96]. In some cases these glycosyltransferases are tethered to the membrane with additional TM helices, as it for example has been observed for DPMS[87] and as it has been predicted for ALG1, ALG2 and ALG11[88]. Both superfamilies contain enzymes that catalyze inverting and retaining glycosyltransfer reactions[96]. These are reactions that either invert or retain the stereochemistry at the anomeric center of the transferred sugars. GT-A enzymes bear a characteristic DXD motif and employ divalent cations for catalysis[96]. GT-B enzymes on the other hand have been proposed to neither require divalent cations nor do they have a DXD motif[96].

Whereas the GT-A and GT-B superfamilies have been well characterized with a large number of representative crystal structures, much less is known about the GT-C superfamily[96]. This superfamily describes transmembrane proteins with a core region, that is composed of transmembrane alpha-helices and a catalytic site which is facing the ER-lumen in eukaryotes, or the periplasm in prokaryotes. Some GT-C members have been shown to have an additional ER-luminal respectively periplasmic, soluble domain[81,97,98]. Also, all of these proteins require isoprenoid-linked donor substrates, possibly for the evolutionary reasons described in the previous section. Since only a small number of structures from different subfamily members of the GT-C superfamily have been determined, no common motifs in the fold have been described yet.

The first reported structure from a member of the GT-C superfamily was that of the bacterial OST PglB from *C. lari,* a member of GT66[81]. Subsequently reported structures of GT66 were that of archaeal[82], yeast[83,84,99] and human OST. Other known structures of GT-C members are that of the bacterial arabinosyltransferase ArnT[98] from GT83 and yeast PMT1 and PMT2[97] from GT39. All members of the ER-luminal ALG pathway are members from the GT-C[65] superfamily but come from different GT families, that all have not been structurally characterized yet. According to predictions, there are conserved motifs in certain ER-luminal loops of ALG enzymes, but the number of TM helices may vary[66,67].
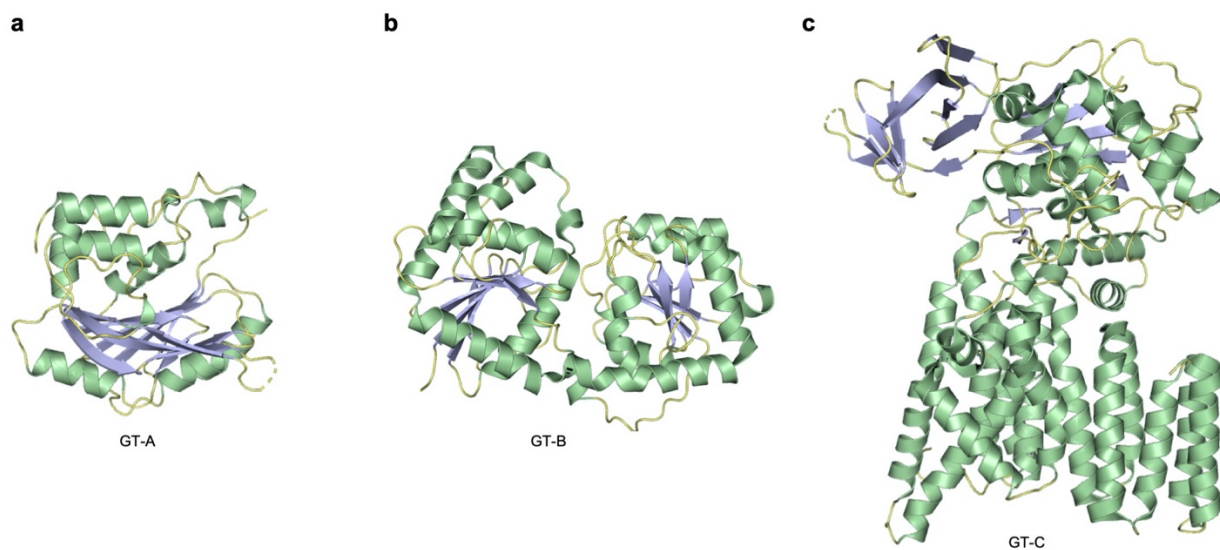
Figure 8 Structural comparison of glycosyltransferase superfamily folds. Representative GT superfamily structures are shown in cartoon representation with alpha-helices in green, beta strands in blue and loops in yellow. **a,** The GT-A fold is composed of a single Rossman fold type domain. The representative structure is from the inverting gylcosyltransferase SpsA from B. subtilus (PDB ID 1QGQ[100]). **b,** The GT-B fold is composed of two Rossman fold type domains. The representative structure is from the retaining glycosyltransferase PglH from C. jejuni (PDB ID 6EJI[101]. **c,** The GT-C fold consists of a bundle of TM-helixes with variable topology and a catalytic center at the periplasmic, respectively ER-liminal side of the protein. Some GT-C fold members have a large soluble periplasmic, respectively ER-luminal domain. The representative structure is from the oligosaccharyltransferase PglB from C. lari (PDB ID 3RCE[81]). Figure and legend adapted form[71].

1.3.4 ER-luminal glucosyltransferase ALG6

ALG6 is an ER-luminal glycosyltransferase, which is part of the enzymatic pathway of LLO-biosynthesis[57,80] (Fig. 7). Like the other ER-luminal glycosyltransferases of this pathway, ALG6 is a member of the GT-C superfamily[65]. ALG6 is an inverting glycosyltransferase, that adds the first glucose to the emerging glycan in an alpha-1-3-linkage. This particular glucose is important for protein folding and quality control, as it later plays an important role in the calnexin/calreticulin cycle[102]. The locus of

ALG6 was first identified in 1984, by P. Robbins and coworkers[80]. A number of genetic mutations in the *ALG6* gene have been associated with congenital disorders of glycosylation (CDG)[103]. As a consequence of these CDGs the patients suffer from hypoglycosylation of glycoproteins [104,105]. However, without any structural information these defects cannot be rationalized. There have been efforts in the past to utilize ALG6 for biotechnological applications in glycan engineering. N. Callewaert and coworkers have used as strain of *Y. lipolytica* with a knockout of ALG3 to generate $GlcNAc_2Man_5Glc_1$ glycans which were enzymatically converted into $GlcNAc_2Man_3$ glycans. The obtained glycoproteins were intended as a starting point for the engineering of recombinantly expressed, humanized glycoproteins [106].

1.4 Aims of this thesis

Part 1:

Humans depend on dietary uptake of vitamin B12 from animal sources[10]. The uptake is mediated by a complex multistep pathway[3]. Malfunction of any of the proteins involved in vitamin B12 uptake can lead to vitamin B12 deficiency, which in turn will impair the development of the individual[107,108].

Unlike HC, TC is highly specific for the binding to vitamin B12[3]. However, it also binds to modified B12 molecules such as antivitamin B12[34,109,110]. Antivitamin B12 is a non-reactive form of B12 that has been developed to quickly induce B12 deficiency in a patient or animal for clinical studies[111]. This would be useful because due to storage in the liver this depletion of vitamin B12 can take months[111].

In the first chapter, I explore the substrate specificity of TC. As it is virtually impossible to obtain apo TC, without denaturing the protein, I express one subunit of TC and study its binding to different substrates including antivitamin B12 by X-ray crystallography. Moreover, I test the whether the assembly of the alpha and the beta subunit of TC is triggered by addition of B12. In the second chapter I study the details of the TC:CD320 binding interaction by co-crystallizing the complex with conformational nanobodies.

As vitamin B12 is vital for DNA synthesis it is essential cell for division and cancer growth. The expression of the TC receptor CD320 is cell cycle associated and is upregulated in a number of in cancer cells[25,26]. It therefore has been a long-standing effort to target the cellular vitamin B12 uptake route for cancer therapy[23,26,27].

In the second chapter, I describe an approach to target TC:CD320 mediated vitamin B12 uptake for the destruction of cancer cells. I generate conformational nanobodies against TC and against the complex of TC:CD320. I then test their potential in detecting and entering highly dividing cells, by fusing them to a toxin. The nanobody drug conjugates are then then tested for their effects on highly dividing human cells.

Part 2:

N-glycosylation is one of the most common prost translational modifications of proteins[48,49]. However, only very little is known about the mechanistic basis of the LLO-biosynthesis. In fact, apart from the very first enzyme in the pathway ALG7[86] (which is not even considered to be a glycosyltransferase[65]) not a single glycosyltransferase that is directly involved in the LLO assembly has been studied structurally.

In the third chapter, I study the functions of ER-luminal ALG enzymes that are involved in the assembly of the LLO. With the help of synthetic donor substrates, provided by the Reymond group I reconstitute a part of the ER-luminal ALG pathway to generate intermediates of the enzymatic pathway of LLO-biosynthesis. To study the structure of the ER-luminal glucosyltransferase ALG6 I generate a Fab binder in collaboration with the Kossiakoff group. The Fab is then used to study nanodisc reconstituted ALG6 by cryo-EM. With the help of a synthetic substrate form the Reymond group I further study the substrate binding interactions of ALG6 and validate my findings with site-directed mutant generation. As ALG6 is a member of the poorly understood GT-C superfamily, the structure is analyzed for common motifs with other GT-C members of known structure in order to rationalize the fold of GT-C enzymes.

1.5 References

1   Callaway, E. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature* **525**, 172-174, (2015).
2   Stubbe, J. Binding site revealed of nature's most beautiful cofactor. *Science* **266**, 1663-1664, (1994).
3   Nielsen, M. J., Rasmussen, M. R., Andersen, C. B. F., Nexo, E. & Moestrup, S. K. Vitamin B-12 transport from food to the body's cells-a sophisticated, multistep pathway. *Nat Rev Gastro Hepat* **9**, 345-354, (2012).
4   Kamper, M. J. & Hodgkin, D. C. Some observations on the crystal structure of a chlorine-substituted vitamin B12. *Nature* **176**, 551-553, (1955).
5   Hodgkin, D. C. *et al.* Structure of vitamin B12. *Nature* **178**, 64-66, (1956).
6   Hodgkin, D. C. X-ray analysis and the structure of vitamin B12. *Fortschr Chem Org Naturst* **15**, 167-220, (1958).
7   Eschenmoser, A. & Wintner, C. E. Natural product synthesis and vitamin B12. *Science* **196**, 1410-1420, (1977).
8   Martens, J. H., Barg, H., Warren, M. J. & Jahn, D. Microbial production of vitamin B12. *Appl Microbiol Biotechnol* **58**, 275-285, (2002).
9   LeBlanc, J. G. *et al.* Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current opinion in biotechnology* **24**, 160-168, (2013).
10   Subar, A. F., Krebs-Smith, S. M., Cook, A. & Kahle, L. L. Dietary sources of nutrients among US adults, 1989 to 1991. *J Am Diet Assoc* **98**, 537-547, (1998).
11   Allen, R. H., Seetharam, B., Podell, E. & Alpers, D. H. Effect of proteolytic enzymes on the binding of cobalamin to R protein and intrinsic factor. In vitro evidence that a failure to partially degrade R protein is responsible for cobalamin malabsorption in pancreatic insufficiency. *J Clin Invest* **61**, 47-54, (1978).
12   Gordon, M. M., Hu, C., Chokshi, H., Hewitt, J. E. & Alpers, D. H. Glycosylation is not required for ligand or receptor binding by expressed rat intrinsic factor. *Am J Physiol* **260**, G736-742, (1991).
13   Birn, H. *et al.* Characterization of an epithelial approximately 460-kDa protein that facilitates endocytosis of intrinsic factor-vitamin B12 and binds receptor-associated protein. *J Biol Chem* **272**, 26497-26504, (1997).
14   Kozyraki, R. *et al.* The human intrinsic factor-vitamin B12 receptor, cubilin: molecular characterization and chromosomal mapping of the gene to 10p within the autosomal recessive megaloblastic anemia (MGA1) region. *Blood* **91**, 3593-3600, (1998).
15   Beedholm-Ebsen, R. *et al.* Identification of multidrug resistance protein 1 (MRP1/ABCC1) as a molecular gate for cellular export of cobalamin. *Blood* **115**, 1632-1639, (2010).
16   Morkbak, A. L., Poulsen, S. S. & Nexo, E. Haptocorrin in humans. *Clin Chem Lab Med* **45**, 1751-1759, (2007).
17   Hall, C. A. The carriers of native vitamin B12 in normal human serum. *Clin Sci Mol Med* **53**, 453-457, (1977).
18   Quadros, E. V. Advances in the understanding of cobalamin assimilation and metabolism. *Brit J Haematol* **148**, 195-204, (2010).

19   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. The protein and the gene encoding the receptor for the cellular uptake of transcobalamin-bound cobalamin. *Blood* **113**, 186-192, (2009).

20   Rudenko, G. *et al.* Structure of the LDL receptor extracellular domain at endosomal pH. *Science* **298**, 2353-2358, (2002).

21   Rutsch, F. *et al.* Identification of a putative lysosomal cobalamin exporter altered in the cblF defect of vitamin B12 metabolism. *Nat Genet* **41**, 234-239, (2009).

22   Coelho, D. *et al.* Mutations in ABCD4 cause a new inborn error of vitamin B12 metabolism. *Nat Genet* **44**, 1152-1155, (2012).

23   Clardy, S. M., Allis, D. G., Fairchild, T. J. & Doyle, R. P. Vitamin B-12 in drug delivery: breaking through the barriers to a B-12 bioconjugate pharmaceutical. *Expert Opin Drug Del* **8**, 127-140, (2011).

24   Amagasaki, T., Green, R. & Jacobsen, D. W. Expression of transcobalamin II receptors by human leukemia K562 and HL-60 cells. *Blood* **76**, 1380-1386, (1990).

25   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. Targeted Delivery of Saporin Toxin by Monoclonal Antibody to the Transcobalamin Receptor, TCblR/CD320. *Mol Cancer Ther* **9**, 3033-3040, (2010).

26   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. Saporin Conjugated Monoclonal Antibody to the Transcobalamin Receptor TCblR/320 Is Effective in Targeting and Destroying Cancer Cells. *Journal of cancer therapy* **4**, 1074-1081, (2013).

27   Waibel, R. *et al.* New derivatives of vitamin B12 show preferential targeting of tumors. *Cancer Res* **68**, 2904-2911, (2008).

28   Furger, E., Frei, D. C., Schibli, R., Fischer, E. & Prota, A. E. Structural Basis for Universal Corrinoid Recognition by the Cobalamin Transport Protein Haptocorrin. *J Biol Chem* **288**, 25466-25476, (2013).

29   Mathews, F. S. *et al.* Crystal structure of human intrinsic factor: Cobalamin complex at 2.6-A resolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 17311-17316, (2007).

30   Wuerges, J. *et al.* Structural basis for mammalian vitamin B12 transport by transcobalamin. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4386-4391, (2006).

31   Andersen, C. B. F., Madsen, M., Storm, T., Moestrup, S. K. & Andersen, G. R. Structural basis for receptor recognition of vitamin-B-12-intrinsic factor complexes. *Nature* **464**, 445-U147, (2010).

32   Alam, A. *et al.* Structural basis of transcobalamin recognition by human CD320 receptor. *Nat Commun* **7**, 12100, (2016).

33   Fedosov, S. N. *et al.* Assembly of the intrinsic factor domains and oligomerization of the protein in the presence of cobalamin. *Biochemistry-Us* **43**, 15095-15102, (2004).

34   Bloch, J. S., Ruetz, M., Krautler, B. & Locher, K. P. Structure of the human transcobalamin beta domain in four distinct states. *Plos One* **12**, e0184932, (2017).

35   Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* **294**, 5181-5197, (2019).

36   Pfeffer, S. *et al.* Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nat Commun* **5**, 3072, (2014).

37   Merz, F. *et al.* Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *Embo J* **27**, 1622-1632, (2008).

38   Boehringer, D., Ban, N. & Leibundgut, M. 7.5-A cryo-em structure of the mycobacterial fatty acid synthase. *J Mol Biol* **425**, 841-849, (2013).

39   Chang, L. F., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Molecular architecture and mechanism of the anaphase-promoting complex. *Nature* **513**, 388-393, (2014).

40   Nguyen, T. H. *et al.* The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47-52, (2015).

41   Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **343**, 1443-1444, (2014).

42   Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485-1489, (2014).

43   Greber, B. J. *et al.* The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283-286, (2014).

44   Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S. & Carragher, B. 2.8 A resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *Elife* **4**, (2015).

45   Chen, J. Z. *et al.* Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10644-10648, (2009).

46   Wu, S. *et al.* Fabs enable single particle cryoEM studies of small proteins. *Structure* **20**, 582-592, (2012).

47   Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-112, (2013).

48   Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* **21**, 576-582, (2011).

49   Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**, 765-778, (2010).

50   Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8, (1999).

51   Sharon, N. in *Complex Carbohydrates, their Chemistry, Bio-synthesis and Functions.*   33-35 (Addison Wesley, Reading, MA, 1975).

52   Gahmberg, C. G. & Tolvanen, M. Why mammalian cell surface proteins are glycoproteins. *Trends in biochemical sciences* **21**, 308-311, (1996).

53   Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54**, 631-664, (1985).

54   Chavan, M. & Lennarz, W. The molecular basis of coupling of translocation and N-glycosylation. *Trends in biochemical sciences* **31**, 17-20, (2006).

55   Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nat Rev Nephrol* **15**, 346-366, (2019).

56   Kaltner, H., Abad-Rodriguez, J., Corfield, A. P., Kopitz, J. & Gabius, H. J. The sugar code: letters and vocabulary, writers, editors and readers and biosignificance of functional glycan-lectin pairing. *Biochem J* **476**, 2623-2655, (2019).

57   Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1426**, 239-257, (1999).

58   Bushkin, G. G. *et al.* Suggestive evidence for Darwinian Selection against asparagine-linked glycans of Plasmodium falciparum and Toxoplasma gondii. *Eukaryot Cell* **9**, 228-241, (2010).

59   Aebi, M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* **1833**, 2430-2437, (2013).

60   Helenius, J. *et al.* Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415**, 447-450, (2002).

61   Frank, C. G., Sanyal, S., Rush, J. S., Waechter, C. J. & Menon, A. K. Does Rft1 flip an N-glycan lipid precursor? *Nature* **454**, E3-E4, (2008).

62   Helenius, J. *et al.* Does Rft1 flip an N-glycan lipid precursor? Reply. *Nature* **454**, E4-E5, (2008).

63   Sanyal, S., Frank, C. G. & Menon, A. K. Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum. *Biochemistry-Us* **47**, 7937-7946, (2008).

64   Rush, J. S., Gao, N. G., Lehrman, M. A., Matveev, S. & Waechter, C. J. Suppression of Rft1 Expression Does Not Impair the Transbilayer Movement of Man(5)GlcNAc(2)-P-P-Dolichol in Sealed Microsomes from Yeast. *J Biol Chem* **284**, 19835-19842, (2009).

65   Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495, (2014).

66   Oriol, R., Martinez-Duncker, I., Chantret, I., Mollicone, R. & Codogno, P. Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. *Mol Biol Evol* **19**, 1451-1463, (2002).

67   Albuquerque-Wendt, A., Hutte, H. J., Buettner, F. F. R., Routier, F. H. & Bakker, H. Membrane Topological Model of Glycosyltransferases of the GT-C Superfamily. *Int J Mol Sci* **20**, (2019).

68   Waechter, C. J. & Lennarz, W. J. The role of polyprenol-linked sugars in glycoprotein synthesis. *Annu Rev Biochem* **45**, 95-112, (1976).

69   Aebi, M.   (2018).

70   Parker, J. L. & Newstead, S. Gateway to the Golgi: molecular mechanisms of nucleotide sugar transporters. *Curr Opin Struct Biol* **57**, 127-134, (2019).

71   Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555, (2008).

72   Bloch, J. S. *et al.* Structure and mechanism of the ER-luminal glucosyltransferase ALG6. *Nature*, In press, (2020).

73   Reiss, G., te Heesen, S., Zimmerman, J., Robbins, P. W. & Aebi, M. Isolation of the ALG6 locus of Saccharomyces cerevisiae required for glucosylation in the N-linked glycosylation pathway. *Glycobiology* **6**, 493-498, (1996).

74   Imbach, T. *et al.* A mutation in the human ortholog of the Saccharomyces cerevisiae ALG6 gene causes carbohydrate-deficient glycoprotein syndrome type-Ic. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6982-6987, (1999).

75   Stagljar, I., te Heesen, S. & Aebi, M. New phenotype of mutations deficient in glucosylation of the lipid-linked oligosaccharide: cloning of the ALG8 locus. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5977-5981, (1994).

76   Verostek, M. F., Atkinson, P. H. & Trimble, R. B. Glycoprotein biosynthesis in the alg3 Saccharomyces cerevisiae mutant. I. Role of glucose in the initial

glycosylation of invertase in the endoplasmic reticulum. *J Biol Chem* **268**, 12095-12103, (1993).

77   Aebi, M., Gassenhuber, J., Domdey, H. & te Heesen, S. Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae. *Glycobiology* **6**, 439-444, (1996).

78   Henquet, M. *et al.* Identification of the gene encoding the alpha1,3-mannosyltransferase (ALG3) in Arabidopsis and characterization of downstream n-glycan processing. *Plant Cell* **20**, 1652-1664, (2008).

79   Sharma, C. B., Lehle, L. & Tanner, W. N-Glycosylation of yeast proteins. Characterization of the solubilized oligosaccharyl transferase. *Eur J Biochem* **116**, 101-108, (1981).

80   Runge, K. W., Huffaker, T. C. & Robbins, P. W. 2 Yeast Mutations in Glucosylation Steps of the Asparagine Glycosylation Pathway. *J Biol Chem* **259**, 412-417, (1984).

81   Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355, (2011).

82   Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17868-17873, (2013).

83   Bai, L., Wang, T., Zhao, G., Kovach, A. & Li, H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555**, 328-333, (2018).

84   Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **359**, 545-550, (2018).

85   Ramirez, A. S., Kowal, J. & Locher, K. P. Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science* **366**, 1372-1375, (2019).

86   Yoo, J. *et al.* GlcNAc-1-P-transferase-tunicamycin complex structure reveals basis for inhibition of N-glycosylation. *Nat Struct Mol Biol* **25**, 217-224, (2018).

87   Gandini, R., Reichenbach, T., Tan, T. C. & Divne, C. Structural basis for dolichylphosphate mannose biosynthesis. *Nat Commun* **8**, 120, (2017).

88   Ramirez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology*, 1-8, (2017).

89   Ramirez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27**, 525-535, (2017).

90   Napiorkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat Struct Mol Biol* **24**, 1100-1106, (2017).

91   Napiorkowska, M., Boilevin, J., Darbre, T., Reymond, J. L. & Locher, K. P. Structure of bacterial oligosaccharyltransferase PglB bound to a reactive LLO and an inhibitory peptide. *Sci Rep* **8**, 16297, (2018).

92   Huffaker, T. C. & Robbins, P. W. Yeast mutants deficient in protein glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 7466-7470, (1983).

93   Huffaker, T. C. & Robbins, P. W. Temperature-sensitive yeast mutants deficient in asparagine-linked glycosylation. *J Biol Chem* **257**, 3203-3210, (1982).

94   Sharma, C. B., Knauer, R. & Lehle, L. Biosynthesis of lipid-linked oligosaccharides in yeast: the ALG3 gene encodes the Dol-P-Man:Man5GlcNAc2-PP-Dol mannosyltransferase. *Biol Chem* **382**, 321-328, (2001).

95   Li, S. T. *et al.* Reconstitution of the lipid-linked oligosaccharide pathway for assembly of high-mannose N-glycans. *Nat Commun* **10**, 1813, (2019).

96   Moremen, K. W. & Haltiwanger, R. S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat Chem Biol* **15**, 853-864, (2019).

97   Bai, L., Kovach, A., You, Q., Kenny, A. & Li, H. Structure of the eukaryotic protein O-mannosyltransferase Pmt1-Pmt2 complex. *Nat Struct Mol Biol*, (2019).

98   Petrou, V. I. *et al.* Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351**, 608-612, (2016).

99   Braunger, K. *et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215-219, (2018).

100   Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. *Biochemistry-Us* **38**, 6380-6385, (1999).

101   Ramirez, A. S. *et al.* Structural basis of the molecular ruler mechanism of a bacterial glycosyltransferase. *Nat Commun* **9**, 445, (2018).

102   Williams, D. B. Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. *J Cell Sci* **119**, 615-623, (2006).

103   Haeuptle, M. A. & Hennet, T. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum Mutat* **30**, 1628-1641, (2009).

104   Jaeken, J., Lefeber, D. & Matthijs, G. Clinical utility gene card for: ALG6 defective congenital disorder of glycosylation. *Eur J Hum Genet* **23**, (2015).

105   Morava, E. *et al.* ALG6-CDG: a recognizable phenotype with epilepsy, proximal muscle weakness, ataxia and behavioral and limb anomalies. *J Inherit Metab Dis* **39**, 713-723, (2016).

106   De Pourcq, K. *et al.* Engineering Yarrowia lipolytica to produce glycoproteins homogeneously modified with the universal Man3GlcNAc2 N-glycan core. *Plos One* **7**, e39976, (2012).

107   Reynolds, E. Vitamin B12, folic acid, and the nervous system. *Lancet Neurol* **5**, 949-960, (2006).

108   Kozyraki, R. & Cases, O. Vitamin B12 absorption: Mammalian physiology and acquired and inherited disorders. *Biochimie* **95**, 1002-1007, (2013).

109   Mutti, E., Ruetz, M., Birn, H., Krautler, B. & Nexo, E. 4-Ethylphenyl-Cobalamin Impairs Tissue Uptake of Vitamin B-12 and Causes Vitamin B-12 Deficiency in Mice. *Plos One* **8**, (2013).

110   Ruetz, M., Salchner, R., Wurst, K., Fedosov, S. & Krautler, B. Phenylethynylcobalamin: a light-stable and thermolysis-resistant organometallic vitamin B(12) derivative prepared by radical synthesis. *Angew Chem Int Ed Engl* **52**, 11406-11409, (2013).

111   Ruetz, M. *et al.* Antivitamin B12 Inhibition of the Human B12-Processing Enzyme CblC: Crystal Structure of an Inactive Ternary Complex with Glutathione as the Cosubstrate. *Angew Chem Int Ed Engl* **56**, 7387-7392, (2017).

2. Structure of the human transcobalamin beta domain in four distinct states

Publication I - Structure of the human transcobalamin beta domain in four distinct states

The contents of this chapter were previously published in PLOS One:

Statement of Contribution

I performed all experiments in this study, except for the synthesis of 4-Ethylphenyl-Cobalamin which was performed by Dr. Markus Ruetz under the supervision of Prof. Bernhard Kräutler. I drafted the manuscript for publication and Prof. Kaspar Locher contributed to the final version of the paper.

Structure of the human transcobalamin beta domain in four distinct states

Joël. S. Bloch[1], Markus Ruetz[2], Bernhard Kräutler[2], Kaspar P. Locher[1]*

[1]Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland
[2]Institute of Organic Chemistry and Center of Molecular Biosciences, University of Innsbruck, Innrain, Innsbruck, Austria

*Corresponding author: locher@mol.biol.ethz.ch

Abstract

Vitamin B12 (cyanocobalamin, CNCbl) is an essential cofactor-precursor for two biochemi- cal reactions in humans. When ingested, cobalamins (Cbl) are transported via a multistep transport system into the bloodstream, where the soluble protein transcobalamin (TC) binds Cbl and the complex is taken up into the cells via receptor mediated endocytosis. Crystal structures of TC in complex with CNCbl have been solved previously. However, the initial steps of holo-TC assembly have remained elusive. Here, we present four crystal structures of the beta domain of human TC (TC-beta) in different substrate-bound states. These include the apo and CNCbl-bound states, providing insight into the early steps of holo-TC assembly. We found that in vitro assembly of TC-alpha and TC-beta to a complex was Cbl-dependent. We also determined the structure of TC-beta in complex with cobinamide (Cbi), an alternative substrate, shedding light on the specificity of TC. We finally determined the structure of TC-beta in complex with an inhibitory antivitamin B12 (anti-B12). We used this structure to model the binding of anti-B12 into full-length holo-TC and could rule out that the inhibitory function of anti-B12 was based on an inability to form a functional complex with TC.

Introduction

Vitamin B12 (cyanocobalamin, CNCbl) is essential for humans, as at least two biochemical pathways are B12 dependent[1-4]: It is a coenzyme precursor for cytoplasmic methionine synthase and mitochondrial methylmalonyl-CoA mutase. Because only certain prokaryotes are able to synthesize cobalamins (Cbls), humans are dependent on dietary intake of vitamin B12[5]. When ingested, the water-soluble vitamin is transported via a multistep transport system through the ileum into the bloodstream[6]. There, 10–30% of total Cbls are bound to transcobalamin (TC). The remaining Cbl and Cbl analogs are bound to haptocorrin (HC)[7]. Despite having a higher affinity for Cbls[8], HC has a lower substrate specificity than TC and is able to bind various Cbl analogs with similar affinities[9]. Cellular uptake of Cbls occurs via receptor-mediated endocytosis of holo-TC by means of the receptor CD320[10]. The structure of the extracellular domain of CD320 in complex with holo-TC was recently determined[11]. Cancer cells, being in elevated need of Cbls, overexpress the CD320 protein[12]. This makes TC:CD320-mediated uptake of Cbls an interesting target for

24

oncological research. TC consist of two domains (alpha and beta) that are connected by a flexible linker. Both domains contribute to Cbl binding, resulting in sub-pico-molar affinity (Fig 1A)[8,13]. It has been hypothesized that the two domains only assemble upon binding of Cbls to either one of the two domains (Fig 1B)[13,14].

To address the question of how Cbl mediated TC assembly proceeds, we determined crystal structures of apo TC-beta and TC-beta in complex with Cbl, putatively representing the first steps of TC assembly. Moreover, we determined the structure of TC-beta in complex with the substrate analog Cbi.

Antivitamins B12 are modified Cbls that are structurally similar to CNCbl, with the β-axial ligand being replaced by inert ligands such as difluorophenylethinyl (S1 Fig)[15-17]. When injected into mice, the anti-B12 compound 4-Ethylphenyl-Cobalamin (EtPhCbl) induced a pronounced Cbl-deficiency[18]. The use of anti-B12 molecules therefore provides a more efficient approach to study the effects of Cbl deficiency than feeding laboratory animals with Cbl- free nutrition, because the latter depends on the slow depletion of the Cbl storage in the liver[19]. Elevated Cbl levels in the blood of mice that have been supplied with EtPhCbl suggest an interference with the cellular uptake of Cbls. The molecular basis of this interference is unknown because it was demonstrated that anti-B12 can bind to TC and can be taken up into cells[18]. To test whether anti-B12 affects Cbl uptake via its interaction with TC, we deter- mined a co-crystal structure of TC-beta bound to Co-β-[2-(2,4- difluorophenyl)ethinyl]cobalamin (F2PhEtyCbl) at high resolution.
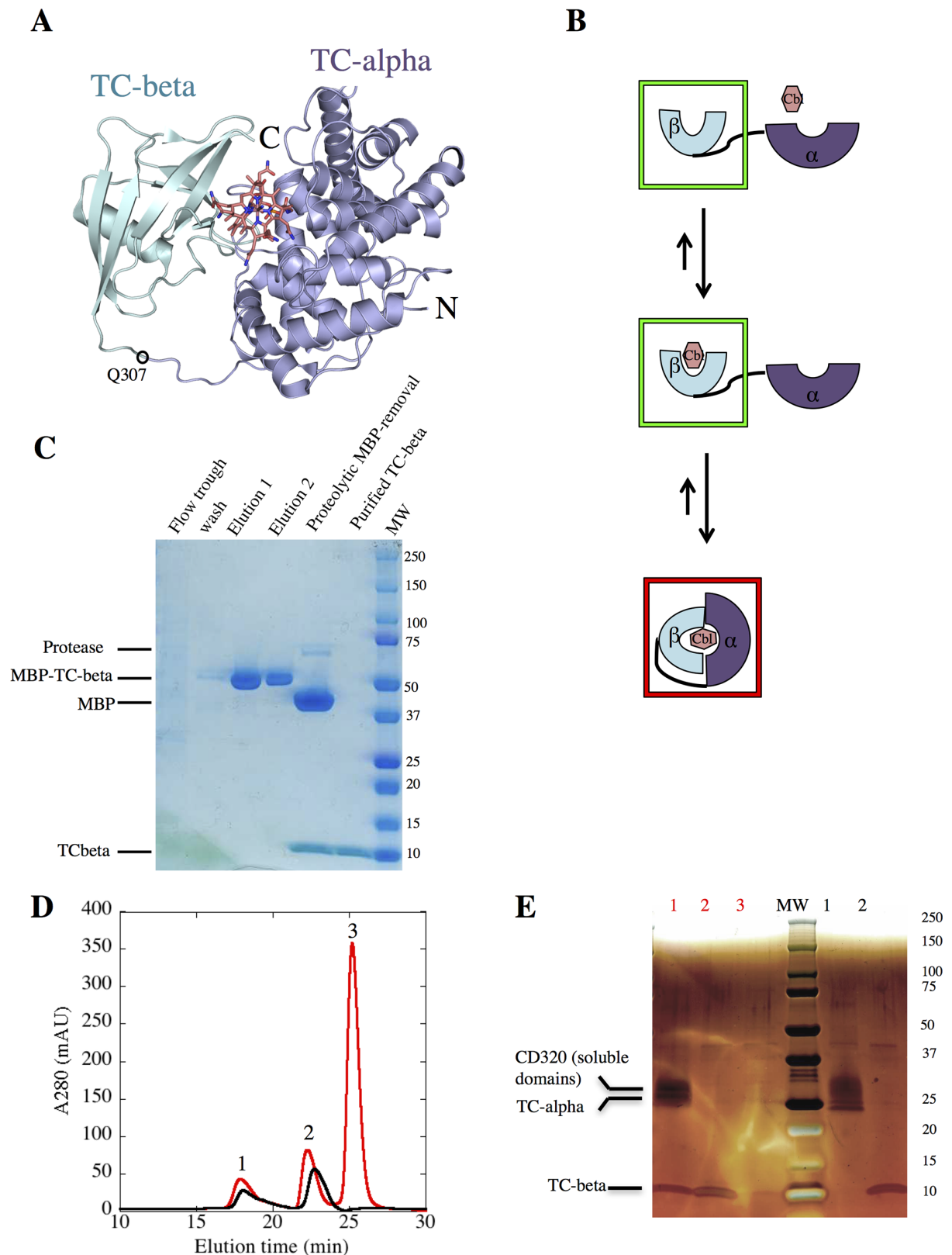
Figure 1 Cbl induced assembly of individually expressed TC domains. (A) PDB entry 4ZRP, holo-TC with the alpha domain (blue) and the beta domain (cyan) in cartoon representation and CNCbl in stick representation (carbons salmon) (receptor CD320 not shown). The TC-beta construct used in this study begins at Q307 and contains additional N-terminal residues from the TEV-cleavage site. (B) Hypothetical stages of TC-alpha, TC-beta assembly upon Cbl addition, green boxes indicate new, red boxes previously published crystal structures. (C) Coomassie stained SDS-PAGE analysis of TC-beta purification. (D) SEC (size exclusion chromatography) profile (TSK-G3000 column) of a mixture of TC-alpha:CD320, TC-beta in the presence (red) or absence (black) of CNCbl of a ratio of 1:4:16 (TC-alpha:TC-beta:CNCbl). (E) Silver stained SDS-PAGE of peak fractions from SEC.

Results

Expression of human TC-beta and Cbl-dependent assembly of holo-TC

The sub-picomolar affinity of TC to Cbl[6,20] and the presence of Cbls in eukaryotic expression media made it challenging to produce apo TC without denaturing the protein. We expressed the two domains of TC individually in insect cells, which lowered the affinity for Cbls while retaining the structural integrity of the domains, as no denaturing was required to generate the Cbl-free domains. TC-beta was expressed as a fusion protein with maltose binding protein (MBP). Extensive washing of the immobilized fusion protein was required to completely remove all bound Cbl. After proteolytic cleavage of the MBP and further affinity purification we obtained pure, substrate free TC-beta (Fig 1C). TC-alpha was co-expressed with the soluble/external domains of CD320, as this increased the stability and the yield of the pure protein. TC-alpha:CD320 did not contain bound Cbls after purification, as evidenced by the absence of a peak at 361 nm in the vis-spectrum.

In full-length TC, the alpha- and the beta domains are connected by a flexible linker. In the structure of CNCbl-bound TC, the alpha- and beta domains share a very small binding interface[13]. We therefore wanted to test the hypothesis that the two domains would only assemble upon Cbl binding (Fig 1B). A similar dependence has been observed previously with the alpha- and beta domains of the structurally similar protein intrinsic factor[21].

We incubated TC-beta with TC-alpha:CD320 in the presence or absence CNCbl. Subsequent size exclusion chromatography and SDS-PAGE analysis of the peak fractions revealed that TC-beta and TC-alpha:CD320 only co-eluted when CNCbl was present(Fig 1D and 1E). In contrast, the domains remained separated when no CNCbl was added.

Crystal structures of TC-beta in apo- and CNCbl-bound form

We first determined the crystal structure CNCbl-bound TC-beta at 1.43 Å resolution (Fig 2A, Table 1). It revealed a highly similar structure compared to the beta domain in full-length holo-TC[11] (rmsd = 0.39 Å for 97 Cα-atoms, Table 2). However, we observed that CNCbl was rotated out of the binding pocket by 9.2˚ compared to holo-TC (Fig 3A). As a consequence, not all H-bonds observed in CNCbl-bound TC were

27

formed. For example, the H-bond from the phosphate to the backbone of L358 was missing in the CNCbl-bound TC-beta structure.
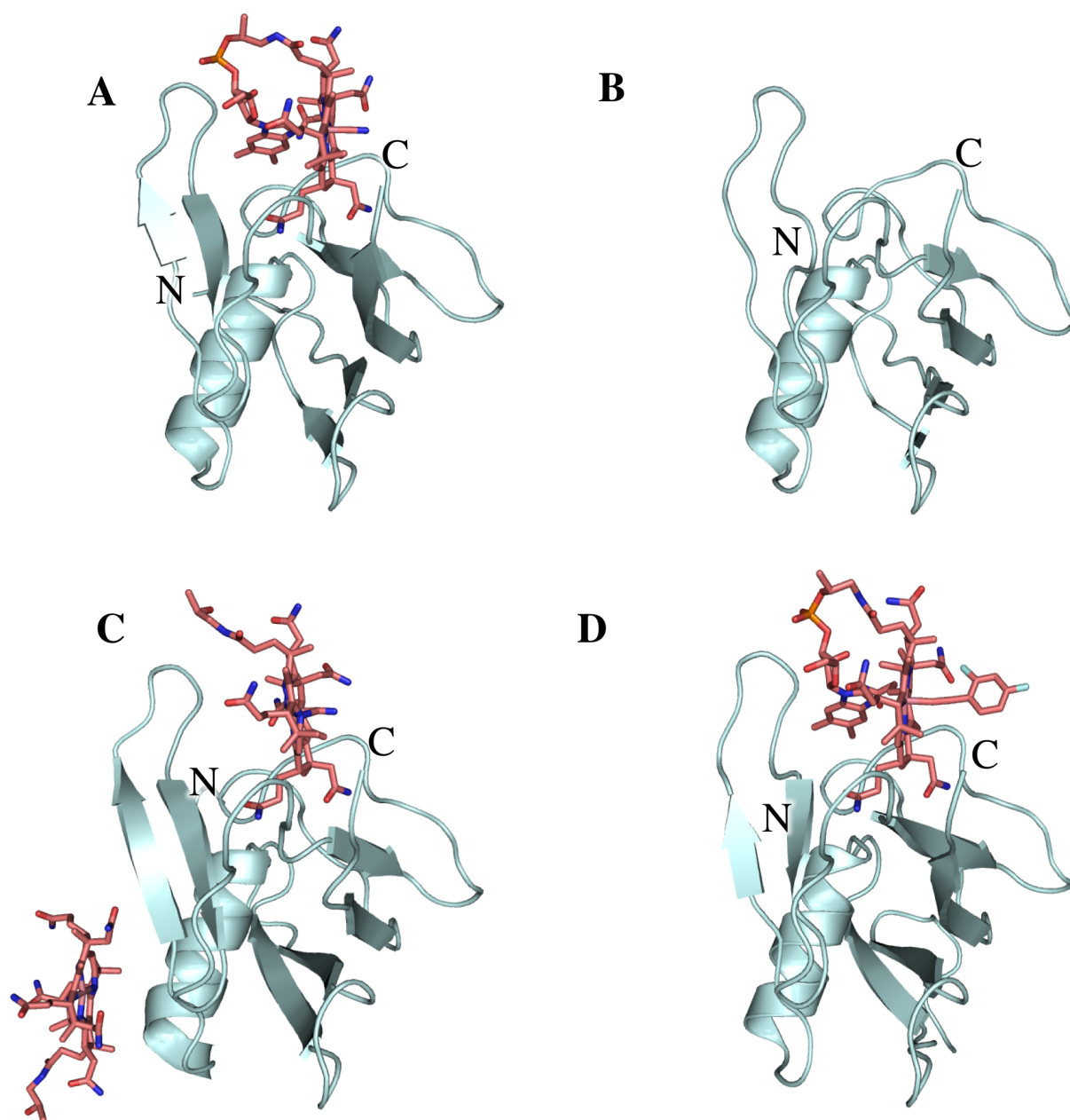
Figure 2 Crystal structures of TC-beta bound to different substrates and in apo form. Crystal structures of TC-beta in four different conformations such as CNCbl-bound TC-beta (A), apo TC-beta (B), Cbi-bound TC-beta (C) and anti-B12-bound TC-beta (D). All 3 ligands bind at the same site of TC-beta in roughly in the same orientation. In the structure of Cbi-bound TC-beta, an additional Cbi is bound to the surface of TC-beta.

We next determined the crystal structure of apo TC-beta at 1.57 Å resolution (Fig 2B, Table 1), which represents to our knowledge the first structure of TC resembling a putative partial apo state of TC in the apo form (i.e. TC-alpha and TC-beta are not associated with each other). The structure was highly similar to the Cbl bound state of

TC-beta (rmsd = 0.41 Å for 97 Cα-atoms, Table 2), with minor changes evident in several side chains compared to CNCbl- bound TC-beta.

Crystal structure of Cbi-bound TC-beta

Co-crystallization of TC-beta with Cbi yielded crystals diffracting to 1.57 Å resolution (Fig 2C, Table 1). The compound used was dicyano-Cbi, which, when dissolved in water, is converted to the stable monocyano-Co(III)cobinamide. To our surprise, the structure revealed two bound Cbi molecules (Cbi1 and Cbi2). Cbi1 was located in the binding pocket, was bound in a similar way as CNCbl in holo-TC. Notably, the corrin ring/plane of Cbi1 was not rotated compared to the structure of CNCbl-bound full-length TC. We observed a large electron density peak (S2 Fig) in the space occupied by the 5,6-dimethylbenzimidazole (DMB) moiety in Cbl-bound TC. This density was attributed to the N-terminus of a symmetry-related TC-beta molecule in the crystal lattice, with the central cobalt atom of Cbi1 coordinating the side chain of His305. A cyanide ligand bound to the central cobalt atom of Cbi1 could also be modeled (opposite side of His305).

The second Cbi molecule (Cbi2) was found attached to the surface of TC-beta. Its central cobalt atom was coordinated to the side chain of the surface-exposed His345. Here, the histidine was coordinated to the opposite side of the corrin ring compared to Cbi1. Apart from an H-bond to the backbone carbonyl of Tyr352, this was the only contact of Cbi2 to TC-beta. However, Cbi2 was also in contact with a TC-beta symmetry mate. The additional interactions probably caused Cbi2 to be ordered and well-resolved. For Cbi2, no electron density for a bound cyanide ligand was observed, suggesting that it had dissociated. To our knowledge this is the first report of a Cbi molecule non-specifically bound to the surface of a protein.

Table 1. Crystallographic data

Structure	Apo TC-beta	TC-beta:Cbl	TC-beta:anti-B12	TC-beta:Cbi
Crystals	1	1	1	1
Wavelength (Å)	1.0	1.0	1.0	1.0
Resolution range	41.72–1.57 (1.63–1.57)	30.12–1.43 (1.49–1.43)	43.85–1.27 (1.33–1.27)	41.68–1.57 (1.63–1.57)
Space group	P 32 2 1	P 32 2 1	P 21 21 2	P 32 2 1
Unit cell: a, b, c (Å)	59.98 59.98 70.04	60.24 60.24 69.57	61.72 62.30 33.76	59.86 59.86 70.11
α, β, γ (°)	90 90 120	90 90 120	90 90 90	90 90 120
Total reflections	203359 (18492)	257023 (22478)	137130 (7405)	203737 (17928)
Unique reflections	20772 (2015)	27022 (2563)	33543 (1161)	19877 (1790)
Multiplicity	9.8 (9.2)	9.5 (8.8)	6.3 (6.4)	10.2 (10.0)
Completeness (%)	99.76 (98.97)	99.23 (96.61)	96.19 (68.73)	95.53 (87.45)
Mean I/sigma(I)	16.80 (0.90)	22.25 (1.03)	24.40 (12.16)	16.79 (0.82)
R-merge[a]	0.075 (2.06)	0.046 (1.99)	0.043 (0.11)	0.05742 (2.31)
CC1/2	1 (0.33)	1 (0.61)	1 (1.00)	1 (0.52)
CC*	1 (0.70)	1 (0.87)	1 (1.00)	1 (0.83)
R-work (%)	15.5 (29.3)	15.6 (32.8)	14.4 (46.9)	17.9 (43.5)
R-free (%)	19.9 (32.0)	19.1 (35.6)	17.9 (48.9)	20.8 (45.5)
Number of non-hydrogen atoms	1017	1163	1231	1075
macromolecules	868	909	892	856
ligands	6	99	102	138
solvent	143	155	236	81
Protein residues	105	106	106	106
RMS(bonds) (Å)	0.005	0.008	0.017	0.014
RMS(angles) (°)	0.84	1	2.42	1.09
Ramachandran favored[b] (%)	97.09	99.04	99	96.15
Ramachandran allowed[b] (%)	1.94	0.96	0.96	2.88
Ramachandran outliers[b] (%)	0.97	0	0	0.96
Rotamer outliers[b] (%)	1.03	0.99	2.04	4.21
Clashscore[b]	4.55	10.39	5.64	16.19
Average B-factor	36.48	31.65	22.90	46.95
macromolecules	32.51	28.05	18.70	43.90
ligands	189	40.94	17.60	62.66
solvent	54.12	46.87	41.30	52.45

Highest resolution shell in parentheses.

[a] Rmerge = ∑hkl |I(hkl)–< I(hkl) > |/ ∑hkl(hkl), where < I(hkl) > is the mean of the symmetry equivalent reflections of I(hkl).

[b] As defined in MolProbity.

Table 2. RMSD values of TC-beta structures aligned to holoTC.

Protein	Rmsd (Å)
TC-beta:Cbl	0.39
Apo TC-beta	0.47
TC-beta:anti-B12	0.34
TC-beta:Cbi	0.48

Alignment of the TC-beta crystal structures to structure of human TC:CD320 (4KRP) for the residue range 313–409.
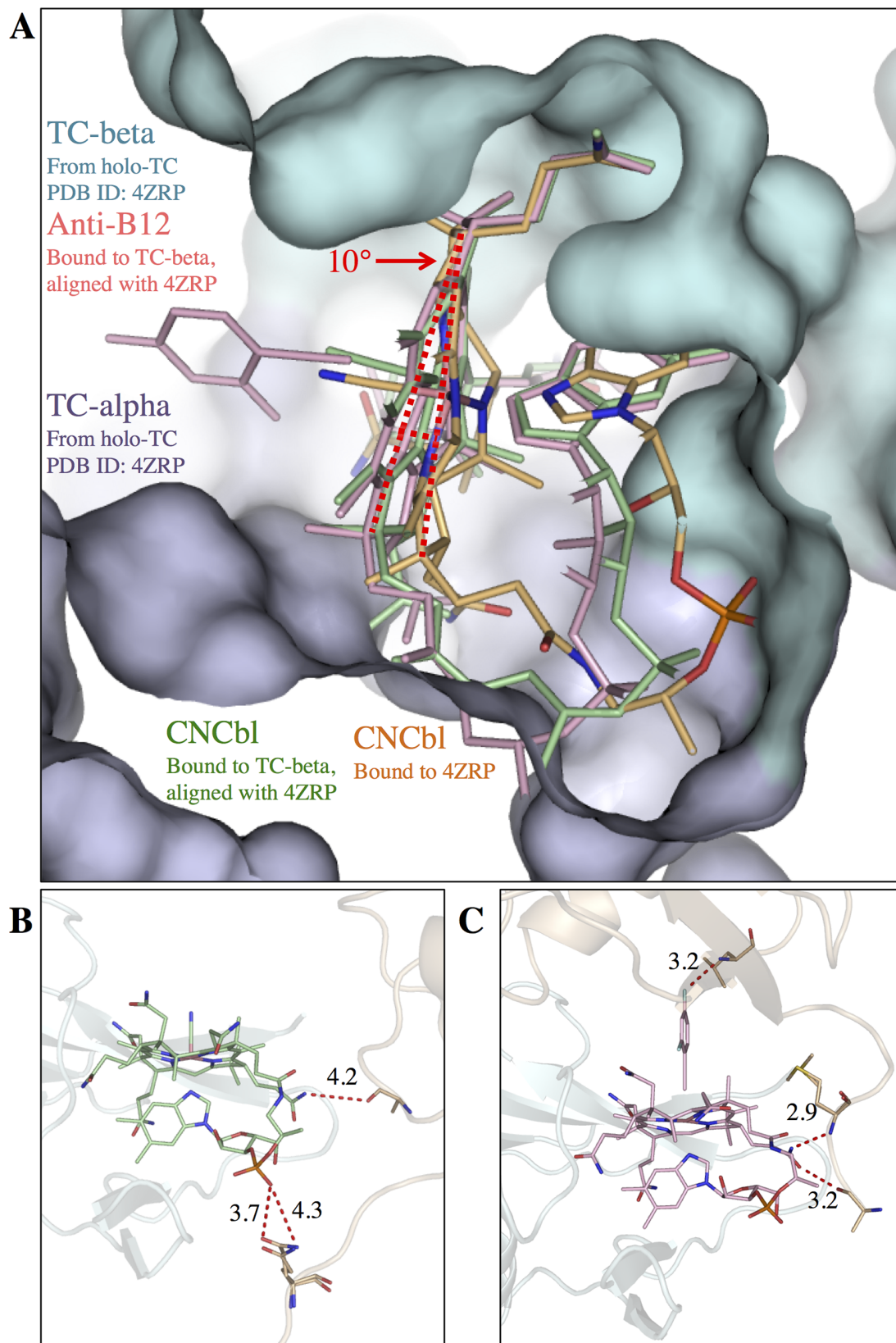
Figure 3 Comparison of ligand positioning in CNCbl-bound TC-beta and full-length TC. (A) Superimposed structures of Cbl-bound TC:CD320 (PDB entry 4ZRP, alpha-domain blue, beta-domain cyan, CNCbl orange, CD320 not shown), CNCbl-bound TC-beta (green, only Cbl shown) and anti-B12-bound TC-beta (pink, only anti-B12 shown). In the absence of TC-alpha, CNCbl and anti-B12 are rotated both by about 10˚ (indicated by dotted red lines). (B, C) Comparison of the structures of CNCbl-bound TC-beta (B) (cyan, green) and anti-B12-bound TC-beta (C) (cyan, pink) with highlighted polar crystal contacts (dotted red lines) for CNCbl and anti-B12 (symmetry mate in orange).

Crystal structure of anti-B12-bound TC-beta

We next determined the crystal structure of TC-beta in complex with the inhibitory B12 analog F2PhEtyCbl (S1 Fig)[17] at 1.27 Å resolution (Fig 2D, Table 1). As apparent in the statistics (high I/σ but low completeness) the crystal would have likely diffracted to higher resolution. However, the data collection was limited by the minimum detector distance of the beamline. The structure and binding mode of anti-B12-bound TC-beta were similar to CNCbl-bound TC-beta (i.e. the anti-B12 was rotated by 11.4° in the same way as CNCbl in TC-beta) (Fig 3A). The F2Ph-group of anti-B12 was clearly visible in the electron density. It was probably well- resolved due to interactions with a symmetry mate in the crystal (S3 Fig). The F2Ph-group was rotated by 90° compared to the small molecule crystal structure of F2PhEty-Cbl[17]. Superposition of the anti-B12-bound TC-beta structure with the crystal structure of CNCbl bound TC: CD320[11] suggested that anti-B12 could bind to TC in the same way as CNCbl and would not generate steric clashes with the F2Ph-β-coaxial ligand (Fig 4).

Discussion

When we purified full-length TC expressed in insect cells, the protein remained Cbl-bound even after extensive washing on affinity columns (result not shown). For structural studies of the apo state, denaturing poses a potential threat of losing structural integrity. In a previously reported crystal structure of human TC, the disulfide-bridge between Cys65 and Cys78 in the alpha domain was missing[13]. It is conceivable that the denaturing protocol in this study may have resulted in disulfide breakage. To avoid such complications, TC-alpha:CD320 and TC- beta were here expressed individually in insect cells.

Our experiments demonstrate that holo-TC could only be assembled from separated alpha- and beta domains upon addition of a Cbl. This is in line with a previously proposed mechanism of holo-TC assembly[13,14] and with an earlier finding that the structurally related protein intrinsic factor (IF) only assembles upon CNCbl addition[21]. The structures of apo TC-beta and CNCbl-bound TC-beta provide insight into distinct states of holo-TC assembly. They demonstrate the structural integrity of apo TC-beta and Cbl- bound TC-beta and therefore further consolidate the hypothesized mechanism of Cbl dependent holo-TC assembly. Virtually no structural changes occur

32

in TC-beta upon Cbl binding, arguing against an induced fit mechanism with regards to this domain of TC.

It was shown previously that Cbi can bind to TC and can be transported into cells via CD320-dependent internalization[22]. HC is structurally related to TC and binds Cbi with a much higher affinity: The dissociation rate constants were reported as k-Cbi > $5*10-2$ s-1 for TC and k-Cbi < $1*10-5$ s-1 for HC[8]. In HC, the missing DMB moiety is replaced by Arg-357[23]. Our structure demonstrates that monocyano-Cbi binds to TC-beta in the same way as CNCbl. The lower affinity of TC for Cbi[20] is most likely caused by the reduced number of protein-ligand interactions, given that cobinamides do not contain a DMB moiety.

In Cbi bound TC-beta, both Cbi molecules are coordinated to a histidine, leading to an arrangement where the histidine side chain assumes the role and position of the DMB moiety in Cbl. We hypothesize that cobinamides have the potential to nonspecifically attach to many proteins, as long as they contain solvent-exposed histidines.

When we superimposed the structure of anti-B12-bound TC-beta and Cbl-bound full-length TC[11], it became evident that the anti-B12 would not clash with TC-alpha (Fig 4). There is no structure of CNCbl or anti-B12 bound mouse TC, however the Cbl binding site of human and mouse TC are highly conserved (Fig 4B). Since the receptor-binding site is remote from the substrate-binding site, we believe that the anti-B12 cannot interfere with CD320-mediated uptake of TC. However, anti-B12 will competitively inhibit Cbl binding to TC, which should lead to decreased cellular Cbl uptake. Anti-B12 molecules have been synthesized with an "upper" axial substituent, that is inert to the Cbl-reducing enzyme methylmalonic aciduria and homocystinuria type C protein (CblC)[17,24,25]. It remains unclear whether the chemical inertness of anti-B12 molecules has an effect on cellular Cbl uptake, such as for example the downregulation of CD320 expression. The 90° rotation of the F2Ph-group in our anti-B12-TC-beat structure compared to the conformation observed in the pure compound suggests that this ligand freely rotates and that non-covalent contacts, as provided by a symmetry mate of TC-beta in our structure, determine its conformation (S2 Fig).

In the structures of CNCbl- and anti-B12-bound TC-beta, the corrin planes of the ligands are rotated relative to CNCbl in full-length holo-TC by ~ 10°. We hypothesize that during holo-TC assembly, TC-alpha pushes the substrate deeper into the binding pocket.

33
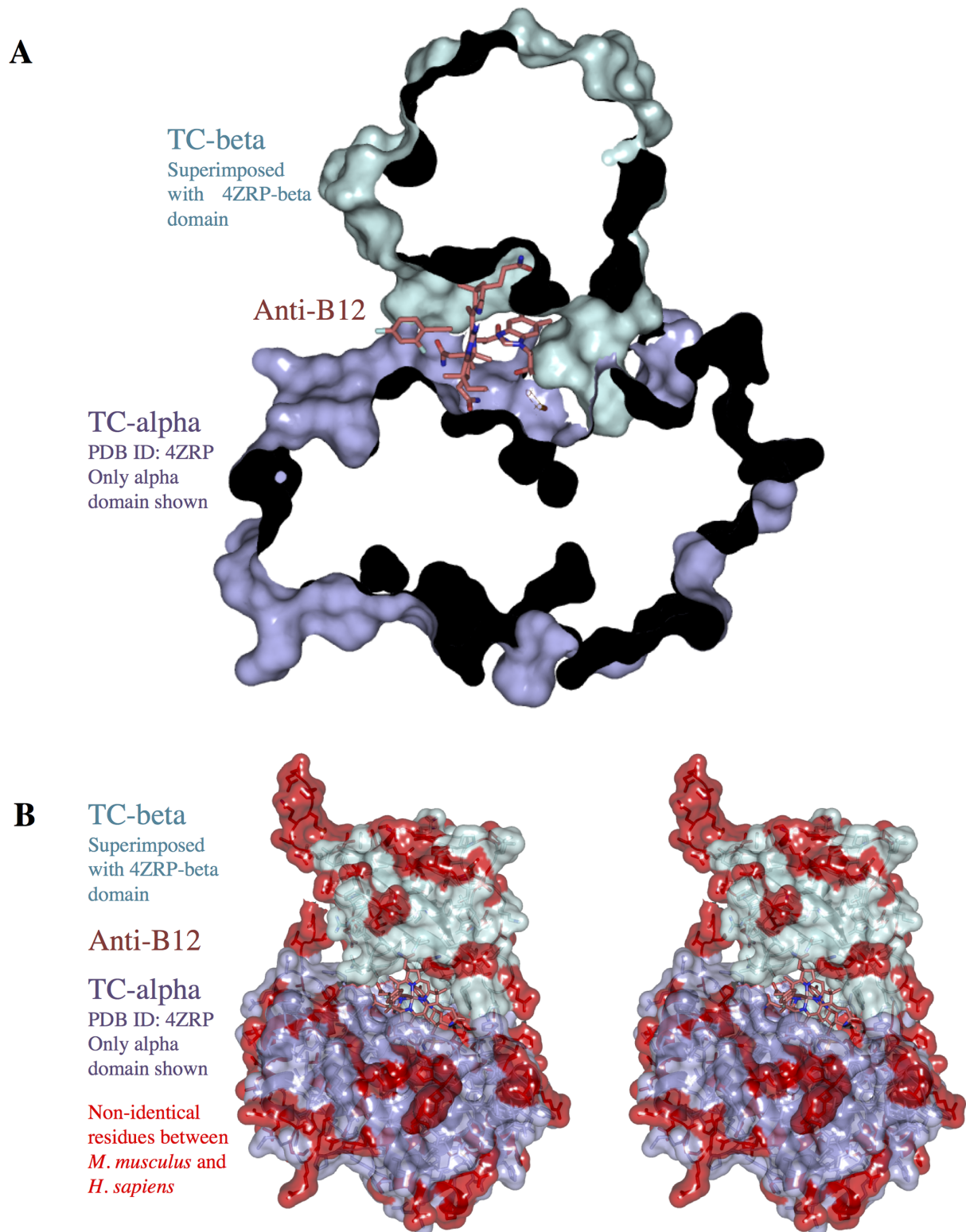
Figure 4 Model of anti-B12 bound to TC. Composite model of the structure of anti-B12-bound TC-beta (salmon, green) superimposed onto CNCbl-bound TC:CD320 (PDB entry 4ZRP, only alpha domain shown, blue). (A) The model shows that anti-B12 bound to TC would not clash with the protein. (B) Stereo representation (wall-eyed), with residues that are not identical between human and mouse TC highlighted in red.

A crystal contact bias of these structures is unlikely because although both ligands are involved in lattice contacts, the two protein complexes have crystallized in different space groups and feature different lattice contacts (Fig 3B and 3C, Table 1).

Finally, our crystal structures provide a visualization of the putative early steps of holo-TC assembly and into the substrate affinity of TC. With the structure of anti-B12-bound TC-beta we were able to shed light on the mechanism of anti-B12-mediated inhibition of Cbl-uptake. Lastly, we demonstrated that the production of apo TC-beta is a robust method to probe the interaction of TC with alternative substrates, as it is easy to produce and to co-crystallize with various substrates. This could be of great use for the development of drugs targeting TC: CD320-mediated uptake of Cbl. For example, Technetium-99m labeled Cbl analogs were synthesized in the past to target TC mediated Cbl uptake in mice and binding of the compounds to TC as well as tumor labeling was demonstrated successfully[26].

Experimental procedures

Expression and purification of TC-beta

All proteins where recombinantly expressed in SF21 cells as described previously for TC: CD320[11]. TC-beta was fused to an N-terminal 9-histidine tag, followed by maltose-binding- protein and a TEV-cleavage site (N-9His-MBP-TEV_cleavage_site-ValAspHisMet-TC-beta (Gln307-W409)). The proteins were purified as described for holo-TC:CD320 previously[11]. For TC-beta an additional washing step during the Ni-NTA immobilization was introduced to remove all bound Cbl: After washing the immobilized 9His-MBP-TC-beta with 10 column volumes of washing buffer (300 mM NaCl, 40 mM imidazole pH 8.0, 20 mM HEPES pH 7.0, 0.5 mM CaCl2) the protein was washed with "desalting buffer" (500 cv of 150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.5 mM CaCl2).

Assembly of TC-alpha:CD320 and TC-beta

10 µM TC-alpha:CD320 in desalting buffer was mixed with 40 µM apo TC-beta with or with- out 160 µM CNCbl. All components were kept in desalting buffer. The mixtures were incubated overnight at 4 ℃. Size exclusion chromatography of the mixtures was performed on a TSK-G3000 column using a 1260 Infinity HPLC System from *Agilent Technologies*. Peak fractions were subsequently analyzed by SDS-PAGE and Silver Stain Plus staining from *Biorad*.

Crystallization, data collection and processing

TC-beta was concentrated in desalting buffer and subsequently crystallized using vapor diffusion in sitting drops. Crystals were cryo protected in 25% glycerol. Diffraction data were collected at the Swiss Light Source (SLS) at the beamlines X06SA (apo TC-beta, Cbi-bound TC- beta, detector EIGER 16M) and X10SA (CNCbl bound TC-beta and anti-B12 bound TC-beta, detector PILATUS 6M) at 1.0 Å. Data were processed using XDS[27]. Model building was done in Coot[28] and refinements were performed in iterative rounds using PHENIX[29]. Initially, XYZ coordinates, Occupancies and Individual B-factors were refined. After every round of refinement the structures were manually improved in Coot[28]. Waters were added initially automatically and afterwards manually. One glycerol molecule was added manually to the structures TC-beta:B12 and apo TC-beta. In the final stages of the refinement, anisotropic B- factor refinement was performed with all structures. Moreover, the X-ray/stereochemistry weight, as well as the X-ray/ADP weights were optimized.

Apo TC-beta was concentrated to 5 mg/ml and was mixed 2:1 with 0.2 M sodium chloride, 20% w/v polyethylene glycol 3,350. The structure was solved by molecular replacement (MR) with PHENIX Phaser-MR[29] using a truncated version of TC (PDB ID: 4ZRP) as search model. Hereby, the residues 1 to 306 as well as the protein CD320 were removed from the model.

For the crystallization of TC-beta:CNCbl, TC-beta was mixed with a 1.5 molar excess of CNCbl to a final concentration of 5 mg/ml and was subsequently mixed 1:1 with 0.15 M Mg formate dihydrate, 22.5% w/v polyethylene glycol 1,000. The structure was solved by MR using the same truncated version of TC that was used for the structure solving of apo TC-beta as search model. The ligand, as well as the geometry-restraints replaced with the files from PDB ID 1CCW[30] [30].

For the crystallization of TC-beta:Cbi2, TC-beta was mixed with a 1.5 molar excess of dicyano-Cbi to a final concentration of 5 mg/ml and was subsequently mixed 1:1 with 0.1 M sodium malonate pH 7.0, 12% w/v polyethylene glycol 3,350. The structure was solved by MR using the same truncated version of TC that was used for the structure solving of apo TC-beta as search model. The ligand structure and geometry-restraints files were downloaded from PDB ID 5M29[31]. In the structure of TC-beta:AntiB12 one calcium molecules was placed, coordinated by one His305 and its backbone amide, which is part of the flexible N-terminus, as well as by two water molecules and the surface exposed His345 of a symmetry mate. A strong electron density peak that is

still visible at >18 σ indicated need to place an ion at this position. All other ions that were present in the buffer did either clash with the coordinating atoms due to a too large atomic radius or contained too many electrons which was observed by a large negative electron density peak after refinement.

For the crystallization of TC-beta:anti-B12, TC-beta was mixed with a 1.5 molar excess of F2PhEtyCbl to a final concentration of 5 mg/ml and was subsequently mixed 1:1 with 0.3 M potassium formate, 17% w/v polyethylene glycol 3,350. The structure was solved by MR using the crystal structure apo TC-beta (PDB ID: 5NO0) as search model. The ligand structure and geometry-restraints files were downloaded from PDB ID 5UOS[17].

All superposition calculations were performed using CCP4 Superpose[32] with the option "superpose usi[17]ng secondary structure matching". For the superposition and rmsd calculations we always used residues 312 to 409 (human TC and TC-beta have the same numbering).

Acknowledgments

Supporting information
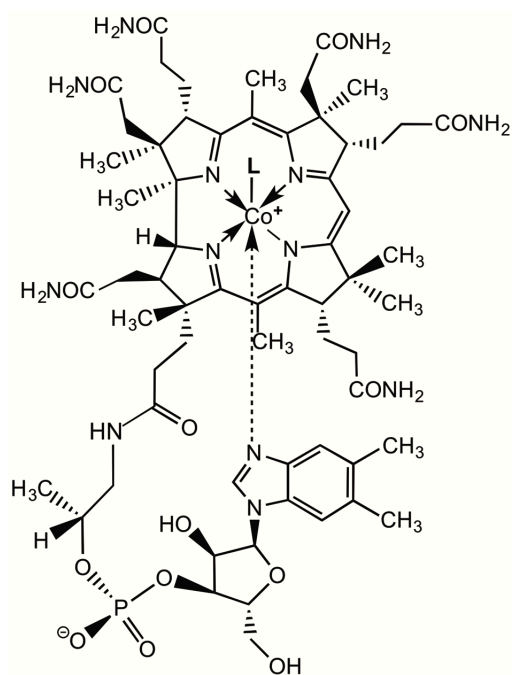
vitamin B$_{12}$:
L = CN: cyanocobalamin

antivitamins B$_{12}$ ('anti-B$_{12}$'):
L = 2-(2,4-difluorophenyl)-ethynyl: F2PhEtyCbl
L = 4-ethylphenyl: EtPhCbl

Supplementary figure 1 Chemical formula of cobalamins (vitamin B12 and antivitamins B12).

"L" indicates the ligand in the structure.
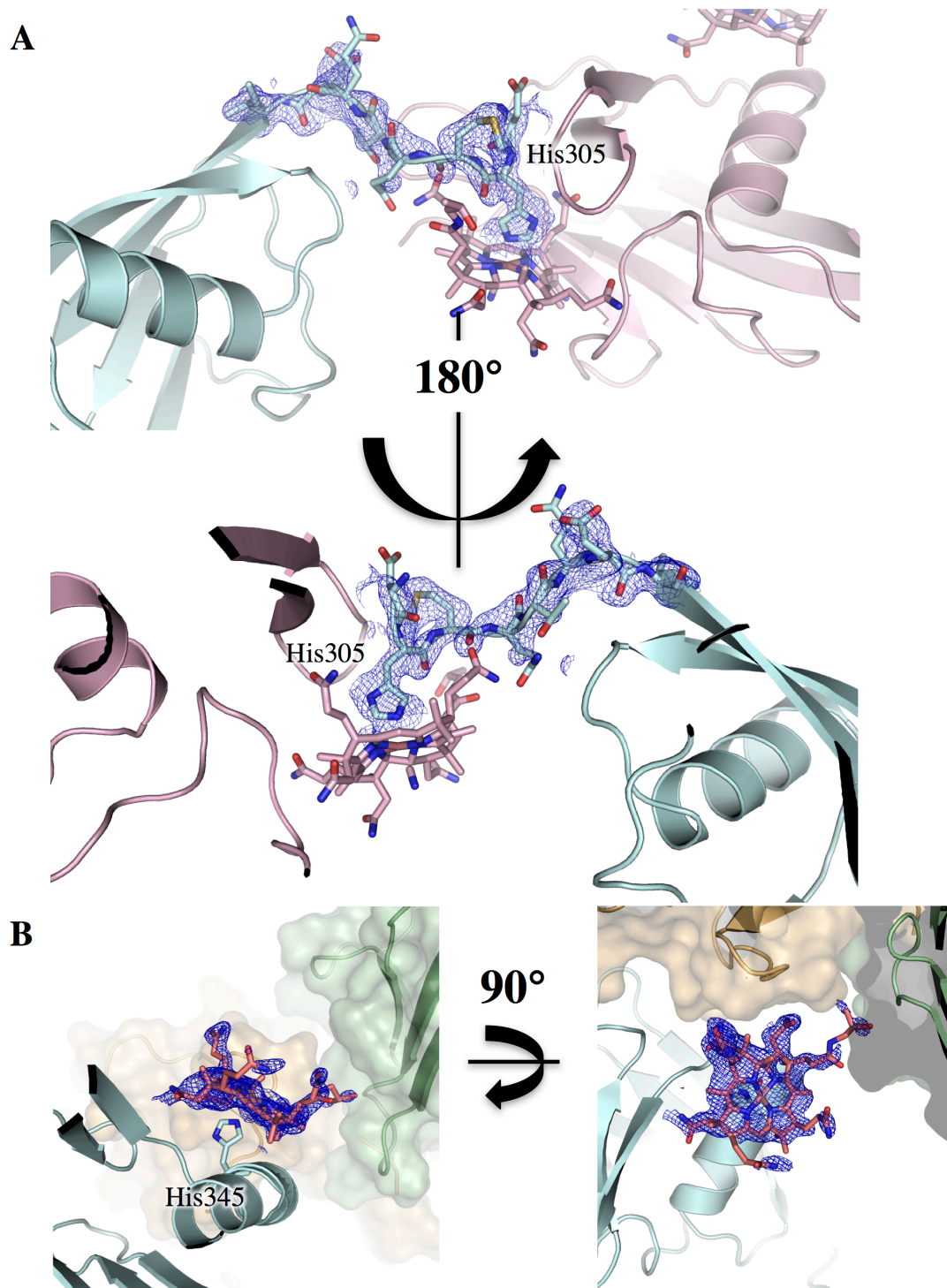
Supplementary figure 2 Cbi binding to histidines in TC-beta. TC-beta in complex with Cbi interacting with symmetry mates in the crystal lattice. The blue mesh represents a composite omit map contoured at 1 σ. (A) Cbi1 (pink) in the TC-beta binding pocket (pink), being coordinated to the side-chain of His305 of a symmetry mate (cyan). (B) Cbi2 (salmon) attached to the surface of TC-beta (cyan), being coordinated to the side-chain of His345. Symmetry mates contacting Cbi2 are shown in green and orange.
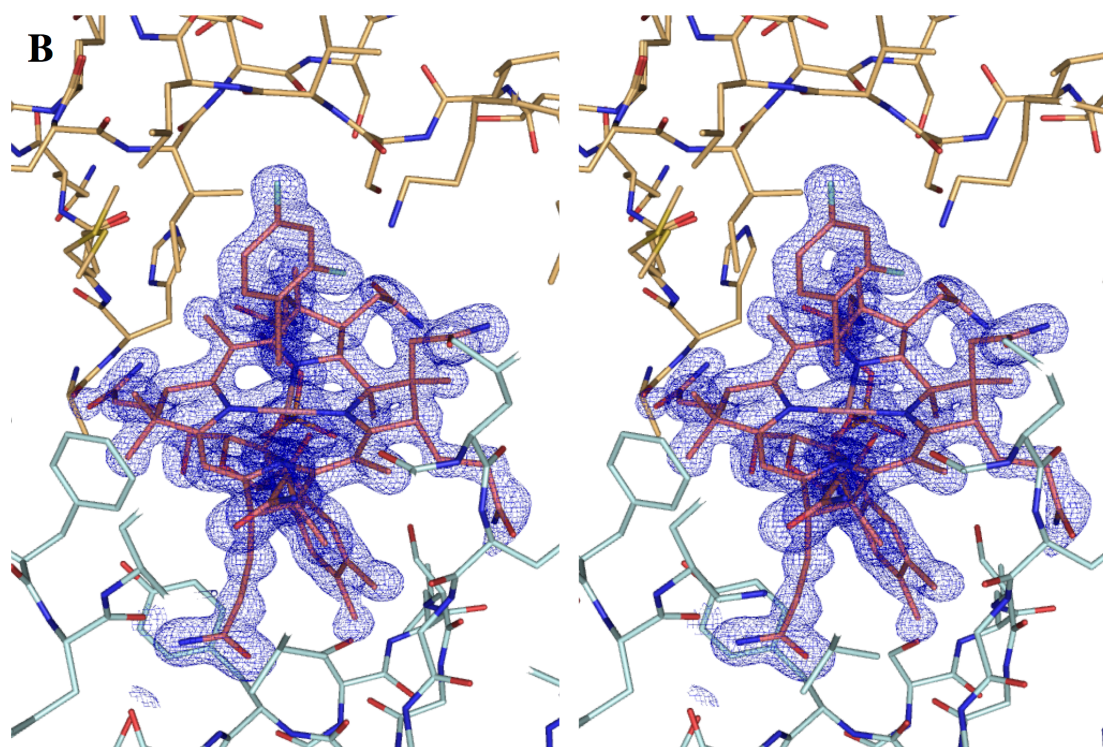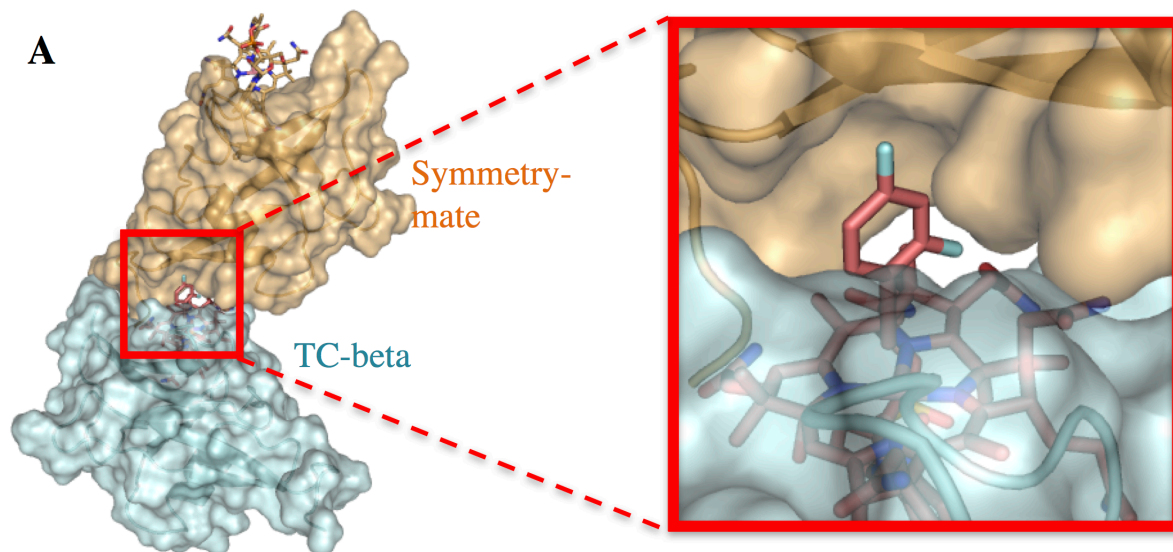
Supplementary figure 3 Crystal contacts of anti-B12 bound to TC-beta. (A) Interaction of TC-beta:anti-B12 (cyan, salmon) with a symmetry-mate (orange) in the crystal. The β-coaxial ligand of anti-B12 is involved in a crystal contact. This interaction probably causes the DFP-group to be ordered in the crystal. (B) Wall-eyed stereo representation of TC-beta:anti-B12. Two TC-beta molecules are shown as sticks, the view is as in A. The blue mesh represents a composite 2Fo-Fc electron density omit map contoured at 2 σ and is shown only around the anti-B12 molecule

References

1   Banerjee, R. & Ragsdale, S. W. The many faces of vitamin B-12: Catalysis by cobalamin-dependent enzymes. *Annu Rev Biochem* **72**, 209-247, (2003).

2   Cannata, J. J., Focesi, A., Jr., Mazumder, R., Warner, R. C. & Ochoa, S. Metabolism of Propionic Acid in Animal Tissues. Xii. Properties of Mammalian Methylmalonyl Coenzyme a Mutase. *J Biol Chem* **240**, 3249-3257, (1965).

3   Matthews, R. G., Koutmos, M. & Datta, S. Cobalamin-dependent and cobamide-dependent methyltransferases. *Curr Opin Struct Biol* **18**, 658-666, (2008).

4   Taylor, R. T. & Weissbach, H. N5-methyltetrahydrofolate-homocysteine transmethylase. Propylation characteristics with the use of a chemical reducing system and purified enzyme. *J Biol Chem* **242**, 1509-1516, (1967).

5   Martens, J. H., Barg, H., Warren, M. J. & Jahn, D. Microbial production of vitamin B12. *Appl Microbiol Biotechnol* **58**, 275-285, (2002).

6   Nielsen, M. J., Rasmussen, M. R., Andersen, C. B. F., Nexo, E. & Moestrup, S. K. Vitamin B-12 transport from food to the body's cells-a sophisticated, multistep pathway. *Nat Rev Gastro Hepat* **9**, 345-354, (2012).

7   Quadros, E. V. Advances in the understanding of cobalamin assimilation and metabolism. *Brit J Haematol* **148**, 195-204, (2010).

8   Fedosov, S. N., Berglund, L., Fedosova, N. U., Nexo, E. & Petersen, T. E. Comparative analysis of cobalamin binding kinetics and ligand protection for intrinsic factor, transcobalamin, and haptocorrin. *J Biol Chem* **277**, 9989-9996, (2002).

9   Kolhouse, J. F. & Allen, R. H. Absorption, plasma transport, and cellular retention of cobalamin analogues in the rabbit. Evidence for the existence of multiple mechanisms that prevent the absorption and tissue dissemination of naturally occurring cobalamin analogues. *J Clin Invest* **60**, 1381-1392, (1977).

10   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. The protein and the gene encoding the receptor for the cellular uptake of transcobalamin-bound cobalamin. *Blood* **113**, 186-192, (2009).

11   Alam, A. *et al.* Structural basis of transcobalamin recognition by human CD320 receptor. *Nat Commun* **7**, 12100, (2016).

12   Fisher, T. S. *et al.* Effects of pH and low density lipoprotein (LDL) on PCSK9-dependent LDL receptor regulation. *J Biol Chem* **282**, 20502-20512, (2007).

13   Wuerges, J. *et al.* Structural basis for mammalian vitamin B12 transport by transcobalamin. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4386-4391, (2006).

14   Fedosov, S. N., Fedosova, N. U., Krautler, B., Nexo, E. & Petersen, T. E. Mechanisms of discrimination between cobalamins and their natural analogues during their binding to the specific B12-transporting proteins. *Biochemistry-Us* **46**, 6446-6458, (2007).

15   Krautler, B. Antivitamins B12 -A Structure- and Reactivity-Based Concept. *Chemistry*, (2015).

16   Ruetz, M., Salchner, R., Wurst, K., Fedosov, S. & Krautler, B. Phenylethynylcobalamin: a light-stable and thermolysis-resistant organometallic vitamin B(12) derivative prepared by radical synthesis. *Angew Chem Int Ed Engl* **52**, 11406-11409, (2013).

17   Ruetz, M. *et al.* Antivitamin B12 Inhibition of the Human B12 -Processing Enzyme CblC: Crystal Structure of an Inactive Ternary Complex with Glutathione as the Cosubstrate. *Angew Chem Int Ed Engl* **56**, 7387-7392, (2017).

18   Mutti, E., Ruetz, M., Birn, H., Krautler, B. & Nexo, E. 4-Ethylphenyl-Cobalamin Impairs Tissue Uptake of Vitamin B-12 and Causes Vitamin B-12 Deficiency in Mice. *Plos One* **8**, (2013).

19   Rappazzo, M. E., Salmi, H. A. & Hall, C. A. The content of vitamin B12 in adult and foetal tissue: a comparative study. *Br J Haematol* **18**, 425-433, (1970).

20   Fedosov, S. N., Petersen, T. E. & Nexo, E. Binding of cobalamin and cobinamide to transcobalamin from bovine milk. *Biochemistry-Us* **34**, 16082-16087, (1995).

21   Fedosov, S. N. *et al.* Assembly of the intrinsic factor domains and oligomerization of the protein in the presence of cobalamin. *Biochemistry-Us* **43**, 15095-15102, (2004).

22   Lildballe, D. L., Mutti, E., Birn, H. & Nexo, E. Maximal load of the vitamin B12 transport system: a study on mice treated for four weeks with high-dose vitamin B12 or cobinamide. *Plos One* **7**, e46657, (2012).

23   Furger, E., Frei, D. C., Schibli, R., Fischer, E. & Prota, A. E. Structural Basis for Universal Corrinoid Recognition by the Cobalamin Transport Protein Haptocorrin. *J Biol Chem* **288**, 25466-25476, (2013).

24   Gherasim, C., Lofgren, M. & Banerjee, R. Navigating the B(12) road: assimilation, delivery, and disorders of cobalamin. *J Biol Chem* **288**, 13186-13193, (2013).

25   Ruetz, M. *et al.* Access to organometallic arylcobaltcorrins through radical synthesis: 4-ethylphenylcobalamin, a potential "antivitamin B(12)". *Angew Chem Int Ed Engl* **52**, 2606-2610, (2013).

26   Waibel, R. *et al.* New derivatives of vitamin B12 show preferential targeting of tumors. *Cancer Res* **68**, 2904-2911, (2008).

27   Kabsch, W. Xds. *Acta Crystallogr D* **66**, 125-132, (2010).

28   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132, (2004).

29   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **66**, 213-221, (2010).

30   Reitzer, R. *et al.* Glutamate mutase from Clostridium cochlearium: the structure of a coenzyme B12-dependent enzyme provides new mechanistic insights. *Structure* **7**, 891-902, (1999).

31   Mireku, S. A. *et al.* Conformational Change of a Tryptophan Residue in BtuF Facilitates Binding and Transport of Cobinamide by the Vitamin B12 Transporter BtuCD-F. *Sci Rep* **7**, 41575, (2017).

32   Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography* **60**, 2256-2268, (2004).

3. A versatile set of nanobodies targeting the human, transcobalamin mediated vitamin B12 uptake route

Statement of Contribution

I performed all experiments in this study, except for the cellular uptake experiments with fluorescently labelled TCNB4 which were performed by Prof. Edward V. Quardos[1] and Jeffrey M. Sequeira[1]. E.V.Q. and J.M.S. also wrote the figure legends and the methods section for their experiments. I wrote the chapter.

[1]Departments of Medicine and Cell Biology, SUNY-Downstate Medical Center, Brooklyn, NY 11203, USA.

Abstract

Cellular uptake of vitamin B12 in humans occurs via CD320 receptor mediated endocytosis of the soluble B12 carrier protein transcobalamin (TC). The expression of CD320 is associated with the cell cycle and is upregulated in highly proliferating cells, such as cancer cells[1-3]. This makes the uptake route of vitamin B12 a potential target for cancer therapy[4]. Here we present a set of four camelid nanobodies, binding to TC and the TC:CD320 complex, with nanomolar affinities. Three of these nanobodies, when conjugated to saporin, are capable of inhibiting the growth of highly proliferating human embryonic kidney (HEK293T) cells and therefore bear the potential to inhibit the growth of cancer cells. In particular, one of these nanobodies (TCNB4), which specifically recognizes the formed TC:CD320 complex, led to complete inhibition of cellular growth. Using fluorescent light microscopy, we were able to demonstrate the cellular uptake of this nanobody. We have determined X-ray crystal structures of the nanobodies in complex with TC:CD320 nanobody complexes, and we could map the binding sites to three distinct epitopes. One of the nanobody co-crystal structures led to a new structure of the TC:CD320 complex at higher resolution than previous structures. The structure shows the complete interface of TC and the LDLR-A1 domain of CD320, revealing additional pH sensitive salt bridges and ordered water molecules. Furthermore, for the first time the disease related residue Glu88 of CD320 could be resolved at sidechain resolution revealing the structural importance of this residue for the stability of the receptor.

Introduction

In humans the final stage of cellular uptake of vitamin B12 is mediated by endocytosis of the soluble B12 carrier protein transcobalamin via binding to the transcobalamin receptor CD320[5]. As vitamin B12 is necessary for DNA synthesis, the need for it is elevated in highly proliferating cells such as cancer cells, which often results in the overexpression of CD320[3]. In spite of over 40 years of research no drug targeting the B12 uptake route has reached the market[6]. In previous studies different groups have attempted to target the cellular uptake route for vitamin B12 for cancer therapy and diagnosis[4]. The strategies included the derivatization of vitamin B12 with cytotoxic drugs[7] or the coupling of an antibody targeting CD320 with saporin[8,9]. It could be demonstrated that *in vitro*, the saporin-conjugated antibody was capable of efficiently destructing a broad spectrum of cancer cell lines[9]. Camelid single chain VHH antibodies respectively nanobodies are highly stable, less immunogenic than therapeutic antibodies, able to penetrate tissues, and can easily be produced at large scale in *E. coli*[10]. For these reasons, a nanobody drug conjugate would be a desirable tool for cancer therapy by targeting the cellular vitamin B12 uptake route.

An open question regarding TC:CD320 mediated uptake are the detailed molecular interactions between TC and CD320. These have been structurally adressed[11], but still leave open questions such as a the loop from Cys74 to Cys89, which is close to the binding interface with LDLR-A1, but could not be resolved in the published structure[11]. Also, the region around the disease implicated residues Glu88 in CD320 could only partially be resolved.

In order to develop a tool for nanobody mediated drug delivery via TC:CD320 mediated B12 uptake we aimed to develop a set of nanobodies targeting the TC:CD320 complex. In addition, we aimed at using these nanobodies as crystallization chaperones for the further structural characterization of the complex.

Results

Production and characterization of camelid nanobodies against TC:CD320

A *Vicugna pacos* was immunized against purified TC:CD320 complex. Using phage display of the cDNA library against immobilized biotinylated TC:CD320, we obtained a set of 20 unique nanobody binders, of which 14 successfully bound in pulldown and SEC complex formation experiments. The four nanobodies TCNB4, TCNB11,

TCNB26 and TCNB34 were chosen for further characterization (Fig, 1a). From these four nanobodies, three nanobodies were suitable for amine-reactive labelling. The labelled nanobodies and were tested for their affinity to TC:CD320 by MicroScale thermophoresis (MST). We measured affinities of 5.2 nM for TCNB4, 87.4 nM for TCNB26 and 12.8 nM for TCNB34 (Fig. 1b). All four nanobodies were co-crystallized with TC:CD320. The resulting X-ray crystal structures (Fig. 1c, Extended Data Table 1) enabled us to identify the binding sites of the nanobodies to 3 distinct binding epitopes. TCNB11 and TCNB34 have overlapping binding epitopes. Notably, TCNB4 is the only nanobody that binds simultaneously to TC and CD320 and therefore specifically recognizes the formation of the TC:CD320 complex.
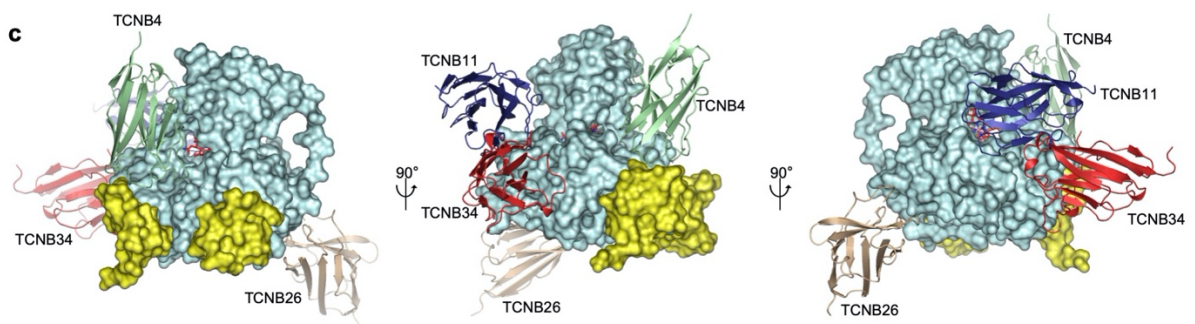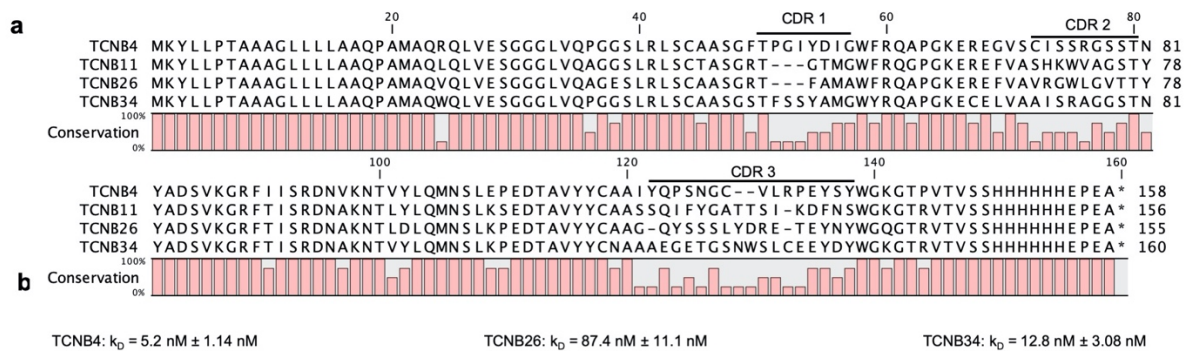
Figure 1 Structural and functional characterization of nanobodies raised against TC:CD320. a, Sequence-alignment of conformational nanobody binders against the TC:CD320 complex. Red bars below the sequence indicate the degree of sequence conservation. CDRs are indicated above the sequences. **b**, MicroScale thermophoresis affinity measurements of nanobodies TCNB4, TCNB26 and TCNB34 binding to TC:CD320. Note that TCNB11 was not tested, as it was not suitable for the required fluorescent labelling. **c**, Overlaid crystal structures of TC:CD320:nanobody complexes with TC (cyan) and CD320 (yellow) in surface representation and TCNB4 (green), TCNB11 (blue), TCNB26 (brown) and TCNB34 (red) in surface representation.

46

Cellular uptake of TCNB4

To test whether TCNB4, for which we had measured the highest affinity for TC:CD320, could enter human cell by hijacking CD320 receptor mediated endocytosis of holo TC (i.e. B12 bound TC), we used fluorescently labelled TCNB4 for cellular uptake studies. By incubating the nanobody with HEK293 cells that constitutively overexpress GFP-tagged CD320, as well as holo TC, we observed colocalization of TCNB4 and CD320 in a fluorescent light microscopy experiment (Fig 2). We further observed that the nanobody was endocytosed by the cells together with the receptor. To test whether the nanobody would get degraded inside the cells we repeated the experiment with dsRed-tagged TC[12] where, over the time course of 4 h, we would only observe the degradation of dsRed-TC however not of TCNB4 (Extended Data Figure 1). One caveat is, that in this experiment one can only observe the lifetime and stability of the fluorophore, which may differ from that of the protein to which the fluorophore is fused to.
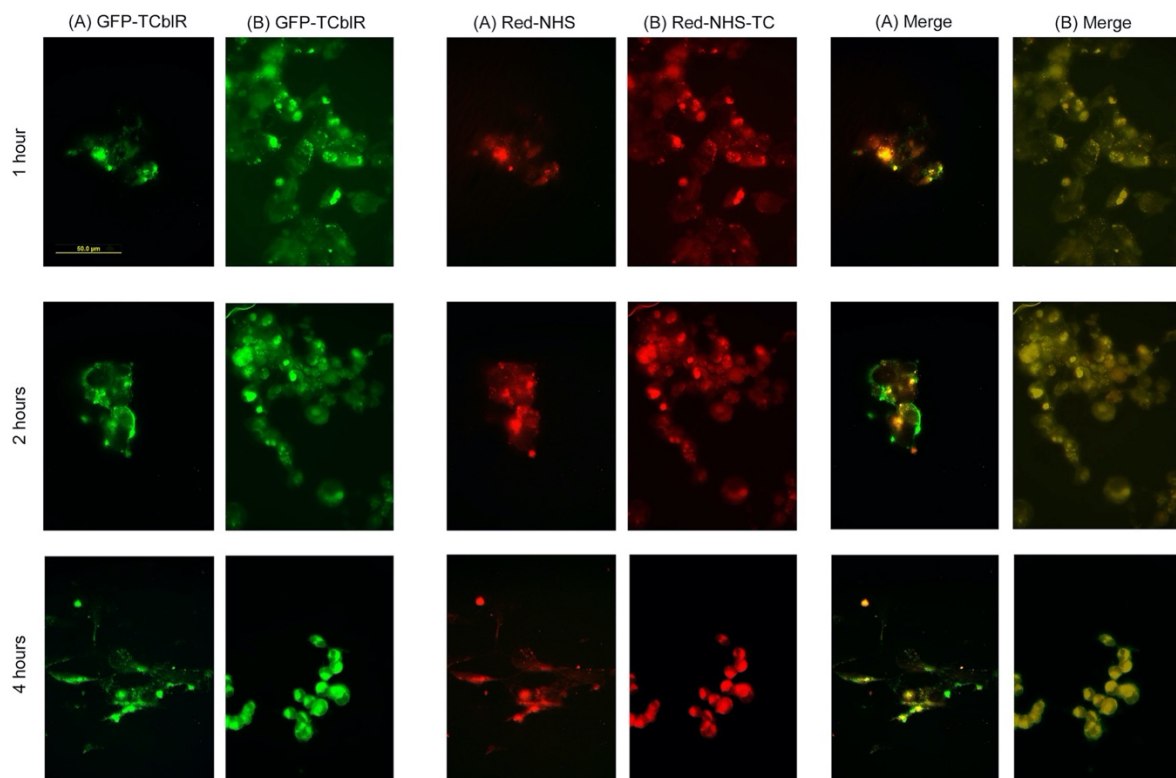
Figure 2 Internalization of nanobody uptake by HEK 293 cells expressing TCblR-GFP. Cells were incubated with fluorescently labelled TCNB4 (A) or with labelled TCNB4 + TC-Cbl for 1, 2 or 4 h at 37 °C. Over the time period tested, TCNB4 without TC appears to bind to the cell surface likely due to nonspecific surface binding (panels A) but this binding is not associated with TCblR (the merge panels A)). Nanobody preincubated with TC-Cbl appears to bind specifically to TCblR-GFP (panels B) and is associated with TCblR-GFP (merge panels B). However, the TCblR-GFP does not appear to degrade at 4h.

Cellular inhibition of growth by saporin-conjugated nanobodies

Knowing about the potential of the TC:CD320 binding nanobodies to be endocytosed by human cells, we wanted to test their applicability in the targeting and the destruction of cancer cells. We therefore generated nanobody drug conjugates which would, if being endocytosed successfully, destroy or inhibit the respective cells. As a drug, respectively toxin we chose the plant enzyme saporin, which deglycosylates and thereby inactivates eukaryotic ribosomes[13]. In order to generate the nanobody drug conjugates we used commercially available, biotin reactive saporin (Streptavidin-ZAP). In order to biotinylate the nanobodies, we engineered enzymatic biotinylation recognition sites into the nanobodies and biotinylated them using the enzyme BirA. Subsequently, the biotinylated nanobodies and the biotin reactive saporin were coupled used for cell-cytotoxicity assays.
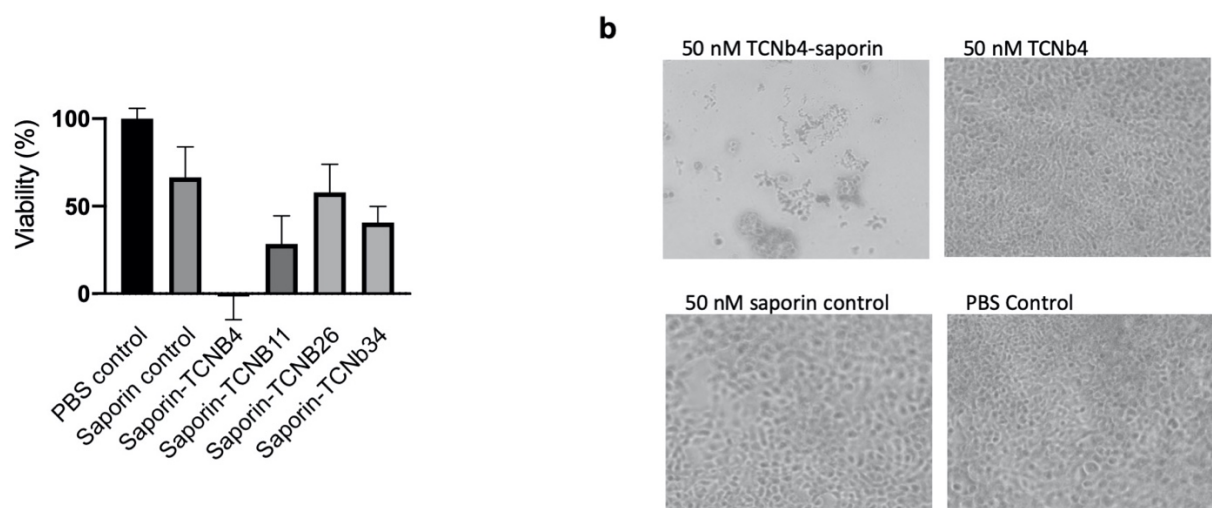
Figure 3 Cell cytotoxicity assays for saporin-conjugated nanobodies targeting cellular vitamin B12 uptake. a, Cell viability after incubation of HEK293 cells with saporin-conjugated nanobodies for 8 days, defined as reduction of MTS tetrazolium reagent by viable mammalian cells, generating a colored formazan dye. 100% viability correlates to the conversion by untreated cells (PBS control) and 0% correlates to the conversion in wells without cells. **b**, Light microscopy images of HEK293T cells after incubation with saporin-conjugated TCNB4, TCNB4 alone, saporin alone or PBS (control) for 8 days.

We seeded HEK293T cells at low density in order to obtain highly proliferating cells, which in turn would have a high expression of CD320. When adding the saporin-conjugated nanobodies at a concentration of 50 nM to these cells, a complete inhibition of growth was observed for the addition of saporin-TCNB4 (Fig. 3a). Control experiments, where either only TCNB4 or saporin-ZAP was added, did either not or significantly less affect the growth of the cells (Fig. 3b). Compared to TCNB4,

TCNB11 and TCNB34 showed lower but still significant inhibition of growth. The TCNB26 saporin conjugate did not inhibit significantly more than the saporin-ZAP control at the tested concentration.

Structural characterization of nanobody binding

TCNB4, which binds its target with high affinity (k_D = 5.2 nM), interacts with CD320 via complementarity-determining region 1 (CDR1) and a non-variable loop between beta-strands 5 and 6 of the nanobody (Extended Data Figure 2). Interactions with TC are mediated mostly by CDR3 and in some part by CDR1 and CDR2. Although CDR2 only interacts via one residue (Arg75) with TC, the loop appears to have a sterically important role in positioning CDR3. This finding is further substantiated by the fact that CDR2 is linked via a disulfide bond to CDR3. Unexpectedly, we observed a Ca^{2+} ion coordinated by three backbone carbonyl groups of CDR3 and by one glutamate sidechain of CDR1. The Ca^{2+} ion seems to further stabilize the conformation of the CDR3 loop.

While TCNB11 binds its target with CDR2 and CDR3, CDR1 does not interact with TC (Extended Data Figure 3). CDR3, which binds close to the B12 recognition site of TC, introduces a twist into a flexible loop of TC, flipping the backbone of the residues in the range of Phe168 to His173 by approximately 180°. The flexibility of this loop has been observed previously, where His173 can serve as upper coaxial ligand of B12[14], but can also be displaced by more stable ligands such as the cyano-group in cyanocobalamin[11]. Nonetheless, this structure serves as a prime example how nanobody- or antibody bound structures should always be interpreted with a grain of salt, especially at the binding interface, where a binder might have altered the structure of the target protein.

TCNB26 binds TC with all three CDRs (Extended Data Figure 4). However, its rather loosely packed binding interface might be the reason for the comparably lower binding affinity of the nanobody. In the crystal structure of the complex of TCNB26 with TC:CD320, the crystal packing was such that there was only enough space for one TCNb26 per two TC molecules.

TCNB34 binds TC mostly via CDR1 and CDR3 (Extended Data Figure 5). CDR2 is further away from the target and only Arg53 would be in reach to potentially form a salt bridge with Glu210 of TC, which however was not observed in the crystal

structure. Nonetheless, CDR2 appears to play a role in sterically positioning CDR3, as it forms a disulfide bond with CDR3.

New features in the structure of the TC:CD320 complex

TCNB34 turned out to be a helpful crystallization chaperone, improving the resolution of the TC:CD320 complex from previously reported 2.1 Å[11] to 1.85 Å (Extended Data Table 1). Apart from the overall better resolved maps, two regions in particular are better resolved in the TCNB34:TC:CD320 co-crystal structure (Fig. 4): The loop in TC from Cys65 to Cys78, which form a disulfide bond, was visible in earlier structures of TC alone[14]. However, in these structures the disulfide bond was not formed, which may have affected the fold of the loop. In the published structure of TC:CD320[11] the disulfide bond was formed, but the loop was disordered in the crystal structure. In the TCNB34:TC:CD320 co-crystal structure the disulfide is formed, and the loop is clearly visible. The structure reveals previously unknown binding interactions between TC and CD320. Glu73 and Asp74 of TC from salt-bridges with Lys58 of CD320. As such salt-bridges are sensitive to changes in pH[15], this finding is in line with the observation that TC and CD320 dissociate at low pH after endocytosis of the complex[11]. In addition, several ordered water molecules at the TC:CD320 binding interface could be resolved. The structure also reveals the complete interface between TC and LDLR-A1 of CD320. Another feature revealed by the TCNB34:TC:CD320 co-crystal structure is the region around Glu88 in CD320. The structure shows that the adjacent residue Glu87 forms a salt bridge with Arg68 via its head group and with Arg73 via its backbone carbonyl. Deletion of either one of the residues Glu86, Glu87 or Glu88 has been associated with vitamin B12 deficiency in patients carrying the mutation[16]. After initial findings that the loss of one of these residues would lead to steric constrictions in the LDLR-A1 domain[11], we now see in detail the structural importance of the residue for the overall stability of the receptor. Notably the interactions of Glu87 are salt-bridges and therefore also pH-dependent. Loss of these interactions at low pH might destabilize the receptor and thereby aid the release of TC from CD320 after endocytosis. A genetic deletion of Glu86, Glu87 or Glu88 however would also destabilize the receptor, which could explain the lower affinity of the ΔE88 mutant receptor to TC and also the decreased resolution of the crystal structure of the TC:CD320ΔE88 complex[11].
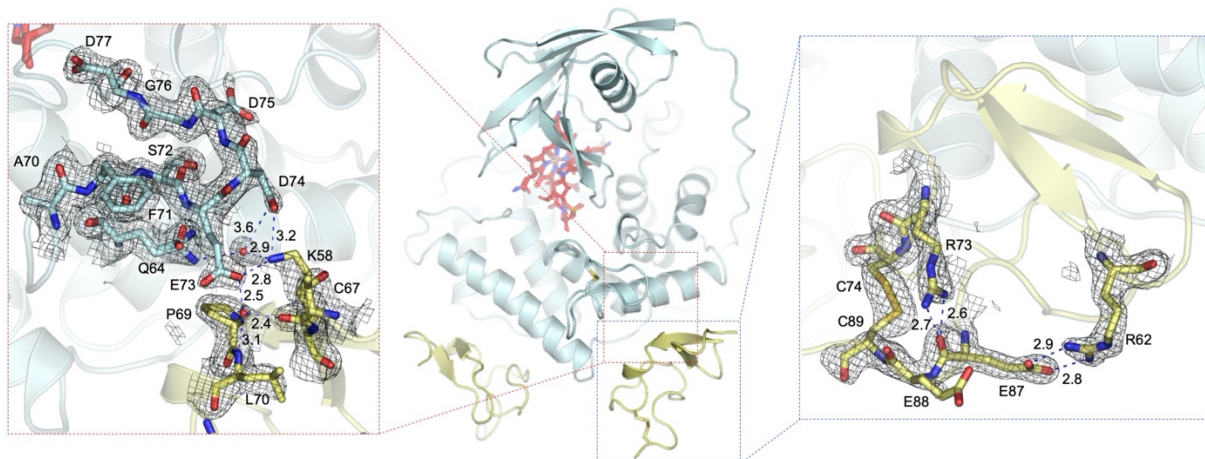
Figure 4 New structural insights into transcobalamin recognition by CD320. In the cocrystal structure of TCNB34:TC:CD320 TC (cyan) and CD320 (yellow) are shown in cartoon representation and B12 (red), as well as disulfide bonds forming the loops of interest in stick representation. TCNB34 has been removed for clarity. The red inset (left) shows a loop in TC which was disordered in previous crystal structures[11,14]. Polar contacts between TC and CD320 are indicated with blue dashed lines with distances indicated in Å. The blue inset shows Glu88 and the surrounding residues of CD320. Salt bridges involving Glu87 are indicated with blue dashed lines. Note that a genomic deletion of Glu88, which would be equal to a deletion of Glu87, has been reported to lead to B12 deficiency[16].

Although the structure and position of the EGF-like domain still remains elusive, the TCNB11:TC:CD320 co-crystal structure reveals 8 previously unresolved residues reaching from LDLR-A1 towards the EGF-like domain, revealing Cys94, the first cysteine residue, and therefore the to some degree, the location of the missing domain (Fig 4).

Conclusions

We have obtained a set of nanobody binders which bind to TC or TC:CD320 at high affinity. The binders have shown to be able to enter human cells and to inhibit growth of these cells when connected to a toxin. Therefore, these nanobodies bear the potential to be further developed as anti-cancer drugs, specifically targeting the cancer marker CD320. As three of the four nanobodies bind to distinct, non-overlapping epitopes, binding affinity and specificity could potentially be improved by engineering bi- or tripartite nanobody fusion proteins. Combining TCNB4 with a nanobody that solely binds to TC might increase the chances in successful assembly of TC, CD320 and the nanobody. Such a bipartite nanobody could bind to soluble TC in the bloodstream already before reaching CD320, rather than only being able to bind to the pre-formed complex of TC:CD320, as it likely is the case for TCNB4.

Fusing TCNB4 to other cancer cell marker binding proteins could further increase its specificity for cancer cells.

To determine the applicability of our nanobody drug conjugates, further studies in closer to real life scenarios will be necessary. Such experiments could involve more tissue like cell cultures, real cancer cell lines and in a later stage cancer tissue from patients. Also fusing the nanobodies to other toxins or radionuclides for imaging studies could be tested.

The crystal structure of TC:CD320 bound to TCNB34 revealed new features of the TC:CD320 complex, such as the full TC:CD320-LDLR-A1 interface and the region around Glu88 in CD320. These findings have helped to further rationalize the dissociation of the TC:CD320 complex at low pH and have provided a deeper explanation for the vitamin B12 deficiency causing mutation ∆E88 in TC. A structural feature of the TC:CD320 complex that still remains elusive is the role of the EGF-like domain in CD320. Although functional data has suggested that the domain ensures the correct stoichiometric binding stoichiometry of TC:CD320[11], no structural information about the domain could be retrieved yet, apart from the TCNB11 bound structure of TCCD320 which shows the first cysteine of CD320. This seems to be caused by the mobility of the domain, which is tethered between by two flexible linkers between LDLR-A1 and LDLR-A2, could possibly be resolved by using cryo-EM. A large dataset of the TC:CD320 complex could be collected and one could potentially classify distinct states of the complex with the EGF-like domain in different positions. This could also help to resolve the structural ensemble of specific orientations in which the EGF-like domain can be present. A similar study has been conducted recently on the flexible ABC transporter protein TmrAB, where multiple high-resolution structures of different states of the same protein could be obtained by 3D classification of a large set of data[17]. As the mass of TC:CD320 with 60 kDa is rather low for cryo-EM, one could utilize a combination of 3 nanobodies (e.g. TCNB4, TCNB34, TCNB26) to increase the mass of TC by another ~45 kDa, which could assist in particle alignment. Also, the TC binding antibody A-5 (Santa Cruz Biotechnology, Inc, Catalogue # sc-137017), if it should bind rigidly, could help with this problem with alignment in cryo-EM.

Methods

Expression and purification of TC:CD320

TC:CD320 was expressed and purified as described previously[11].

Nanobody selection, expression and purification

Nanobodies were generated expressed and purified as described previously[18]. with the following modifications: For the panning TC:CD320 was immobilized via biotinylation of CD320 as described previously[11]. Purified nanobodies were desalted into 20 mM Tris pH 7.5, 0.5 mM CaCl2 and 150 mM NaCl.

In order to generate nanobody-saporin conjugates a C-terminal Avi Tag (GLNDIFEAQKIEWHE), preceding the His6 Tag and flanked by flexible GGGS linkers, was fused to the nanobodies. Biotinylation was performed as described previously[11]. For the saporin-fusion, the biotinylated samples were desalted into PBS (Gibco) and were mixed at a concentration of 5 µM with a 2.4-fold molar excess of Streptavidin-ZAP (Advanced Targeting Systems). The mixture was incubated for 1h on ice before diluting it in PBS to a working concentration of 500 nM.

MicroScale Thermophoresis affinity measurements

MST affinity measurements for the binding of TCNB4, TCNB26 and TCNB34 to TC:CD320 were performed on a monolith NT.115 instrument (NanoTemper)[19]. Nanobodies were labeled using the lysine reactive Monolith NT Protein Labeling kit RED-NHS (NanoTemper). As TCNB11 has a lysine residue in CDR2, it was not suitable for this labeling method and was therefore not analyzed by MST. Labelled nanobodies at concentrations ranging from 50 nM to 1.6 nM were pre-incubated with TC:CD320 at either 18 nM (for TCNB4) or 50 nM (for TCNB26 and TCNB34) for 20 min at room temperature. Subsequently the samples were loaded into glass capillaries (NanoTemper) and MST measurements were performed at various values of LED power and MST power, reaching the clearest results at 100% LED power and 80% MST power for TCNB4, at 90% LED power and 80% MST power for TCNB26 and at 90% MST power and 40% LED power for TCNB34. NanoTemper software was used to calculate the respective K_D values.

Cellular nanobody uptake experiments (Written and performed by Prof. Edward Quadros)

TCNB4 was covalently labeled with Red-NHS using the Monolith NT kit (Nanotemper technologies) and used in the cellular binding and uptake studies. HEK 293 cells transfected with the CD320 cDNA fused to a green fluorescent protein (GFP) expressing protein in plasmid pEGFP[20]. Cells were selected for stable expression of TCblR-GFP by culturing in DMEM containing G418 antibiotic and used in the present experiments. Cells were seeded at a density of 0.1 x 106 cells per well in 60mm dishes, cultured for 48 h in DMEM containing 10% FBS and 200ug/ml G418.

For Fig. 2, cells were washed 3 times with PBS and incubated in DMEM with labeled TCNB4 alone or an equivalent amount preincubated for 1hr with 200ul of human serum + cobalamin as a source of holoTC. At 1, 2 or 4 h cells were washed three times with PBS and fixed in 4% paraformaldehyde for 30 min. The cell layer was washed 3 times with PBS and prepared for fluorescent microscopy. Cells incubated on ice for 1-2 h and cells incubated at 37 °C with free dye served as controls.

For Extended Data Figure 1, cells were washed 3 times with PBS and incubated with DMEM with dsRed TC-Cbl or TCNB4+dsRed TC-Cbl. At 1, 2 or 4 h cells were washed with PBS and fixed with 4% paraformaldehyde for 30 min.

Cell cytotoxicity Experiments

HEK293T cells in DMEM medium supplemented, with 10% v/v FBS and 200 µg/ml G418, were seeded in 96 well sterile cell culture plates at 2×10^3 cells per well. After letting the cells attach for 1-2h at 37 °C. The medium was replaced with the same medium, containing 5% v/v human serum and 5 µM cyanocobalamin as source of holoTC, as well as either 50 nM TCNB4, saporin conjugated TCNBs 4, 11, 26 or 34, unconjugated Streptavidin-ZAP as saporin control, or PBS as negative control. After incubation for 8 days at 37 °C the medium was exchanged to DMEM medium supplemented, with 10% v/v FBS and 200 µg/ml G418, to avoid bias by color changes in the medium and inhibition of growth respectively total cellular viability was assayed using the MTS Assay Kit (Abcam). 100% viability was defined as the readout from non-treated cells (addition of PBS) and 0% viability was defined by the readout performed on wells where no cells where seeded. Measurements were

conducted in triplicates and positive and negative controls were performed in quintuplicates.

Crystallization and data collection

After preincubating TC:CD320 at 10 mg/ml with a 1.2-fold molar excess of TCNB4 for 1h on ice, crystals of the TCNB4:TC:CD320 complex were grown by sitting drop vapor diffusion in 100 mM Bis-Tris pH 5.5, 25% w/v PEG 3350, supplemented with 10% (0.2% w/v Betaine anhydrous, 0.2% w/v L-Glutamic acid, 0.2% w/v L-Proline, 0.2% w/v Taurine, 0.2% w/v Trimethylamine N-oxide dihydrate, 0.02 M HEPES sodium pH 6.8 from the Silver Bullets Screen (Hampton) at room temperature. The crystals were cryo-Protected by addition of a final concentration of 25% glycerol. Diffraction data was collected from a single crystal at Swiss Light Source (SLS) at the beamline X06SA. For the data collection increments of 0.1° per 0.1 s were collected for 360° at a beam transmission of 10% and a beam size of 10 μM x 10 μM.

After preincubating TC:CD320 at 10 mg/ml with a 1.2-fold molar excess of TCNB11 for 1h on ice, crystals of the TCNB11:TC:CD320 complex were grown by sitting drop vapor diffusion in 0.2 M sodium malonate pH 6.0, 20% w/v PEG 3350. The crystals were cryo-Protected by addition of a final concentration of 25% glycerol. Diffraction data was collected from a single crystal at Swiss Light Source (SLS) at the beamline X06SA. For the data collection increments of 0.1° per 0.1 s were collected for 360° at a beam transmission of 10% and a beam size of 10 μM x 40 μM.

After preincubating TC:CD320 at 10 mg/ml with a 1.2-fold molar excess of TCNB26 for 1h on ice, crystals of the TCNB26:TC_2:$CD320_2$ complex were grown by sitting drop vapor diffusion in 150 mM Ammonium citrate tribasic pH 7.0, 21% w/v PEG 3350. The crystals were cryo-Protected by addition of a final concentration of 25% glycerol. Diffraction data was collected from a single crystal at Swiss Light Source (SLS) at the beamline X06SA. For the data collection increments of 0.1° per 0.1 s were collected for 360° at a beam transmission of 10% and a beam size of 20 μM x 20 μM.

After preincubating TC:CD320 at 10 mg/ml with a 1.2-fold molar excess of TCNB34 for 1h on ice, crystals of the TCNB34:TC:CD320 complex were grown by hanging drop vapor diffusion in 200 mM Ammonium Iodide, 16% w/v PEG 3350 and 1.4 x

10^5-fold diluted seed stocks from crystals of the same protein, grown in similar conditions. The crystals were cryo-Protected by soaking in a drop containing 25% ethylene glycol. Seeds were generated using a Seed Bead™ Kit (Hampton). Diffraction data was collected from a single crystal at Swiss Light Source (SLS) at the beamline X06SA. For the data collection increments of 0.1° per 0.1 s were collected for 360° at a beam transmission of 10% and a beam size of 20 µM x 40 µM.

Data processing, structure determination, model building and refinement

Crystallographic data were indexed, and reduced by XDS[21], Structures were solved by molecular replacement in Phaser MR[22] using a single monomer of TC:CD320 (PDB ID 4ZRP)[11] as input model. Nanobodies were fitted based on an existing model of a nanobody (PDB ID 3P0G)[23]. Model building was done in Coot[24] and Refinement was performed in Phenix[25]. To avoid bias from the comparably low-resolution map of TCNB26:TC$_2$:CD320, for this structure only rigid body and B-factor refinement was performed.

Figure preparation and data analysis

Protein sequence alignments were performed in CLC Genomics Workbench. Cell viability assay data was analyzed and plotted in GraphPad Prism. Protein structure images were generated in PyMol[26].

Acknowledgements

Extended Data

Extended Data Figure 1 Internalization of dsRed TC-Cbl and dsRed TC-Cbl–TCNB4 by HEK293 cells expressing TCblR-GFP. Cells were incubated with dsRedTC-Cbl (A) or with nanobody TCNB4 +dsRed TC-Cbl for 1, 2 or 4 h at 37 °C. Over time the dsRed-TC-Cbl is internalized and is degraded (panels A), however, the TCNB4 + dsRed TC-Cbl is internalized but is not degraded (panels B).
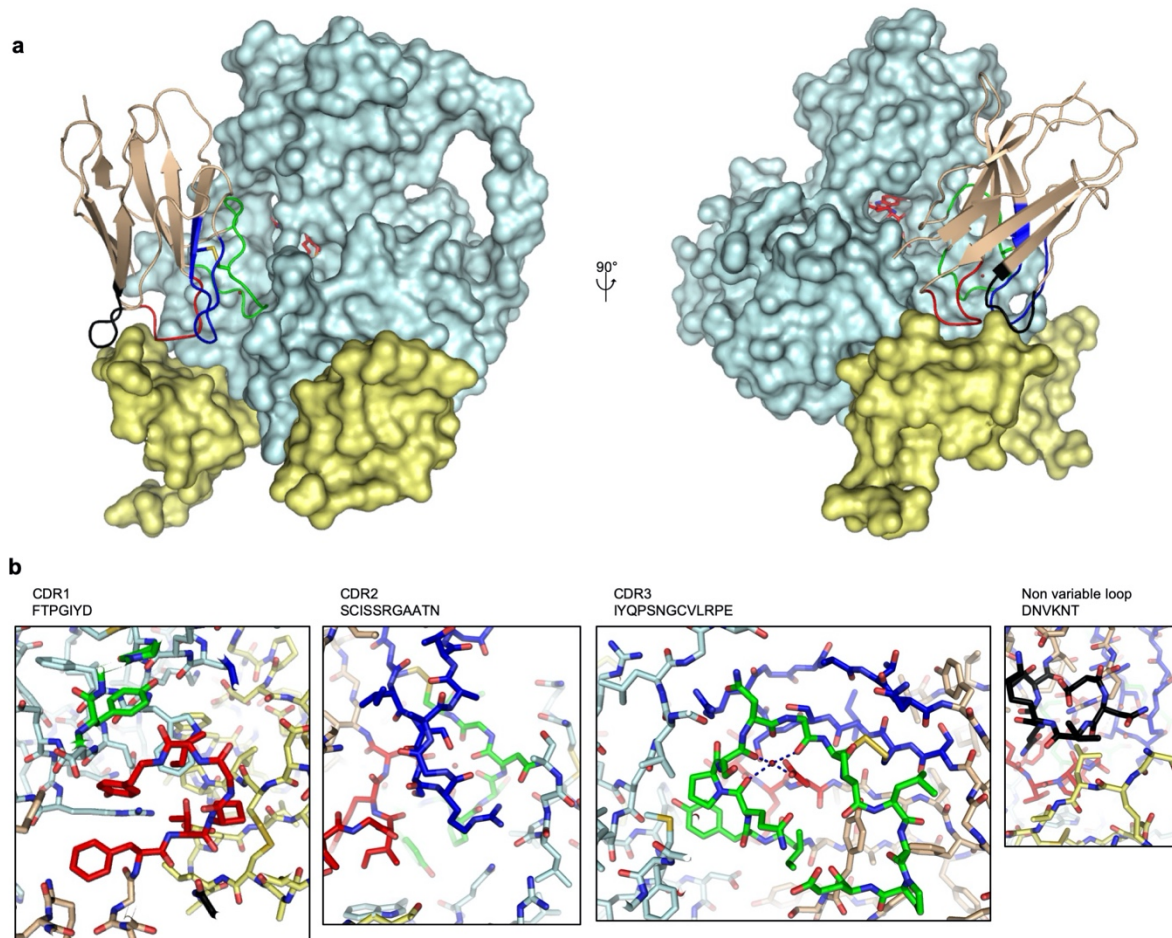
Extended Data Figure 2 Crystal structure of TC:CD320 bound to TCNB4. a, TC (cyan) and CD320 (yellow) are shown in surface representation. Cyanocobalamin (red) is shown in stick representation. TCNB4 (brown) is shown in cartoon representation with CDR1 (red), CDR2 (blue), CDR3 (green) as well as a non-variable loop (black) that is interacting with CD320 are highlighted with colors. A disulfide bond between CDR2 and CRD3 is shown in stick representation. Note that the nanobody is interacting with both TC as well as the receptor CD320. **b**, Close up view on molecular interaction between different CDR loops of the nanobody and TC:CD320 shown in stick representation, using the same colors as in a. Note that in CDR3 there seems to be a Ca^{2+} ion coordinated by three backbone carboxyl groups of CDR3 and by one aspartate of CDR1.

58
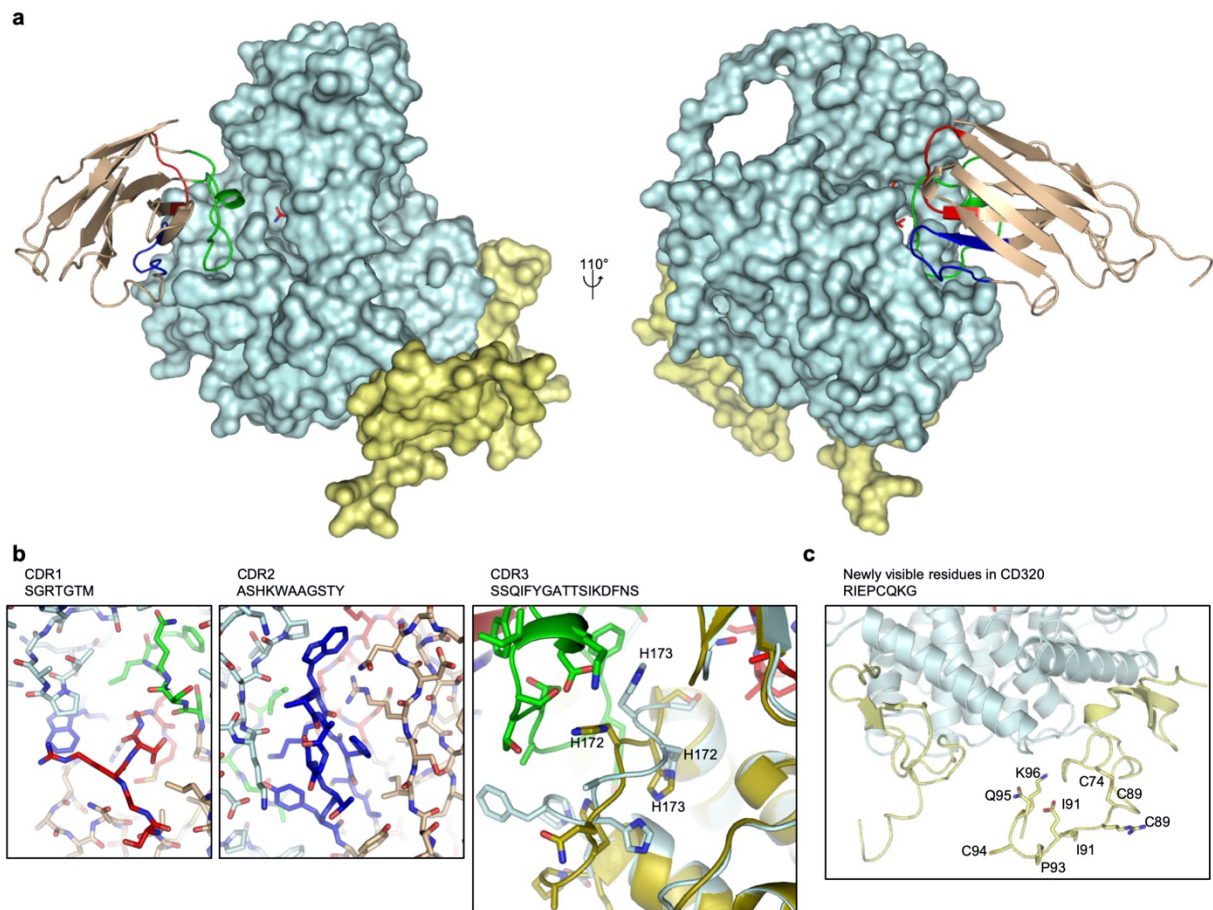
Extended Data Figure 3 Crystal structure of TC:CD320 bound to TCNB11. a, TC (cyan) and CD320 (yellow) are shown in surface representation. Cyanocobalamin (red) is shown in stick representation. TCNB11 (brown) is shown in cartoon representation with CDR1 (red), CDR2 (blue), CDR3 (green) are highlighted with colors. **b,** Close up view on molecular interaction between different CDR loops of the nanobody and TC shown in stick representation, using the same colors as in a. In the panel for CDR3 the structure of TC has been superposed with a nanobody free structure of TC (gold) (4ZRP[11]) and for clarity the structures are shown in cartoon representation and only selected residues and B12 are shown in stick representation. Note that binding of TCNB11 induces a twist in TC in a loop at the binding interface. **c,** Residues of CD320 extending LDLR-A1 that were not resolved in previous structures, as well as the disulfide-bond forming resides Cys74 and Cys89 are shown in stick representation**.**
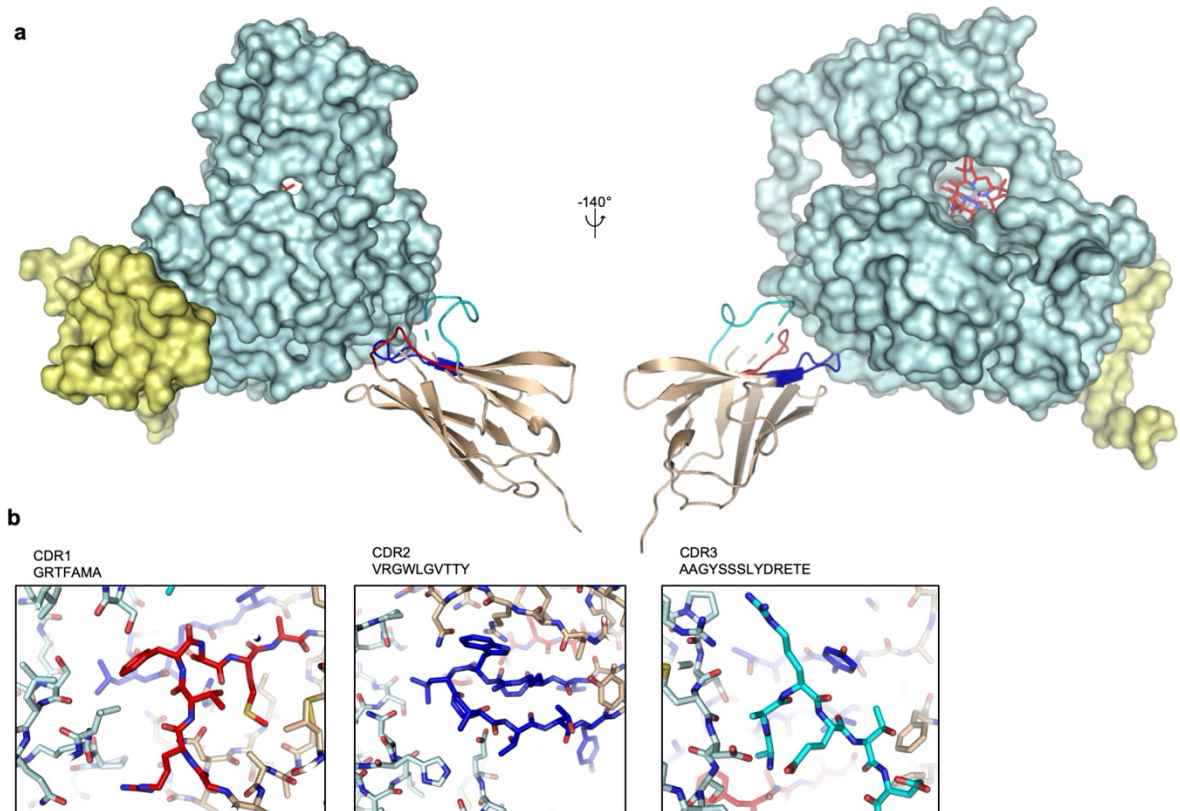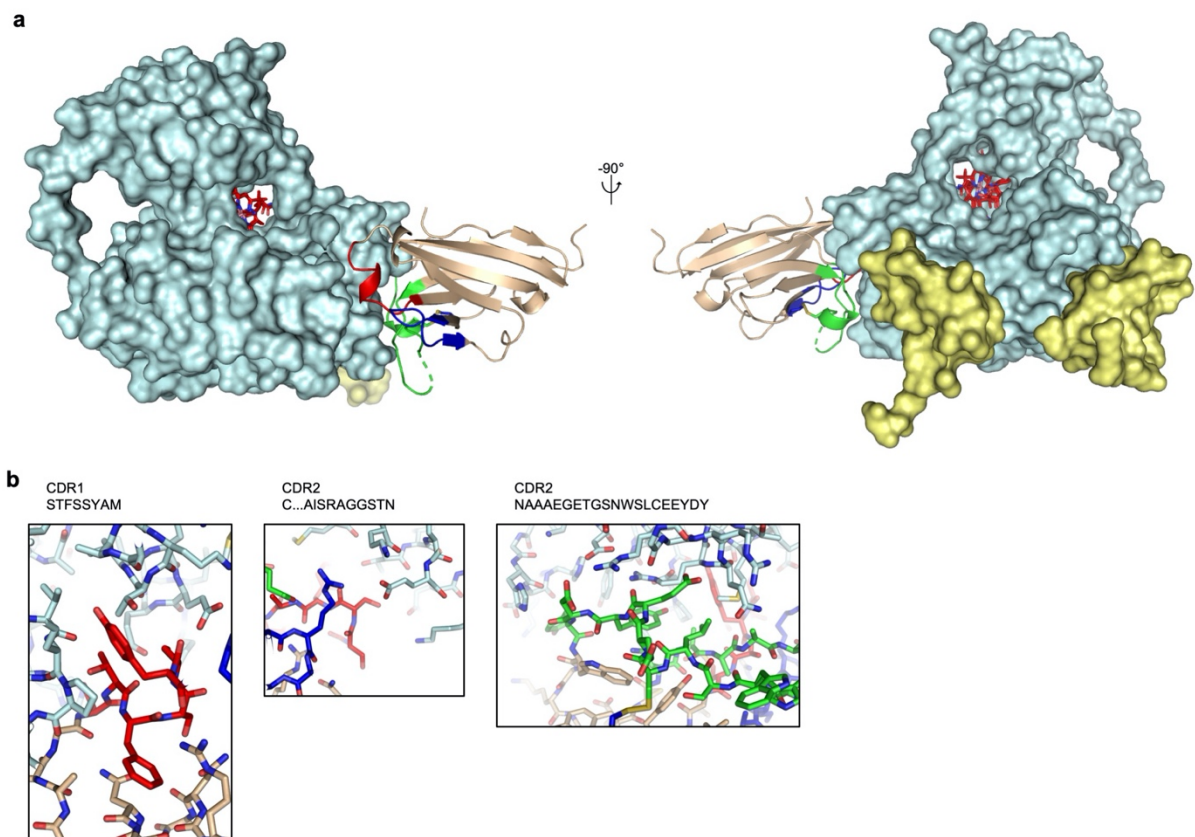
Extended Data Figure 4 Crystal structure of TC:CD320 bound to TCNB26. a, TC (cyan) and CD320 (yellow) are shown in surface representation. Cyanocobalamin (red) is shown in stick representation. TCNB26 (brown) is shown in cartoon representation with CDR1 (red), CDR2 (blue), CDR3 (green) highlighted with colors. **b**, Close up view on molecular interaction between different CDR loops of the nanobody and TC, shown in stick representation, using the same colors as in **a**.

Extended Data Figure 5 Crystal structure of TC:CD320 bound to TCNB34. a, TC (cyan) and CD320 (yellow) are shown in surface representation. Cyanocobalamin (red) is shown in stick representation. TCNB34 (brown) is shown in cartoon representation with CDR1 (red), CDR2 (blue), CDR3 (green) highlighted with colors. A disulfide bond between CDR2 and CRD3 is shown in stick representation. **b**, Close up view on molecular interaction between different CDR loops of the nanobody and TC shown in stick representation, using the same colors as in **a.**

Extended Data Table 1

	TCNB4:TC:CD320	TCNB11:TC:CD320	TCNB26:TC$_2$:CD320$_2$	TCNB34:TC:CD320
Wavelength	0.979502	1	0.979502	0.979502
Resolution range	48.92 - 2.692 (2.789 - 2.692)	48.96 - 3.0 (3.107 - 3.0)	49.14 - 4.184 (4.334 - 4.184)	49.27 - 1.854 (1.92 - 1.854)
Space group	P 21 21 21	C 2 2 21	P 43 21 2	P 43 21 2
Unit cell	105.339 116.418 125.498 90 90 90	103.717 195.831 198.385 90 90 90	102.129 102.129 360.998 90 90 90	104.822 104.822 167.46 90 90 90
Total reflections	84243 (7177)	81479 (8030)	28382 (1406)	158598 (15319)
Unique reflections	42564 (3745)	40763 (4011)	14241 (752)	79459 (7693)
Multiplicity	2.0 (1.9)	2.0 (2.0)	2.0 (1.9)	2.0 (2.0)
Completeness (%)	97.98 (87.42)	99.84 (99.80)	94.99 (51.80)	99.79 (98.15)
Mean I/sigma(I)	7.97 (1.17)	11.45 (2.30)	10.17 (1.59)	19.37 (1.89)
Wilson B-factor	56.39	68.96	164.21	31.94
R-merge	0.08061 (0.6841)	0.05523 (0.371)	0.04472 (0.4819)	0.0272 (0.4422)
R-meas	0.114 (0.9675)	0.07811 (0.5247)	0.06324 (0.6815)	0.03847 (0.6254)
R-pim	0.08061 (0.6841)	0.05523 (0.371)	0.04472 (0.4819)	0.0272 (0.4422)
CC1/2	0.993 (0.459)	0.997 (0.791)	0.999 (0.851)	0.999 (0.66)
CC*	0.998 (0.793)	0.999 (0.94)	1 (0.959)	1 (0.892)
Reflections used in refinement	42558 (3745)	40739 (4008)	14207 (748)	79457 (7693)
Reflections used for R-free	2128 (187)	2037 (201)	712 (37)	3973 (384)
R-work	0.2287 (0.3664)	0.2057 (0.3629)	0.2666 (0.3837)	0.1731 (0.2875)
R-free	0.3015 (0.3973)	0.2731 (0.4116)	0.2866 (0.3419)	0.2044 (0.3046)
CC(work)	0.933 (0.671)	0.948 (0.792)	0.843 (0.681)	0.957 (0.824)
CC(free)	0.896 (0.494)	0.895 (0.695)	0.694 (0.644)	0.938 (0.784)
Number of non-hydrogen atoms	9598	9548	8372	5545
macromolecules	9352	9354	8182	4826
ligands	193	191	190	312
solvent	53	3	1070	407
Protein residues	1217	1217	0.008	612
RMS(bonds)	0.016	0.018	1.24	0.012
RMS(angles)	1.73	1.75	93.89	2.58
Ramachandran favored (%)	86.11	84.04	5.15	98.17
Ramachandran allowed (%)	10.29	10.36	0.95	1.83
Ramachandran outliers (%)	3.6	5.6	1.36	0
Rotamer outliers (%)	0.1	0	8.9	0.76
Clashscore	29.25	20.96	213.36	9.14
Average B-factor	74.07	74.18	214.3	40.38
macromolecules	74.34	74.33	172.98	40.28
ligands	66.01	66.99		34.87
solvent	54.3	57.6		45.8

References

1   Hall, C. A., Colligan, P. D. & Begley, J. A. Cyclic activity of the receptors of cobalamin bound to transcobalamin II. *J Cell Physiol* **133**, 187-191, (1987).

2   Lindemans, J. *et al.* Uptake of transcobalamin II-bound cobalamin by HL-60 cells: effects of differentiation induction. *Exp Cell Res* **184**, 449-460, (1989).

3   Amagasaki, T., Green, R. & Jacobsen, D. W. Expression of transcobalamin II receptors by human leukemia K562 and HL-60 cells. *Blood* **76**, 1380-1386, (1990).

4   Nielsen, M. J., Rasmussen, M. R., Andersen, C. B. F., Nexo, E. & Moestrup, S. K. Vitamin B-12 transport from food to the body's cells-a sophisticated, multistep pathway. *Nat Rev Gastro Hepat* **9**, 345-354, (2012).

5   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. The protein and the gene encoding the receptor for the cellular uptake of transcobalamin-bound cobalamin. *Blood* **113**, 186-192, (2009).

6   Clardy, S. M., Allis, D. G., Fairchild, T. J. & Doyle, R. P. Vitamin B-12 in drug delivery: breaking through the barriers to a B-12 bioconjugate pharmaceutical. *Expert Opin Drug Del* **8**, 127-140, (2011).

7   Waibel, R. *et al.* New derivatives of vitamin B12 show preferential targeting of tumors. *Cancer Res* **68**, 2904-2911, (2008).

8   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. Targeted Delivery of Saporin Toxin by Monoclonal Antibody to the Transcobalamin Receptor, TCblR/CD320. *Mol Cancer Ther* **9**, 3033-3040, (2010).

9   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. Saporin Conjugated Monoclonal Antibody to the Transcobalamin Receptor TCblR/320 Is Effective in Targeting and Destroying Cancer Cells. *Journal of cancer therapy* **4**, 1074-1081, (2013).

10   De Meyer, T., Muyldermans, S. & Depicker, A. Nanobody-based products as research and diagnostic tools. *Trends Biotechnol* **32**, 263-270, (2014).

11   Alam, A. *et al.* Structural basis of transcobalamin recognition by human CD320 receptor. *Nat Commun* **7**, 12100, (2016).

12   Jiang, W., Nakayama, Y., Sequeira, J. M. & Quadros, E. V. Characterizing monoclonal antibodies to antigenic domains of TCblR/CD320, the receptor for cellular uptake of transcobalamin-bound cobalamin. *Drug Deliv* **18**, 74-78, (2011).

13   Polito, L., Bortolotti, M., Mercatelli, D., Battelli, M. G. & Bolognesi, A. Saporin-S6: a useful tool in cancer therapy. *Toxins (Basel)* **5**, 1698-1722, (2013).

14   Wuerges, J. *et al.* Structural basis for mammalian vitamin B12 transport by transcobalamin. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4386-4391, (2006).

15   Kishan, K. V., Newcomer, M. E., Rhodes, T. H. & Guilliot, S. D. Effect of pH and salt bridges on structural assembly: molecular structures of the monomer and intertwined dimer of the Eps8 SH3 domain. *Protein Sci* **10**, 1046-1055, (2001).

16   Quadros, E. V. *et al.* Positive Newborn Screen for Methylmalonic Aciduria Identifies the First Mutation in TCblR/CD320, the Gene for Cellular Uptake of Transcobalamin-bound Vitamin B-12. *Hum Mutat* **31**, 924-929, (2010).

17   Hofmann, S. *et al.* Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580-583, (2019).

18   Mireku, S. A., Sauer, M. M., Glockshuber, R. & Locher, K. P. Structural basis of nanobody-mediated blocking of BtuF, the cognate substrate-binding protein of the Escherichia coli vitamin B12 transporter BtuCD. *Sci Rep* **7**, 14296, (2017).

19   Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* **1**, 100, (2010).

20   Jiang, W. X., Nakayama, Y., Sequeira, J. M. & Quadros, E. V. Mapping the functional domains of TCblR/CD320, the receptor for cellular uptake of transcobalamin-bound cobalamin. *Faseb J* **27**, 2988-2994, (2013).

21   Kabsch, W. Xds. *Acta Crystallogr D* **66**, 125-132, (2010).

22   Mccoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, (2007).

23   Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **469**, 175-180, (2011).

24   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132, (2004).

25   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **66**, 213-221, (2010).

26   The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

4. Structure and mechanism of the ER-based glucosyl-transferase ALG6

The contents of this chapter have been published in Nature:

Statement of Contribution

K.P.L. and J.S.B. conceived the project. J.S.B. cloned, screened, expressed and purified proteins, reconstituted ALG6 into lipid nanodiscs, performed biochemical experiments and produced Dol25-PP-GlcNAc2Man9 using synthetic precursor substrates. G.P. synthesized Dol25-P-Glc and derivatives. J.B. synthesized Dol25-P-Man and Dol25-PP-GlcNAc2. K.N. and J.S.B. performed synthetic antibody generation. J.S.B. prepared grids. R.N.I. collected cryo-EM data. J.S.B. assisted cryo-EM data collection, processed data, built the ALG6 models and refined the structures. T.D. and J.L.R. supervised the synthesis of LLO substrates. A.A.K. provided the Chaperone-Assisted Structure Determination phage display pipeline and supervised synthetic antibody generation. M.A. analyzed data and contributed to paper writing. J.S.B. and K.P.L. analyzed data and wrote the paper.

Structure and mechanism of the ER-based glucosyltransferase ALG6

Joël S. Bloch[1], Giorgio Pesciullesi[2], Jérémy Boilevin[2], Kamil Nosol[1,3], Rossitza N. Irobalieva[1], Tamis Darbre[2], Markus Aebi[4], Anthony A. Kossiakoff[3], Jean-Louis Reymond[2], Kaspar P. Locher[1]*

[1]Institute of Molecular Biology and Biophysics, ETH Zürich, Otto-Stern-Weg 5, 8093 Zürich, Switzerland.
[2]Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland.
[3]Department of Biochemistry and Molecular Biology, University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA.
[4]Institute of Microbiology, ETH Zürich, Vladimir-Prelog-Weg 1-5/10, 8093 Zürich, Switzerland.

In eukaryotic protein N-glycosylation, a series of glycosyltransferases catalyze the biosynthesis of a dolichylpyrophosphate-linked oligosaccharide prior to its transfer onto acceptor proteins[1]. The final seven steps occur in the lumen of the endoplasmic reticulum (ER) and require dolichylphosphate-activated mannose and glucose as donor substrates[2]. The responsible enzymes ALG3, ALG9, ALG12, ALG6, ALG8, ALG10 are glycosyltransferases of the C-superfamily (GT-Cs), which is loosely defined as containing membrane-spanning helices and processing an isoprenoid-linked carbohydrate donor substrate[3,4]. Here we present the cryo-EM structure of yeast ALG6 at 3.0 Å resolution, which reveals a new transmembrane protein fold. A comparison with reported GT-C structures suggests that GT-C enzymes contain a modular architecture with a conserved and a variable module, each with distinct functional roles. The activity of ALG6 was recapitulated *in vitro* using synthetic analogs of dolichylphosphate- and dolichylpyrophosphate-linked sugars and enzymatic glycan extension to generate donor and acceptor substrates using purified enzymes of the ALG pathway. A second cryo-EM structure of ALG6 bound to an analog of dolichylphosphate-glucose at 3.9 Å resolution revealed the active site of the enzyme. Functional analysis of ALG6 variants identified a catalytic aspartate residue likely acting as a general base. This residue is conserved in the GT-C superfamily. Our results define the architecture of ER-luminal GT-C enzymes and provide a structural basis for understanding their catalytic mechanisms.

In the eukaryotic cell, N-glycosylation of secretory proteins is an essential process that involves the transfer of a high-mannose glycan ($GlcNAc_2Man_9Glc_3$) from a dolichylpyrophosphate (DolPP) carrier, catalyzed by oligosaccharyltransferase[1]. The biosynthesis of this donor substrate is a sequential process that is initiated in the cytoplasm, requires the flipping of the $GlcNAc_2Man_5$-containing intermediate, and is completed on the luminal side of the ER, where the transfer of the final seven hexoses occurs[2] (Fig. 1a). For the cytoplasmic reactions, the donor substrates are soluble, nucleotide-activated sugars. In contrast, the ER-luminal reactions depend on the membrane-embedded donor substrates dolichylphosphate-mannose (DolP-Man) and dolichylphosphate-glucose (DolP-Glc)[2]. ALG6 transfers the first of three glucose moieties onto the pre-assembled $GlcNAc_2Man_9$ glycan. Following the identification[5] and cloning[6] of the *ALG6* locus, it was shown that deficiencies in ALG6 are a frequent cause of congenital disorders of glycosylation (CDG)[7], where patients generally suffer from hypo-glycosylated serum glycoproteins[8,9]. To understand the mechanism of ALG6, we sought to obtain high-resolution structural insight.
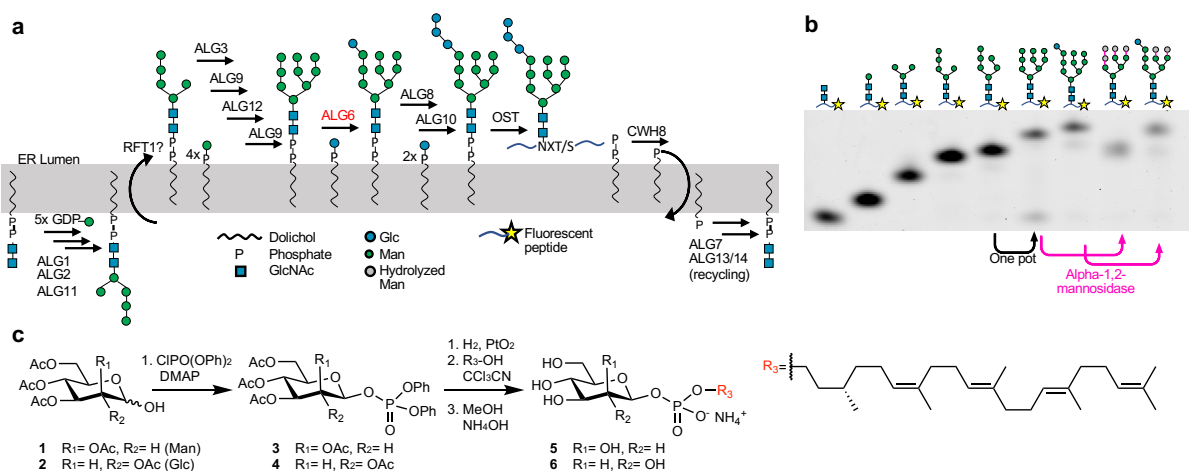

Figure 1 Functional analysis of ALG6. a, Schematic of eukaryotic LLO biosynthesis and glycan transfer by oligosaccharyltransferase (OST)2,36. Enzyme names are indicated above reaction arrows. Curved arrows across the ER membrane indicate flipping reactions. **b**, Tricine gel-based analysis of chemo-enzymatic LLO extension and of ALG6-catalyzed glucose transfer. Glycans were transferred onto a fluorescent peptide using purified T. brucei STT3A enzyme11. The glycan structures are indicated schematically above the lanes. Note that alpha-1,2-mannosidase is unable to remove mannoses from the A-branch of GlcNAc2Man9Glc, demonstrating the presence of a glucose moiety. **c**, Synthesis scheme of dolichylphosphate-hexose analogs including Dol25-P-Man (5) and Dol25-P-Glc (6). Note that the citronellyl-farnesyl moiety of Dol25 faithfully reflects the stereochemistry of native dolichol.

In vitro activity of purified yeast ALG6

Neither the donor nor the acceptor substrate of ALG6 is commercially available. To generate the acceptor substrate, we started from our previously developed[10] method

for chemo-enzymatic synthesis and analysis of Dol25-PP-GlcNAc$_2$Man$_5$, where Dol25 refers to a citronellyl-farnesyl moiety. The lipid-linked glycans were analyzed by tricine gel analysis following a transfer onto a fluorescent acceptor peptide catalyzed by the eukaryotic, single-subunit oligosaccharyltransferase STT3A from *T. brucei*[10] (Fig. 1b). To append mannose moieties to Dol25-PP-GlcNAc$_2$Man$_5$, we heterologously expressed and purified the membrane-integral enzymes ALG3, ALG9 and ALG12 (Supplementary Information Fig. 1a) and synthesized the required mannose donor analog Dol25-P-Man (Fig. 1c). Using a stepwise procedure, we extended Dol25-PP-GlcNAc$_2$Man$_5$ to the ALG6 acceptor substrate Dol25-PP-GlcNAc$_2$Man$_9$ (Fig. 1b). We also synthesized the required donor substrate Dol25-P-Glc *de novo* (Fig. 1c). Incubation of ALG6 with Dol25-P-Glc and Dol25-PP-GlcNAc$_2$Man$_9$ led to complete conversion to Dol25-PP-GlcNAc$_2$Man$_9$Glc (Fig. 1b). To confirm the physiologically relevant glycan structure, the peptide-attached glycans were incubated with alpha-1,2-mannosidase, which showed that ALG6-catalyzed glucose addition protected the A-branch of the Man$_9$-containing glycan from mannosidase activity (Fig. 1b). ALG6 was found to be active in different detergents as well as reconstituted in lipid nanodiscs (Extended Data Fig. 1a).

Notably, our chemo-enzymatic approach enabled us to synthesize every lipid-linked oligosaccharide (LLO) intermediate of the biosynthetic pathway (Extended Data Fig. 1b). Our LLO analogs are readily accepted both by the single subunit OST enzyme from *Trypanosoma brucei*[11] and the octameric yeast OST enzyme[12].

Structure determination of ALG6

We used phage display and a synthetic library of antigen-binding Fab fragments[13] to identify a conformational binder (6AG9-Fab) that increased the size and thermostability of detergent-solubilized ALG6 (Fig. 2a. Supplementary Information Fig. 2a, b) but did not interfere with its catalytic activity (Fig. 2b). The increased particle mass and size (Extended Data Fig. 2) aided particle alignment, and the EM reconstruction of nanodisc-reconstituted apo-ALG6 bound to 6AG9-Fab had an overall resolution of 3.0 Å (Fig. 2c, Extended Data Fig. 2, Extended Data Table 1, Supplementary Information Fig. 3). This allowed us to build an atomic model of ALG6 lacking only 37 residues at the presumably flexible N-terminus and 28 residues in the cytoplasmic loop connecting TM12 and TM13, a segment that contains several short

linear motifs (SLiMs) including a di-arginine ER-retention motif[14]. 6AG9-Fab bound to a conformational, ER-luminal epitope of ALG6 and all interacting CDR-loops of the Fab fragment are well-resolved (Extended Data Fig. 3). The EM maps revealed bound lipids and cholesterol hemisuccinate molecules in contact with transmembrane helices of ALG6 (Extended Data Fig. 4a).
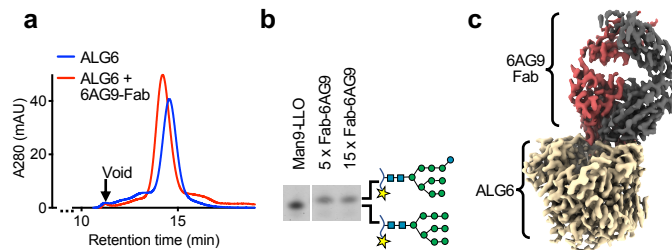

Figure 2 ALG6-Fab interactions. a, Size exclusion chromatography of purified ALG6 (64 kDa) in detergent in the presence and absence of 6AG9-Fab (49 kDa). **b**, Analysis of ALG6 activity as described in Fig. 1b but in the presence of 6AG9-Fab. 6AG9-Fab was incubated with ALG6 at a 5- or 15-fold molar excess to test inhibition. The lane labeled Man9-LLO is a control sample for size comparison. **c**, Refined and B-factor sharpened EM map of nanodisc-reconstituted ALG6 bound to 6AG9-Fab colored by proximity to protein chains of ALG6 (light brown), 6AG9-Fab heavy chain (red), and 6AG9-Fab light chain (grey).

ALG6 topology and GT-C architecture

ALG6 has 14 transmembrane helices and two long loops (EL1 and EL4) that form helices in the ER lumen (Fig. 3a-c). A disulfide bridge covalently links the ER-luminal end of TM14 with the external loop EL4 that connects TM7 and TM8 (Fig. 3b, c). While the overall fold and membrane topology of ALG6 is novel, we noticed that the TM arrangement and topology of the first half of the protein resembles partial structures of previously reported GT-C enzymes. We superimposed the structure of ALG6, which belongs to the GT57 subfamily according to the CAZy data base[15], to PglB[16] (GT 66), AglB[17] (GT66), STT3[12,18] (GT66), ArnT[19] (GT83), and PMT1/PMT2[20] (GT39). We found that the first seven TM helices (including the luminal helices EL1-h1 and EL1-h2) are structurally similar, whereas the remainder of the proteins are distinct in fold, topology and number of TM helices (Fig. 3d, Extended Data Fig. 5). This suggests that GT-C enzymes contain a structurally conserved, N-terminal module consisting of TM1-TM7, and a structurally variable, C-terminal module (Fig. 3b-d). The structural similarity of the conserved module is not reflected in sequence conservation, as the primary sequences of the compared proteins are dissimilar[21]. However, the first external loop contains the catalytically essential aspartate residue Asp69 (see below).

Our interpretation of a modular architecture is supported by the observation that in all published structures of GT-C enzymes, the active site and the substrate-binding cavities are located at the interface of the conserved and variable modules (purple asterisk and red arrows in Fig. 3d). The architectural similarity further extends to how GT-Cs bind their dolichyl(pyro)phosphate-linked donor substrates: The dolichol moiety invariably interacts with TM6 of the conserved GT-C module, whereas the distinct attached carbohydrates, which include single sugars or complex glycans, interact with the variable module. The donor substrates reach their respective active sites by "diving under" arch-like structures formed by the external (or ER-luminal) loops that either link the conserved to the variable modules, or are part of the variable modules (Fig. 3d). These loops contain at the minimum a helix running parallel to the membrane but may contain entire domains (Fig. 3d). We conclude that the conserved module of GT-Cs likely served as a platform that allowed distinct functionalities to evolve and that the modularity has allowed GT-C enzymes to accommodate different donor and acceptor substrates and generate diverse active sites for distinct catalytic mechanisms during glycan transfer reactions.
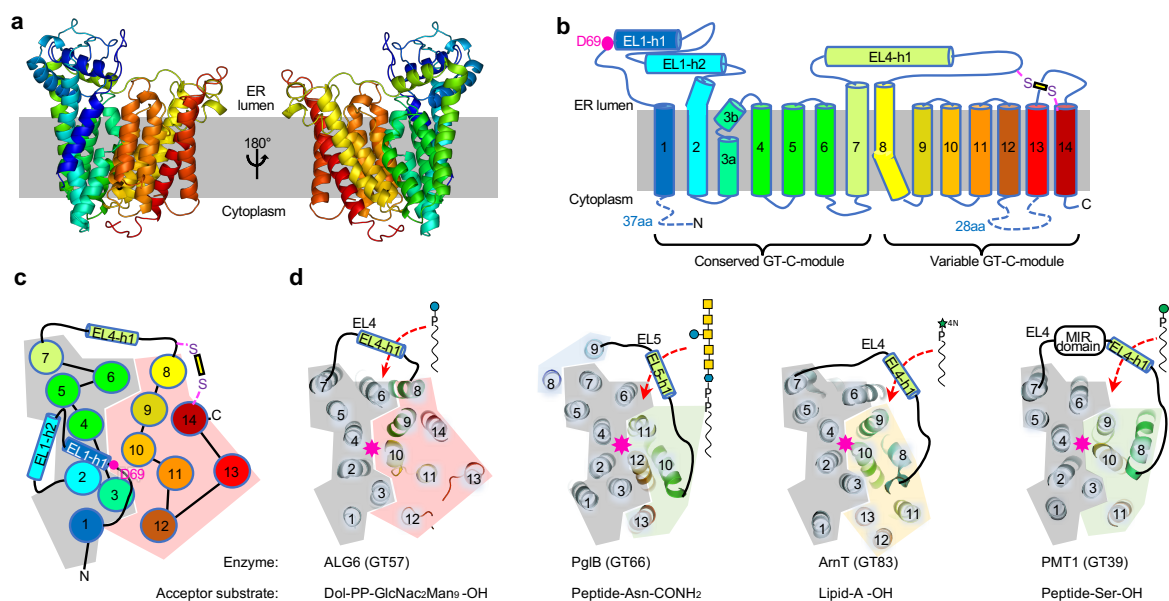

Figure 3 ALG6 structure and GT-C architecture. a, Cartoon representation of ALG6 in rainbow coloring (blue at N-terminus, red at C-terminus). **b**, Schematic representation ALG6 topology, with TM helices and external loops (EL) numbered and colored as in **a**. The pink dot indicates the proposed catalytic residue Asp69. The yellow bar depicts a disulfide bond. **c**, Top view of schematic representation of ALG6 colored as in b, with numbered circles depicting TM helices and cylinders depicting helices in ER luminal loops. **d**, comparison of the topologies of ALG6 and other selected GT-C enzymes of known structure. Ribbon diagrams of top views (as in **c**) of the structures are shown, and the constant GT-C modules are shaded grey, whereas the variable GT-C-modules are shaded in different colors for each structure. TM helices are numbered. Note that GT39 and GT66 contain similar variable modules (shaded green). Purple asterisks indicate the locations of the catalytic sites. Dashed red arrows indicate the entry site for the dolichol-linked donor substrates, which are depicted using standard glycobiology symbols.

An analysis of residues reported to be implicated in ALG6-associated human CDGs[7,22-25] revealed that none of them directly point at the presumed active site. Rather, they are second shell-residues with respect to the active site or are located even further away (Extended Data Fig. 6). It was indeed shown that such mutations do not fully abolish ALG6 function but rather reduce the enzymatic activity to a level that is not yet lethal[22].

Structure of Dol25-P-Glc-bound ALG6

ALG6 forms a large, hydrophilic cavity facing the ER lumen and a groove-shaped cavity facing the lipid bilayer. The residues lining these cavities are not only highly conserved among ALG6 homologs, but also in the functionally related glucosyltrasferase ALG8[2,26], suggesting that they are involved in substrate binding or form the active site of the enzyme (Extended Data Fig. 7). To investigate the roles of these cavities, we collected a second cryo-EM dataset of the ALG6-Fab complex (Extended Data Fig. 8, Extended Data Table 1). Here, ALG6 was solubilized in digitonin and pre-incubated with the synthetic donor substrate Dol25-P-Glc. Although the overall resolution of the structure is 3.9Å and thus lower than that of nanodisc-reconstituted apo-ALG6, well-resolved density for bound donor substrate was observed in a lipid-exposed groove formed by TM6, TM7 and TM8 (Fig. 4a, b. Extended Data Fig. 4b). The shape of the density can clearly accommodate a Dol25-P-Glc molecule. The citronellyl-farnesyl moiety is located in the conserved, lipid-facing groove, interacts mostly with hydrophobic residues from TM6. The loop EL4 forms an arch above the groove on the ER-luminal side, resulting in a funnel-like entrance, and the Dol25 moiety bends under this arch at the level of the membrane boundary. This causes the phosphate group, which represents the leaving group of the glucose transfer reaction, to be lodged in a slightly positively charged surface region of ALG6 (Extended Data Fig 4b).

The EM density revealed a feature large enough to accommodate the glucose moiety of Dol25-P-Glc, but the quality of the map does not allow exact positioning. This is in part due to the fact that this density also covers the adjacent side chain of the conserved histidine residue His378 (Fig. 4a, b). We could nevertheless assign the orientation of the glucose moiety, with one face packing against the surface of ALG6 and the other pointing towards the solvent. We built the model such that the α-

anomeric position of the C1 carbon is accessible for a nucleophilic attack by the mannose C3 hydroxyl group of the acceptor glycan. This binding mode can ensure the specificity of ALG6 for DolP-Glc over DolP-Man, as the axial C2 hydroxyl group of a mannose would cause a steric clash with the enzyme surface (Fig. 4a). The specificity of ALG6 can be demonstrated *in vitro*, where only glucose, but not mannose, is transferred from a Dol25-P carrier (Extended data Fig. 9a).
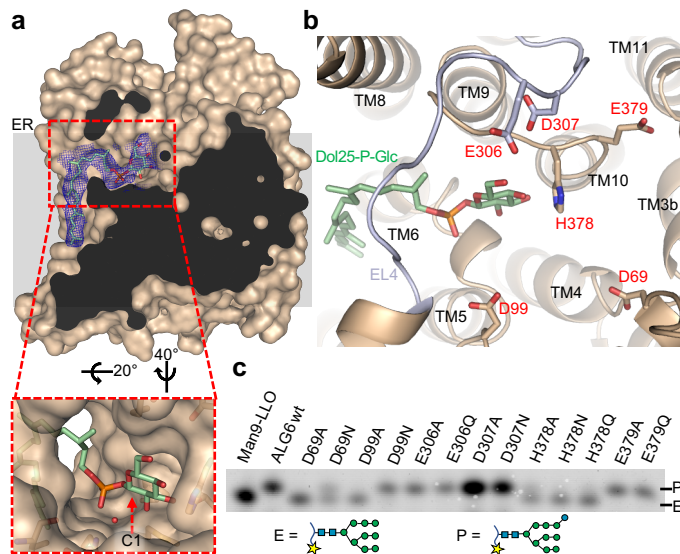

Figure 4 Donor substrate binding and catalytic site of ALG6. a, Surface representation of substrate-bound ALG6 in light brown with EM density map for Dol25-P-Glc in blue mesh carved to 3Å, with bound Dol25-P-Glc in stick representation. The front of ALG6 was clipped for clarity. The inset shows a zoomed-in view. A dashed, red arrow indicates the anomeric C1 of the glucose moiety. An ordered water molecule close to the C1 carbon is shown as a red sphere. **b**, Structure of the active site, with ALG6 shown in light brown ribbon representation with EL4 colored in light blue. Acidic residues in the active site are shown in stick representation and labeled, bound Dol25-P-Glc is depicted in stick representation, with carbon atoms colored green. **c**, Analysis of substrate conversion by ALG6 variants, with mutations indicated above the lanes.

We could not unambiguously identify a second dolichol-binding groove in the ALG6 structure, and multiple smaller grooves are potential candidates for binding the acceptor substrate DolPP-GlcNAc$_2$Man$_9$ (Extended Data Fig, 4c). Because several of the TM helices of ALG6 are not perpendicular to the lipid bilayer, the protein appears to induce a distortion of the lipid bilayer of the nanodisc, with the lipid bilayer appearing to thin out at the side of the cavity opposite to the "arch" generated by EL4 (Extended Data Fig. 4d). Such membrane thinning might facilitate the binding of the dolichol-linked donor substrate form the membrane. A deformation of a membrane caused by a protein has been reported previously in the TMEM16 scramblase, where it has been proposed to facilitate or assist in lipid flipping[27].

Catalytic mechanism

Given that ALG6 is an inverting glycosyltransferase, its reaction is thought to occur via an S_N2 mechanism[28]. This implies that the C3 hydroxyl group of the terminal A-branch mannose of Dol-PP-GlcNAc$_2$Man$_9$ has to be activated for a nucleophilic attack on the anomeric C1 of the glucose moiety to form an α-1-3-glycosidic bond. This is generally achieved by deprotonation of the attacking hydroxyl group by an aspartate or glutamate side chain acting as a general base, either directly or by forming catalytic dyads or triads with suitable groups such as imidazole rings (histidine), hydroxyl groups of protein sidechains or ordered water molecules[29-31]. In our Dol25P-Glc-bound ALG6-structure, we identified five acidic residues in external loops that could act as a general base (Fig. 4b): Asp69 and Asp99 from EL1, Glu306 and Asp307 from EL4 and Glu379 from EL5. We mutated all five residues to alanine and, in the case of aspartates or glutamates, to asparagine or glutamine. We also mutated His378, which is in direct contact with the glucose moiety, to investigate its potential function in substrate binding. We expressed and purified the resulting ALG6 variants (Supplementary Information Fig. 1b, c) and tested their *in vitro* activity using our coupled assay (Fig. 4c). We found that activity was unaffected by mutation of Glu306, Asp307, and Glu379. Mutating Asp99 to an alanine abolished ALG6 function, whereas mutation to an asparagine did not reduce enzymatic activity. This suggests that Asp99 has a functional role, possibly in acceptor substrate binding, but is not the essential catalytic base. Finally, mutating Asp69 to an alanine abolished ALG6 function, and mutating it to an asparagine strongly reduced activity. We conclude that the structurally conserved Asp69 is the catalytic base in ALG6. The corresponding aspartates (Extended Data Fig. 5b) have been demonstrated to be catalytically essential in oligosaccharyltransferases and the arabinosyltransferase ArnT[12,16,17,19] and proposed to be essential in the mannosyltransferase Pmt1-Pmt2[20]. Hence, an aspartate or glutamate at the N-terminal end of the helix EL-h1 may be catalytically essential in all GT-C enzymes. Mutating His378, which is in the immediate vicinity of the glucose moiety of Dol25-P-Glc, to alanine, asparagine or glutamine, strongly reduced the activity of ALG6 but did not fully abolish it. Given the location of this residue, it is most likely involved in the binding and orientation of the donor substrate. Although it was shown that at least one of the ER-luminal mannose transfer reactions of the LLO biosynthesis pathway is metal-dependent[32], our EM map showed no evidence for a

bound divalent metal ion (M^{2+}), nor does ALG6 contain a DXD sequence motif similar to those shown to be essential for M^{2+} binding in oligosaccharyltransferase[12,16,20]. Purified ALG6 indeed retains activity in the presence of EDTA (Extended data Fig. 9a), demonstrating that the reaction is metal ion-independent.

To probe the reaction mechanism further, we tested whether ALG6 could process synthetic Dol25P-Glc analogs where the glucose moiety carried distinct substituents originally designed for labeling studies (Extended data Fig. 9a). We found that substituting the C4 hydroxyl group with a 2-azido-N-ethylacetamido group abolished glucose transfer, which can be rationalized by the structural data as this would lead to a clash with the side chain of His378 (Fig. 4b). Substituting the C6 hydroxyl group of glucose with an azido group did not impair glucosyl transfer, which is in line with the structural data given that this hydroxyl group is facing the solvent and can probably freely rotate (Fig. 4a). ALG6 activity was abolished when a larger substituent (PEG$_4$-linked Rhodamine110) replaced the C6 hydroxyl group (Extended data Fig. 9a), where it probably hinders access of the acceptor substrate to the active site (Fig. 4a, Extended Data Fig. 4b).

Based on our structural and functional data, we propose a three-state mechanism for ALG6 function (Extended Data Fig. 9b). The structure of substrate-bound ALG6 suggests that DolP-Glc binds before the Man$_9$-containing acceptor substrate, because the glucose moiety is at the bottom of the active site cavity. We therefore propose that donor and acceptor substrates bind sequentially and that Asp69 acts as a general base that abstracts the proton of the 3-hydroxyl group of the terminal A-branch mannose of the acceptor substrate in order to activate it for a nucleophilic attack. This may require a conformational change during which the glucose moiety of the donor substrate moves closer to the acceptor substrate and to the catalytic residue Asp69, akin to induced fit. An analogous conformational change was indeed described in another GT-C enzyme, the bacterial oligosaccharyltransferase PglB[33]. In ALG6 an induced fit mechanism would allow the enzyme to prevent futile donor substrate hydrolysis. Upon formation of a ternary complex, the glucose moiety may not only become better ordered, but might deviate from the chair conformation, as was previously observed in other GTs and glycosyl hydrolases (GHs)[34,35].

Conclusions

Our ALG6 structure not only defines the architecture of eukaryotic ALG enzymes involved in ER-luminal N-glycan synthesis, but also suggest an unexpected structural modularity of GT-C enzymes. The conserved GT-C module acts both as a scaffold and as a substrate activator by providing a catalytically essential general base residue, whereas the variable module allows different GT-Cs to catalyze diverse glycosyl transfer reactions and bind distinct substrates. We identified key residues involved in ALG6 catalysis and provide the basis for future studies of the catalytic mechanism of this and related GT-C enzymes. Our approach of combining synthetic chemistry with chemo-enzymatic glycan extension presents a highly efficient and controlled way to generate LLO analogs that are not only essential for studying ALG6-substrate interactions but may allow all enzyme-catalyzed reactions of the ER-based protein *N*-glycosylation pathway to be studied in molecular detail.

Methods

Overexpression and purification of ALG3

A synthetic gene (GenScript) of full-length ALG3-PreScission-EYFP-1D4 from *S. cerevisiae* (UniProtKB – P38179) was optimized for the expression in *S. frugiperda* (GeneArt, Thermofisher). The gene was cloned into a pOET1 vector (Oxford Expression Technologies) and was expressed in *S. frugiperda* (Sf9) cells transfected with baculovirus that was generated with FlashBAC DNA (Oxford Expression Technologies). Cells were cultured in serum free SF-4 Baculo Express ICM medium (AMIMED) at 27 °C. Cells were transfected at a density of $1x10^6$ cells/ml and were harvested after three days. For purification, the cells were resuspended 150 mM NaCl, 50 mM HEPES pH 7.4 with 0.1 mg/ml DNAseI, 1:100 protease inhibitor cocktail (Sigma), 0.1 mg/ml PMSF and were lysed by dounce homogenization before solubilization by addition of, 1% n-dodecyl-β-D-maltopyranosinde (DDM, Antrace) 0.2% cholesteryl hemisuccinate (CHS, Anatrace) and 10% glycerol. After 1h of solubilization cell debris were pelleted by centrifugation at 30,000 rpm in a type T45 Ti rotor (Beckmann). The supernatant was added to anti-1D4 antibody coupled beads and was incubated for 1h. The beads then were washed with 2 x 20 column volumes of washing buffer (150 mM NaCl, 40 mM HEPES pH 7.4, 0.03% DDM, 0.006% CHS).

Then the protein was eluted by incubation of washing buffer, supplemented with 0.5 mg/ml 1D4 peptide for 1h.

Overexpression and purification of ALG6, ALG9 and ALG12

Synthetic genes (GenScript) of Flag-EYFP-PreScission-full length ALG6 (UniProtKB – Q12001), ALG9 (UniProtKB - P53868) or Alg12 (UniProtKB - P53730) from *S. cerevisiae* were optimized for the expression in *H. sapiens* (GeneArt, Thermofisher). Proteins were expressed, and cells were lysed, solubilized and pelleted as described above for ALG3. The supernatant was added to anti-Flag antibody coupled resin and was incubated for 1h. The beads then were washed with 2 column volumes of washing buffer (150 mM NaCl, 40 mM HEPES pH 7.4, 0.03% DDM, 0.006% CHS). Then the protein was eluted by incubation of washing buffer, supplemented with 0.2 mg/ml Flag peptide for 1h. For functional assays ALG6 was eluted with washing buffer supplemented with 3C-protease. For the substrate bound structure, the immobilized ALG6 was solubilized, immobilized to Flag resin and washed as described above. Subsequently the immobilized protein was washed with 10 cv washing buffer supplemented with 0.1% digitonin (Huberlab) and with 2 x 10 cv digitonin washing buffer (150 mM NaCl, 40 mM HEPES pH 7.4, 0.1% digitonin). The sample was eluted with 3C-protease and was subsequently purified by size exclusion chromatography (SEC) in the same buffer.

ALG6 mutant generation

ALG6 mutants were generated by site directed mutagenesis PCR and were expressed and purified by immobilization to Flag-resin and elution with peptide as describe above for wild type ALG6. The proteins were further purified by SEC.

Nanodisc reconstitution of ALG6

A mixture of yeast polar lipids (Yeast Extract Polar, Avanti Polar Lipids) and CHS (80:20, w:w) was diluted to a concentration of 10 mM in buffer (150 mM NaCl, 40 mM HEPES pH 7.4). Lipids were solubilized by addition of 10 mM DDM and sonication for 20 min. Purified EYFP-PreScission-ALG6 in DDM:CHS supplemented buffer, purified MSP1D1[36] protein and solubilized lipids were mixed in a ratio of 1:6:390. After incubation for 10 min at 4 °C, the mixture was incubated for 20 min at room

temperature. Detergent was removed by addition of 0.8 mg activated Biobeads (BIO-RAD) per ml of reconstitution mixture and subsequent incubation at 4 °C overnight. Biobeads were removed and the reconstitution mixture was 10 x diluted in washing buffer (150 mM NaCl, 40 mM HEPES pH 7.4). To remove empty nanodiscs, the reconstitution mixture was incubated with anti-Flag antibody coupled resin for 1h. After washing the resin with 2 x10 column volumes of wash buffer, the sample was eluted by incubation with 3C-protease.

Expression and purification of EYFP

His6-EYFP-1D4, was cloned into a modified pET19b vector a (Novagen) and overexpressed in BL21 gold cells (DE3) (Stratagene, San Diego, CA). Cells were grown in Terrific Broth medium at 37 °C to an OD_{600} of 3.0 before induction with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). For protein expression cells were then grown overnight at 25 °C and harvested by centrifugation at 10'000 x g.

For purification cells were resuspended in PBS (Gibco) supplemented with 1 mM PMSF and 0.025 mg/ml DNAseI. Cells were lysed by three cycles of sonication (50% amplitude and 50% duty cycle). Cell debris were pelleted by ultracentrifugation at 35,000 rpm, in a type T45 Ti rotor, for 30 min. The supernatant was incubated with Ni-NTA resin for 1h and was subsequently washed with 10 cv 150 mM NaCl, 50 mM Tris-HCl pH 7.5 and 10 cv 300 mM NaCl, 40 mM Tris-HCl pH 8.0 and 50 mM imidazole, before elution with 5 cv 300 mM NaCl, 40 mM Tris-HCl pH 8.0 and 400 mM imidazole. Immediately afterwards, the protein buffer was exchanged into 150 mM NaCl, 50 mM Tris-HCl pH 7.5.

Enzymatic biotinylation of MSP1D1

MSP1D1 fused to a C-terminal Avitag was biotinylated in biotinylation buffer (75 mM NaCl, 50 mM bicine pH 8.3, 10 mM Mg acetate, 10 mM ATP, 0.25 mM biotin) with 2 µM BirA protein, overnight at 4 °C. 2.5 µM. 3C protease was added for 6h at 4 °C. 3C-Protease as well as BirA were removed by reverse binding to Ni-NTA. Subsequently, the biotinylated MSP1D1 protein was desalted into 150 mM NaCl, 50 mM HEPES pH 7.4 and subsequently used for nanodisc reconstitution as described above.

Synthetic antibody generation

For phage display experiments enzymatically biotinylated MSP1D1 was used for the nanodisc reconstitution and no reverse binding or cleavage after nanodisc reconstitution was performed. Aggregates were removed by SEC after the nanodisc reconstitution.

To validate the immobilization efficiency of the final sample, pull-down assays of the target proteins immobilized on Streptavidin MagneSphere® Paramagnetic Particles (Promega) were performed, followed by biopanning using Fab Library E[13,37]; DNA was kindly provided by Shohei Koide and phage library was prepared by Somnath Mukherjee. Five subsequent rounds of selection rounds were performed in 25 mM HEPES, pH 7.4 and 150 mM NaCl (Selection Buffer) supplemented with 1% bovine serum albumin (BSA)[38,39].

The phage display selection was performed according to published protocols[38,40]. In brief, in the first round, 400 nM Flag-EYFP-Prescission-ALG6 in nanodiscs was immobilized onto paramagnetic particles for biopanning, followed by three washes and resuspension of the particles directly in log-phase *E. coli* XL-1 Blue cells for infection. Phage were amplified in 2xYT media supplemented with ampicillin (100 µg/mL) and M13-KO7 helper phage (10^9 pfu/mL). From the second to fifth round sorting was performed using a Kingfisher magnetic beads handler (ThermoFisher) with decreasing concentration of target protein: 200nM for 2nd round, 100 nM for 3rd round, 60 nM for 4th round and 30 nM for the last round. Phage elution was done by disrupting of immobilized nanodiscs with 1% Fos-choline 12 in Selection Buffer. Also, for rounds 2-5, each phage pool was negatively selected against streptavidin-coated particles and then used as an input for the next round. Throughout the entire selection process soluble competitors were included in excess: 2 µM non-biotinytaled empty nanodiscs (same reconstitution procedure as for ALG6, without the addition of ALG6 and using non-biotinylated MSP1D1) and 2 µM purified EYFP protein. Phage pools from the 4th and 5th round were used for binder screening.

Synthetic antibody screening

Single-point phage ELISA was performed to screen and validate the binding affinities of selected binders[38]. Briefly, *E. coli* XL-1 Blue were transfected with selected phage pools and plated on LB/agar plates with Ampicillin. Colonies were used to inoculate

2xYT media with 100 μg/mL ampicillin and 10^9 pfu/ml M13-KO7 helper phage. Overnight amplified cultures were pelleted by centrifugation and 10-fold diluted supernatants were used for ELISA. For this purpose, 96-well plates coated with 2 μg/ml Neutravidin were blocked with Selection Buffer with 1% BSA and Flag-EYFP-Prescission-ALG6 protein reconstituted in biotinylated nanodiscs (50nM) was immobilized in each well. Subsequent ELISA assays were performed in Selection Buffer with 2% BSA. Diluted phage particles were assayed against the target protein by using an HRP-conjugated anti-M13 monoclonal antibody (GE Healthcare). Detection was done with TMB substrate (ThermoFisher) quenched with 1.0M HCl after 3 minutes and subsequent absorbance measurement at 450 nm. Biotinylated empty nanodiscs were used as a control and only target-specific binders were selected.

Selected binders were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing facility and cloned into a pRH2.2 vector (kind gift of Sachdev Sidhu) using the In-Fusion Cloning kit (Takara). The antigen-binding fragments were expressed in *E. coli* BL21-Gold cells and purified by Protein A chromatography and ion exchange chromatography as described previously[41].

Subsequently multi-point protein ELISA was performed to estimate the binding affinity. ELISA plate preparation and target immobilization was done similar to single-point ELISA, followed by purified Fabs in 3-fold dilutions (starting from 3 μM) assayed against target protein by HRP-conjugated mouse anti-human IgG F(ab')2 monoclonal antibody (Jackson). Detection was done as described above. Data points were plotted assuming sigmoidal dose response and EC50 values were calculated for affinity estimation.

Fab binding characterization using size exclusion chromatography

For SEC binding analysis 3C-protease cleaved ALG6 in DDM:CHS supplemented buffer was incubated with 4-fold molar excess of Fab for 30 min on ice and subsequently analyzed by SEC. Peak fractions were subsequently analyzed by SDS-PAGE.

Thermostability experiments were performed as described previously[42], with 3C-protease cleaved ALG6 in DDM:CHS supplemented buffer pre-incubated with three fold molar excess of Fab for 30 min on ice and by measuring A280 instead of fluorescence during SEC.

EM sample preparation

For the apo structure, ALG6 in MSP1D1 lipid nanodiscs was incubated with Fab 6AG9 in 1:2 molar ratio for 1h at 4 °C and subsequently purified by SEC. Pooled peak fractions were concentrated to 0.33 mg/ml. Quantifoil holey carbon grids, Cu, R 1.2/1.3, 300 mesh were glow discharged for 45s, 25 mA using a PELCO easiGLOW glow discharger. 4 µl of sample were applied to the cryo-EM grids and blotted for 1-2 s before plunge freezing in a liquid ethane-propane mixture with a Vitrobot Mark IV (Thermo Fischer Scientific) operated at 4 °C and 100% humidity.

For the substrate bound structure ALG6 in digitonin supplemented buffer at 7.2 mg/ml was incubated with 1.5-fold excess of Fab 6AG9 for 1h, at 4 °C. Subsequently, the sample was purified by SEC and peak fractions were pooled and were concentrated to 6.3 mg/ml. The sample was supplemented with a final concentration of 150 µM Glc-P-dol25 and incubated for 1.5 h before grid preparation. Grids were prepared the same way as for the nanodisc sample, but with 3.5 s blotting time.

EM data collection

Data was recorded on a Titan Krios electron microscope (Thermo Fischer Scientific) operated at 300 kV, equipped with a Gatan BioQuantum 1967 filter with a slit width of 20 eV and a Gatan K3™ camera. Movies were collected semi-automatically using SerialEM[43] at a nominal magnification of 105,000 and a pixel size of 0.42 Å per pixel in super resolution mode. The defocus range was -0.6 to -2.8 µm. Each movie contained 40 images per stack with a dose per frame of 2.3 electrons/Å^2.

EM data processing, model building and refinement

For the apo structure of ALG6 in nanodiscs, 6,801 movies were corrected for beam induced motion using MotionCor2[44]. The micrographs were visually inspected and 5,143 were used for further processing in RELION 3.0[45]. Contrast transfer function (CTF) was estimated using gCTF[46]. 3,437,475 particles were auto-picked and extracted with 3-fold binning to a pixel size of 2.52 Å/pixel. After two subsequent rounds of 2D classification 445,702 particles were used to generate an initial model. The particles were subjected to two subsequent rounds of 3D classification and the remaining 171,764 particles were subjected to 3D refinement. Subsequently, the particles were re-extracted at 0.84 Å/pixel and subjected to 3D refinement yielding a

map at 3.2 Å resolution. Subsequent particle polishing and 3D refinement lead to a map at 3.1 Å resolution. Applying of a tight mask that masked out most of the flexible constant domain yielded a map at 3.0 Å resolution.

For the substrate-bound structure in digitonin, 6,187 movies corrected for beam induced motion using MotionCor2[44]. The micrographs were visually inspected and 4,000 were used for further processing in RELION 3.0[45]. Contrast transfer function (CTF) was estimated using gCTF[46]. 995,878 particles were auto-picked and extracted with 3-fold binning to a pixel size of 2.52 Å/pixel. After two subsequent rounds of 2D classification 287,977 particles were selected and were used to calculate an initial model. After one round of 3D classification 115,590 particles were selected for one round of 3D refinement and subsequent re-extraction at 0.84 Å/pixel. Subsequent 3D refinement yielded a final map at 3.9 Å.

Model building was performed in Coot[47]. Most of the Fab, except for the binding interface, was built based on a preexisting model at high resolution (PDB ID 5UCB). The structures were refined in PHENIX[48] and ligand files were generated in PHENIX eLBOW[49]. The correlation between model and map was validated using PHENIX mtriage[50].

Figure preparation and data analysis

Local resolution estimations were calculated in ResMap.[51] Figures were prepared in PyMol[52] and UCSF Chimera[53] and UCSF ChimeraX[54]. Graphs were generated in GraphPad Prism 8. Protein sequence alignments were performed using CLUSTAL W[55], Clustal Omega[56] and EMBOSS Needle[57]. Figure 3d was generated using the following structures: *Saccharomyces cerevisiae* ALG6 (this study), *Campylobacter lari* PglB (PDB 3RCE), *Cupriavidus metallidurans* ArnT (PDB 5F15), *S. cerevisiae* PMT1 (PDB 6P25).

Chemical synthesis of substrate analogs

β-mannosyl (3) and β-glucosyl (4) triesters were synthesized starting from 2,3,4,6-tetracetylated glycosides with and optimized procedure based on a previously described method[58] Diphenyl protected glycosyl phosphates were readily hydrogenated and coupled to farnesylcitronellol in presence of trichloroacetonitrile. Acetyl deprotection gave β-glycosyl phosphoisoprenoids Dol25-P-Man (5) and Dol25-

P-Glc (6). To generate potentially clickable analogues of 6, an azide was introduced in 6 position following known procedure[59], and in 4 position introducing a short linker via O-alkylation of orthogonal protected glucopyranoside[60] with bromoacetonitrile followed by reduction of the nitrile with BH3•SMe2 (ref.[61]) and acylation of the primary amine. β-phosphorylation of the azide bearing lactols was performed using diallylphosphoryl chloride[62] to avoid hydrogenation steps. The desired configuration was confirmed by ROESY (Supplementary Information methods). Diallyl phosphates were deprotected, converted to the phosphoric acid and coupled with farnesylcitronellol[11], as described above. More detailed procedures are described in Extended Data methods.

Synthesis and purification of Dol25-PP-GlcNAc$_2$Man$_9$

Dol25-PP-GlcNAc$_2$Man$_5$ was produced as previously described[10]. 20 µM Dol25-PP-GlcNAc$_2$Man$_5$ and 120 µM Dol25-P-Man were mixed in 150 mM NaCl, 50 mM HEPES pH 7.4, 0.03% DDM, 0.006% CHS. 1.5 µM purified ALG3 was added and the mixture was incubated overnight at 4 °C. The reaction was inactivated by heating for 10 min at 95 °C. After cooling down to 4 °C, the reaction was supplemented with 10 mM MgCl$_2$ and 1.5 µM ALG9 and 1.5 µM ALG12 were added and incubated overnight at 4 °C. For purification of Dol25-PP-GlcNAc$_2$Man$_9$, the reaction was lyophilized followed by a first extraction with CHCl$_3$:MeOH (2:1, v/v) to remove contaminants of smaller LLO species. The remaining pellet was then subjected to extraction with CHCl$_3$:MeOH:H$_2$O (10:10:3, v/v/v). Subsequently the extracted substrate was dried under N$_2$ and was resuspended in 150 mM NaCl, 40 mM HEPES pH 7.4. The yield was quantified as described previously[10].

Note that a similar synthetic route for *in vitro* synthesis of an LLO analog was previously reported[63]. However, the lipid carrier in that study was a phytanol moiety instead of dolichol.

In vitro glycosyl transfer assays

Glycosyl transfer reactions for ALG1, ALG2 and ALG11 were performed as previously described[10]. Glycosyl transfer reactions for ALG3, ALG9, ALG12 and were performed in reaction buffer (150 mM NaCl 50 mM HEPES pH 7.4, 0.03% DDM, 0.006% CHS) supplemented with 10-20 µM acceptor substrate (Starting from purified Dol25-PP-

GlcNAc$_2$Man$_5$ as acceptor substrate with sequential addition and heat inactivation of ALG enzymes) and 50-120 µM donor substrate (Dol25-P-Man). Reactions for ALG9 and ALG12 were supplemented with 10 mM MgCl$_2$. Protein concentrations were ranging from 0.5 - 2 µM. Reactions were carried out overnight at 4 °C. Reactions were stopped by heating at 98 °C for 10 min. This ensured that during the ALG mediated glycosyl transfer reactions, no other active ALG enzymes from previous reactions were present and eliminated fluorescence from EYFP fusion proteins, which otherwise would affect subsequent tricine gel analysis.

Reactions for ALG6 were carried out in reaction buffer (150 mM NaCl 50 mM HEPES pH 7.4, 0.03% DDM, 0.006% CHS) supplemented with 10 µM purified Dol25-PP-GlcNAc$_2$Man$_9$ acceptor substrate, 50 µM of the respective donor substrate and 0.2 µM tag-cleaved ALG6. For reactions in presence of EDTA or Fab, the entire reaction mixture was preincubated for 1h on ice, before adding donor and acceptor substrates. The reactions were incubated for 6 h at 10 °C. Subsequently the reactions were stopped by heating at 98 °C for 10 min. Reactions of ALG6 mutants were carried out as described above, at 10 °C for 24 h. All ALG reaction products were analyzed by TbSTT3A mediated oligosaccharyl-transfer and subsequent tricine gel analysis as previously described[10]. Due to lower substrate specificity of TbSTT3A for Glc1Man9GlcNAc2-PP-dol25, oligosaccharyl-transfer reactions for ALG6 reactions were performed for 12 h at 30 °C. For the oligosaccharyl-transfer reactions for ALG6 reactions that were carried out in presence of EDTA 30 µM of MnCl$_2$ was added, to avoid inhibition of TbSTT3A.

Statistics and Reproducibility

Unless otherwise stated, ALG6 assays and tricine gel-based analyses were conducted once as shown in the figures.

and purification, J. Kowal and I. Manolaridis for help with EM data collection, and A. Ramirez and A. Alam for helpful discussions.

References

1   Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54**, 631-664, (1985).
2   Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1426**, 239-257, (1999).
3   Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci* **12**, 1418-1431, (2003).
4   Moremen, K. W. & Haltiwanger, R. S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat Chem Biol* **15**, 853-864, (2019).
5   Runge, K. W., Huffaker, T. C. & Robbins, P. W. 2 Yeast Mutations in Glucosylation Steps of the Asparagine Glycosylation Pathway. *J Biol Chem* **259**, 412-417, (1984).
6   Reiss, G., te Heesen, S., Zimmerman, J., Robbins, P. W. & Aebi, M. Isolation of the ALG6 locus of Saccharomyces cerevisiae required for glucosylation in the N-linked glycosylation pathway. *Glycobiology* **6**, 493-498, (1996).
7   Haeuptle, M. A. & Hennet, T. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum Mutat* **30**, 1628-1641, (2009).
8   Jaeken, J., Lefeber, D. & Matthijs, G. Clinical utility gene card for: ALG6 defective congenital disorder of glycosylation. *Eur J Hum Genet* **23**, (2015).
9   Morava, E. *et al.* ALG6-CDG: a recognizable phenotype with epilepsy, proximal muscle weakness, ataxia and behavioral and limb anomalies. *J Inherit Metab Dis* **39**, 713-723, (2016).
10   Ramirez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology*, 1-8, (2017).
11   Ramirez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27**, 525-535, (2017).
12   Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **359**, 545-550, (2018).
13   Fellouse, F. A. *et al.* High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol* **373**, 924-940, (2007).
14   Gouw, M. *et al.* The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res* **46**, D428-D434, (2018).
15   Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495, (2014).
16   Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355, (2011).
17   Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17868-17873, (2013).

18 Bai, L., Wang, T., Zhao, G., Kovach, A. & Li, H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555**, 328-333, (2018).

19 Petrou, V. I. *et al.* Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351**, 608-612, (2016).

20 Bai, L., Kovach, A., You, Q., Kenny, A. & Li, H. Structure of the eukaryotic protein O-mannosyltransferase Pmt1-Pmt2 complex. *Nat Struct Mol Biol*, (2019).

21 Albuquerque-Wendt, A., Hutte, H. J., Buettner, F. F. R., Routier, F. H. & Bakker, H. Membrane Topological Model of Glycosyltransferases of the GT-C Superfamily. *Int J Mol Sci* **20**, (2019).

22 Imbach, T. *et al.* A mutation in the human ortholog of the Saccharomyces cerevisiae ALG6 gene causes carbohydrate-deficient glycoprotein syndrome type-Ic. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6982-6987, (1999).

23 Imbach, T. *et al.* Multi-allelic origin of congenital disorder of glycosylation (CDG)-Ic. *Hum Genet* **106**, 538-545, (2000).

24 Westphal, V., Schottstadt, C., Marquardt, T. & Freeze, H. H. Analysis of multiple mutations in the hALG6 gene in a patient with congenital disorder of glycosylation Ic. *Mol Genet Metab* **70**, 219-223, (2000).

25 Dercksen, M. *et al.* ALG6-CDG in South Africa: Genotype-Phenotype Description of Five Novel Patients. *JIMD Rep* **8**, 17-23, (2013).

26 Runge, K. W. & Robbins, P. W. A New Yeast Mutation in the Glucosylation Steps of the Asparagine-Linked Glycosylation Pathway - Formation of a Novel Asparagine-Linked Oligosaccharide Containing 2 Glucose Residues. *J Biol Chem* **261**, 5582-5590, (1986).

27 Lee, B. C. *et al.* Gating mechanism of the extracellular entry to the lipid pathway in a TMEM16 scramblase. *Nat Commun* **9**, 3251, (2018).

28 Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555, (2008).

29 Albesa-Jove, D., Cifuente, J. O., Trastoy, B. & Guerin, M. E. Quick-soaking of crystals reveals unprecedented insights into the catalytic mechanism of glycosyltransferases. *Methods in enzymology* **621**, 261-279, (2019).

30 Notenboom, V. *et al.* Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants. *Nat Struct Biol* **5**, 812-818, (1998).

31 Chang, A., Singh, S., Phillips, G. N., Jr. & Thorson, J. S. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Current opinion in biotechnology* **22**, 800-808, (2011).

32 Sharma, C. B., Knauer, R. & Lehle, L. Biosynthesis of lipid-linked oligosaccharides in yeast: the ALG3 gene encodes the Dol-P-Man:Man5GlcNAc2-PP-Dol mannosyltransferase. *Biol Chem* **382**, 321-328, (2001).

33 Napiorkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat Struct Mol Biol* **24**, 1100-1106, (2017).

34 Davies, G. J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Acc Chem Res* **45**, 308-316, (2012).

35 Ardevol, A. & Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab

Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J Am Chem Soc* **137**, 7528-7547, (2015).

36   Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc* **126**, 3477-3487, (2004).

37   Fellouse, F. A., Wiesmann, C. & Sidhu, S. S. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12467-12472, (2004).

38   Dominik, P. K. & Kossiakoff, A. A. Phage display selections for affinity reagents to membrane proteins in nanodiscs. *Methods in enzymology* **557**, 219-245, (2015).

39   Dominik, P. K. *et al.* Conformational Chaperones for Structural Studies of Membrane Proteins Using Antibody Phage Display with Nanodiscs. *Structure* **24**, 300-309, (2016).

40   Hornsby, M. *et al.* A High Through-put Platform for Recombinant Antibodies to Folded Proteins. *Mol Cell Proteomics* **14**, 2833-2847, (2015).

41   Borowska, M. T., Dominik, P. K., Anghel, S. A., Kossiakoff, A. A. & Keenan, R. J. A YidC-like Protein in the Archaeal Plasma Membrane. *Structure* **23**, 1715-1724, (2015).

42   Hattori, M., Hibbs, R. E. & Gouaux, E. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure* **20**, 1293-1299, (2012).

43   Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* **152**, 36-51, (2005).

44   Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, (2017).

45   Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530, (2012).

46   Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, (2016).

47   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132, (2004).

48   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **66**, 213-221, (2010).

49   Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D* **65**, 1074-1080, (2009).

50   Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol* **74**, 814-840, (2018).

51   Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65, (2014).

52   The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

53   Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, (2004).

54   Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, (2018).

55   Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680, (1994).

56   Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145, (2018).

57   Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443-453, (1970).

58   Sabesan, S. & Neira, S. Synthesis of Glycosyl Phosphates and Azides. *Carbohyd Res* **223**, 169-185, (1992).

59   Maunier, V., Boullanger, P., Lafont, D. & Chevalier, Y. Synthesis and surface-active properties of amphiphilic 6-aminocarbonyl derivatives of D-glucose. *Carbohyd Res* **299**, 49-57, (1997).

60   Williams, R. J. *et al.* Combined inhibitor free-energy landscape and structural analysis reports on the mannosidase conformational coordinate. *Angew Chem Int Ed Engl* **53**, 1087-1091, (2014).

61   Malet, C. & Hindsgaul, O. Versatile Functionalization of Carbohydrate Hydroxyl Groups through Their O-Cyanomethyl Ethers. *J Org Chem* **61**, 4649-4654, (1996).

62   Li, T., Tikad, A., Pan, W. & Vincent, S. P. beta-Stereoselective phosphorylations applied to the synthesis of ADP- and polyprenyl-beta-mannopyranosides. *Org Lett* **16**, 5628-5631, (2014).

63   Li, S. T. *et al.* Reconstitution of the lipid-linked oligosaccharide pathway for assembly of high-mannose N-glycans. *Nat Commun* **10**, 1813, (2019).
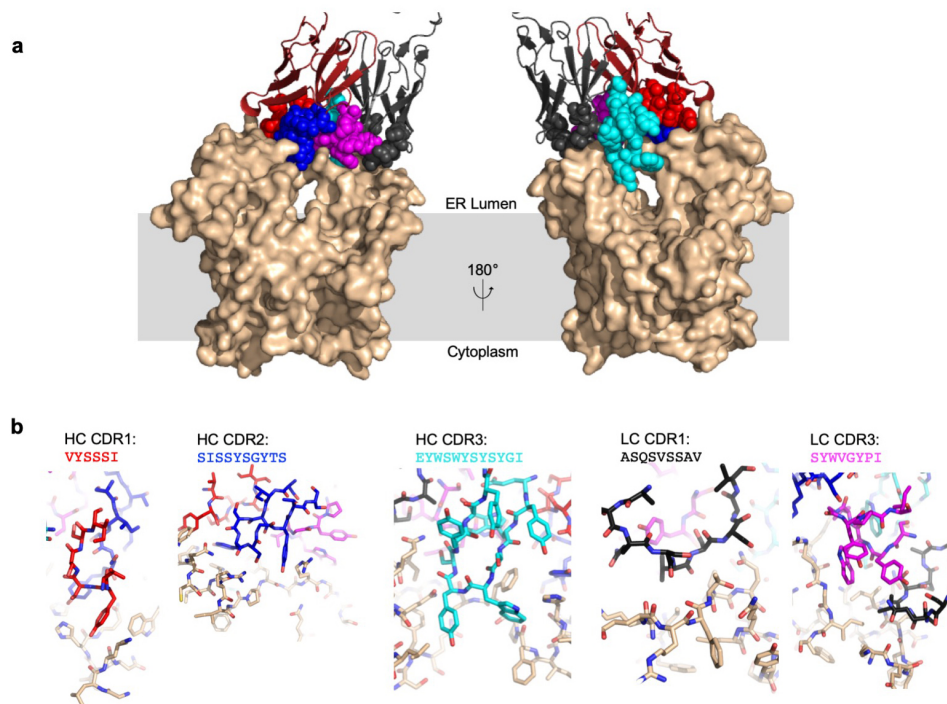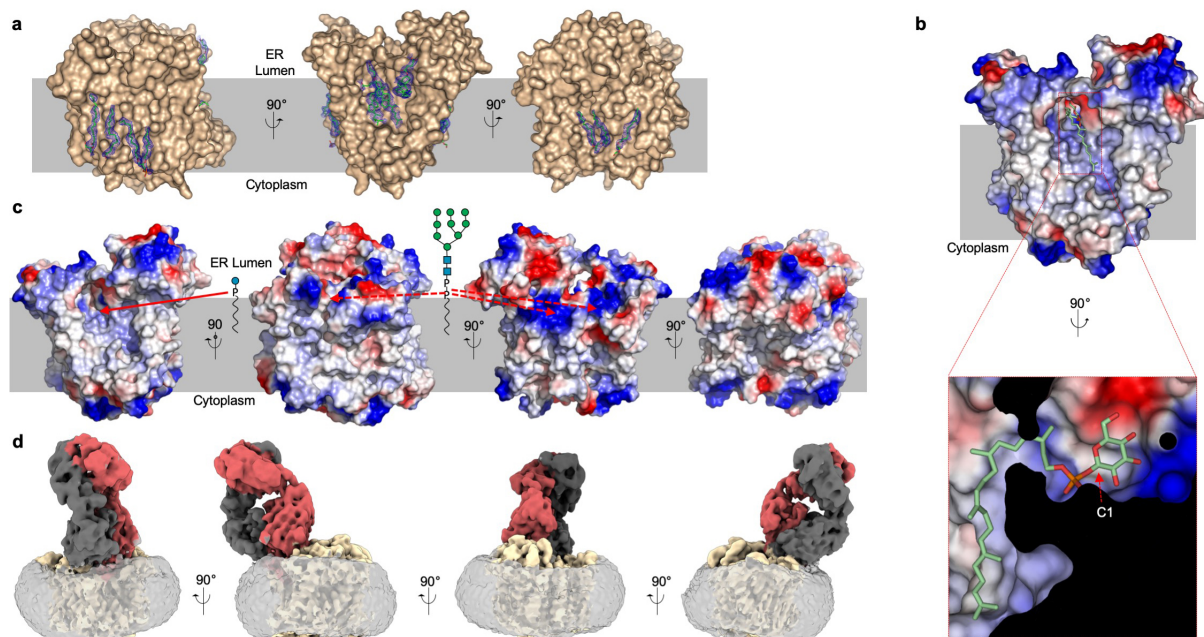
Extended Data
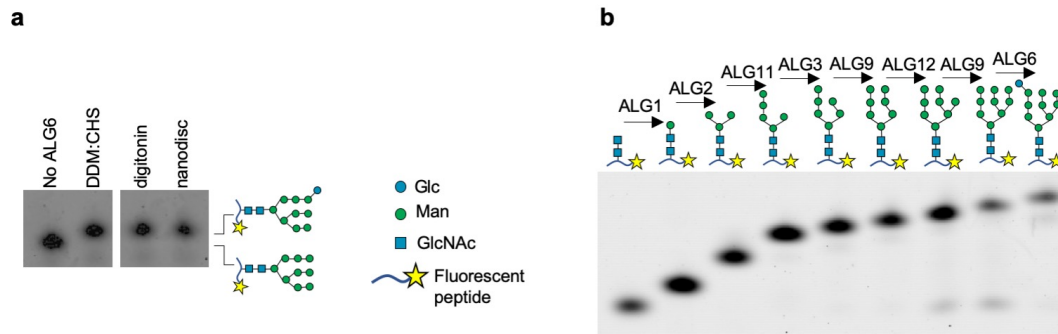
Extended Data Figure 1 ALG6 activity in distinct lipidic environments and chemo-enzymatic synthesis of LLO intermediates. a, Tricine gel-based analysis of ALG6-catalyzed reaction using Dol25-PP-GlcNAc2Man9 and Dol25-P-Glc substrates. The reactions were carried out in distinct detergents or with ALG6 reconstituted in lipid nanodiscs, as indicated above the gel lanes. The glycans were analyzed as described in Fig. 1b. **b**, Tricine gel-based analysis of ALG reaction cascade intermediates with reactions analyzed as described in Fig. 1b. Structures of glycopeptides are schematically shown above the lanes and respective glycosyltransferase enzymes indicated above the arrows.
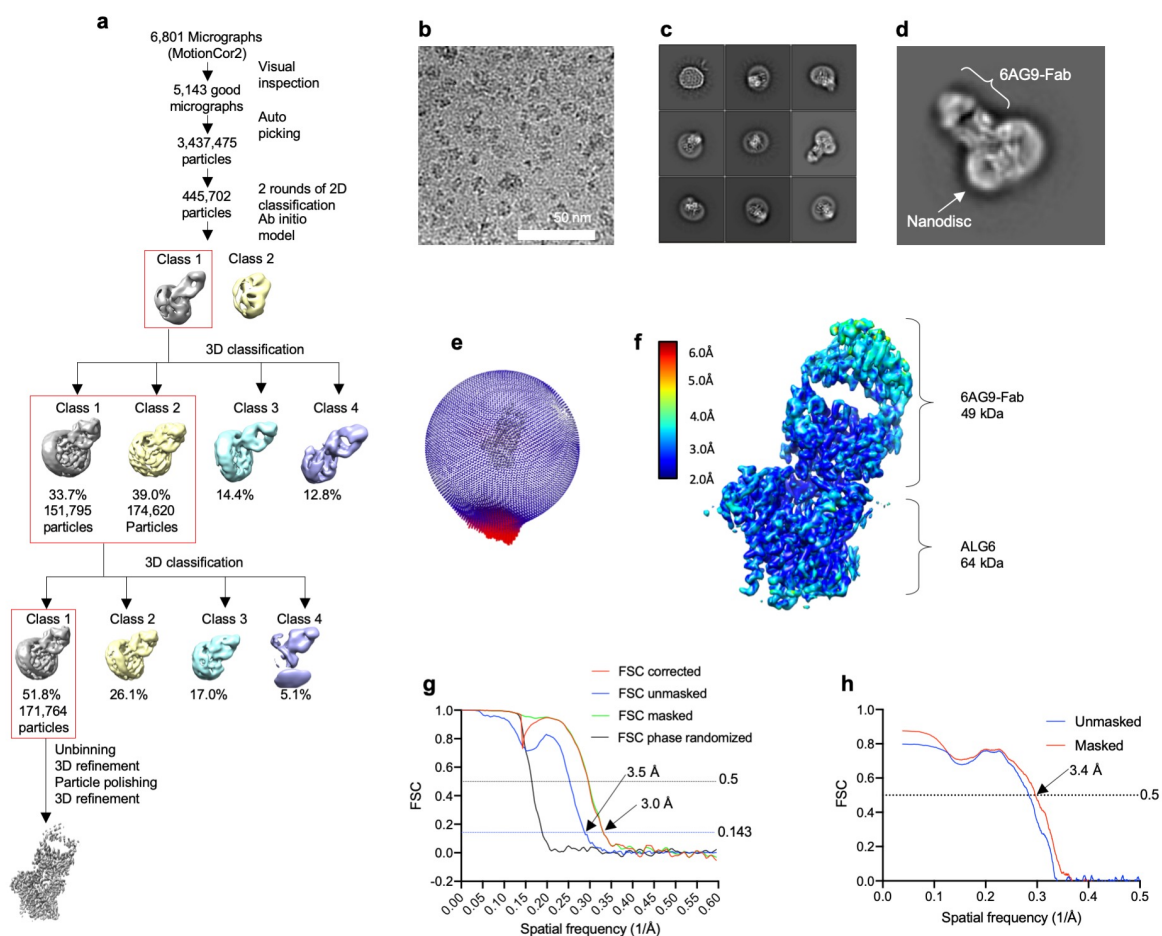
Extended Data Figure 2 Data processing and structure determination of substrate-free ALG6-6AG9-Fab complex in lipid nanodiscs. a, Overview of the EM data processing and structure determination pipeline using RELION 3.0[46]. **b**, Representative cryo-EM micrograph. **c**, Representative 2D classes. **d**, Representative 2D class with 6AG9-Fab and nanodisc belt highlighted. **e**, Spatial distribution of particles in the final iteration of 3D refinement. **f**, Refined and B-factor sharpened EM map, colored by local resolution estimation as calculated in ResMap[52]. **g**, Resolution estimation of final map via Fourier shell correlation, as calculated in RELION 3.0. **h**, Fourier shell correlation (FSC) between model and map, as calculated in PHENIX mtriage[51].

Extended Data Figure 3 Molecular interactions at the interface of ALG6 and 6AG9-Fab. a, Surface representation of ALG6 in light brown and cartoon representation of the variable fragment of 6AG9-Fab colored red (heavy chain) and black (light chain). The 5 CRD-loops of 6AG9-Fab interacting with ALG6 are shown in sphere representation and make contact mostly with the ER-luminal "arch" formed by EL4 of ALG6. **b**, Detailed view of the individual CDR:ALG6 interactions, with 6AG9-Fab CDRs and ALG6 in stick representation and colored as in a. The CDR-sequences are indicated above the panels. Note that CDR1 of the light chain (LC) was not randomized.

**Extended Data Figure 4 a,** Distinct lipid binding sites were identified in the structure of nanodisc-reconstituted ALG6, shown in light brown surface representation. Ordered phospholipids and cholesteryl-hemisuccinate molecules are shown in stick representation (green), and the EM density map was contoured at 5 rmsd and carved to 1.6 Å. **b**, Electrostatic surface potential of ALG6, with negative charges colored in red, neutral charges in white and positive charge in blue and bound Dol25-P-Glc depicted in stick representation with carbons in green. The inset shows a zoomed-in view of the kink induced in the dolichol moiety. **c**, Electrostatic surface potential of ALG6 shown from four angles to indicate potential binding sites for lipid-linked substrates of ALG6 based on the presence of suitable surface cavities. A solid red arrow indicates the binding site of the observed donor substrate Dol25-P-Glc. Dashed red arrows indicate potential binding sites of the acceptor substrate Dol-PP-GlcNAc$_2$Man$_9$. These sites were chosen based on groove-like features in the TM region that might bind dolichol and positively charged patches at the ER-luminal membrane boundary that could help bind the pyrophosphate moiety of the acceptor LLO. Binding to these sites would allow the terminal mannose of the A-branch of the acceptor substrate to reach the active. **d**, Membrane deformation by ALG6 shown from distinct views. The EM map of the apo-ALG6 complex with 6AG9-Fab complex in a lipid nanodisc is colored yellow for ALG6, red (heavy chain) and black (light chain) for 6AG9-Fab, and shown transparent grey for density indicating the lipid nanodisc. Note that the lipid bilayer of the nanodisc surrounding the ALG6 protein is twisted and thinner in the region opposite the arch formed by EL4.

94
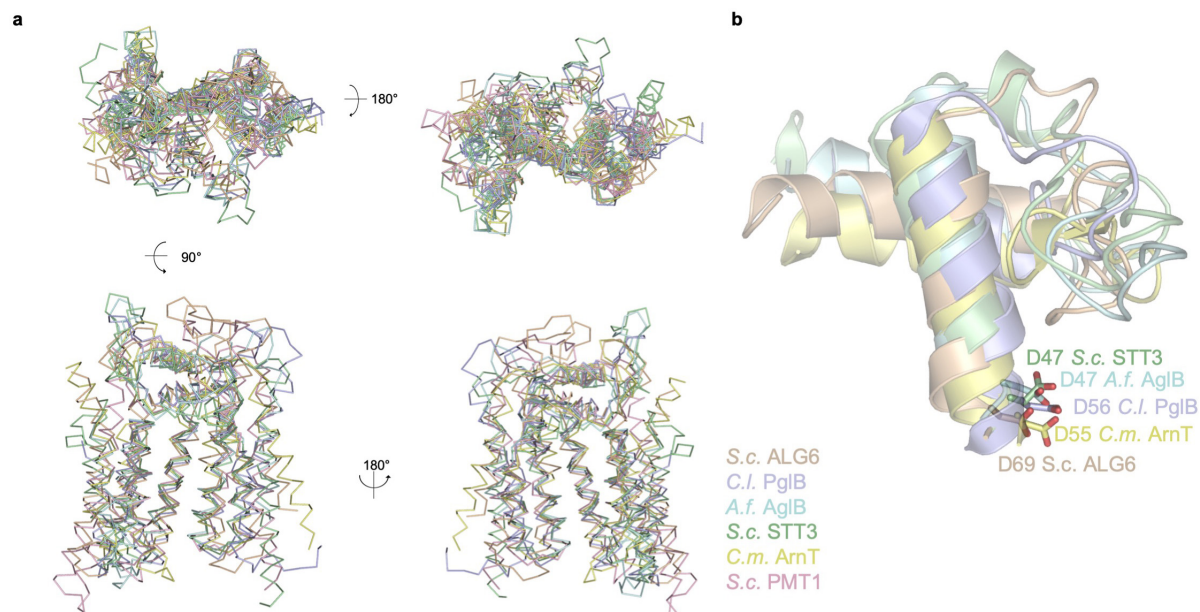
Extended Data Figure 5 Structural similarity of conserved module of GT-C enzymes. a, Cα traces of the first 7 TM helices and the first external loop EL1 of selected structures are shown following superposition. These following GT-C family members of known structure were used, with GT family number indicated: S. cerevisiae PMT1 (GT39, PDB-ID 6P25), S. cerevisiae ALG6 (GT57, this study), C. lari PglB (GT66, PDB-ID 3RCE), A. fulgidus AglB (GT66, PDB-ID 3WAJ), S. cerevisiae STT3 (GT66, PDB-ID 6EZN), C. metallidurans ArnT (GT83, PDB 5F15). **b**, Close-up view of EL-h1, EL-h2 and connecting loop of GT-Cs where the catalytic role of the conserved aspartate, shown in stick representation, was experimentally demonstrated.
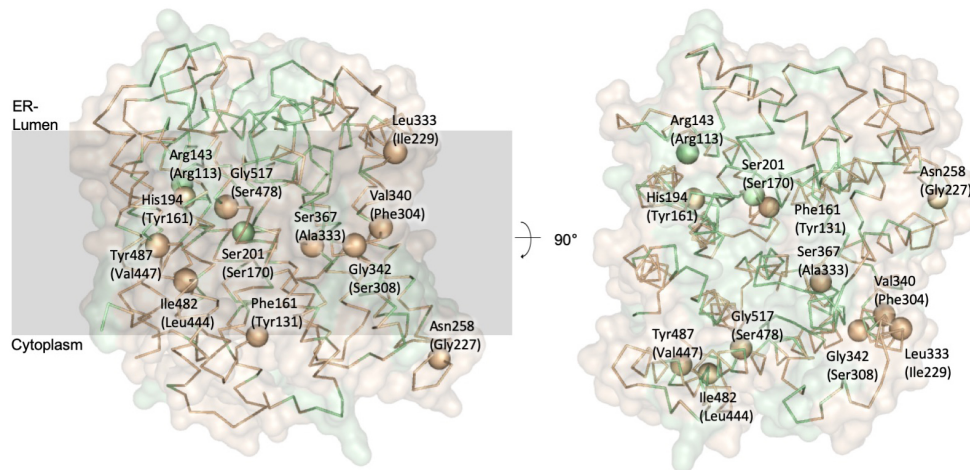
Extended Data Figure 6 Residues involved in congenital disorders of glycosylation (CDGs) caused by *ALG6* mutations. The structure of yeast ALG6 is depicted as a Cα trace and as a transparent surface representation colored in light brown. The surface is colored green for residues conserved between yeast and human ALG6. Cα atoms of residues that have been shown to cause congenital disorders of glycosylation (CDG) in the human homolog of ALG6 are shown in sphere representation. Residue numbers in brackets indicate the corresponding residues of human ALG6 according to pairwise protein sequence alignment with EMBOSS needle[58].
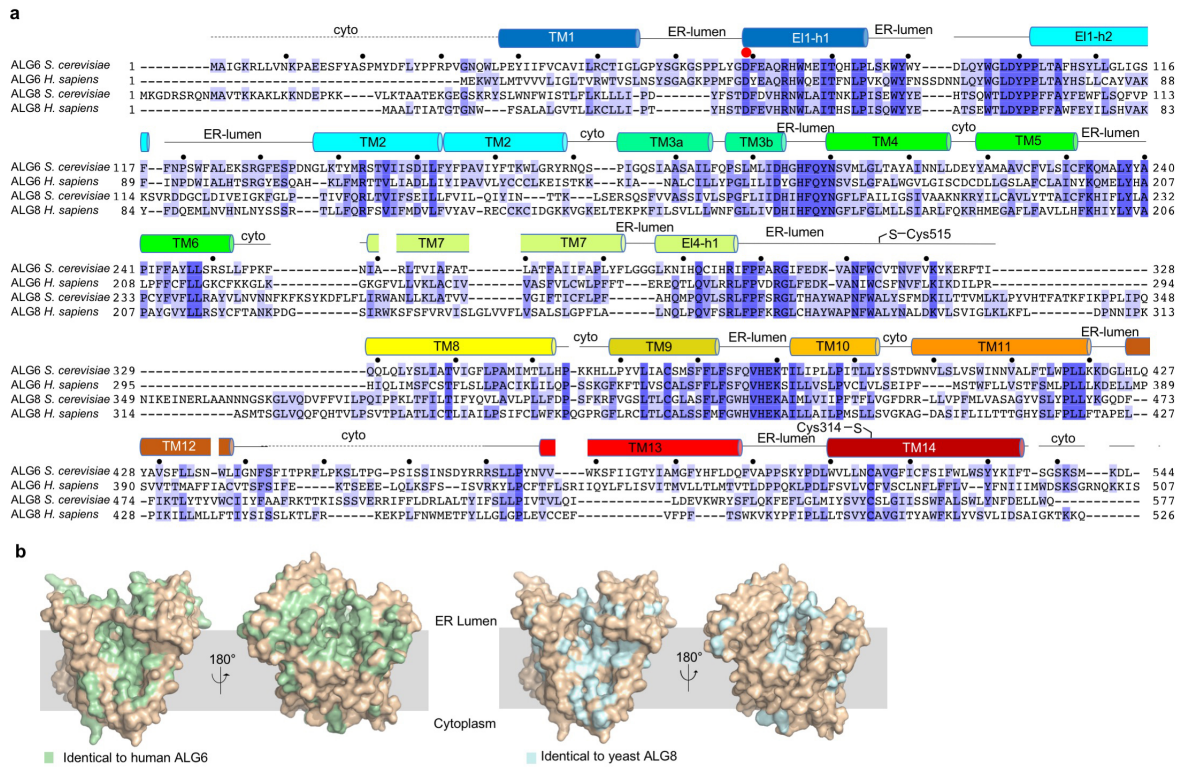
Extended Data Figure 7 Sequence alignment of ALG6 homologs and orthologs. **a**, Alignment of amino acid sequences of yeast and human ALG6 and ALG8 proteins, generated with Clustal Omega[57] (Uniprot identifiers: Q12001, Q9Y672, P40351, Q9BVK2). Secondary structure elements of yeast ALG6 are depicted and labeled above the sequence. Dashed lines indicate regions that are disordered in the ALG6 structures. Cytosolic regions are labelled "cyto", ER-luminal regions are labelled "ER-lumen". A red dot indicates the catalytic base Asp69, which is a structurally conserved aspartate residue that has been observed or proposed to be the catalytic base in all GT-C enzymes of known structures. **b**, Surface representation of yeast ALG6 (light brown) with residues identical to human ALG6 highlighted in green (left) and residues identical to yeast ALG8 highlighted in blue (right).
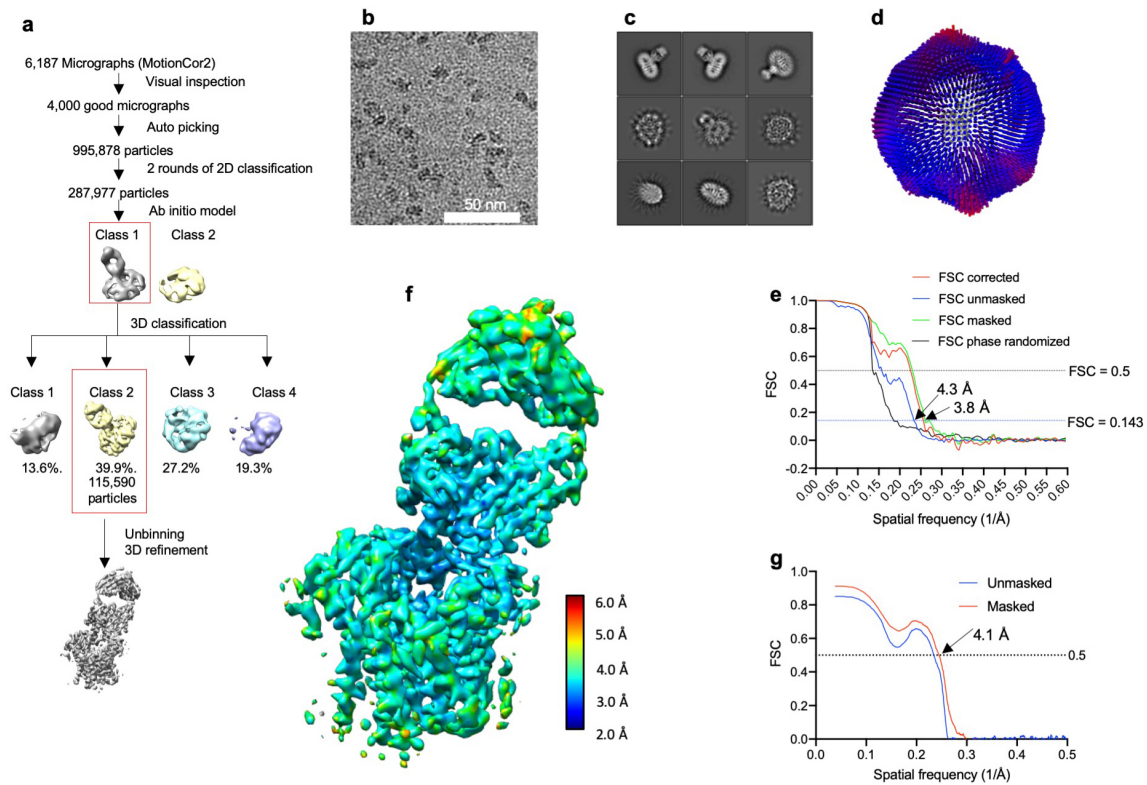
Extended Data Figure 8 Data processing and structure determination of substrate-bound ALG6-6AG9-Fab complex in detergent solution. a, Overview of the EM data processing and structure determination pipeline using RELION 3.0[46]. **b**, Representative cryo-EM micrograph. **c**, Representative 2D classes. **d**, Spatial distribution of particles in the final iteration of 3D refinement. **e**, Resolution estimation of final map via Fourier shell correlation, as calculated in RELION 3.0. **f**, Final refined and B-factor sharpened map, colored by local resolution estimation in as calculated in ResMap[52]. **g**, Fourier shell correlation (FSC) between model and map, as calculated in PHENIX mtriage[51].
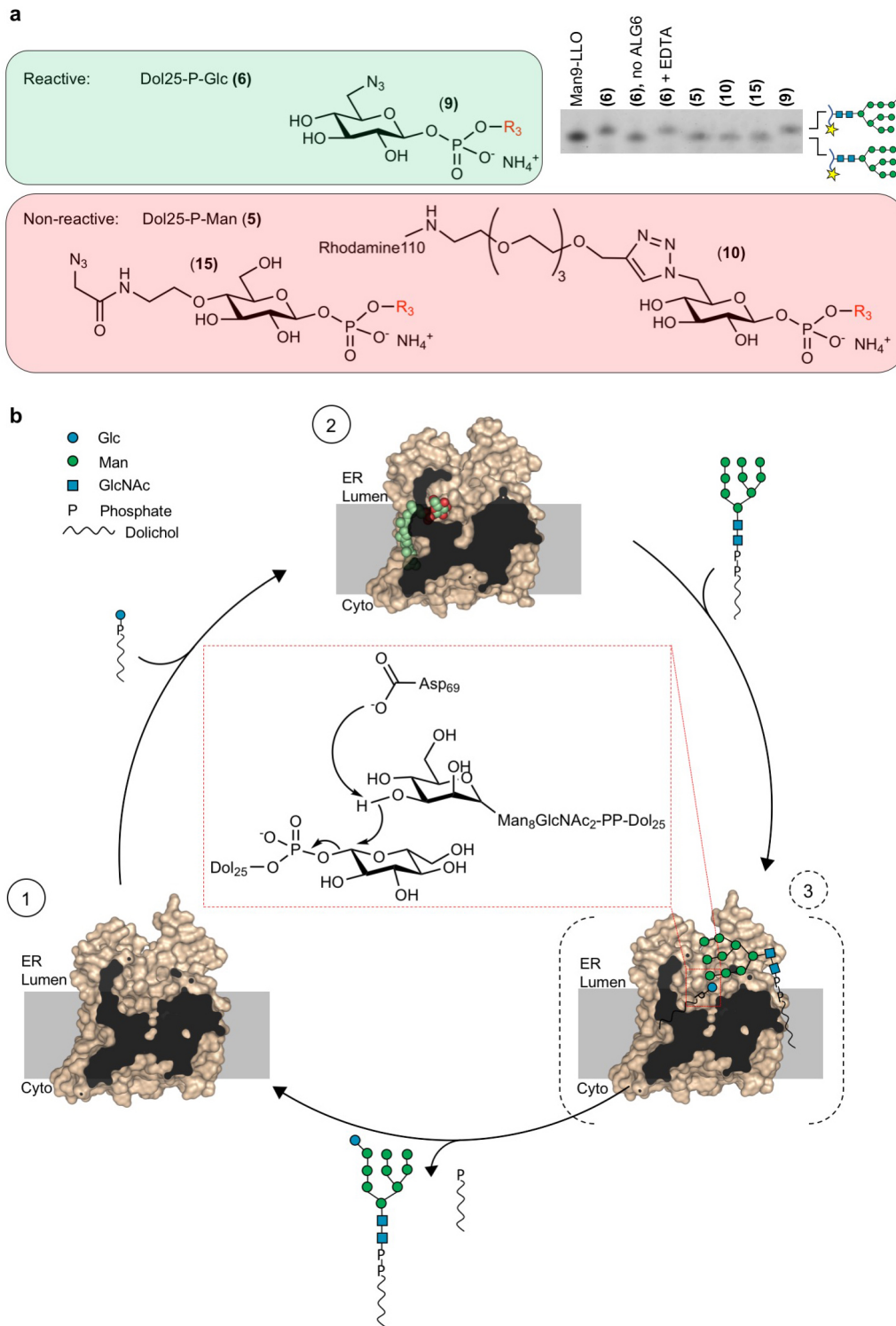
Extended Data Figure 9 Mechanism of ALG6-catalyzed glucosyl transfer. a, Chemical structures of synthetic donor substrate analogs (left) and their functional analysis (right). Compound numbers are indicated in bold and in parentheses. Analysis of ALG6 activity as described in Fig. 1b but in the presence of different substrate analogs as indicated above the lanes. The lane labeled Man9-LLO is a control sample for size comparison. In one of the ALG6 reactions, a pre-incubation of ALG6 with 10 mM EDTA was used to remove any divalent ions from the solution. **b**, Proposed three-state mechanism of ALG6, intermediate states are numbered and encircled. State 1 represents the apo state of ALG6 based on our EM structure, with ALG6 in surface representation (brown) and clipped for better visualization of the substrate binding pocket. State 2 represents ALG6 bound to donor substrate based on our EM structure, with Dol25-P-Glc shown in sphere representation with carbon atoms colored green. State 3 represents a putative ternary complex based on the apo structure of ALG6, with substrates drawn manually. The inset depicts the proposed reaction mechanism, with Asp69 acting as a general base that deprotonates the C3-hydroxyl group of the attacking mannose moiety.

Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics

	#1 ALG6 apo (EMDB-10258) (PDB 6SNI)	#2 ALG6 Dol25-P-Glc (EMDB-10257) (PDB 6SNH)
Data collection and processing		
Magnification	105.000	105,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	2.3	2.3
Defocus range (μm)	-0.6 to 0.8	-0.6 to 0.8
Pixel size (Å)	0.42	0.42
Symmetry imposed	C1	C1
Initial particle images (no.)	3,437,374	995,878
Final particle images (no.)	171,764	115,590
Map resolution (Å)	3.03	3.94
FSC threshold	0.143	0.143
Map resolution range (Å)	25-3.04	25-3.89
Refinement		
Initial model used (PDB code)	5UCB (Fab)	6SNI
Model resolution (Å)	2.96	3.90
FSC threshold	0.143	0.143
Model resolution range (Å)[a]	2.96-3.04	3.74-3.89
Map sharpening B factor (Å^2)	-68.9	-178.4
Model composition		
Non-hydrogen atoms	7,510	7,315
Protein residues	918	918
Ligands	8	1
B factors (Å^2)		
Protein	19.57	48.23
Ligand	42.76	84.99
R.m.s. deviations		
Bond lengths (Å)	0.006	0.008
Bond angles (°)	0.848	0.912
Validation		
MolProbity score	3.09	2.49
Clashscore	23.86	19.09
Poor rotamers (%)	8.07	1.49
Ramachandran plot		
Favored (%)	90.55	88,90
Allowed (%)	9.23	10.99
Disallowed (%)	0.22	0.11

[a] as defined in PHENIX[49]
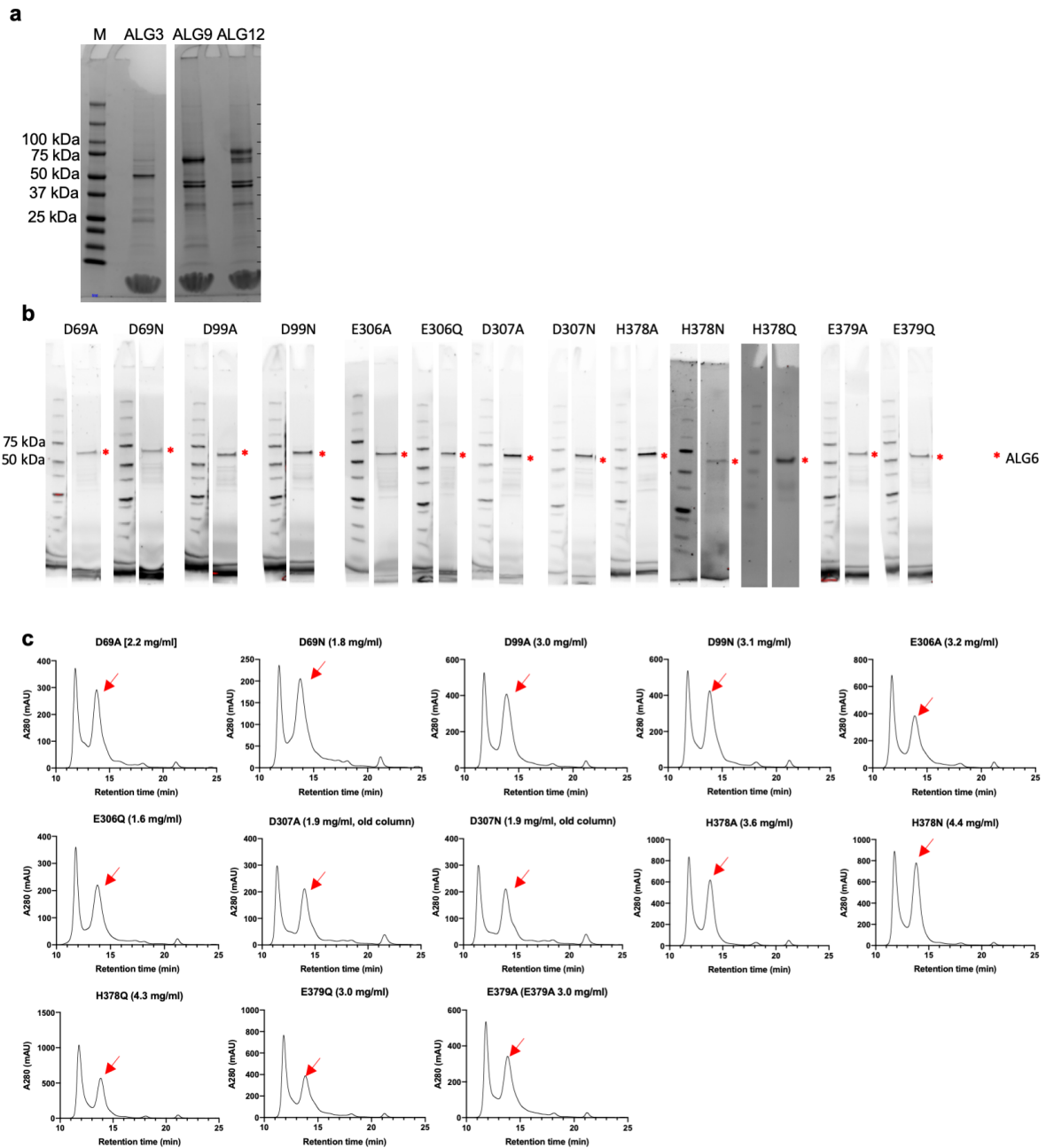
SI Figure 1 Purification of ALG3, ALG9, ALG12 and ALG6 variants. **a**, SDS-PAGE analysis of purified enzymes ALG3, ALG9 and ALG12. **b**, SDS-PAGE analysis of purified ALG6 variants. The red asterisks indicate ALG6 **c**, Size exclusion chromatography analysis (TSKgel G3000WXL, 0.5 ml/min) of purified ALG6 variants in detergent. The red arrows indicate the peak fractions that were used for functional assays. Note that ALG6 variants D307A and D307N were purified with a different column, leading to a slightly shifted elution peak.
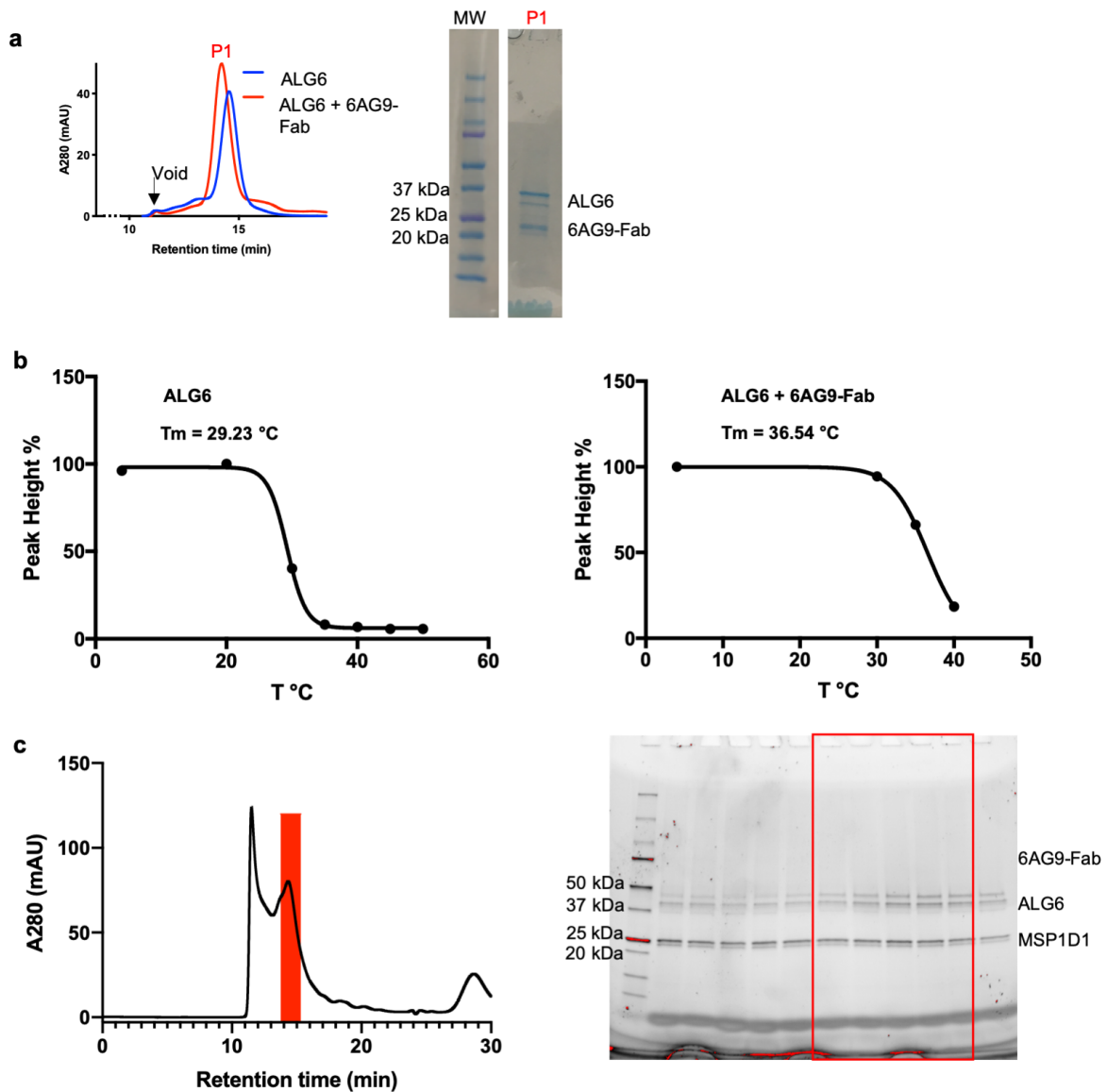
SI Figure 2 Characterization on of 6AG9-Fab binding interactions with ALG6 and sample preparation for cryo-EM analysis. a, SEC profile of ALG6 in the presence and absence of 6AG9-Fab, as shown in Fig. 2a in the main manuscript (left) and reducing SDS-PAGE analysis of SEC elution peak of the ALG6-6AG9-Fab complex (right). **b**, Thermostability analysis[42] of wild type ALG6 in detergent solution and in the presence or absence of 6AG9-Fab, performed by comparing the heights of SEC elution peaks after heating samples for 10 min at the respective temperatures prior to SEC analysis. **c**, SEC profile (TSKgel G3000WXL, 0.5 ml/min) of ALG6 reconstituted in lipid nanodiscs and in complex with 6AG9-Fab (left) and non-reducing SDS-PAGE analysis of corresponding peak fractions (right). The red box indicates the fractions that were pooled and used for cryo-EM analysis.
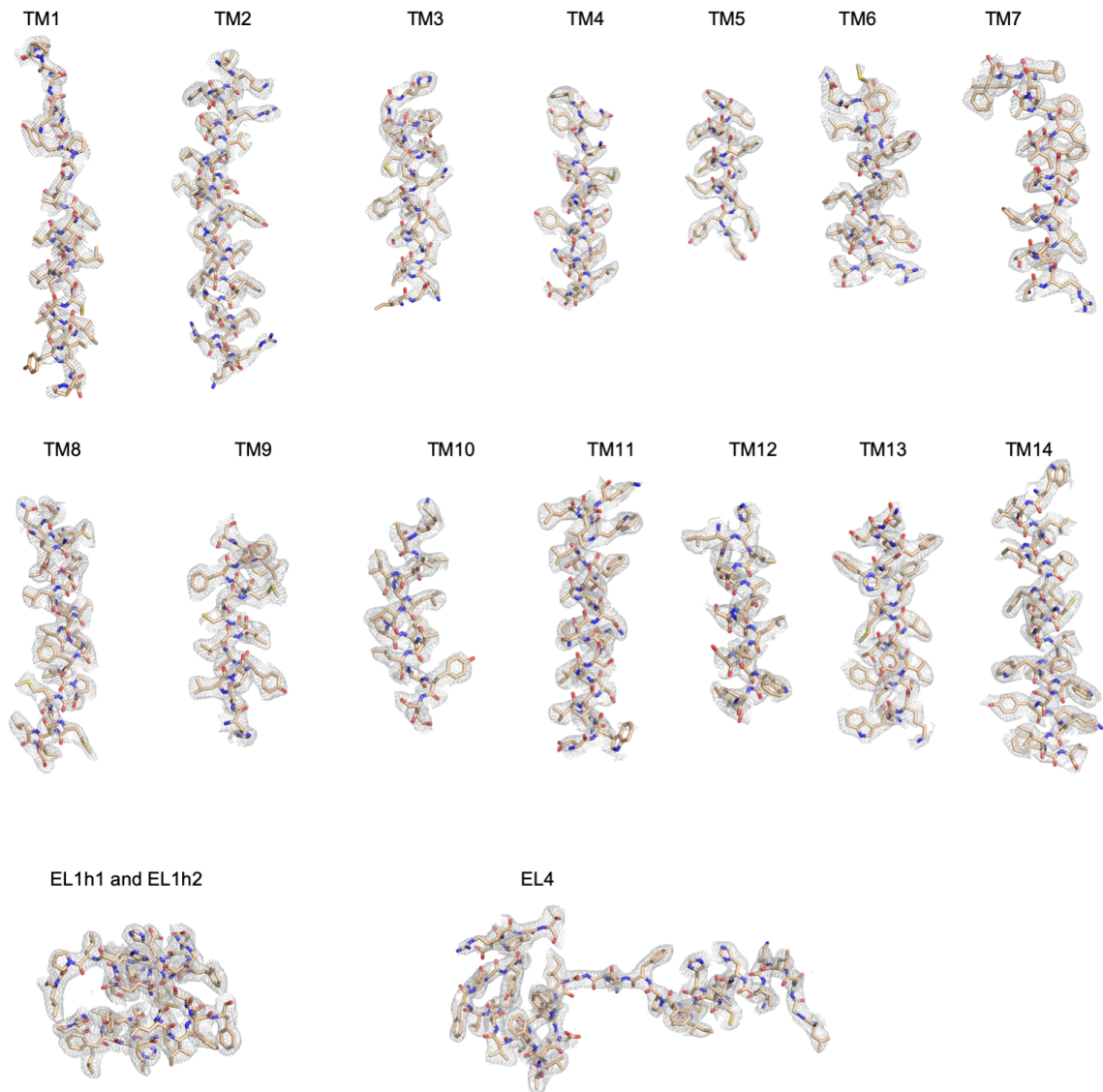
SI Figure 3 EM density for different segments of the structure of substrate-free ALG6-6AG9-Fab complex in lipid nanodiscs at 3.0 Å resolution. ALG6 is depicted in stick representation. Maps are displayed at a contour level of 7.0 rmsd and were carved at 1.6.

5. Conclusions and outlook

Part 1:

Vitamin B12 is an essential nutrient for humans and the deficiency can lead to severe developmental defects[1,2]. Ativitamin-B12 is a tool dedicated to study the short-term effects of vitamin B12 deficiency[3-6]. With our structural and mechanistic studies of the beta subunit of transcobalamin we were able to provide insights into the substrate specificity of transcobalamin. We were also able to elucidate a two-step binding mechanism, in which cobalamins bind first to the beta subunit of TC, and a second step in which the alpha domain assembles with the cobalamin bound beta domain. The construct of the beta domain of TC is well expressed and is crystallizes readily. The construct therefore provides a formidable tool for the fast and efficient study of the binding mode of TC to different ligands.

As vitamin B12 is required for DNA synthesis[7] the vitamin is crucial for the growth of carcinogenic tissue[8]. Cancer cells have been shown to have an upregulated expression of the TC receptor CD320[8]. Targeting the TC:CD320 mediated cellular uptake of B12 for cancer treatment has been a long-standing effort[8-11]. With our saporin conjugated nanobodies, we provide a new, promising tool for the destruction of cancer cells. Our in vitro experiments we demonstrate the efficient destruction of HEK293T cells by saporin conjugated TCNB4. Further experiments will be necessary to compare the killing efficiency among different cell lines at different stages of the cell cycle to make sure that the nanobody is only targeting cancer cells. Also, cell culture models that are closer to real cancer tissues, as well as actual cancer tissue would need to be tested. A drawback is that the nanobodies, which are specific for human transcobalamin cannot be tested in mice. However, the biotinylated nanobodies provide a versatile platform to derivatize the nanobodies with different probes. To assess the localization of the nanobody in a real cancer patient one could for example conjugate the nanobody with a fluorescent marker or with a non-invasive radio marker such as TC99m. Cancer therapy is one of the biggest challenges that medical research faces today and it might never be completely solved due to the extreme diversity and evolvability of the diseases. However, I do sincerely hope that my results help pushing the frontier in the battle against cancer a little bit further.

Part 2:

N-glycosylation is one of the most common post translational modifications[12,13]. The structural basis of biosynthesis of the LLO which constitutes the basis of N-glycosylation has only been poorly understood. In this thesis we were able to express and purify a number of enzymes from the ER-luminal part of the biosynthetic pathway. With the help of synthetic substrate analogs from the Reymond group and building on the previously established chemoenzymatic synthesis of Dol-PP-GlcNAc$_2$Man$_5$[14] we were able to generate high-mannose LLO intermediates. These molecules were successfully used to study the reactivity of the ER-luminal glycosyltransferases ALG3, ALG9, ALG12 and ALG6. Gaining structural insights into any of the ER-luminal glycosyltransferases was another daunting challenge. With these eukaryotic proteins being limited to low yields and being very unstable when being reconstituted in detergent they were not suitable for crystallization. On the other hand, with sizes ranging from ~50 kDa to ~70 kDa these proteins were by far out of the current realm of cryo-EM for membrane proteins. In a collaborative effort and a research visit with the Kossiakoff group we managed to obtain a non-inhibitory, conformationally binding Fab, that would not only bind to ALG6 but did also thermostabilize the protein. With the help of the Fab we managed to increase the size of ALG6 sufficiently to obtain cryo-EM structures at side-chain resolution. With the help of a synthetic donor molecule provided by the Reymond group, we were also able to identify the donor substrate binding site of ALG6. Modified substrate analogs provided by the Reymond group enabled the probing of the substrate specificity of ALG6. Finally, with the help of side directed mutagenesis we were able to identify the catalytic residue Asp69 of ALG6 and were able to speculate about a potential catalytic mechanism.

The established chemoenzymatic synthesis of high-mannose LLO intermediates provides a great tool for studying the enzymatic reactions of the pathway. By scaling up the reactions it should also be possible to generate large amounts of purified LLO intermediates which could be used for more extensive biochemical or even structural studies of enzymes of the ALG-pathway. One remaining question concerning ALG6 is the binding site of the acceptor substrate Dol-PP-GlcNAc$_2$Man$_9$Glc$_3$. With sufficient amounts of the donor substrate, it might be possible to determine the structure of acceptor substrate bound ALG6, or even of ALG6 in a ternary complex with both substrates. For this however, one will also have to find a way to trap the substrates in a state close to the transition state, but without hydrolyzing the donor substrate. An

extant challenge is the biosynthesis of the final Dol-PP-GlcNAc$_2$Man$_9$Glc$_3$ glycan, which will require biochemical control over the enzymes ALG8 and ALG10. Such a compound may become very useful to study substrate interactions of eukaryotic oligosaccharyltransferases.

The Fab binder has proven to be an extremely helpful tool for the structure determination of small membrane proteins by cryo-EM. To our knowledge, at this time, the apo structure of the nanodisc reconstitute ALG6:6AG9-Fab complex at 3.0 Å resolution is the highest resolved cryo-EM structure of a membrane embedded complex at this size. To fully understand the biosynthetic pathway of LLOs it will be inevitable to determine more structures of ALG enzymes and conformationally binding Fabs might be a great help to achieve this goal.

Perhaps the biggest insight of this study was the discovery of an evolutionary conserved GT-C fold. With the ALG6 structure, it was finally possible to find a common pattern among GT-C enzymes. The GT-C fold, which is conserved among all kingdoms of life, consists of two membrane embedded domains which together form the active site of the protein. Whereas one of these domains is structurally conserved, the other domain is highly variable. The substrate binding sites, and mechanisms of catalysis also seem to be conserved. With this newly described fold we were able to rationalize the common architecture of GT-C enzymes. In the future knowing about the variability of this fold might become useful for the directed evolution or de-novo design of new GT-C enzymes.

References

1   Reynolds, E. Vitamin B12, folic acid, and the nervous system. *Lancet Neurol* **5**, 949-960, (2006).

2   Kozyraki, R. & Cases, O. Vitamin B12 absorption: Mammalian physiology and acquired and inherited disorders. *Biochimie* **95**, 1002-1007, (2013).

3   Mutti, E., Ruetz, M., Birn, H., Krautler, B. & Nexo, E. 4-Ethylphenyl-Cobalamin Impairs Tissue Uptake of Vitamin B-12 and Causes Vitamin B-12 Deficiency in Mice. *Plos One* **8**, (2013).

4   Ruetz, M., Salchner, R., Wurst, K., Fedosov, S. & Krautler, B. Phenylethynylcobalamin: a light-stable and thermolysis-resistant organometallic vitamin B(12) derivative prepared by radical synthesis. *Angew Chem Int Ed Engl* **52**, 11406-11409, (2013).

5   Krautler, B. Antivitamins B12 -A Structure- and Reactivity-Based Concept. *Chemistry*, (2015).

6   Ruetz, M. *et al.* Antivitamin B12 Inhibition of the Human B12 -Processing Enzyme CblC: Crystal Structure of an Inactive Ternary Complex with Glutathione as the Cosubstrate. *Angew Chem Int Ed Engl* **56**, 7387-7392, (2017).

7   Nielsen, M. J., Rasmussen, M. R., Andersen, C. B. F., Nexo, E. & Moestrup, S. K. Vitamin B-12 transport from food to the body's cells-a sophisticated, multistep pathway. *Nat Rev Gastro Hepat* **9**, 345-354, (2012).

8   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. Targeted Delivery of Saporin Toxin by Monoclonal Antibody to the Transcobalamin Receptor, TCblR/CD320. *Mol Cancer Ther* **9**, 3033-3040, (2010).

9   Waibel, R. *et al.* New derivatives of vitamin B12 show preferential targeting of tumors. *Cancer Res* **68**, 2904-2911, (2008).

10   Clardy, S. M., Allis, D. G., Fairchild, T. J. & Doyle, R. P. Vitamin B-12 in drug delivery: breaking through the barriers to a B-12 bioconjugate pharmaceutical. *Expert Opin Drug Del* **8**, 127-140, (2011).

11   Quadros, E. V., Nakayama, Y. & Sequeira, J. M. Saporin Conjugated Monoclonal Antibody to the Transcobalamin Receptor TCblR/320 Is Effective in Targeting and Destroying Cancer Cells. *Journal of cancer therapy* **4**, 1074-1081, (2013).

12   Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**, 765-778, (2010).

13   Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* **21**, 576-582, (2011).

14   Ramirez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology*, 1-8, (2017).

CURRICULUM VITAE

PERSONAL DATA

Name	Joël Sharon Bloch
Date of Birth	02.03.1989
Nationality	Swiss and German
Address	Peter-Debye-Weg 13.508
	CH-8049 Zürich ZH, Switzerland
E-Mail	joel.bloch@mol.biol.ethz.ch

EDUCATION AND TRAINING

Nov 2014 – present

ETH Zürich, PhD student
Advisor: Prof. Kaspar Locher
Cryo-EM and X-Ray studies of eukaryotic membrane proteins and receptor-ligand complexes

May 2019 – Jun 2019

The University of Chicago – Prof. Anthony Kossiakoff
PhD internship: Generation of Fab binders by phage display.

Sep 2009 – Jul 2014

ETH Zürich, BSc and MSc in biology
MSc Thesis Advisor: Prof. Donald Hilvert
Major in structural biology and biophysics

Jun 2014 – Sep 2014

ETH Zürich – Prof. Cornelia Halin-Winter
Research assistant: Computational analysis of confocal microscopy data on lymphocyte migration in mice.

Feb 2013 – May 2013

The Weizman Institute of Science – Prof. Ada Yonath
MSc research project: X-ray crystallography of ribosomes.

PUBLICATIONS

1.   **Bloch, J.S.**, Pesciullesi, G., Boillevin, J., Nosol K., Irobalieva, R.N. Aebi, M., Kossiakoff, A.A., Reymond, J.L. and Locher, K.P. (2020) *Structure and Mechanism of ER-based glucosyltransferase ALG6*, **Nature**, published online 26.02.2020

2.   **Bloch, J.S**., Ruetz, M., Kräutler, B., and Locher, K.P. (2017) *Structure of the human transcobalamin beta domain in four distinct states.* **Plos One** 12, e0184932

3.   Alam, A., Woo, J.S., Schmitz, J., Prinz, B., Root, K., Chen, F., **Bloch, J.S**., Zenobi, R., and Locher, K. P. (2016) *Structural basis of transcobalamin recognition by human CD320 receptor.* **Nat Commun** 7, 12100

4.   Auerbach-Nevo, T., Baram, D., Bashan, A., Belousoff, M., Breiner, E., Davidovich, C., Cimicata, G., Eyal, Z., Halfon, Y., Krupkin, M., Matzov, D., Metz, M., Rufayda, M., Peretz, M., Pick, O., Pyetan, E., Rozenberg, H., Shalev-Benami, M., Wekselman, I., Zarivach, R., Zimmerman, E., Assis, N., **Bloch, J**., Israeli, H., Kalaora, R., Lim, L., Sade-Falk, O., Shapira, T., Taha-Salaime, L., Tang, H., and Yonath, A. (2016) *Ribosomal Antibiotics: Contemporary Challenges.* **Antibiotics** (Basel) 5

5.   Russo, E., Teijeira, A., Vaahtomeri, K., Willrodt, A.H., **Bloch, J.S**., Nitschke, M., Santambrogio, L., Kerjaschki, D., Sixt, M., and Halin, C. (2016) *Intralymphatic CCL21 Promotes Tissue Egress of Dendritic Cells through Afferent Lymphatic Vessels.* **Cell Rep** 14, 1723-1734

SUBMITTED MANUSCRIPT

2.   Kries H*., **Bloch, J.S***., Bunzel*, H.A. Pinkas, D.M. Hilvert, D., *Contribution of oxyanion stabilization to Kemp eliminase efficiency,* **under revision in ACS Catalysis** on 07.02.2020
***denotes co-first authorship**

CONFERENCE ATTENDANCE

1.   EUROCARB XX in Leiden, NL 2019. →POSTER AWARD WINNER, FIRST PLACE
2.   Glycobiology Conference GRC in Lucca IL, 2019. →POSTER AWARD WINNER
3.   Glycobiology Seminar GRS in Lucca IL, 2019. →SELECTED SPEAKER
4.   ETH Zürich D-Biol Symposium, Davos CH, 2016