Division of Molecular Structural Biology
Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

Structural studies of the ERGIC-53/MCFD2 glycoprotein transport receptor complex

Edvard Wigren

Karolinska Institutet

Stockholm 2012

Published by Karolinska Institutet. Printed by Larserics Digital Print AB

ISBN 978-91-7457-755-6

ABSTRACT

The secretory pathway defines and maintains the intracellular architecture of the eukaryotic cell. Proteins targeted to either the plasma membrane, the extracellular medium or to specific organelles within the cell are dependent on the correct transfer along this pathway. The importance of its function has become increasingly clear and several proteins in this pathway have been linked to human diseases. One disease that is genetically associated with transport processes in the early secretory pathway is the combined blood coagulation factor V and VIII deficiency (F5F8D). It has been established that F5F8D is caused by mutations in the genes that code for the membrane bound glycoprotein receptor ERGIC-53 or its co-receptor protein MCFD2. Together these two proteins form a calcium dependent complex with 1:1 stoichiometry that specifically interacts and assist transport of the glycosylated blood coagulation proteins FV and FVIII from the endoplasmic reticulum.

In this thesis, NMR-spectroscopy and X-ray crystallography have been applied in order to clarify the organisation of the ERGIC-53/MCFD2 glycoprotein transport receptor complex. The work resulted in the three-dimensional structure of MCFD2 in solution determined by NMR and the crystal structure of MCFD2 in complex with the carbohydrate recognition domain of ERGIC-53. These structures gave a first molecular view of the organisation of a cargo receptor complex in the early secretory pathway and revealed that MCFD2 undergoes significant conformational changes upon complex formation.

NMR and CD-spectroscopy analysis showed MCFD2 to be disordered in the absence of Ca^{2+} ions, but to adopt a predominantly ordered structure upon binding Ca^{2+} ions. Hence, these data suggest that calcium binding and consequent folding of MCFD2 to be the underlying mechanism for the previously observed calcium dependence of the MCFD2/ERGIC-53 interaction. Moreover, the consequences of all known F5F8D causing missense mutations found in MCFD2 could be explained at a molecular level. The results highlight the importance of intact calcium binding EF-hand motifs for the structural stability of MCFD2 and point toward disruption of the ERGIC-53/MCFD2 interaction as the underlying mechanism for these mutations in causing F5F8D.

Overall, this thesis work has provided new insights into the structural organisation of the ERGIC-53/MCFD2 transport receptor complex and highlighted the importance of calcium for the regulation of its function. By studying this complex the hope is that we have shed some light, not only on the mechanisms behind transport of FV and FVIII by ERGIC-53 and MCFD2, but also on general processes underlying the assisted glycoprotein transport in the early secretory pathway.

LIST OF PUBLICATIONS

I. Guy JE*, **Wigren E***, Svärd M, Härd T, Lindqvist Y. New insights into multiple coagulation factor deficiency from the solution structure of human MCFD2. J Mol Biol. 2008; 381(4): 941-55.

II. **Wigren E**, Bourhis JM, Kursula I, Guy JE, Lindqvist Y. Crystal structure of the LMAN1-CRD/MCFD2 transport receptor complex provides insight into combined deficiency of factor V and factor VIII. FEBS Lett. 2010; 584(5): 878-82.

III. Elmahmoudi H*, **Wigren E***, Laatiri A, Jlizi A, Elgaaied A, Gouider E, Lindqvist Y. Analysis of newly detected mutations in the MCFD2 gene giving rise to combined deficiency of coagulation factors V and VIII. Haemophilia. 2011;17(5):e923-7.

*Equally contributing authors

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CD	circular dichroism
COPII	coat protein complex II
CRD	carbohydrate recognition domain
ER	endoplasmic reticulum
ERES	ER exit sites
ERGIC	ER Golgi intermediate compartment
ERGIC-53	ER-Golgi intermediate compartment 53 kDa protein
ERGL	ERGIC-53-like protein
F5F8D	combined deficiency of factor V and factor VIII
GlcNAc	N-acetyl glucosamine
HSQC	heteronuclear single quantum coherence
LMAN1	lectin mannose-binding 1
L-type lectins	leguminous-type lectins
MCFD2	multiple coagulation factor deficiency protein 2
N-linked	asparagine linked
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
PDB ID	protein data bank identifier
R.M.S.D.	root mean square deviation
SEC	size exclusion chromatography
SPR	surface plasmon resonance
VIP36	vesicular integral-membrane protein of 36 kDa
VIPL	VIP36-like protein

1 INTRODUCTION

1.1 GENERAL INTRODUCTION

In the evolution of the eukaryotic cell from its ancestor, the prokaryotic cell, an important step was the development of specialised organelles that perform dedicated functions. This compartmentalisation of the cell, along with the elaborate transport and delivery systems between the compartments, has provided the eukaryotic cell with the capacity to evolve a variety of cellular processes and establish multicellular organisms. Central among these systems is the secretory pathway, which contains organelles that have been optimised in a sequential pathway in order to maturate and sort proteins. This pathway was first described over 40 years ago by the pioneers of modern cell biology Caro and Palade (Caro & Palade 1964). Since then the general structure underlying these processes has become more defined and several of the major protein components and signalling systems involved have been identified and their mechanisms elucidated. It is now established that proteins targeted to either the plasma membrane, the extracellular medium or to specific organelles within the cell are dependent on the correct transfer along this pathway (Mellman & Warren 2000).

Based on proteomic analysis it is estimated that one-third of the ~25,000 translated proteins in the eukaryotic genome enters the secretory pathway (Levine *et al.* 2005). An intriguing question is how all these secretory proteins are deciphered, folded, processed and sorted to their proper location within this machinery. Genetic and biochemical data have greatly increased our understanding of how the proteins within the secretory pathway function and interact to handle this enormous task. Nevertheless, we are still only beginning to grasp the extent of this complex network of interactions and processes. In order to fully understand the underlying molecular mechanisms governing the secretory pathway, a vital key is the determination of three-dimensional structures of the components in this machinery.

The importance of the secretory pathway has become increasingly clear and several of the proteins in this pathway have been linked to human diseases (De Matteis & Luini 2011). One of the first diseases that was genetically associated with processes in the secretory pathway is the combined blood coagulation factor V and VIII deficiency (F5F8D) (Nichols *et al.* 1998). More recently, it has been established that F5F8D is caused by mutations in the genes that codes for the ER-Golgi intermediate compartment 53 kDa protein (ERGIC-53), or the multiple coagulation factor deficiency protein 2 (MCFD2). Together these two proteins form a complex that specifically interacts and assist transport of the glycosylated blood coagulation proteins factor V (FV) and factor VIII (FVIII) (Zhang *et al.* 2003).

In this thesis X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy have been applied to elucidate the structure and function of the ERGIC-53/MCFD2 transport receptor complex at a molecular level. By studying this complex the hope is that we have shed some light, not only on the mechanisms behind transport of FV and FVIII by ERGIC-53 and MCFD2, but also on general processes underlying the assisted glycoprotein transport in the early secretory pathway.

1.2 OVERVIEW OF THE EARLY SECRETORY PATHWAY

After being translated on the ribosome, all proteins that travel along the secretory pathway start their journey in the endoplasmic reticulum (ER), and are subsequently sorted to different organelles, including the ER-Golgi intermediate compartment (ERGIC), the Golgi apparatus, endosomes and lysosomes (figure 1). Each of these organelles maintains a unique environment and protein composition, despite the large exchange of material throughout the pathway. The transmembrane and soluble proteins that traverse through these organelles are collectively termed secretory cargo proteins. During their journey along this pathway the secretory cargo proteins undergo a series of post-translational modifications that promote folding and assist in targeting them to the correct intra- or extra-cellular destination (Anelli & Sitia 2008).

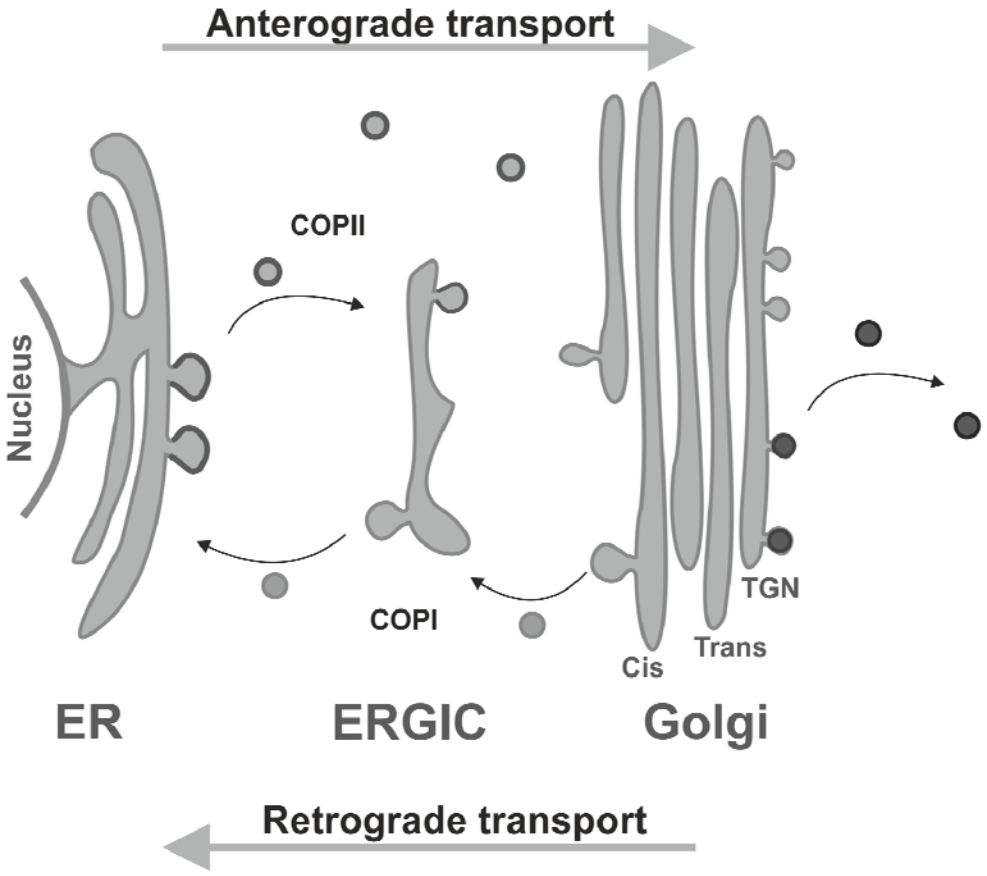


Figure 1. Schematic overwiev of the secretory pathway. Transport steps are indicated by arrows and transport vesicles: COPII in dark grey, COPI light grey and clathrin in black. Cargo destined for secretion leaves the ER at distinct exit sites in COPII-transport vesicles that bud off and are transported to the ERGIC. Cargo proteins are sorted from the ERGIC in anterograde COPII carriers that move them to the Golgi. After passage through the Golgi cargo proteins are delivered further in clathrin coated vesicles. Retrograde-directed COPI coat vesicles bud off from the ERGIC and Golgi compartments to recycle vesicle components and retrieve resident proteins that have escaped the ER.

The general principles for transport between the different organelles are similar, regardless of the destination. Transport vesicles bud off with the cargo proteins from one organelle and transport them to the subsequent organelle where they fuse with the membrane of the organelle and deposit the cargo proteins (Bonifacino & Glick 2004). Three classes of vesicles with different coat proteins, COPI, COPII and clathrin, are involved these transport events at different stages of the secretory pathway (Lee & Goldberg 2010). As a whole, the secretory pathway links organelles together to provide a framework for proteins to reach their proper destination correctly folded.

The secretory pathway is not a fixed isolated system, but is highly dynamic and responds to the surrounding environment and different intracellular contexts. Thus, the amount and variety of proteins expressed and transported can rapidly and dramatically change. As a consequence the secretory pathway continuously remodels its architecture and modulates the secretion capacity in order to adapt to the quantitative and qualitative secretion demand. The secretory pathway also responds to changes during specific stages of development programs by modifying aspects of its functional organisation and protein composition (Kondylis *et al.* 2009).

Although all eukaryotic cells essentially rely on this pathway for synthesis and sorting of secretory proteins there are major differences between different cell types. For instance, specialised secretory cells, such as antibody-producing plasma cells, have increased ER size and enlarged Golgi volume to accommodate the production of high levels of specific proteins (Tagliavacca *et al.* 2003). In contrast, the ER and the Golgi are not as highly developed in non-secretory cells and they produce only basal levels of proteins that reside in the secretory pathway to maintain their cell homeostasis.

The ER

The ER is the entry station for all secretory cargo proteins and the resident proteins of the secretory pathway**.** It consists of a membranous interconnected network of tubules, vesicles and cisternae, which takes up more than half of the total cellular membrane volume in most cells. Compared to the cytosol the ER lumen has a distinctly different milieu. For instance, the total concentration of Ca^{2+} in the ER lumen is normally in the range of 0.5-1 mM, compared to the nanomolar levels found in the cytosol (Prins & Michalak 2011, Pezzati *et al.* 1997). Moreover, the ER lumen has a distinct oxidising environment and chaperone systems that promote disulfide bond formation and protein folding (Margittai & Bánhegyi 2010). This unique ER environment provides the means to create an intracellular maturation playground that prepares secretory proteins to remain stable under challenging extracellular conditions.

The ERGIC

The ERGIC, also known as vesicular-tubular clusters or pre-Golgi intermediates, is a highly dynamic compartment found only in higher eukaryotic cells. As the name implies, this compartment is located at the interface of the ER-Golgi boundary. Although the dynamic nature of the ERGIC has been a matter of controversy, it is now recognised as a structurally distinct compartment with a unique biochemical composition (Appenzeller-Herzog & Hauri 2006). The ERGIC is proposed to have a different pH and a lower calcium concentration, compared to the ER and the Golgi. Although the exact pH of ERGIC still remains elusive, a

progressive acidification from the ER (pH 7.1-7.4) to the Golgi (pH 5.9-6.3) has been established, and the pH of the ERGIC is expected to be somewhere in between the ER and the Golgi (Paroutis *et al.* 2004).

Measurements of calcium concentrations of the secretory pathway have revealed high levels of calcium in the ER (0.5-1 mM) and Golgi (0.5-1 mM), whereas calcium levels in the ERGIC remained below detection levels. (Pezzati *et al.* 1997). This drop in pH and calcium levels from the ER to ERGIC is believed to assist in sorting proteins and allow the ERGIC to function as preliminary sorting station before the Golgi. By this means, the secreted cargo proteins are concentrated in the ERGIC and transported towards the Golgi while the ER resident proteins and the recycling proteins are retrieved back to the ER.

The Golgi

The Golgi consists of a series of discrete compartments, which include the cis- and trans-Golgi, and the trans-Golgi network (TGN) (Glick 2000). These series of compartments play an important role at the crossroads of the secretory pathway by receiving proteins from the ERGIC, modifying them, and distributing them to various destinations. The secretory proteins arriving from the ERGIC enter the Golgi cis-side, travel across the series of intra-Golgi compartments, and leave the Golgi on the trans-face. On their way through the different regions of the Golgi, the secretory proteins are subjected to distinct processing and sorting events. Among these events, glycoproteins are modified by ordered remodelling of their N-linked oligosaccharide side chains alongside biosynthesis and attachment of O-linked glycans. (Helenius & Aebi 2001, Wopereis *et al.* 2006). Finally before exiting the Golgi, the secretory cargo proteins are further sorted at the TGN and then continue their journey to their final destinations.

1.2.1 Folding and quality control in the ER

After being synthesised on the ribosome secretory proteins are targeted to the ER by a signal sequence at their N-terminal (Blobel 1975). After translocation into the ER, the nascent proteins encounter a unique folding environment provided by chaperones which prevent them from aggregating and foldases, which accelerate folding (Gething & Sambrook 1992). This folding machinery distinguishes native from non-native conformations by recognising features such as hydrophobic regions, glycosylation or exposed cysteins. The major portion of the ER chaperones belongs to three families of heat shock proteins (Hsp), Hsp40, Hsp70, and Hsp90, named according to their molecular weight (Cintron & Toft 2006). The Hsp70 family includes the central and most abundant chaperone, binding immunoglobulin protein (BiP) also named glucose-regulated protein of 78 kDa (GRP78) (Ellgaard & Helenius 2003). By successive cycles of binding and release, driven by ATP hydrolysis, BiP assists in folding of surface-exposed hydrophobic patches. This BiP chaperone cycle is tightly regulated and assisted by a set of co-chaperones that together make up a system that promotes folding and assembly by distinguishing and shielding hydrophobic regions (Bukau *et al.* 2006).

While many of the protein folding and maturation processes in the ER are similar to those in the cytosol, the ER also sustains a set of covalent modifications that exclusively occur

in this compartment, and which for many secretory proteins are a necessity to obtain their native conformation. One distinctive covalent modification exclusively occurring in the ER is the formation of disulfide bonds, which assist in folding and stabilise the protein. This involves oxidation of disulfides to form bonds, as well as isomerisation of non-native disulfide bonds. These oxidation and isomerisation processes are assisted and regulated by ER-specific enzymes of the protein disulfide isomerase (PDI) and ERo1 (ER oxidoreductin 1) family members (Margittai & Bánhegyi 2010). Other ER-specific post-translational modifications include proteolytic removal of the signal sequence by the signal peptidase complex (Paetzel *et al.* 2002), GPI anchor addition by the transamidase complex, tethering proteins to membranes (Chatterjee & Mayor 2001), and addition of aspargine-linked glycosylation by the oligosaccharyltransferase (OST) complex (Mohorko *et al.* 2011).

An elaborate ER quality control system works together, inseparable from the folding machinery, to ensure that only native conformers are transported from the ER. The existence of an ER quality control system was first discovered when it was found that misfolded viral membrane proteins were retained in the ER during their transport (Gething *et al.* 1986, Kreis & Lodish 1986). Since then it has been established that secretory proteins undergo multiple rounds of control for proper folding by the ER quality control system (Ellgaard & Helenius 2003). Ultimately, correctly folded and assembled proteins are packaged into COPII-coated vesicles, whereas terminally misfolded proteins are re-translocated from the ER back to the cytosol and degraded by the ubiquitin-proteasome system, by a process known as ER-associated degradation (ERAD) (Bernasconi & Molinari 2011).

1.2.2 Glycoprotein folding

The majority of the proteins that pass through the ER have a glycan attached to an asparagine residue that occurs in an -NXS/T- consensus sequence, where X is any amino acid except proline (Gavel & Von Heijne 1990). This N-linked glycosylation influences the sequence of folding events and acts as a sensor for the quality control machinery. In addition to this, the N-linked glycosylation assists in targeting the secreted proteins to their correct destination in the secretory pathway (Schwarz & Aebi 2011). The proteins that acquire N-linked glycans and pass through this system include most of the proteins present in blood, such as immunoglobulins, coagulation factors and several hormones.

The N-linked glycans are added to polypeptide chains upon entry to the ER lumen. This addition of a branched, triglucosylated core oligosaccharide $Glc_3Man_9GlcNAc_2$, to the nascent protein is catalysed by the OST complex (figure 2) (Mohorko *et al.* 2011). After transfer of the glycan to the polypeptide, glucosidases I and II trim the oligosaccharide moiety by sequentially removing the two terminal glucose residues. This generates $Glc_1Man_9GlcNAc_2$, with only one glucose moiety attached to the folding protein (Grinna & Robbins 1980). This monoglucosylated protein acts as a substrate for the transmembrane lectin calnexin (CNX) and its soluble paralog calreticulin (CRT) (Helenius *et al.* 1997). These two chaperone-like lectins associate with the thiol-disulfide oxidoreductase ERp57, which catalyses the formation of intramolecular disulfide bonds in glycoproteins (Pollock *et al.* 2004). The folding glycoprotein substrates undergo multiple rounds of binding and dissociation from this chaperone system.

This is alternated with cycles of deglucosylation by glucosidase II, and re-glucosylation by UDP-glucose:glycoprotein glucosyltransferase, which recognises unfolded proteins (Helenius 1994, Sousa & Parodi 1995).

If the protein fails to fold correctly after prolonged periods in the CRX/CRT cycle, α1,2-mannosidase I removes the mannose moiety from the middle branch of the core oligosaccharide $Man_9GlcNAc_2$ (Tremblay 1999). This generated $Man_8GlcNAc_2$ oligosaccharide targets the unfolded protein to the ERAD machinery, directed by the ER degradation-enhancing α-mannosidase I-like (EDEM) family of proteins (Kanehara *et al.* 2007). In contrast, glycoproteins that obtain their correct fold can no longer be re-glucosylated and exit the CRX/CRT cycle, and are subsequently transported from the ER. All together, this elaborate system maintains glycoproteins in the ER until they are correctly folded and ensures the quality of the glycoproteins that exit the ER.

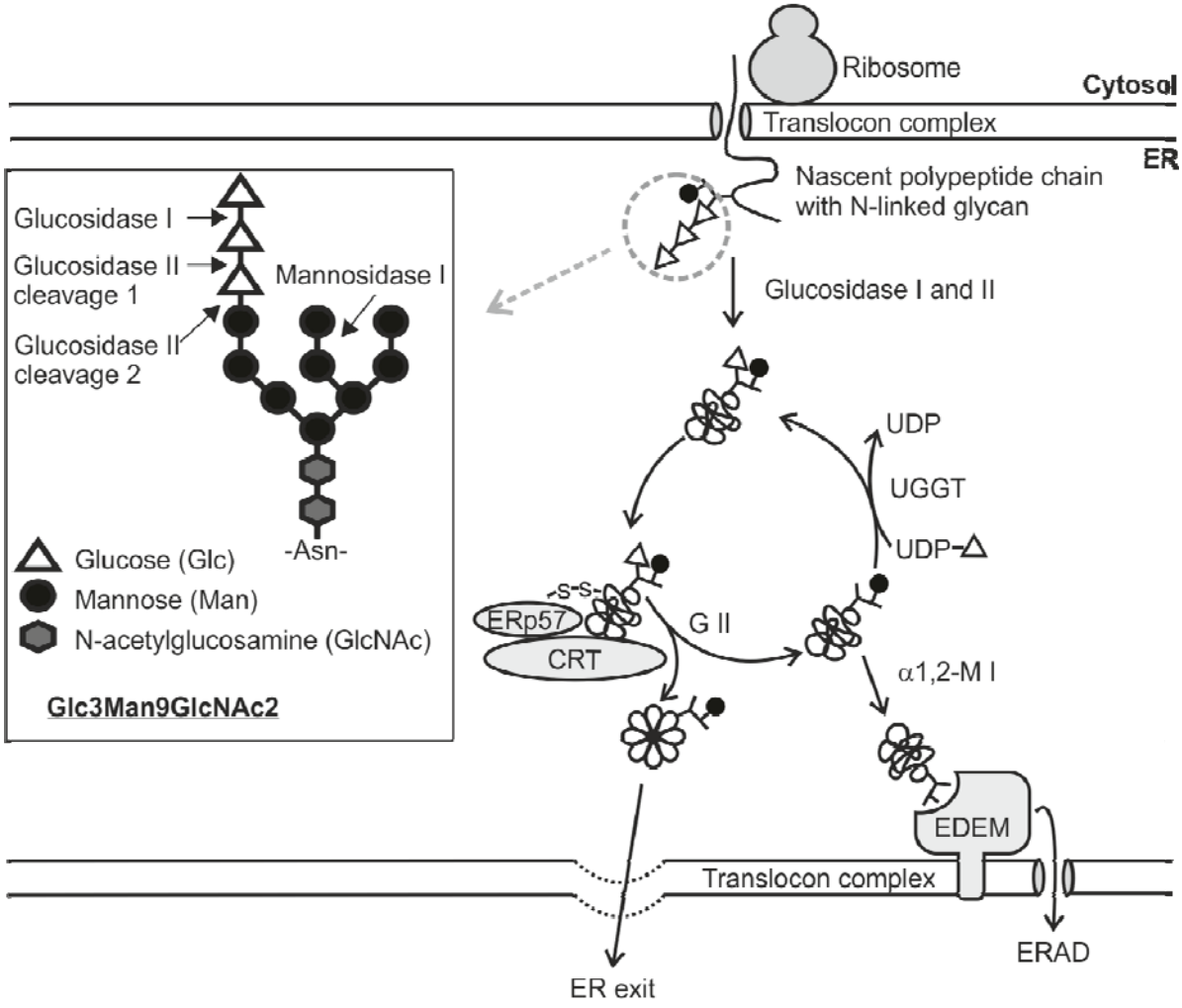


Figure 2. Schematic overview of the Glycoprotein folding pathway. The N-linked oligosaccharide shown in more detail to the left with the cleavage sites of glycosidases indicated. The figure is compiled from articles referred to in the text. The membrane bound CNX has been omitted for clarity and only the folding pathway proceeding through ERp57-CRT complex is shown.

1.2.3 Protein exit from the ER

Properly folded and assembled proteins that have passed the ER quality control are selectively sorted from ER resident proteins and concentrated into specific subdomains, known as ER exit sites (ERES) (Budnik & Stephens 2009). When visualised by electron microscopy these exit sites appear as morphologically characteristic ribosome-free regions on the ER membrane (Bannykh *et al.* 1996). The sorting of cargo proteins at ERES is tightly coupled to transport vesicle formation that is achieved by the components of the coat protein complex II (COPII) (Barlowe *et al.* 1994, Lee & Miller 2007). Although recent findings have shown alternative secretion routes, particularly under cellular stress (Giuliani *et al.* 2011), a majority of the secretory proteins rely on COPII coat transport out of the ER. The minimal machinery to drive COPII coat formation comprises three subunits: the small GTPase Sar1p, the Sec23p–Sec24p dimeric complex and the Sec13p–Sec31p heterotetrameric complex. (Lee & Miller 2007, Zanetti *et al.* 2012).

The vesicle formation process is initiated by activation of Sar1 from the GDP-bound to the GTP-bound state by the membrane-bound nucleotide exchange factor Sec12p (figure 3). Upon binding GTP Sar1 exposes an amphipathic α-helix that is inserted into the ER membrane and induces curvature of the membrane, alongside recruitment of the Sec23p–Sec24p dimer. (Bi *et al.* 2002, Russell & Stagg 2010). This initial complex engages cargo proteins either by direct interaction with transmembrane proteins or by adaptor proteins to soluble proteins. Once assembled onto the ER membrane, the newly formed pre-budding cargo complex recruits the Sec13-Sec31 heterotetramer as an outer layer of the COPII coat. The Sec13-Sec31 outer layer collects pre-budding complexes, which polymerise into a cage-like structure (Russell & Stagg 2010). This polymerisation event induces membrane scission and formation of a COPII coated vesicle that separates from the ER membrane. This results in a spherical membrane vesicle, carrying secretory cargo proteins, which is directed to and fuses with the ERGIC. These transport and docking events are coordinated by a highly conserved machinery involving numerous proteins with the key players belonging to two families: Rab GTPases, which are the main regulators of the events, and SNARE proteins, which drive membrane docking and fusion (Bröcker *et al.* 2010).

1.2.4 Protein sorting

Retrograde transport

A vital aspect for maintaining the ER and ERGIC as organelles is the balance between anterograde (forward) and retrograde (backward) vesicular trafficking (figure 1). Retrograde traffic from the ERGIC and the Golgi is mediated by COPI vesicles. These vesicles assemble as a seven-subunit protein complex on the ERGIC and the Golgi membranes in a process that is activated by the GTPase ADP ribosylation factor 1. Unlike the COPII vesicles, the COPI vesicles recruit their coat protein coatomer in a single step. Vesicle membrane deformation and carrier budding is otherwise analogous to the assembly of the COPII vesicle (Hsu & Yang 2009).

Alongside keeping the membrane balance within the organelles intact, the COPI vesicles also recycle proteins that have escaped the ER retention mechanisms back to the ER. Soluble ER proteins, such as chaperones and components of the quality control machinery, generally contain a C-terminal KDEL sequence that is recognised by the COPI associated KDEL receptor, which binds and retrieves the escaped proteins from the ERGIC and the Golgi (Munro & Pelham 1987). Membrane proteins that contain a cytosolic di-lysine KKXX-motif directly interact with the COPI vesicles coats and are in this way recycled to the ER (Sohn *et al.* 1996).

Anterograde transport

While highly abundant proteins, such as amylase, chymotrypsinogen and albumin, are passively exported by bulk flow (Martinez *et al.* 1999, Mizuno & Singer 1993), growing evidence supports the co-existence of selective transport processes where proteins are actively sorted into COPII vesicles (Malkus *et al.* 2002). This selective capture and sorting relies on specific ER export motifs within the cargo proteins. These cargo proteins can be divided into two classes: transmembrane proteins that are recruited by direct interaction with COPII coats, and soluble cargo proteins that require specific transmembrane receptors (figure 3) (Barlowe 2003).

A number of ER export motifs have been identified in transmembrane proteins in various organisms. A common feature of these ER export motifs is the location at the N- or C-terminal exposed cytoplasmic tail. One example is a category of transmembrane cargo proteins that directly interacts with the Sec23p–Sec24p dimer in the initial COPII complex via an exposed di-acidic motif of the cytoplasmic tail, -D/EXD/E- (where X can be any amino acid). This motif was first revealed by mutational analysis of the cytoplasmic tail on the vesicular stomatitis virus glycoprotein (Nishimura & Balch 1997). Since its first identification, this di-acidic motif has been found to be required for ER export of a number of proteins, among them the mammalian proteins Kir1.1 and Kir2.1 potassium channels (Ma *et al.* 2001), the cystic fibrosis transmembrane conductance regulator (Wang *et al.* 2004), and lysosomal acid phosphatase (Nishimura & Balch 1997).

A di-hydrophobic motif (FF, YY, LL or FY) at the extreme C-terminal cytoplasmic tail defines another class of ER export signal that has been extensively studied in the human protein ERGIC-53 and the yeast homologues Emp46p and Emp47p (Kappeler *et al.* 1997, Sato & Nakano 2002). Other transmembrane cargo proteins that require this motif for efficient transport from the ER include the ER-vesicle protein complex Erv41/Erv46 and the p24 family proteins (Dancourt & Barlowe 2010). Interestingly, proteins with the di-hydrophobic export motifs form oligomeric complexes in order to be efficiently transported. Hence, this suggests that the di-hydrophobic export motif only functions within the context of a larger export signal connected to the oligomeric presentation of this motif.

A recent study demonstrated that a conserved triple arginine (3R) motif, found in the G protein-coupled receptor (GPCR) superfamily, functions as an ER export signal for the multispanning transmembrane $_{\alpha 2B}$-adrenergic receptor ($_{\alpha 2B}$-AR). Furthermore, it was shown that this 3R motif, located in the third intracellular loop, facilitates ER export of $_{\alpha 2B}$-AR when transferred to its C-terminus (Dong *et al.* 2012).

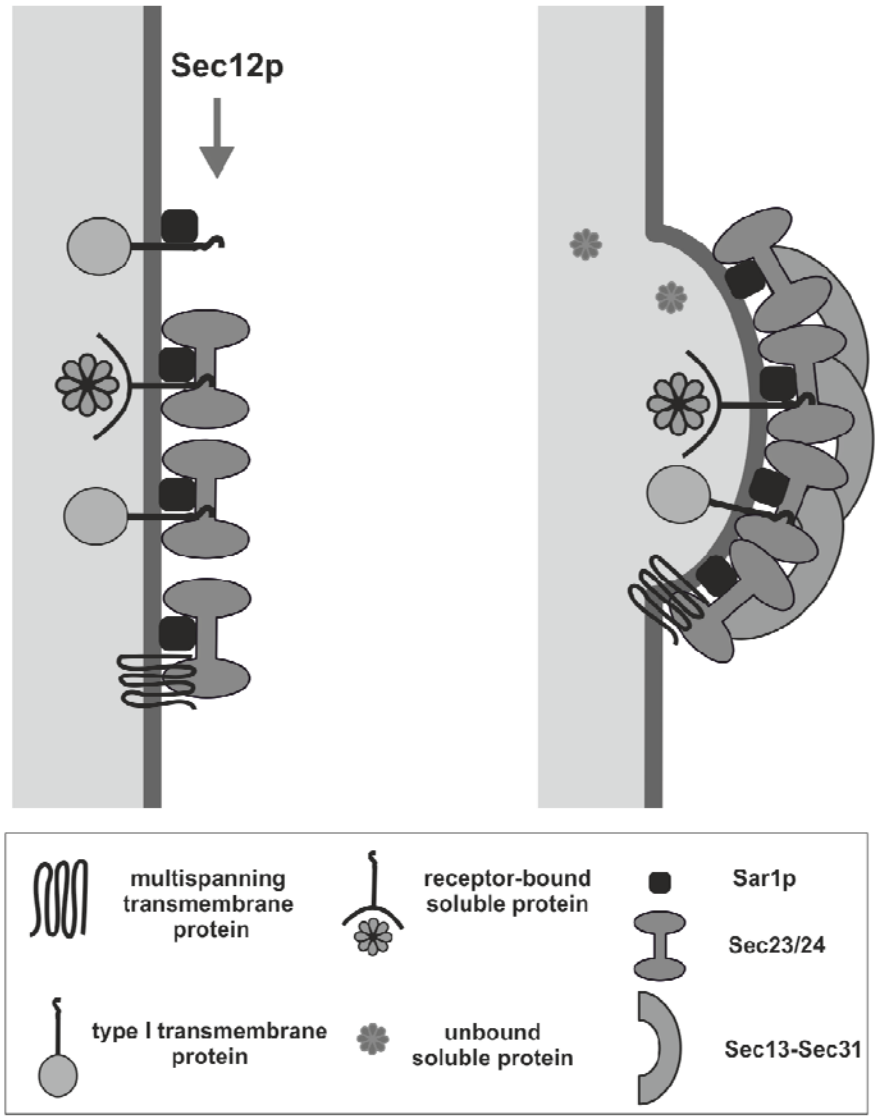


Figure 3. COPII vesicle formation and selective uptake of cargo proteins. **(A)** Vesicle formation is initiated by activation of Sar1p by nucleotide exchange factor Sec12p. Activated Sar1-GTP binds to the ER membrane and recruits the Sec23/24 subcomplex. The cytoplasmically exposed motif on the transmembrane cargo is captured by direct contact with Sec24, while transmembrane cargo receptors mediate the interaction of soluble cargo proteins to the Sec23/24 complex. **(B)** The prebudding cargo complex recruits the outer layer Sec13-Sec31 complex, leading to coat polymerisation and vesicle formation, which induces membrane scission and formation of a COPII coated vesicle loaded with soluble and transmembrane secretory cargo proteins.

In contrast to the transmembrane cargo proteins, selective export of soluble cargo proteins requires specific transmembrane receptors to mediate the interaction with the COPII coats. These receptors recognise and bind to specific export signals within soluble cargo proteins. As such they not only act as transporters, but also as key components of the secretory quality control machinery by recognising correctly folded and assembled proteins. Even though concentrative sorting of soluble proteins was shown in several studies (Salama *et al.* 1993, Mizuno & Singer 1993) and transport receptors were assumed to exist already in the early

1990s, it is only during the past 10 years that experimental evidence supporting selective sorting of soluble cargo proteins has started to emerge. Still growing in numbers, several specific cargo receptor/cargo interactions have been identified in both yeast and mammals to date (table 1).

Table 1. Protein sorting receptors in the early secretory pathway

Receptor	Species	Signals recognised on cargo	Oligomerisation state
Multispanning transmembrane receptors			
Erv29p	Yeast	I-L-V	Oligomer
Erv26p	Yeast	Lumenal determinant	Oligomer
Erv14p	Yeast	Lumenal determinant	Oligomer
KDEL receptor	Mammals	KDEL (HDEL in yeast)	Oligomer
p24 proteins			
Erv25p	Yeast	Unknown	Hetero-oligomer (Emp24p)
Emp24p	Yeast	Unknown	Hetero-oligomer (Erv25p)
L-type lectins			
Emp46p	Yeast	High-mannose N-glycans	Hetero-oligomer (Emp47p)
Emp47p	Yeast	High-mannose N-glycans	Hetero-oligomer (Emp46p)
ERGIC-53	Mammals	High-mannose N-glycans, β-hairpin	Hexameric
ERGL	Mammals	Unknown	Unknown
VIP36	Mammals	High-mannose N-glycans	Monomeric
VIPL	Mammals	High-mannose N-glycans	Monomeric

1.2.5 Transport receptors

A great deal of our current knowledge of transport receptors and their associated cargo proteins has been acquired by the genetic and biochemical analysis of the yeast *Saccharomyces cerevisiae*. The best characterised transport receptor in yeast, the ER-vesicle protein Erv29p, was first identified as a highly abundant vesicle associated integral transmembrane protein. Since its first discovery, Erv29p has been shown to act as cargo receptor required for the export of a set of soluble cargo proteins, including carboxypeptidase Y, the α-factor mating pheromone (gpαf) and proteinase A (Belden & Barlowe 2001, Caldwell *et al.* 2001). Moreover, from studies of the interaction of Erv29p with gpαf a specific motif of three hydrophobic residues (I-L-V) within a segment of the gpαf proregion was shown to be required for Erv29p-dependent recognition. This short hydrophobic sequence motif was also found to function as an ER export motif when fused to the ER resident protein Karp2 (Otte & Barlowe, 2004).

Erv14p and Erv26p are two other integral transmembrane receptors in yeast that have been shown to interact with cargo proteins and dynamically cycle between the ER and Golgi. Erv14p directly interacts with the Sec23-Sec24-Sar1 GTP pre-budding complex and mediates incorporation of the cargo protein Ax12p into COPII vesicles (Powers & Barlowe 2002). The tetra-spanning membrane protein Erv26p is required for efficient transport of the type II membrane proteins, pro-alkaline phosphatase and Ktr3p mannosyltransferase, by linking them to the COPII machinery (Bue *et al.* 2006, Bue & Barlowe 2009).

The p24 proteins comprise a family of putative cargo receptors in the 24 kDa molecular weight range that share a similar structural organisation, with an ER lumen oriented domain and a single transmembrane segment followed by an exposed cytoplasmic tail (Strating & Martens 2009). Members of this family form heteromeric complexes and cycle between the ER and Golgi in yeast. Several functions have been proposed for p24 proteins, including structural and regulating functions of vesicle trafficking in the secretory pathway. There are also strong indications that several p24 proteins are involved in protein sorting by direct interaction with cargo proteins, and it has been shown that two members of this family, Emp24p and Erv25p, form a complex that is required for efficient transport of the glycosylphosphatidylinositol-anchored protein Gas1 (Muniz *et al.* 2000).

Other proposed cargo receptors in yeast include the two type-I membrane proteins Emp46p and Emp47p, which form a hetero-oligomeric complex that interacts with COPII and COPI vesicles. Although no specific cargo proteins have been identified for Emp46p and Emp47p, it has been shown that gene deletion of either *EMP47* or *EMP46* leads to a defect in the secretion of a subset of glycoproteins (Sato & Nakano 2002). Emp46p and Emp47p both share a conserved luminal carbohydrate recognition domain (CRD) that is homologous to leguminous plant lectins (L-type lectins). The recently solved crystal structures of the CRD from both Emp46p and Emp47p (Satoh *et al.* 2006) resemble the mammalian orthologue ERGIC-53 (Velloso *et al.* 2002), which has been confirmed to act as a glycoprotein receptor. In contrast to ERGIC-53, which relies on two bound Ca^{2+} ions to interact with oligosaccharide ligands, Emp46p binds a K^{+} ion, and Emp47p does not bind any metal ion. Although Emp46p and Emp47p have different metal binding properties compared to the canonical L-type lectins, *in vitro* studies have shown that both proteins bind glycoproteins carrying high mannose glycans (Satoh *et al.* 2006), thus strengthening the theory that Emp46p and Emp47p are involved in transporting glycoproteins.

Although most identified transport receptors in yeast are conserved in mammals, it has been challenging to directly link the mammalian orthologues to transport of specific cargo proteins. The molecular-genetic methods applicable in yeast have proven more problematic in mammalian cells. Moreover, analysis of protein transport in yeast, which lacks the ERGIC, can be accommodated within a simple model in which COPII vesicles mediate transport from the ER to the Golgi. The addition of the ERGIC to the secretory pathway in mammals adds an increasing complexity to this model.

Thus far, with the exception of the well characterised glycoprotein receptor ERGIC-53 and its close homologue the vesicular integral-membrane protein of 36 kDa (VIP36), there is no direct evidence for ER-to-Golgi transport receptor functions of yeast orthologues in higher eukaryotes. For example, there is no evidence for transport functions in orthologues of the well characterised transport receptor Erv29p from yeast. Interestingly though, the mammalian orthologue of Erv29p, Surf4, was recently shown to interact with ERGIC-53, and to be essential for maintaining the architecture of the ERGIC and the Golgi (Mitrovic *et al.* 2008). Novel functions have also been proposed for mammalian orthologues of p24 family members and the importance of this family of proteins in mammals has been demonstrated by deletion of a single family member (p23 or p24δ1) in mice, leading to embryonic lethality (Denzel *et al.* 2000). In contrast, the p24 proteins seem to be dispensable in yeast, since deletion of all eight p24s found in yeast leads only to minor defects in cargo transport (Springer *et al.* 2000).

The best characterised transport receptor in mammals, ERGIC-53 belongs to a family of four type-I membrane proteins, which also includes the ERGIC-53-like protein (ERGL), VIP36 and the VIP36-like protein (VIPL) (Hauri *et al.* 2002). In common with their yeast homologues Emp46p and Emp47p, these four proteins all share a conserved luminal CRD. All of these four mammalian lectins have been proposed to assist in glycoprotein sorting and transport in the early secretory pathway. However, their exact roles in these processes remain to be revealed.

An interesting question is how the binding and release of soluble cargo proteins are regulated by their cognate receptors? The differences in calcium and pH levels between the different compartments have been proposed to influence these transport events. Thus far this has only been demonstrated between specific cargo proteins and ERGIC-53 (Itin *et al.* 1996, Appenzeller-Herzog *et al.* 2004). Furthermore, the common feature of cargo receptors to form oligomeric protein complexes is most likely a key event in these transport processes. It is possible that changes in compartment environment would influence the oligomeric state and, as a consequence, regulate binding affinity of the cargo to its receptor.

It is generally assumed that each cargo receptor accounts for efficient ER-exit of a limited set of soluble secretory proteins. What distinguishes these proteins from other secretory proteins and why have specific transport routes evolved for these proteins? The requirement for a combination of transport determinants, such as oligosaccharides and specific structural motifs within the cargo proteins, most likely limits the repertoire of possible cargo proteins for. This requirement for a combination of motifs ensures that only properly folded cargo proteins are transported further and thus functions as a complementary quality control system.

1.2.6 L-type lectins in the early secretory pathway.

The term Lectin from the Latin *legere* "to pick out or choose" was first introduced by William Boyd in 1954 in order to describe plant agglutinins ability to distinguish between erythrocytes of different blood types (Boyd & Shapleigh 1954). Since this first definition the knowledge of this family of non-enzymatic sugar binding proteins has grown considerably. Lectins have been found in all organisms and to serve in a wide variety of different biological functions, such as cell-cell recognition events and in the innate immune system. In general these proteins consist of multiple domains but the carbohydrate binding function can usually be ascribed to a single CRD (Loris 2002).

L-type lectins are a major class of lectins that was first discovered in the seeds of leguminous plants. This family of lectins is primarily distinguished by the structural resemblance of their CRDs, which largely consist of a β-sandwich structure with a concave sheet and a convex sheet connected by short loops and β-bends (figure 4) (Loris 2002). These L-type lectins vary significantly in sugar binding preferences and several structures with different sugar moieties have been determined. Despite this fact, the binding site consisting of a negatively-charged cleft is highly conserved within the L-type lectins. Another conserved feature is the presence of two binding sites for metal ions in close proximity to this sugar binding site. The binding of these two ions, a compulsory Ca^{2+} ion at one site and a transition metal ion, usually Mn^{2+}, at the second site, serves to orient and position the residues in the sugar binding site and consequently enables the recognition of specific sugars (Bouckaert *et al.* 1999).

In the early 1990s, two mammalian proteins with sequence homology to L-type lectins, ERGIC-53 and VIP36, were discovered and linked to the secretory pathway (Schweizer *et al.* 1988, Fiedler & Simons 1994). The crystal structures of the CRDs of these proteins confirmed their relationship and helped to define a new class of animal L-type lectins that reside in the secretory pathway (figure 4) (Velloso *et al.* 2002, Satoh *et al.* 2007). Since then, two related proteins VIPL (Neve *et al.* 2003 Nufer *et al.* 2003), and ERGL (Yerushalmi *et al.* 2001) have been discovered, and today the family of animal calcium-dependent L-type lectins comprises four members which all share an L-type lectin CRD. In contrast to plant L-type lectins, which are soluble secreted proteins, these four animal L-type lectins are all type-I membrane proteins. Although homologous and structurally similar, these four transmembrane lectins are considered to play distinctly different roles in glycoprotein sorting and transport processes in the secretory pathway.

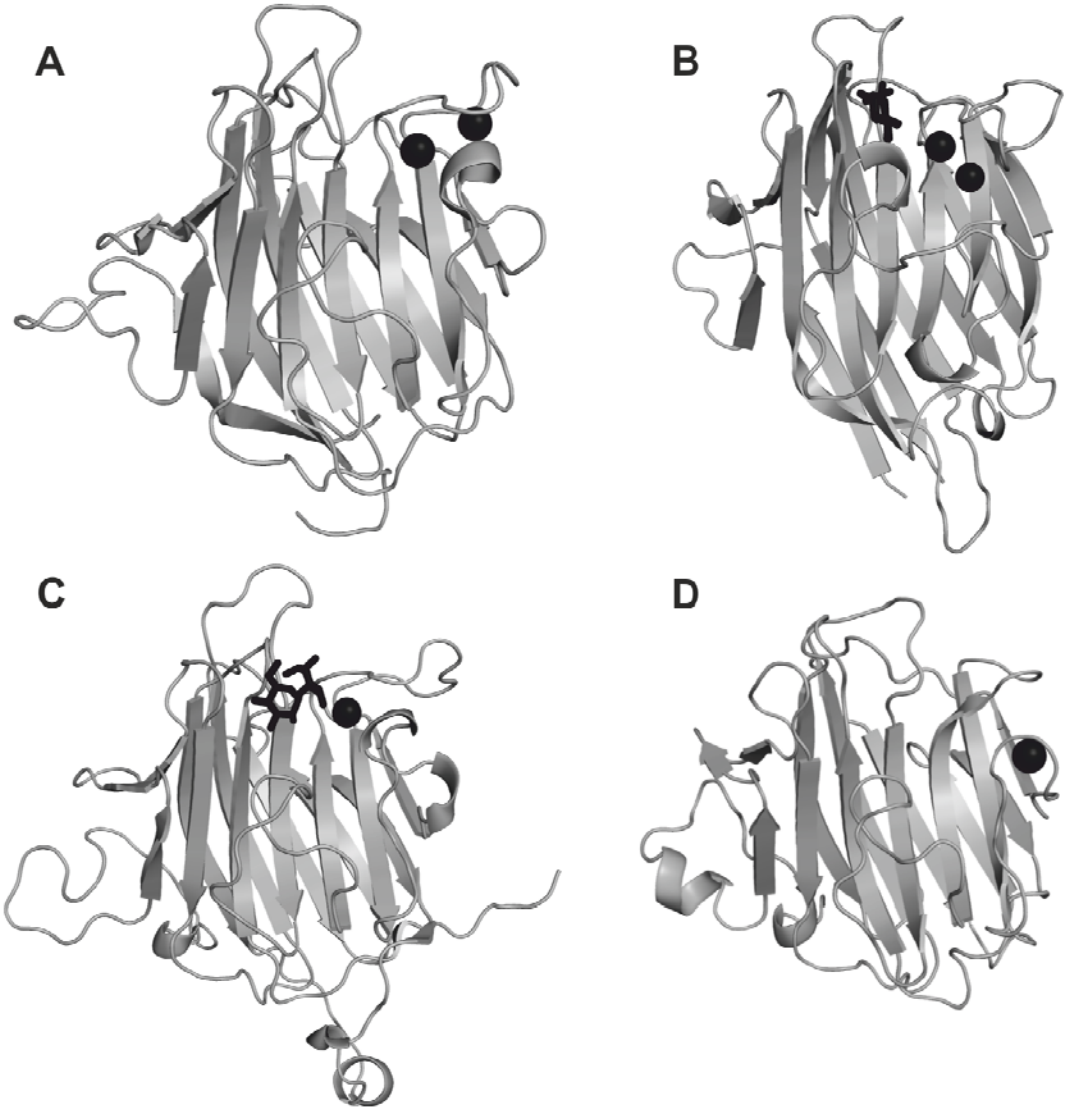


Figure 4. Structures of L-type lectins with bound ions shown as a dark spheres and sugars as stick models in black. PDB IDs in square brackets. **(A)** ERGIC-53 with two bound Ca^{2+} ions [3lcp], **(B)** Pea lectin-D-glucopyranose complex with a Mn^{2+} ion bound at one site and a Ca^{2+} ion at the second site [2bqp], **(C)** VIP36- Man-α-1,2-Man complex with a bound Ca^{2+} ion [2dur], **(D)** Emp46p with a bound K^{+} ion [2a6v].

The sequence and predicted domain organisation is highly similar between ERGIC-53 and ERGL. Both proteins share a coiled-coil domain that connects the CRD to the transmembrane region. This domain has been shown to be involved in the assembly of ERGIC-53 to a homo-oligomeric complex. In VIP36 and VIPL, this domain is absent, and data have suggested that both proteins function as monomers (Fiedler & Simons 1996, Neve *et al.* 2003, Nufer *et al.* 2003).

In all members of this family, the transmembrane segment is followed by a short cytosolic tail which mediates the sorting of ERGIC-53, VIP36 and VIPL within the secretory pathway. ERGL, on the other hand, lacks typical transports motifs in its tail and is suggested to be a non-cycling ER resident protein. VIPL has also been found to be confined to the ER, although it carries the potential ER export motif (FY). However, mutational studies have revealed that VIPL localisation is determined by a di-arginine-based (RKR) ER retention motif in its cytosolic tail (Nufer *et al.* 2003).

In contrast, ERGIC-53 and VIP36 are dynamic proteins and cycle in the secretory pathway. The cycling properties of ERGIC-53 are determined by two motifs in its cytosolic tail: a di-phenylalanine motif (FF) that mediates ER export, and a di-lysine motif (KK) that retrieves ERGIC-53 back from the ERGIC (Hauri *et al.* 2000). While the ER to ERGIC cycling route of ERGIC-53 is well established, there is still some controversy about the cycling properties of VIP36. The cytoplasmic tail of VIP36 contains two non-canonical transport motifs, the hydrophobic (FY) ER export motif, which is also found in VIPL, and a di-basic (KR) retrieval motif. Although post-Golgi cycling routes have been proposed for VIP36 (Fiedler *et al.* 1994), recent data suggest that VIP36 cycles between the ERGIC and the Golgi, and mainly localises to the Golgi (Reiterer *et al.* 2010).

The sugar-binding specificities of ERGIC-53, VIP36 and VIPL were recently studied by frontal affinity chromatography using a pyridylaminated sugar library (Kamiya *et al.* 2008). The results revealed specific characteristic sugar binding profiles for each lectin. ERGIC-53 showed broad specificity, but low affinity towards various high-mannose carbohydrate structures. In accordance with previous findings, binding was most efficient at neutral pH 7.5–8, as found in the ER lumen. However, in contrast to what has been proposed, ERGIC-53 did not discriminate between monoglucosylated glycans and deglucosylated glycans. These results might reflect the fact that the measurements were performed using the monomeric CRD and that oligomerisation of ERGIC-53 is a prerequisite for specificity and affinity.

VIP36 showed highest affinity towards the deglucosylated high-mannose A branch. This finding is supported by the recently determined crystal structures of VIP36-CRD in complex with Man-α-1,2-Man-α -1,2-Man, which corresponds to the A branch. The sugar binding of VIP36 was also shown to be pH-dependent, with an optimum pH of 6.5 corresponding to the pH of the cis-Golgi. Collectively, from cycling and sugar binding properties a model has been proposed, where VIP36 functions in quality control of glycoproteins in the Golgi by recycling inadequately trimmed glycoproteins that have escaped the ER QC machinery (Reiterer *et al.* 2010).

VIPL, the close homologue of VIP36, also shows high affinity for the deglucosylated high-mannose A branch. In contrast to VIP36 though, VIPL binds oligosaccharides optimally at pH 7.5, corresponding to the lumenal pH of the ER. From these data it has been suggested that the ER resident protein VIPL binds deglucosylated high-mannose glycoproteins released

from the CNX/CRT cycle and delivers them to ERGIC-53 for transport from the ER. In this manner, VIPL is predicted to protect correctly folded glycoproteins from demannosylation by α1,2-mannosidase I and subsequently from ERAD.

1.3 THE ERGIC-53/MCFD2 TRANSPORT RECEPTOR COMPLEX

1.3.1 ERGIC-53

ERGIC-53, also known as LMAN1, was originally discovered as a 53 kDa protein by Schweizer *et al*. in 1988 in a monoclonal antibody screen against purified Golgi membrane fractions from human intestinal epithelial cells. Independently, but using a similar approach, Saraste and co-workers identified a 58 kDa protein in Golgi fractions from rat pancreas and named it p58 (Saraste *et al.* 1987). The p58 protein was originally believed to be a glycoprotein localised to the cis-Golgi, but further studies revealed p58 to be non-glycosylated and the sequence 89% identical to ERGIC-53 (Lahtinen *et al.* 1996). Since their discovery, ERGIC-53 and the closely related rat orthologue p58, have been extensively used as model proteins to study mechanisms underlying protein trafficking in the early secretory pathway. ERGIC-53 has mainly served as a marker for the ERGIC, in which it is present at high concentrations. In these studies ERGIC-53 has also been found, although at lower levels, in the ER and in the cis-Golgi (Schweizer *et al.* 1988, Hauri *et al. 2000)*. While ERGIC-53 levels are generally low in the ER, high localised concentrations of ERGIC-53 can be found in the ERES budding structures and associated with the COPII and COPI machinery (Tisdale *et al.* 1997, Wendeler *et al.* 2007). Collectively, these studies have established an ERGIC-53 recycling pathway from the ER to ERGIC accomplished through interactions of the C-terminal, di-phenylalanine (FF) ER exit motif and di-lysine (KK) ER retrieval motif, with the COPII and COPI vesicles.

On the basis of initial sugar binding studies and sequence homology to L-type lectins ERGIC-53 was recognised as a Ca^{2+}-dependent lectin and was proposed to be involved in ER export of mannose-rich glycoproteins. This theory was strengthened when it was shown that secretion of the lysosomal glycoproteins cathepsin C (catC) and cathepsin Z (catZ) was delayed in cells expressing a C-terminal KKAA-mutant variant of ERGIC-53 that is retained in the ER (Vollenweider *et al.* 1998). Subsequent studies revealed that efficient binding of catZ to ERGIC-53 was not only dependent on its N-linked glycan moieties, but also involved an exposed β-hairpin next to the glycan site (Appenzeller-Herzog *et al.* 2005). The interaction was also shown to be pH-dependent, demonstrated by the impaired binding of ERGIC-53 to catZ upon acidification of cells (Appenzeller-Herzog *et al.* 2004). Further support for ERGIC-53 to function as a glycoprotein receptor came from human genetics, when the cause of the bleeding disorder F5F8D was associated with mutations in ERGIC-53 and its transport of the heavily glycosylated coagulation proteins FV and FVIII (Nichols *et al.* 1998).

The list of proteins associated with processes involving ERGIC-53 has become longer since these first studies. In a recent study using a yellow fluorescent protein-based protein complementation assay, the glycosylated serine protease inhibitor α1-antitrypsin was shown to depend on ERGIC-53 for efficient transport from the ER (Nyfeler *et al.* 2008). Two other proteins that have been found to act as cargo for ERGIC-53 are the glycosylated membrane protein Nicastrin, a component of the γ-secretase complex, which has been associated with processes involving Alzheimer's disease (Morais *et al.* 2006), and Sulfatase Modifying Factor 1, a formylglycine generating enzyme that activates sulfatases (Fraldi *et al.* 2008).

In addition to ascribed cargo receptor functions, ERGIC-53 has also been associated with assembly and transport of secretory hexameric IgM and recently it was shown that ERp44, a soluble protein involved in thiol-mediated retention, interacts with ERGIC-53 and assists in these processes (Anelli *et al.* 2007). Furthermore, ERGIC-53 has been shown to interact with the fibroblast growth factor receptor 3 (FGFR3) and to act in pre-assembly and quality control mechanisms of this receptor, findings which have significant implications for understanding the pathogenesis of FGFR3-related disorders (Lievens *et al.* 2008).

ERGIC-53 is a type-1 transmembrane, non-glycosylated protein containing a signal sequence that is cleaved off after mediating translocation into the ER. The mature protein comprises a lumenal domain, a transmembrane domain and a short cytoplasmic tail, which contains sorting motifs for the cycling properties between the ER and the ERGIC. The lumenal domain can be divided into two subdomains, an N-terminal CRD and a membrane-proximal stalk domain. The crystal structures of the CRD, in apo- and Ca^{2+}-bound form, revealed a large carbohydrate-binding site capable of accommodating highly branched mannose glycans (Velloso *et al.* 2002, 2003). Moreover, these structures showed that binding of two Ca^{2+} ions stabilise and coordinate two loop regions in close proximity to the carbohydrate binding site, supporting a calcium dependent sugar binding mechanism.

The membrane proximal stalk domain is predicted to consist of four α-helices that form a coiled coil structure. Two cysteines in this domain, located adjacent to the TM domain, form inter-molecular disulfide bonds, which in conjunction with structural determinants in the TM and lumenal domains, have been shown to mediate oligomerisation of ERGIC-53 (Lahtinen *et al.* 1999, Nufer *et al.* 2003, Neve *et al.* 2005). Morphological and biochemical data have shown that ERGIC-53 exists as homohexamers and that disulfide-linked stabilisation and proper oligomerisation are prerequisites for its cargo receptor function (Nufer *et al.* 2003, Neve *et al.* 2005, Zhang *et al,* 2006, Zheng *et al,* 2010B).

ERGIC-53 represents the best characterised cargo receptor in the early secretory pathway in mammalian cells and a general model for how ERGIC-53 acts a transport receptor for a limited set of glycoproteins has been established (figure 5), although the mechanism for several steps remains elusive. In this model hexameric ERGIC-53 captures correctly folded secretory glycoproteins in the ER and assists them through the COPII pathway to the ERGIC where dissociation occurs, triggered by the more acidic pH and lower calcium concentration of this compartment. After releasing the secretory cargo proteins, ERGIC-53 is recycled back to the ER in COPI vesicles for a new transport cycle, while the released glycoproteins proceed further through the secretory pathway.

Interestingly, only a limited sub-set of all glycoproteins that exit the ER have been found to rely on ERGIC-53 for their transport, thus implying that other determinants alongside N-linked glycans are required for selective cargo recognition by ERGIC-53. The surface exposed β-hairpin loop on pro-cathepsin Z remains the only confirmed determinant in this category today (Appenzeller-Herzog *et al.* 2005). The known cargo proteins differ considerably both structurally and in domain organisation, and interesting challenges for future studies will be to reveal new structural determinants.

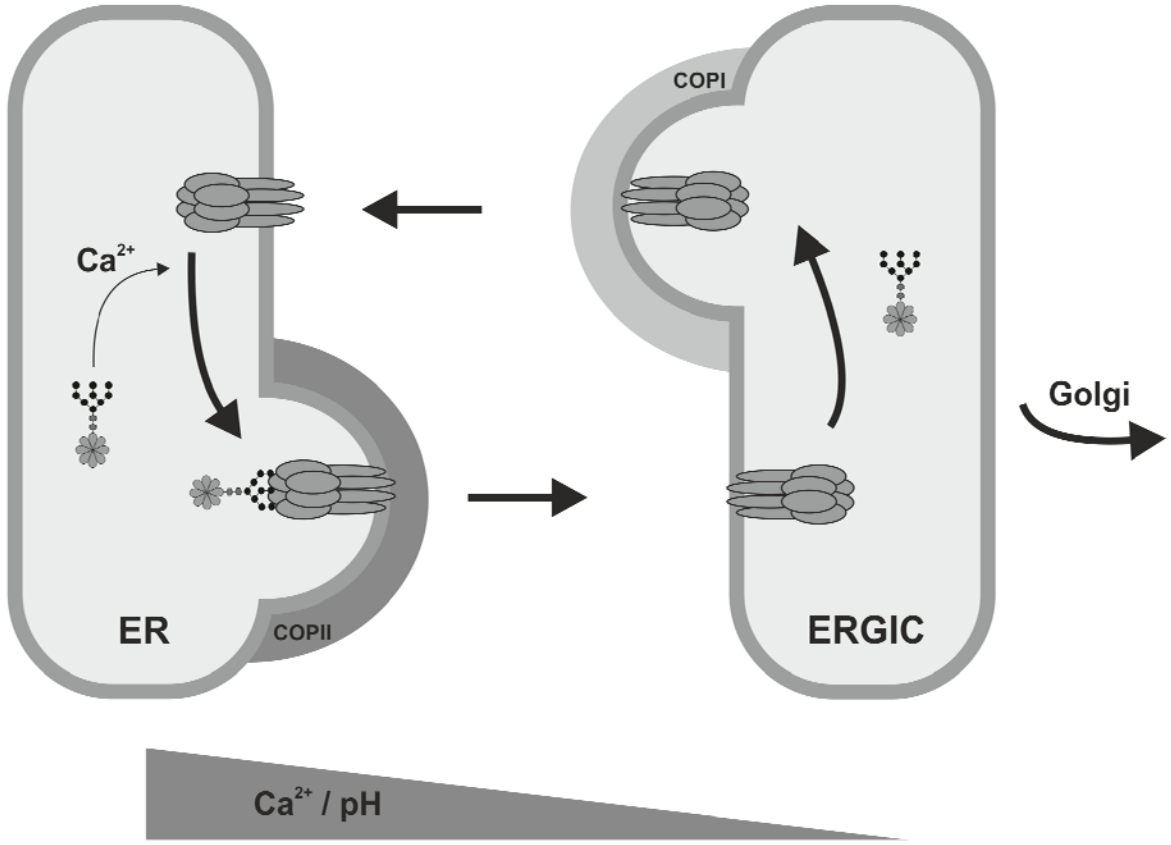


Figure 5. Schematic overview of the cargo receptor and recycling of ERGIC-53. Hexameric ERGIC-53 binds to the $Man_9GlcNAc_2$ glycan on correctly folded glycoproteins in a Ca^{2+} ion dependent manner. The cargo receptor complex is recruited to the COPII vesicle pre-budding complex by direct interaction between, the FF-motif in the cytosolic tail of ERGIC-53 and the Sec23/24 coat proteins. The COPII vesicle are then transported to and fuses with the ERGIC, where the cargo is released from ERGIC-53, as a consequence of the lower pH and Ca^{2+} ion levels of this compartment. The released cargo proteins continue their journey to the Golgi, while free ERGIC-53 is recruited into the COPI vesicle and recycles back to the ER by interaction of its cytosolic KK-signal with the COPI components. In the ER, ERGIC-53 recruits new cargo and next transport cycle is initiated.

1.3.2 MCFD2, an EF-hand protein

The first link between MCFD2 and the secretory pathway was established when it was identified as the second gene location of previously unexplained F5F8D causing mutations (Zhang *et al.* 2003). While the role of ERGIC-53 as a glycoprotein receptor was well established, less was known about the involvement of MCFD2 in transport processes in the secretory pathway when this project was initiated. Still to date, no other proteins except FV and FVIII have been found to require MCFD2 for transport, and it has been shown that knockdown of MCFD2 has no effect on the interaction of CatC or CatZ with ERGIC-53 (Nyfeler *et al.* 2006).

MCFD2 is a small soluble protein of 16 kDa, with an N-terminal ER translocation signal sequence and two C-terminal calcium-binding EF-hand motifs. Studies have shown that MCFD2 is localised to the ER and the ERGIC. However, in contrast to many other soluble ER-resident proteins MCFD2 lacks the C-terminal KDEL retrieval signal that mediates recycling back to the ER, suggesting that localisation of MCFD2 is reliant on its interaction with ERGIC-53. Orthologues of MCFD2 have been found in vertebrates from mammals to fish and the EF-hand motifs in MCFD2 also show high sequence conservation to other EF-hand proteins, such as calmodulin and parvalbumin.

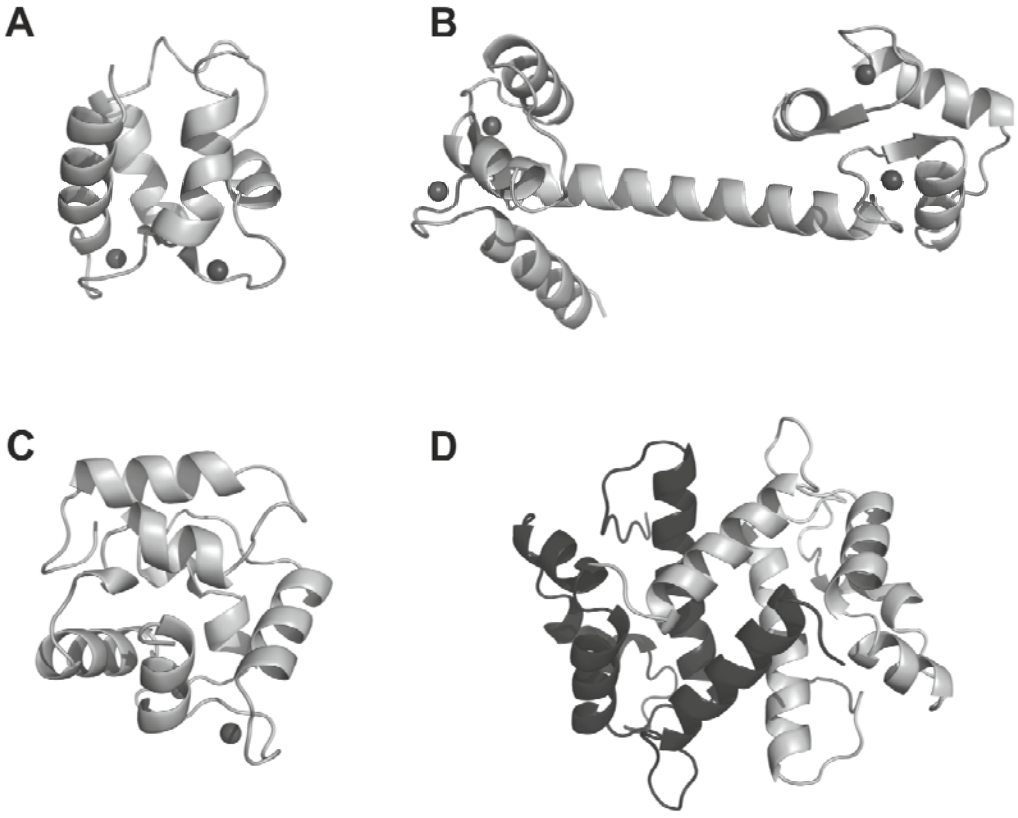


Figure 6. Domain organisation of EF-hand proteins with bound Ca^{2+} ions shown as dark spheres. PDB IDs in square brackets. **(A)** Calbindin D9K containing only a single pair of EF-hands [1b1g]. **(B)** Calmodulin a four EF-hand protein with two independent domains connected by a flexible linker [1cll]. **(C)** Parvalbumin a three-EF-hand protein with a non-functional EF-hand serving to stabilise the Ca^{2+}-binding pair [2pvb]. **(D)** The S100 family member Psoriasin in absence of Ca^{2+} ions, a two EF-hand protein that exists as a homodimer, one of the monomer in colored dark grey in the figure. Structure solved in absence of Ca^{2+} ions [1psr].

The name EF-hand was introduced by Kretsinger and derived from the initial descriptions of the crystal structure of parvalbumin, which was the first solved structure of protein containing the helix-loop-helix Ca^{2+}-binding EF-hand motif (Kretsinger *et al.* 1972). Most EF-hand proteins contain an even number of EF-hands, which occur as interacting pairs. These pairs of EF-hand motifs form a structurally stable EF-hand domain through hydrophobic contacts among the helices flanking the Ca^{2+}-binding loops and a network of hydrogen bonds forming a short antiparallel β-sheet that connects the Ca^{2+}-binding loops (Grabarek 2006). Although sharing highly similar EF-hand domains, EF-hand proteins are very diverse in their domain architecture and in how they use these domains to function (figure 6). Consequently, EF-hand proteins constitute a large and diverse family of proteins that plays essential roles in a variety of cellular processes involving calcium regulation, such as muscle contradiction (Holmes 1996) and signal transduction between cellular organelles (Niki *et al.* 1996).

The canonical EF-hand motif consists of a 12-residue loop in which a Ca^{2+} ion is 7-fold coordinated in a pentagonal-bipyramidal arrangement. The consensus sequence for this motif is commonly described as X*Y*Z*−Y*−X**−Z, where the X, Y, and Z pairs are the Ca^{2+} coordinating ligands named according to their alignment on the approximate axes of a pentagonal bipyramid, and the stars represent intervening residues (figure 7). The first nine residues of this consensus sequence form the loop, and the final residue a Ca^{2+} ligand at position −Z, is the third residue of the following helix (Gifford *et al.* 2007).

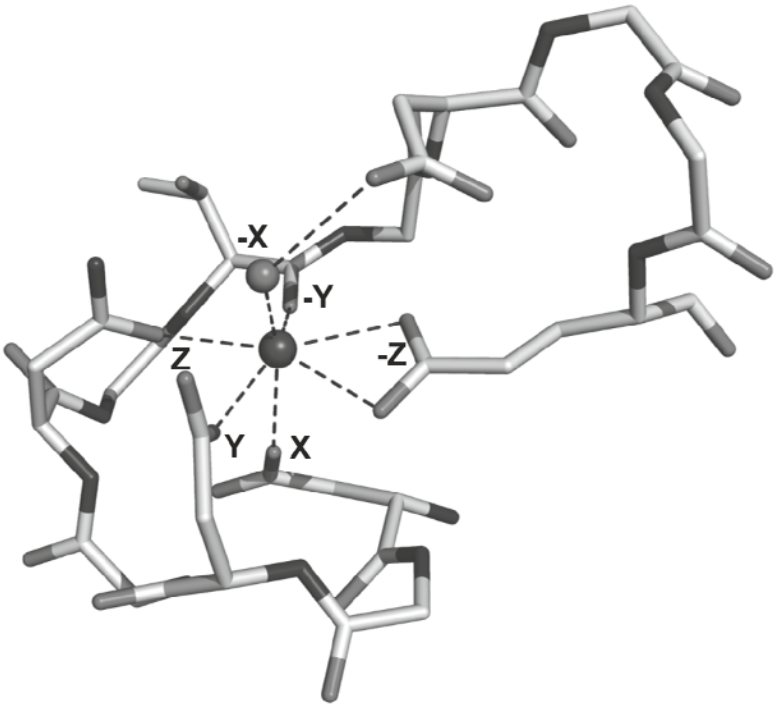


Figure 7. The Ca^{2+} co-ordination by the pentagonal bipyramid illustrated by the canonical EF-hand 1 of calmodulin (PDB ID 1exr). The Ca^{2+} ion is coordinated by side chain oxygen atoms at the X, Y, and Z vertices and by a backbone carbonyl at −Y. At the –X axis, a side chain oxygen coordinates the Ca^{2+} ion through a bridged water molecule and at the −Z axis, a side chain carboxylate makes a bidentate interaction with the Ca^{2+} ion.

The calcium binding affinities of EF-hand proteins range from K_d 10^{-4} to 10^{-9} M and have been shown to be highly dependent on the amino acid sequence in their EF-hand motif. Calcium binding to EF-hand motifs is most commonly connected to conformational changes that are essential for regulatory functions (Yap *et al.* 1999). Accordingly, EF-hand proteins are well adapted to respond to changes in calcium concentrations and act in signalling events and protein-protein interactions.

1.3.3 Combined deficiency of factor V and factor VIII

Combined deficiency of factor V and factor VIII (F5F8D) is an inherited bleeding disorder, which was first described by Oeri *et al.* in 1954. The disorder is characterised by reduced plasma concentrations of blood coagulation factors V and VIII, with both typically in the region of 5–30% of normal levels. F5F8D affects males and females in equal numbers, but is estimated to be extremely rare in the general population (1:1,000,000) (Mannucci *et al.* 2004). Only approximately 150 patients in 90 families have been diagnosed with F5F8D, with half of the families originating from the Mediterranean region (Zhang *et al.* 2008). The high frequency of F5F8D in some of the communities in this region can partly be explained by a high incidence of consanguineous marriages.

In general the symptoms in F5F8D patients are equivalent to those observed in patients with single FV or FVIII deficiencies. Affected patients present with a mild to moderate bleeding tendency with commonly observed mild bleeding symptoms such as easy bruising, gum bleeding and nosebleeds. Excessive bleeding episodes are particularly common after or during, dental extraction, surgery, labour or trauma. Due to these relatively mild bleeding symptoms and similarity in bleeding manifestation to single factor deficiencies, F5F8D may be significantly under-diagnosed or misdiagnosed in countries with limited genetic and haematology infrastructure.

The genetics behind F5F8D was not understood until 1998 when Nichols *et al.* by using positional cloning localised disease-causing mutations to the *LMAN1* gene that encodes the protein ERGIC-53. Although ERGIC-53 was commonly used by cell biologists and had already been proposed to act as a transporter for glycoproteins, this finding was unexpected. However, no mutations could be detected in the *LMAN1* gene for approximately 30 % of the affected patients. The cause for the disorder in this sub-set of patients was later revealed by Zhang *et al.* in 2003, when they showed that mutations in the gene *MCFD2* gave rise to the same clinical phenotype as mutations in *LMAN1*. To date at least 48 different F5F8D causing mutations have been located in the *LMAN1* gene, all of which are null mutants except for two, one that disrupts the disulfide bond formation supposed to be required for proper oligomerisation and another that causes structural destabilisation of the CRD (Zhang *et al.* 2008, Yamada *et al.* 2009). In contrast to in *LMAN1*, the majority of the mutations in MCFD2 are missense mutations. Among the nineteen presently reported mutations in MCFD2, ten have been found to result in substitutions of single of amino acid residues.

A direct molecular interaction between the encoded proteins was revealed by experiments in COS1 cells, where ERGIC-53 and MCFD2 coimmuno-precipitated (Zhang *et al.* 2003). Further experiments have confirmed this initial finding and shown that ERGIC-53 and MCFD2 form a calcium dependent complex with 1:1 stoichiometry that specifically interacts with FV and FVIII and cycles between the ER and the ERGIC (Nyfeler *et al.* 2005, Zhang *et al.* 2006, Kawasaki *et al.* 2008).

FV and FVIII are two large plasma glycoproteins that function as essential cofactors for the proteolytic activation of proteases involved in the blood coagulation cascade (Mann *et al.* 1988, Kane & Davie 1988). These two highly homologous proteins share a common A1-A2-B-A3-C1-C2 domain organisation (figure 8) and undergo similar post-translational modifications, including signal peptide cleavage, formation of conserved disulfide bonds, and addition of different oligosaccharides (Camire & Bos 2009).

Figure 8. Domain organisation of FV and FVIII. Both proteins consist of three A domains, each of ~300 residues, two C domains, each of ~160 residues, and a ~900 residues heavily glycosylated B domain, which is cleaved off by thrombin-mediated proteolysis in the activated protein.

Although FV and FVIII are highly homologous, the amino acid sequence of their B-domains is not conserved. The B-domains of both FV and FVIII, however, share the common feature of being extensively glycosylated, containing 25 and 18 consensus sites for N-linked glycosylation, respectively. It has been shown that efficient transport of FVIII from the ER requires an intact glycosylated B-domain (Miao *et al.* 2004), implying that the interaction between the sugars on the heavily glycosylated B-domain and the CRD of ERGIC-53 are essential for efficient transport of FV and FVIII. A sugar dependent interaction has also been confirmed by mutations in the Ca^{2+}-ion and sugar binding sites of the CRD disrupting the FV and FVIII interaction (Zheng *et al.* 2010A).

The role of MCFD2 in the transport of FV and FVIII is more puzzling. The findings that unglycosylated FVIII can still interact with MCFD2 (Zhang *et al.* 2005), and that MCFD2 mutants, which fail to coimmuno-precipitate with ERGIC-53 retain their ability to interact with FVIII (Zheng *et al.* 2010A), suggests that MCFD2 interacts directly with FV and FVIII via a protein-protein interaction. A common structural motif for this interaction most likely does not involve the B-domain of FV and FVIII, but can be found in one of the other more conserved domains. Although the cargo receptor function of the ERGIC-53/MCFD2 complex today is well established, key questions still remain regarding the different roles of the proteins and how they recognise and interact with FV and FVIII.

2 AIM OF THE THESIS

The overall aim of this thesis project was to gain insights into how selectively mediated glycoprotein traffic in the secretory pathway is organised and controlled. In order to elucidate these processes, the ERGIC-53/MCFD2 glycoprotein transport receptor complex has been structurally characterised by X-ray crystallography and NMR spectroscopy.

The specific aims of the thesis project were:

1. To determine the three-dimensional structure of human MCFD2 and give insights into the observed Ca^{2+} ion dependent MCFD2 binding to ERGIC-53.
2. To determine the structural organisation of the ERGIC-53/MCFD2 glycoprotein transport receptor complex at a molecular level.
3. To facilitate understanding of the function and the mechanism by which mutations in MCFD2 cause F5F8D.

3 RESULTS AND DISCUSSION

3.1 NMR studies and biophysical characterisation of MCFD2 (Paper I)

Since the first link between MCFD2 and the secretory pathway was established (Zhang *et al.* 2003) it has attracted the interest of our group. Although clearly shown to be required for efficient transport of FV and FVIII, the precise role of MCFD2 in these processes was an enigma. It had been suggested that Ca^{2+} binding of MCFD2 was a prerequisite for formation of a functional ERGIC-53/MCFD2 transporter complex, but no structural evidence was available to prove this theory when the work described in paper I of this thesis was initiated. In order to facilitate understanding of the function of MCFD2 we initially aimed to determine the structure of human MCFD2 by X-ray crystallography. However, after extensive crystallisation efforts which did not produce diffracting crystals, NMR studies on MCFD2 was initiated in collaboration with Professor Torleif Härd at the Swedish NMR Centre at the University of Gothenburg.

3.1.1 NMR spectroscopy and structure determination

A construct encoding the complete mature sequence of human MCFD2, excluding the signal sequence, was used to produce native and isotope-labelled (^{15}N and $^{13}C/^{15}N$) protein in *E. coli* for the NMR experiments. Unexpectedly, all spectra collected in traditional calcium free NMR buffers showed the characteristics of an unfolded protein, with no evidence of secondary structure or a hydrophobic core. This is for instance apparent in the ^{15}N-HSQC spectrum of MCFD2, which shows poorly resolved amide resonances in the random coil region with ^{1}H chemical shifts clustered near 8.0 ppm (figure 9a). However, upon addition of calcium ions the spectrum of MCFD2 shows a marked difference with a new set of well-dispersed peaks appearing, indicative of both secondary and tertiary structure (figure 9b).

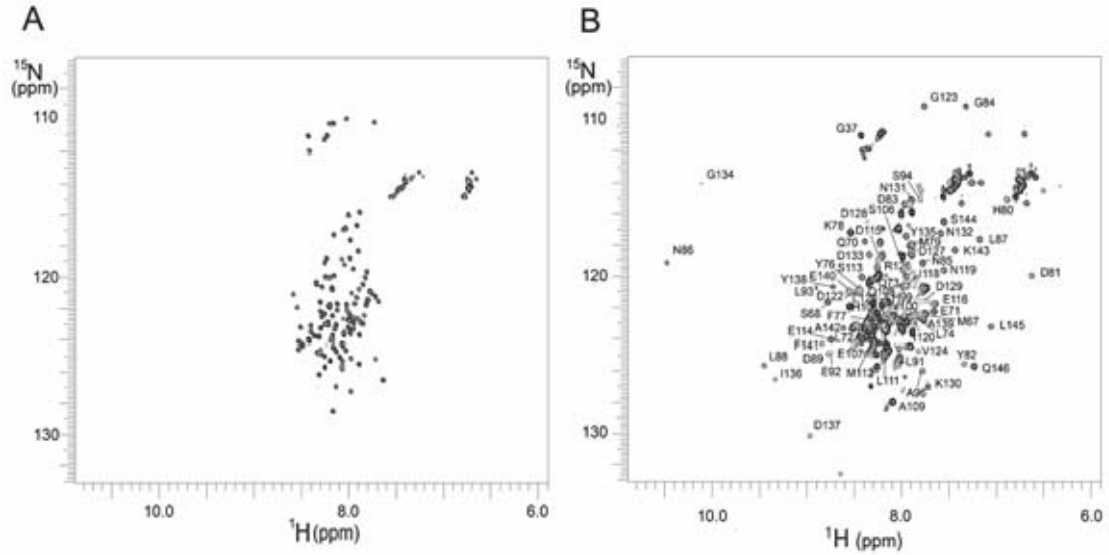


Figure 9. Changes in the ^{15}N-HSQC NMR spectrum of MCFD2 on addition of calcium. **(A)** Spectrum recorded in the absence of calcium and **(B)** after addition of 10 mM $CaCl_2$, with the assignments indicated for all backbone NH resonances in the ordered domain.

Due to the observed calcium-dependent folding of MCFD2, all NMR spectra used in the structure calculations were recorded in the presence of calcium ions. Even in the presence of calcium ions, MCFD2 is not completely folded and some disorder remains; this can be observed in the ^{15}N-HSQC spectrum as intense resonances clustered in or close to the random-coil region (figure 9b).

Several multidimensional heteronuclear NMR spectra were recorded to obtain the resonance assignments of ^{1}H, ^{13}C and ^{15}N nuclei for the structure calculation. Despite the observed disorder in the spectra, nearly complete backbone assignments were achieved with the exception of a limited set of residues in the disordered regions. In the ordered region of the structure almost complete side-chain resonance assignment was accomplished, but due to poorly resolved resonances several side-chains could not be assigned in the disordered regions.

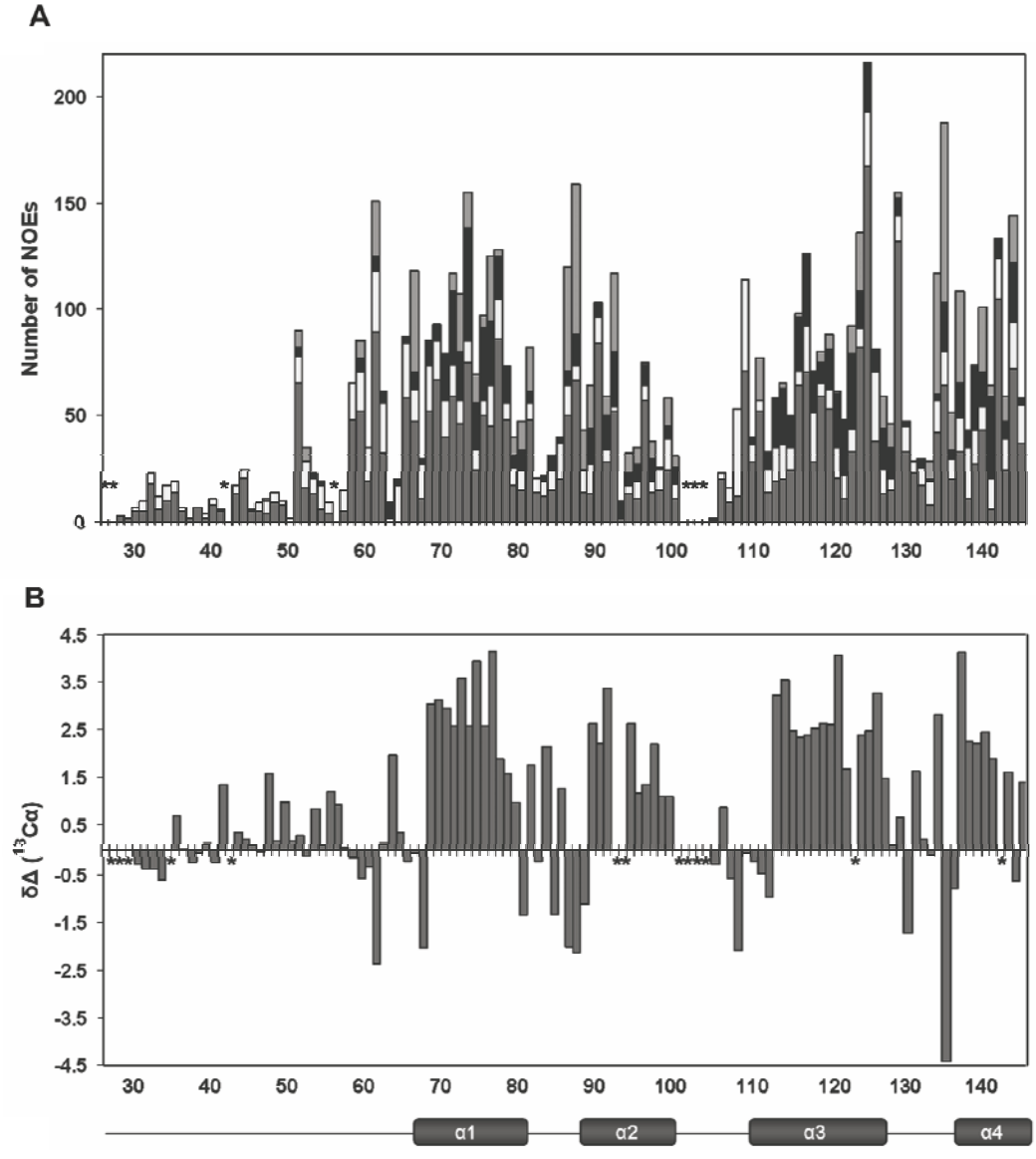


Figure 10. Structural order and disorder in MCFD2 in the calcium-bound state. (A) NOE restraint statistics—the number of NOEs assigned for each residue. Inter-residue NOEs are shown in dark grey, sequential NOEs in white, short nonsequential NOEs in black and long-range NOEs in light grey. (B) Deviation from random-coil chemical shifts (secondary ^{13}Cα chemical shift values; δΔ ^{13}Cα). Unassigned Cα resonances are denoted by *.

A total of 2142 NOE-derived distance restraints were assigned and from the distribution of these the disordered regions could be localised to the N-terminal residues 27–66 and residues 102–112 in a flexible loop region, both of which contain very few medium- or long-range NOE restraints (figure 10a). Providing further support for the NOE data analysis, the deviations from random-coil Cα chemical shift values are small in these regions, in contrast to those in the EF-hand motifs (figure 10b). Based on the assignments secondary structure elements were defined by the chemical shift indices (CSI) of Hα, Cα, Cβ and C′ using the program TALOS and confirmed with HN-HN and HN-Hα NOE connectivities. At the final stage the solution structure of MCFD2 was calculated by automated iterative NOE assignment and structure calculations by simulated annealing (SA).

3.1.2 Description of the overall structure

The final representative ensemble of structures of the complete protein and the ordered domain are shown in figure 11. The first ordered region with secondary structure in MCFD2 begins at residue 69 in helix 1. The structured part of MCFD2 contains four α-helices and two short anti-parallel β-strands. The overall fold is similar to other EF-hand proteins, with the two EF-hands packed against each other to form a single structural domain.

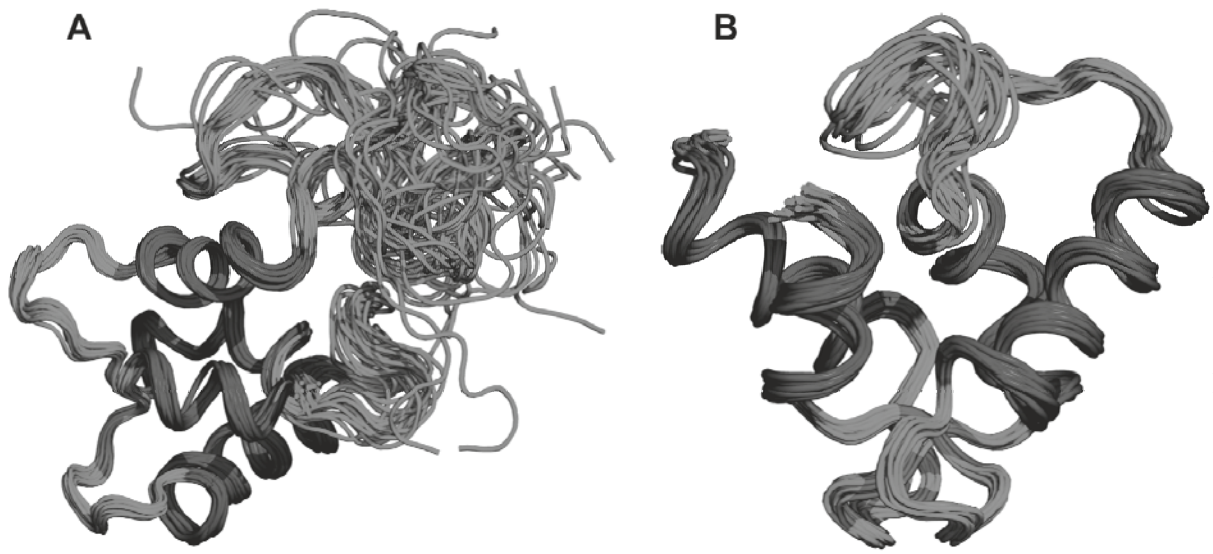


Figure 11. The ensemble of the 20 lowest energy SA structures of MCFD2. **(A)** the complete structure and **(B)** only the ordered region, residues 67–146 (PDB ID 2vrg).

Overall MCFD2 shows a high degree of structural similarity to other EF-hand proteins, with the C-terminal domain of calmodulin being most similar. However, despite this similarity there are differences with the most evident being the significantly longer loop connecting the two EF-hand motifs, and the second helix, which in MCFD2 is slightly bent and packed more tightly into the structure than in structural homologues (figure 12). This second helix is held in this bent conformation by interactions with residues within the C-terminal and the long loop region between the two EF hands. Interestingly, the residues at the C-terminal end of this

second helix and the following loop fragment show backbone amide resonance broadening, indicative of conformational interconversions on the microsecond to millisecond time scales. This suggests that the second helix and the proceeding loop can adopt two or more conformations. This finding was later confirmed in the crystal structure of MCFD2 in complex with ERGIC-53, which revealed significant conformational changes in this region upon complex formation.

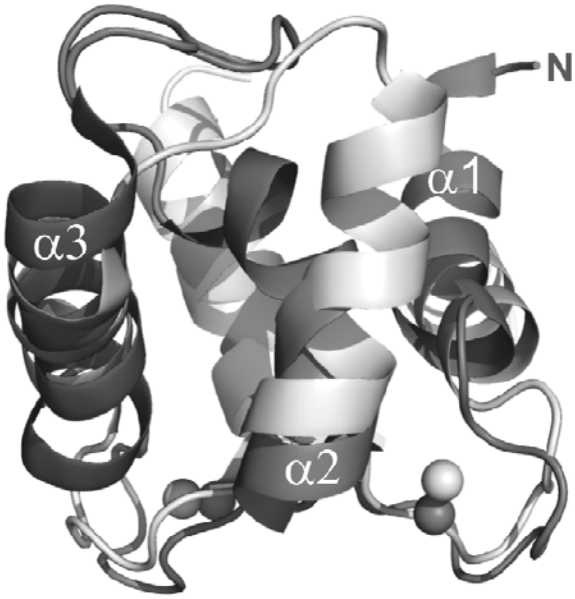


Figure 12. Superposition of the MCFD2 structure determined by NMR (dark grey) and a representative EF-hand domain, the structure of soybean calmodulin (light grey) (PDB2roa)

Another interesting region of MCFD2 is the disordered N-terminus, which is highly conserved within MCFD2 orthologues, but shows very low sequence identity to other proteins. It is unlikely that this conserved region plays no role in the function of MCFD2. Although speculative, a possible scenario is that this region could adopt a more ordered conformation upon binding the cargo proteins FV or FVIII and form part of an interaction surface.

3.1.3 Calcium induced folding of MCFD2

Circular Dichroism (CD) spectroscopy performed in the presence and absence of calcium ions confirmed the apparent calcium-dependent folding of MCFD2 seen in the NMR spectra. The far-UV CD spectrum of MCFD2 in the absence of calcium is typical of an unfolded disordered protein, as shown by the strong contribution from negative ellipticity around 200 nm (figure 13a). Although predominantly disordered, the weak negative ellipticity appearing as a shoulder around 220 nm indicates the presence of some residual secondary structure elements. The CD spectrum in the presence of calcium ions shows a marked difference, with a dominating contribution from α-helical secondary structure with minima at 207 and 222 nm and a change of sign from negative to positive at 197 nm. Collectively, the NMR and CD data suggest that in the absence of calcium MCFD2 has some transient helical structure but no hydrophobic core. Upon addition of calcium, the helical state becomes fully populated and folding of the protein occurs, with formation of the hydrophobic core.

The calcium-dependent folding was further studied by monitoring the CD signal at 222 nm as a function of calcium concentration (figure 13b). The negative ellipticity at this wavelength, which is primarily due to the contribution from α-helical secondary structure, increases sharply between 0 to 0.5 mM calcium and continues to increase up to a concentration of at least 1 mM. Interestingly, the calcium dependent folding profile of MCFD2 observed in this study is strikingly similar to the calcium dependence of the MCFD2/ERGIC-53 interaction shown *in vitro* by Kawasaki *et al.* Taken together, these data strongly imply that calcium binding and consequent folding of MCFD2 are a requirement for the formation of the MCFD2/ERGIC-53 complex to occur.

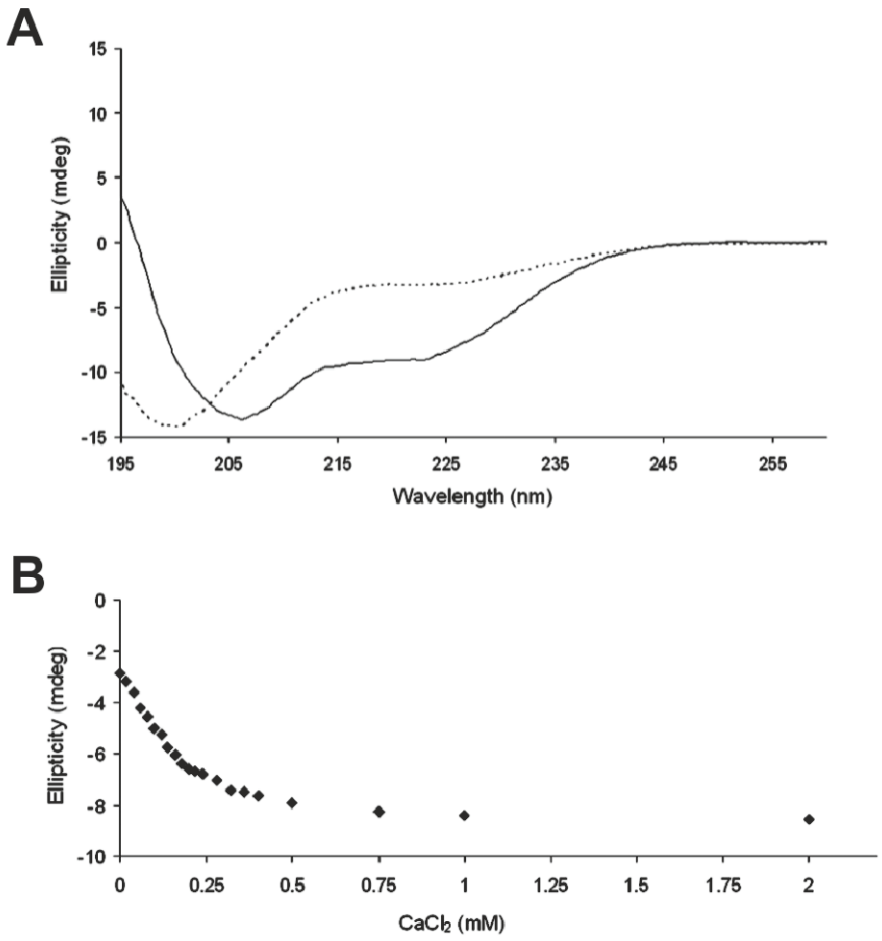


Figure 13. CD measurements of MCFD2. **(A)** Far-UV CD spectra of MCFD2 recorded in the absence of calcium (dotted line) and in the presence of 10 mM calcium (continuous line). **(B)** Calcium titration of the change in ellipticity at 222 nm.

In summary, we determined the first structure of MCFD2 and provided a detailed description of the protein. The results show MCFD2 to be disordered in the absence of calcium ions and to adopt a structure consisting of two EF-hand motifs on binding calcium ions. However, even in the presence of calcium ions the N-terminus remains disordered and there is large conformational flexibility in the extended loop connecting the two EF-hand motifs. Overall, these results gave a first insight into Ca^{2+} as an allosteric regulator of MCFD2 binding to ERGIC-53, by revealing the Ca^{2+} dependent mechanism behind the interaction**.**

3.2 The ERGIC-53/MCFD2 transport receptor complex (Paper II)

3.2.1 Interaction analysis and construct design

Although the cargo receptor function of the ERGIC-53/MCFD2 complex in the transport of FV and FVIII was already well established, the structural organisation of this complex remained elusive when the work described in paper II of this thesis was initiated. From cross-linking experiments in COS-1 cells, MCFD2 and ERGIC-53 were known to form a 1:1 Ca^{2+}-dependent complex (Zhang *et al.* 2006). We quantitatively measured this interaction *in vitro* using surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC). The results confirmed that MCFD2 specifically binds to ERGIC-53 in the presence of calcium, with a K_d in the lower nM range, and that the interaction is disrupted in the absence of calcium. SPR data also demonstrated that F5F8D causing mutant variants of MCFD2 have affinities for ERGIC-53 that are reduced by 2–3 orders of magnitude (table 2). These data are consistent with the findings of Kawasaki *et al.* that were published at the same time as our study. By performing SPR assays on constructs of different length of ERGIC-53 and MCFD2 we mapped the interaction between the proteins to the CRD of ERGIC-53 (ERGIC-53-CRD, residues 58–146) and the two EF-hand motifs at the C-terminus of MCFD2 (MCFD2-ΔN, residues 32–277). These two truncated protein variants showed an equally strong interaction, with a K_d value of 35 nM, as the longer constructs in the SPR measurements (table 2).

Table 2. SPR analysis of the interaction between ERGIC-53 and MCFD2.

A

MCFD2 variants	K_d (nM)
Wild-type	30
MCFD2-ΔN	35
D122V	5900
D129E	7700
I136T	6000
ΔSLQ	3300

B

ERGIC-53 variants	K_d (nM)
CRD	10
CRD-2H	35

Measurements were performed in the presence of 5 mM $CaCl_2$ on a Biacore 3000 instrument. **(A)** K_d values from analysis of MCFD2 variants, including wild-type MCFD2, MCFD2-ΔN and F5F8D causing mutant variants, injected over a sensor chip with immobilised ERGIC-53-CRD. **(B)** K_d values from analysis of immobilised wild-type MCFD2 on a sensor chip and injected variants of ERGIC-53, including the isolated CRD and the CRD and two proximal helices of the stalk domain.

On the basis of the SPR analysis, a protocol for *E. coli* expression and purification of the complex of MCFD2-ΔN and the ERGIC-53-CRD was established. The proteins were purified separately by nickel- and glutathione affinity chromatography, subsequently mixed and applied to a size exclusion chromatography (SEC) column. Analytical SEC, cross-linking experiments and mass spectrometry confirmed a homogeneous 1:1 complex between MCFD2-ΔN and ERGIC-53-CRD. Crystallisation trials of this complex initially yielded clusters of small plate like hexagonal crystals in a single condition. Further optimisation of this condition resulted in single crystals that were used for data collection.

3.2.2 Crystallisation and structure determination

The complex of MCFD2-ΔN and ERGIC-53-CRD crystallised in space group $P6_1$ with two molecules in the asymmetric unit. The thin plate-like morphology of the crystals, in combination with merohedral twinning and a long cell axis of 396.9 Å, complicated data collection and processing. Despite these complications, the structure of the complex could be solved by molecular replacement and refined to 2.45 Å resolution by using the CRD of rat ERGIC-53 as a search model (table 3). The electron density map is essentially complete from residue 41–274 in the CRD and, with the exception of a few solvent exposed side chains, all residues are well-defined. MCFD2 is less well ordered, and density is missing for the N- and C- terminal residues, as well as for the loop connecting the two EF-hand motifs, thus only residues 66–98 and 110–144 of MCFD2 could be modelled.

Table 3. Data collection and refinement.

Data collection	
Space group	$P6_1$
Unit cell	a = 56.6 Å, c = 396.9 Å
Resolution (Å)	2.45
Unique reflections	26465
Redundancy	11.8 (7.3)
R_{sym} (%)	12 (30)
I/σ	17.6 (5.8)
Completeness (%)	98.9 (85.6)
‹B› Wilson plot (Å^2)	43.0

Refinement and model building statistics	
Reflections in work set	25039 (1672)
Reflections in test set	1426 (87)
R_{factor}	19.6 (26.1)
R_{free}	24.7 (40.5)
Number of protein atoms	4695
Average B-factors (Å^2)	
ERGIC-53-CRD	23.6
MCFD2 -ΔN	27.8
Bond length R.M.S.D. (Å)	0.016
Bond angle R.M.S.D. (°)	1.47
Ramachandran plot	
Percentage in preferred regions	93.9
Percentage in disallowed regions	0.8
Twin fraction	
h, k, l	0.732
-h - k, k, - l 0	0.268

Numbers in parenthesis refer to highest resolution shell (2.51–2.45 Å).

In the crystal, two different packing interactions occur between the ERGIC-53-CRD and MCFD2-ΔN molecules: first a major interaction burying a surface area of 850 Å^2, and secondly an interaction, likely to be crystal contact, which buries a smaller surface area of 590 Å^2 and includes only two hydrogen bonds and one salt bridge. A structure of the ERGIC-53-CRD/MCFD2 complex determined from un-twinned crystals in another space group, which was published at the same time as ours (PDB ID 3a4u) (Nishio *et al.* 2009), confirmed the first of these to be the correct biologically relevant interaction. Superposition shows the two complex structures to be essentially the same, with the major interaction surface common to both. Recent *in vivo* experiments further confirmed the first interaction surface to be correct, by demonstrating that mutation of residues in this interaction surface abolishes the binding of MCFD2 to ERGIC-53 (Zheng *et al.* 2010A).

3.2.3 Overall structure of the complex

The CRD of ERGIC-53 comprises a globular β-sandwich fold consisting of a concave β-sheet and a convex β-sheet (figure 14). On the concave face of the CRD two calcium ions are bound and form the presumed carbohydrate ligand binding site. Overall the CRD in complex with MCFD2 is highly similar to the crystal structure of the uncomplexed ERGIC-53-CRD from rat and superposition of the two structures gives an R.M.S.D. of 0.32 Å for 227 Cα-atoms. The only minor structural changes that occur upon complex formation are at the interface with MCFD2, found on the side opposite to the carbohydrate recognition site. The location of the interface, and fact that only localised conformational changes occur in this region suggest that the cargo carbohydrate binding of ERGIC-53 is neither directly nor allosterically influenced by the interaction with MCFD2.

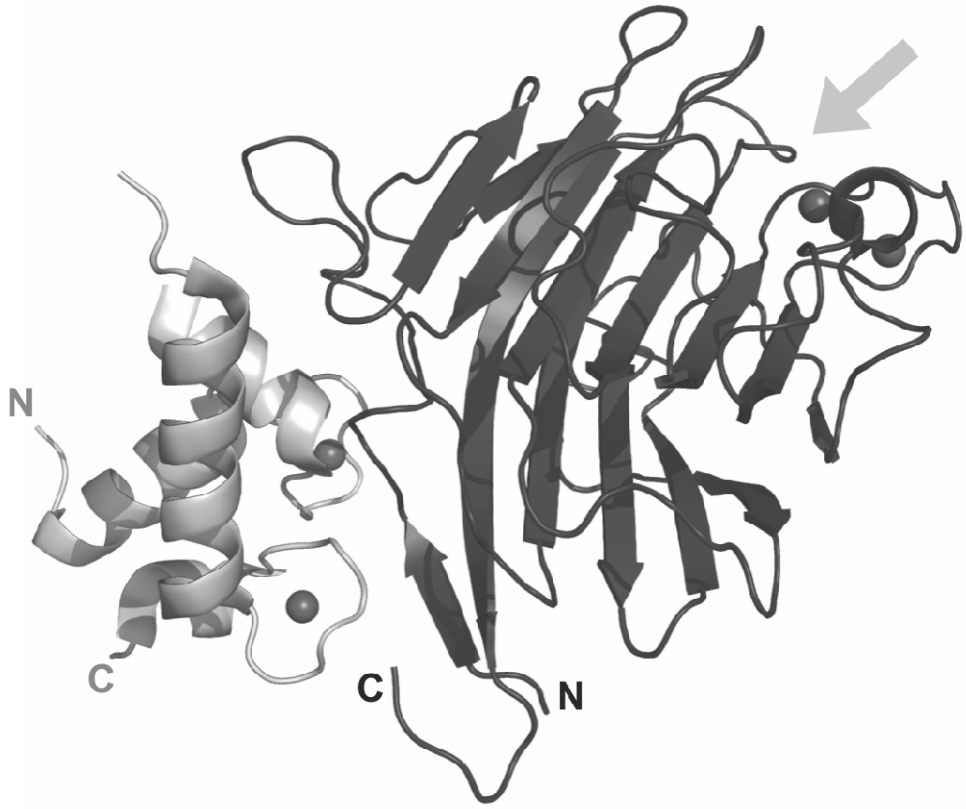


Figure 14. Structure of the complex. MCFD2 (grey) and the CRD (dark grey) (PDB ID 3lcp). The presumed carbohydrate binding site on the CRD is indicated with an arrow.

In contrast to the CRD, MCFD2 undergoes significant conformational changes upon complex formation. A comparison between MCFD2 in the complex and the lowest energy NMR structure of free MCFD2 gives an R.M.S.D. of 2.3 Å for 59 Cα-atoms. Unexpectedly, and in contrast to most EF-hand proteins which commonly interact with a hydrophobic cleft opposite to the Ca^{2+} binding EF-hand motif, MCFD2 binding to ERGIC-53 involves the two Ca^{2+}-binding loops. Overall, the organisation of the ERGIC-53-CRD/MCFD2-ΔN complex is unexpected and raises new questions on how the complex interacts with its cargo proteins FV and FVIII.

The previous findings that mutations in the sugar binding site of the CRD disrupt the binding to FV and FVIII, and that unglycosylated FVIII still can interact with MCFD2, suggest the existence of two separable interactions between the ERGIC-53/MCFD2 complex and the cargo proteins FV and FVIII; a glycosylation dependent interaction mediated by the ERGIC-53-CRD, and protein-protein interaction mediated by MCFD2. Interestingly, binding of the cargo protein simultaneously to the carbohydrate site on the CRD and to MCFD2 would require the cargo to reach over a distance of more than 45 Å on the CRD (figure 14). Considering that ERGIC-53 assembles into hexamers during cargo transport *in vivo*, a perhaps more likely scenario is that while the sugar moieties on a cargo protein binds to the CRD, the protein component of the cargo interacts with MCFD2 bound to a neighbouring CRD. Analysis of the packing interactions in the crystal does not reveal any assemblies corresponding to an oligomeric receptor complex. Future studies including the stalk domain of ERGIC-53, which have shown to mediate the oligomerisation *in vivo*, will be of high interest in order to reveal exactly how the ERGIC-53/MCFD2 complex interacts with FV and FVIII.

3.2.4 The interface between ERGIC-53 and MCFD2

The interface between the ERGIC-53-CRD and MCFD2 buries a surface area of 830 $Å^2$, which is surprisingly small considering that it mediates an interaction with a K_d value of ~35 nM. The interaction involves a combination of hydrophobic and charge interactions, including six hydrogen bonds, three ionic interactions and a few water-mediated hydrogen bonds (figure 15). The interaction area on the CRD is limited to one region and mainly involves residues from the first β1 strand and the following loop (H43-F66). In contrast, 17 % of the surface area of the ordered part of MCFD2 is buried upon complex formation with residues from both of the Ca^{2+}-binding loops and helix 3 mediating interactions to the CRD.

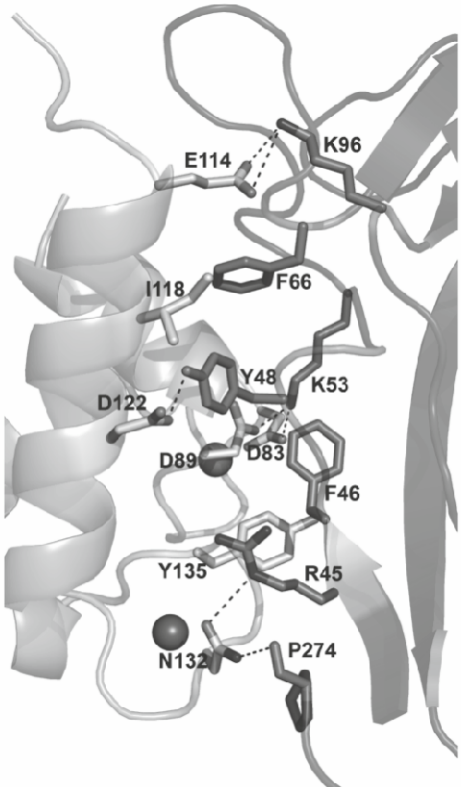


Figure 15. The interface between MCFD2 (grey) and the CRD (dark grey). Hydrogen bonds are shown as dashed lines and interacting residues are shown as sticks and labelled.

Interestingly, the interaction surface involves a number of previously reported F5F8D-causing missense mutations, and several of these contribute directly to the interaction (D89A, D89N, D122V, D129A and Y135N). For instance D89, positioned in the first Ca^{2+} binding loop, forms a salt bridge to K53 in the CRD upon complex formation. D122, positioned in the third helix which is tilted towards the CRD in the complex, forms a hydrogen bond with the side chain of Y48 on the CRD. From the second Ca^{2+} binding loop, D129 forms a water mediated bond to the side chain of R45 and Y135 packs against residue F46 in the CRD. Additional residues in MCFD2 which contribute to the interaction include N132, which forms a hydrogen bond with the carbonyl of P274, and E114, at the end of the third helix, that forms a salt bridge with K96 in the CRD.

3.2.5 Structural changes in MCFD2 upon complex formation

The major conformational changes which occur in MCFD2 upon complex formation are located outside of the Ca^{2+} binding EF-hand motif (figure 16). In comparison to the NMR structure of free MCFD2, the third helix is tilted towards the CRD by approximately 20° in the complex. Furthermore, the second helix, which is slightly bent in the NMR structure of MCFD2, is straightened out in the complex. The loop following this helix, which was shown to be highly flexible in the NMR structure, is not visible in the electron density, indicating that this region remains flexible in the complex. Electron density is also missing for the last three C-terminal residues of MCFD2 in the complex. These residues are part of the core structure and interact with bent second helix in the NMR structure of MCFD2, thus the missing electron density implies that these intramolecular interactions are lost upon complex formation.

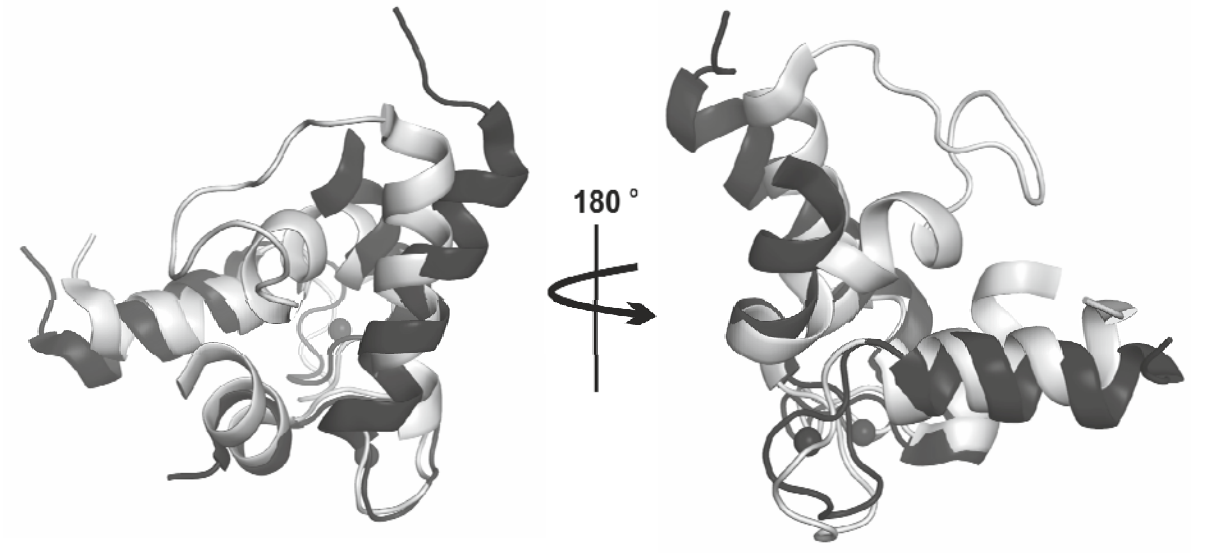


Figure 16. Superposition of the MCFD2 structure determined by NMR (light grey) and MCFD2 in the complex (dark grey).

Collectively, these conformational changes result in a more open conformation, which exposes the hydrophobic core of MCFD2 on the opposite side of the Ca^{2+}-binding EF-hand motif (figure 16). A common feature of EF-hand proteins is that Ca^{2+} binding induces the exposure of the hydrophobic core and provides the interaction site for target molecules (Yap *et al.* 1999). It is plausible that the hydrophobic surface area exposed on MCFD2 upon complex formation with ERGIC-53 could act in an analogous fashion, and provide an interaction surface for the cargo proteins FV and FVIII. Interestingly, several residues in the Ca^{2+} binding EF-hand motif interact with residues in the CRD in the complex. One could imagine a mechanism where calcium levels regulate the interaction between these residues, and thus the exposure and closure of the hydrophobic core. However, the exact mechanism for such a regulatory process involving the hydrophobic interaction surface on MCFD2 in the transport of FV and FVIII is yet to be proven.

In summary, the structure of the ERGIC-53-CRD/MCFD2-ΔN complex provided the first molecular view of the organisation of a cargo receptor complex in the secretory pathway. Overall the organisation of the complex was unexpected and the structure revealed that MCFD2 undergoes significant conformational changes upon complex formation, while only localised changes occur in ERGIC-53.

3.3 Characterisation of F5F8D causing mutations in MCFD2 (Papers I, II and III)

To date, nineteen mutations in MCFD2 have been reported in patients with F5F8D. Among these, ten are deletion or splicing mutations predicted to result in truncated protein products, including one which creates a stop codon at position 122, effectively deleting the last three amino acids of the protein (ΔSLQ) (Nyfeler et al. 2008). The additional nine are missense mutations, all of which alter amino acid residues located within the EF-hand domain of MCFD2. All of these missense mutations have been shown to result in reduction of plasma levels of FV and FVIII with both typically in the range of 5–15 % of normal levels (table 4). Five of these mutations have been expressed in COS-1 cells and were found to severely impede the binding of MCFD2 to ERGIC-53 (D81Y, D89A, D122V, D129E and I136T) (Zhang *et al.* 2003, 2008). Furthermore, in these studies the expressed MCFD2 mutant variants localised to the ER, whereas wild-type MCFD2 co-localised with ERGIC-53 in the ERGIC, thus confirming both the impaired interaction between the proteins, and that MCFD2 is reliant on ERGIC-53 for its proper intracellular localisation (Zhang *et al.*, 2003, 2008). *In vitro* surface plasmon resonance analysis of recombinant proteins has demonstrated the MCFD2 mutants, D122V, D129E and I136T, to have affinities for the CRD of ERGIC-53 that are reduced by .2–3 orders of magnitude (Kawasaki *et al.* 2008, own unpublished data). Although extensively studied, the molecular mechanism by which mutations in MCFD2 cause F5F8D was poorly understood when the work described in papers I, II and III of this thesis was initiated.

Table 4. F5F8D patients with reported missense mutations in MCFD2, their geographic origin, sex and FV and FVIII levels in % compared to normal values.

Origin	Sex	FV	FVIII	Mutation	Reference
Tunisia	M	8	5	p.D81H	*(Abdallah 2010)*
Tunisia	M	9	7	p.D81H	"
Saudi Arabia	F	8	6	p.D81Y	*(Zhang 2008)*
Saudi Arabia	F	13	12	p.D81Y	"
Tunisia	F	10	19	p.D89N	*(Elmahmoudi 2011)*
Greece	F	7	5	p.D89A	*(Zhang 2006)*
India	F	13	14	p.D122V	*(Jayandharan 2007)*
Venezuela	M	9	7	p.D129E	*(Zhang 2003)*
Poland	F	11	17	p.Y135N	*(Ivaskevicius 2008)*
Poland	M	13	38	p.Y135N	"
Kosovo	F	4,5	1,5	p.I136T	*(Zhang 2003)*
Venezuela	M	9,4	9	p.I136T	"
Venezuela	F	4,3	7,4	p.I136T	"
Venezuela	F	6,2	8,2	p.I136T	"
S. America	F	11,5	12	p. ΔSLQ	*(Nyfeler 2008)*

*included in table the nonsense mutation ΔSLQ characterised in this study

3.3.1 NMR- and CD-spectroscopy analysis

NMR and CD spectroscopy analysis of two disease causing mutations, D129E and I136T, revealed that both these mutations cause structural destabilisation and prevent folding of MCFD2 (paper I). The ^{15}N-HSQC spectra of these mutant variants demonstrated profound changes compared to the wild-type spectrum of MCFD2. Even though recorded in the presence of calcium, the spectra of each mutant show poorly resolved resonances and resemble the spectrum of the unfolded apo form of MCFD2 (figure 17). CD spectroscopy analysis further confirmed the lower degree of folding, with each mutant resulting in spectrum profiles reminiscent of the calcium-free unfolded form of MCFD2 (figure 17d).

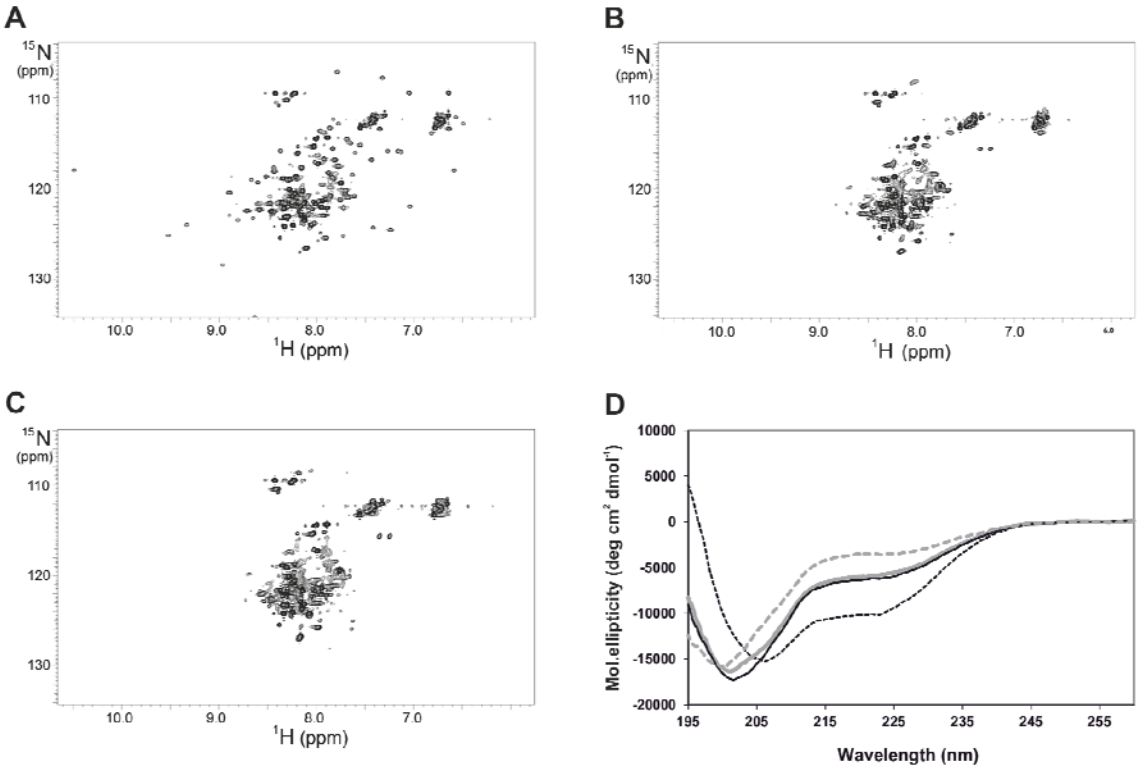


Figure 17. ^{15}N-HSQC NMR spectra recorded in the presence of calcium of the mutant variants. **(A)** wild-type MCFD2, **(B)** D129E and **(C)** I136T. **(D)** The CD spectra of the mutants D129E (grey continuous line) and I136T (black continuous line) recorded in the presence of calcium together with wild-type MCFD2 recorded in the presence (broken black line) and absence of calcium (broken grey line).

The results of this study provided the foundation for our subsequent analysis of F5F8D causing mutations reported in MCFD2 (Paper II and III). In these studies the CD-signal at 222 nm, which is primarily due to contribution from α-helical secondary structure, was monitored as a function of temperature in order to characterise mutants with CD spectra resembling that of the Ca^{2+}-bound form of wild-type MCFD2. The obtained thermal unfolding profiles demonstrated a decrease in secondary structure stability, and yielded lower melting temperatures (T_m) compared to wild-type MCFD2, for all except one of these mutants (figure 18). From this data, and our previous structural characterisation of ERGIC-53 and MCFD2, the consequences of all known F5F8D causing missense mutations found in MCFD2 could be explained at a molecular level.

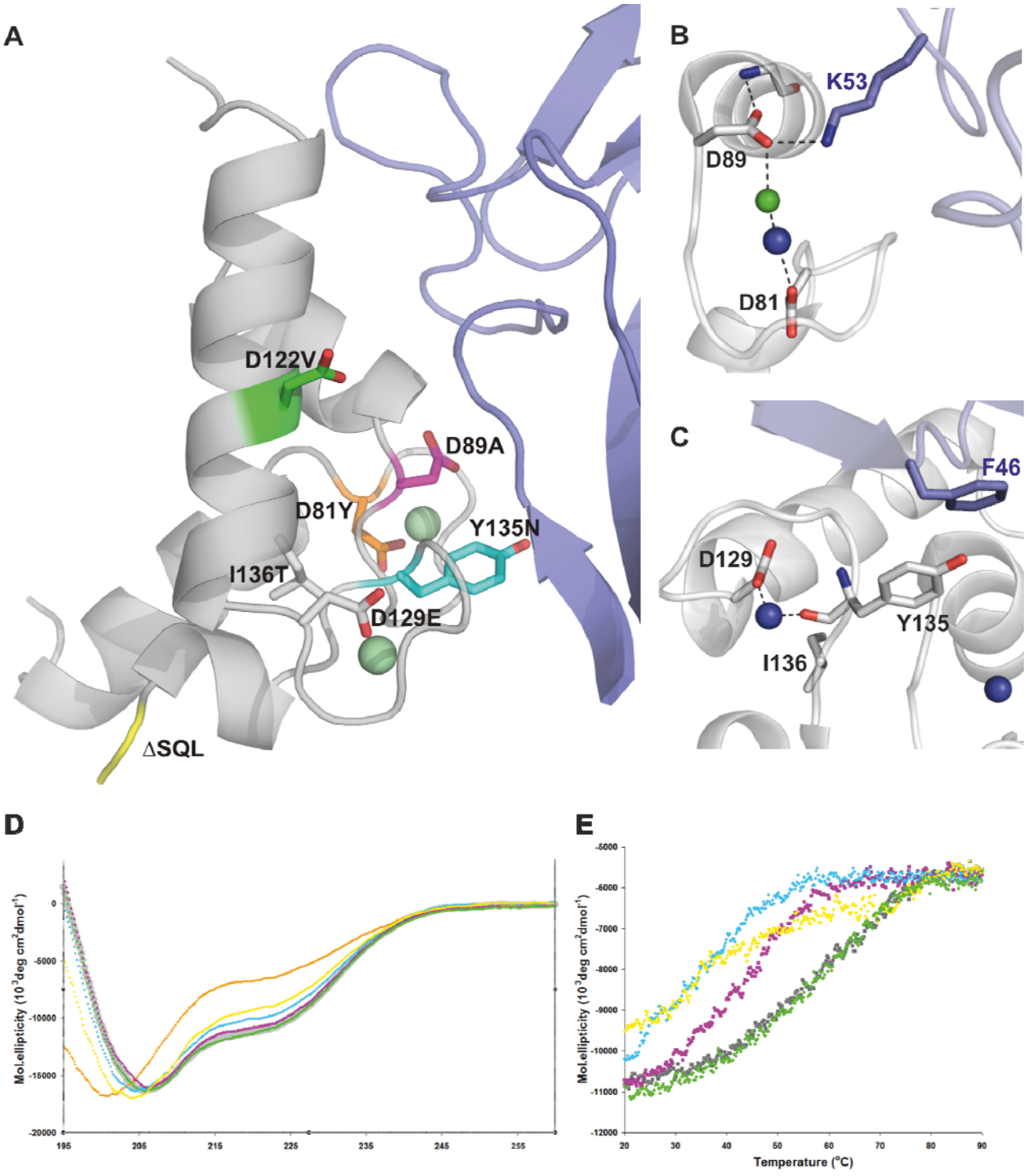


Figure 18. (A) The interface between MCFD2 (grey) and the CRD (blue) showing the positions of the F5F8D-causing mutations on MCFD2. The positions of the mutations analysed by CD in paper 2 shown as sticks and coloured in orange (D81Y), pink (D89A), green (D122V), cyan (Y135N) and yellow (DSLQ). Mutations analysed by NMR in paper 1 shown in grey (D129E and I136T). Close up view of the first **(B)** and second **(C)** Ca^{2+}-binding motif with positions of the mutated residues shown as sticks in grey and with interacting residues in ERGIC-53 in light blue. **(D)** Far-UV CD spectra of wild-type MCFD2 and F5F8D mutant variants recorded in presence of calcium. Mutant data are shown with the same colour coding as in (A) and with folded wild type in grey. **(E)** Thermal unfolding experiments performed monitoring the CD signal at 222 nm, colour coding as in the Far-UV CD spectra.

Table 5. T_m values of MCFD2 variants.

MCFD2	T_m (ºC)
Wild-type	59 (+/-0.2)
D81H	n.d
D81Y	n.d
D89A	43 (+/-0.2)
D89N	45 (+/-0.1)
D129E	n.d
I136T	n.d
Y135N	34 (+/-0.3)
V100D	44 (+/-0.1)
D122V	58 (+/-0.2)
dSLQ	34 (+/-0.6)

3.3.2 Mutations in the first Ca^{2+}-binding motif - D81Y, D81H, D89A and D89N

The CD spectra of the mutants D81Y and D81H both resemble the disordered apo state of wild-type MCFD2, indicating that these two mutants have lost the ability to bind Ca^{2+}-ions and are essentially unfolded (figure 18d). D81 is the first residue of the EF-loop consensus sequence and binds Ca^{2+} through a side-chain oxygen atom. The conservation of aspartate at this position is 100% and this residue forms several intra-loop hydrogen bonds, contributing to the precise arrangement of the loop (Gifford *et al.* 2007) (figure 18b). In light of this, it is not unexpected that substitution of this amino acid prevents correct formation of the EF hand motif and by this means disrupts the core structure of the protein.

In contrast, the CD spectra of the mutants D89A and D89N are highly similar to the spectrum of the calcium bound form of wild-type MCFD2 (figure 18d). However, thermal unfolding profiles demonstrate a decrease in structural stability and yield melting temperatures ~15 °C lower than that of wild-type MCFD2 for both mutants (figure 18e, table 5). The residue D89 occupies the ninth position in the EF-hand loop and contributes to the arrangement of the loop by forming a hydrogen bond to a water ligand of the Ca^{2+} ion to complete the pentagonal bipyramid (figure 18b). Moreover, the carboxyl group of the side chain of D89 caps the N-terminus of the helix following the EF-hand loop. The mutations D89A and D89N result in loss of these two interactions and give rise to structural destabilisation demonstrated by their lower T_m values. In addition, upon complex formation D89 forms a salt bridge with K53 in ERGIC-53, and substitution of this residue most likely disrupts this interaction and results in attenuation of binding affinity.

3.3.3 Mutations in the second Ca^{2+}-binding motif - D129E, Y135N, I136T

NMR and CD spectroscopy demonstrate that the mutations D129E and I136T abolish secondary and tertiary structure in MCFD2 (figure 17). D129 occupies the first highly conserved position of the EF-loop and, as for D81 in the first EF-hand, substitution of this amino acid prevents binding of Ca^{2+} ions and thereby disrupts the core structure of the protein (figure 18c).

I136 occupies the eighth position of the EF-hand loop and does not directly bind calcium, but is situated between two Ca^{2+} ligands. The exact identity of the residue at this position is not conserved, but it is always hydrophobic (Gifford *et al.* 2007). In wild-type MCFD2 the side chain of I136 faces away from the Ca^{2+} ion binding site, and is instead packed with other hydrophobic side chains in the small hydrophobic core of MCFD2. Substitution of this isoleucine to a polar threonine most likely disturbs packing of this small hydrophobic core, and thereby results in a disordered protein.

The observed decrease in secondary structure stability for the Y135N mutation, demonstrated by the significantly lower T_m value of 34 °C compared to 59° C for wild-type MCFD2, is presumably due to the cooperative disturbance of two interactions. This tyrosine occupies the seventh position of the EF-hand loop and binds Ca^{2+} ion through its main-chain oxygen (figure 18c). In addition, the aromatic ring of this residue is involved in hydrophobic interactions with L87 in the first EF-hand. This hydrophobic interaction is most likely disrupted by the Y135N substitution, and as a consequence the specific conformation of Y135 in the Ca^{2+}-binding loop is distorted, resulting in breakage of the bond between the main-chain oxygen and the Ca^{2+} ion. Thus, the loss of these interactions most likely leads to destabilisation of the protein. Furthermore, in the complex of the ERGIC-53-CRD and MCFD2, Y135 packs against F46 in ERGIC-53 through interactions between their aromatic rings and the Y135N substitution therefore weakens this interaction.

Mutations outside the Ca^{2+}-binding motifs - ΔSLQ, V100D and D122V

The observed T_m value of 34 °C for the nonsense mutation that results in deletion of the last three amino acids of MCFD2 (ΔSLQ) suggests that this C-terminal truncation causes severe destabilisation of MCFD2. These deleted residues are part of the core of the MCFD2 structure and NMR data show that they are involved in a large number of long-range NOE interactions. It is therefore not surprising that their removal has severe consequences for the integrity of the structure of MCFD2.

The mutation V100D is not a confirmed disease-causing mutation, but is found in a heterozygous state together with D81H in a homozygote state. The thermal melting profile for V100D, yielding a T_m value of 44 °C, demonstrates that this mutation causes destabilisation of MCFD2. In the NMR structure of un-complexed MCFD2 the side chain of V100 packs tightly with other hydrophobic residues and contributes to the hydrophobic core. Substitution of this valine with a charged aspartate probably disrupts the packing of the hydrophobic core and has a destabilising effect. Hence, this mutation cannot be ruled out as cause of F5F8D, although the mutation D81H, which leads to an essentially unfolded protein, is most likely the main cause in these patients.

In contrast to the other F5F8D causing mutations that result in a disordered or destabilised protein, CD analysis of the mutation D122V results in a spectrum and T_m value identical to wild-type MCFD2 (figure 18), thus implying a fold and stability characteristic of the wild-type protein. Positioned in the middle of the third α-helix, which is tilted towards ERGIC-53 upon complex formation, D122 contributes strongly to the binding of ERGIC-53 by an ionic interaction with K49 and a hydrogen bond to Y48 on ERGIC-53. In light of this, it is

not surprising that substitution of this aspartate with a valine severely affects binding to ERGIC-53 and as a consequence causes F5F8D.

In summary, the results from these studies clearly emphasise the importance of intact calcium binding EF-hand motifs for the structural stability of MCFD2. This is demonstrated by the fact that all, except one, of the missense mutations reported to date are localised within these motifs, and that they give rise to a disordered or severely destabilised MCFD2 protein. None of the characterised mutations can be directly implicated in cargo binding of FV and FVIII. Instead our results suggest disruption of the ERGIC-53/MCFD2 interaction as the underlying mechanism for these mutations in causing F5F8D.

3.4 Oligomerisation properties of ERGIC-53

The influence of oligomerisation on ER export efficiency has been studied in detail for ERGIC-53 (Lahtinen *et al.* 1999, Nufer *et al.* 2003, Neve *et al.* 2005). From these studies a mechanism has been proposed in which lumenal and transmembrane determinants mediate the oligomerisation of ERGIC-53 required for its recruitment into budding COPII vesicles *via* the C-terminal di-hydrophobic motif. In these events two membrane proximal cysteines (C466 and C475) have been found to be important in the formation of disulfide-linked oligomers of ERGIC-53. Recent data have confirmed their significance and suggested that ERGIC-53 can exist in two homo-hexameric forms: the first composed of a covalent disulfide-linked hexamer and the second consisting of three disulfide-linked dimers assembling into a non-covalent complex (Neve *et al.* 2005). Moreover, it has been shown that the transmembrane domain mediates hexamerisation of ERGIC-53 and this has been proposed to be a prerequisite for the oligomerisation (Nufer *et al.* 2003). Recent data has also indicated the presence of a third structural determinant, within the four predicted α-helices of the stalk domain, to be involved in and promote oligomerisation (Neve *et al.* 2005).

In order to investigate the oligomerisation properties of ERGIC-53 within a structural context protein constructs coding for the full-length lumenal part of the protein, the CRD with alternative end points of the stalk domain and the isolated stalk domain were generated and expressed in *E. coli*. Two of these constructs, the CRD including two predicted α-helices of the stalk domain (CRD-2H) and the isolated stalk domain (figure 19), could be further characterised by size-exclusion chromatography (SEC), CD spectroscopy and polyacrylamide gel electrophoresis (PAGE). Characterisations of the other constructs were hampered by low yield and low solubility.

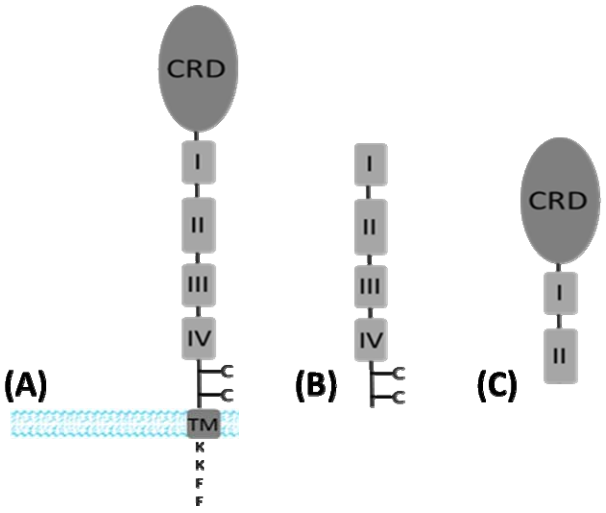


Figure 19. (A) Schematic representation of the structural organisation of ERGIC-53, the carbohydrate recognition domain indicated as an oval (CRD), the predicted helices (I, II, III and IV of the stalk domain) and the transmembrane domain (TM) as squares. The two conserved cysteine residues (C466 and C475) and the transport determinants of the cytoplasmic domain are highlighted. Truncated constructs used in the study, **(B)** the isolated stalk domain, and **(C)** CRD-2H, the CRD and the two proximal helices.

3.4.1 Evidence for disulfide-bond independent oligomerisation

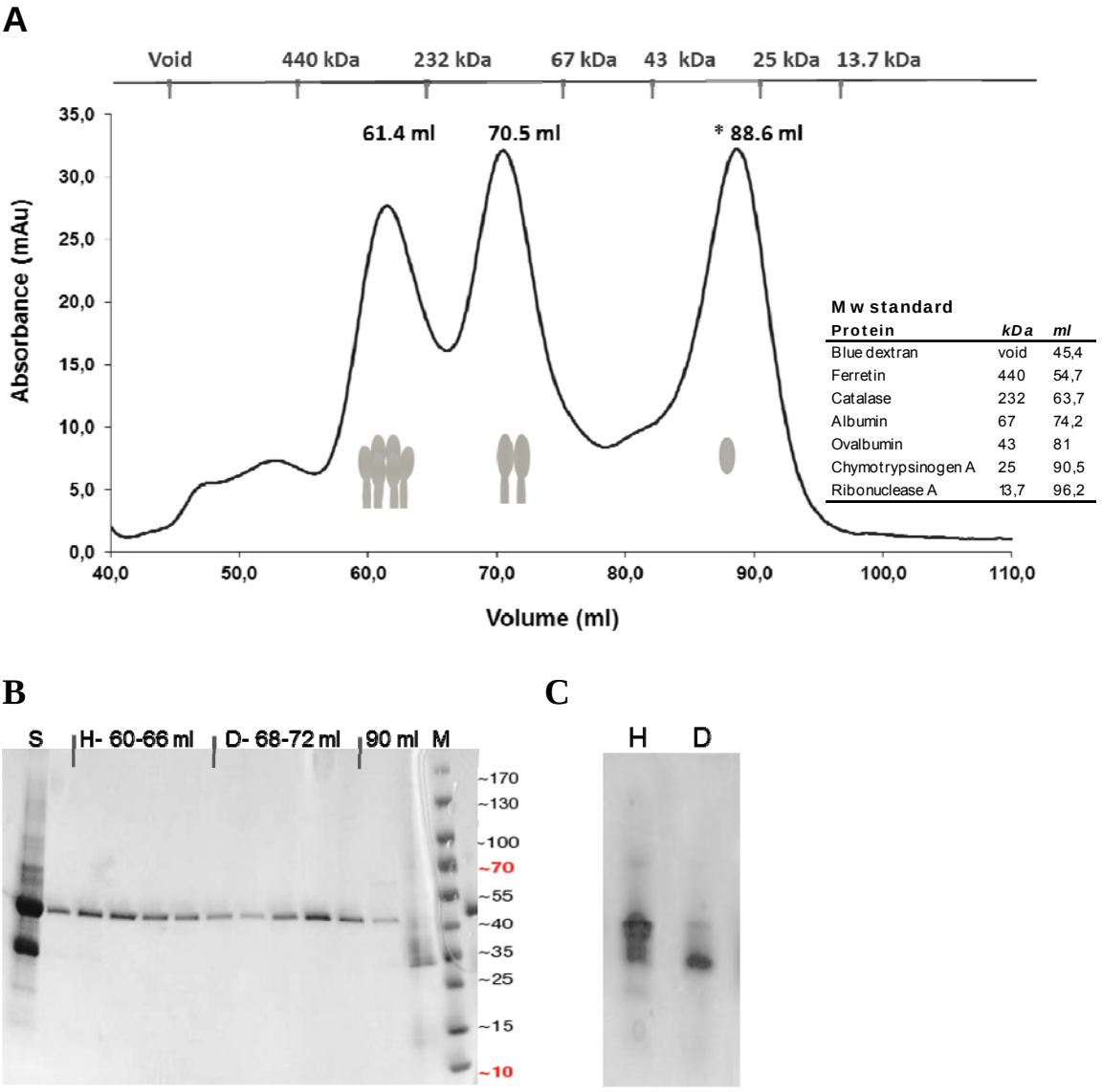


Figure 20. (A) The oligomeric states of the construct CRD-2H (Mw 43279.2 Da) analysed by SEC on a Superdex 200 10/300 column. Apparent elution volume, molecular weights and classification of eluted oligomers have been indicated. **(B)** SDS-PAGE gel analysis of SEC, stained by Coomassie blue with molecular weight of marker indicated to the right. Lane 1, S- sample before SEC, followed by analysed fractions, H-hexameric CRD-2H, 60-66 ml , D-dimeric CRD-2H, 68-72 ml and *CRD, 90 ml. **(C)** Native PAGE gel analysis of pooled fractions corresponding to H-hexameric and D-dimeric CRD-2H.
* CRD of ERGIC-53 (~25 kDa) - proteolytic degradation product from expression in *E. coli.*

During SEC the CRD-2H variant eluted as two distinct oligomeric species with apparent molecular masses of approximately 250 kDa and 80 kDa (figure 20a), which correspond to the hexameric and the dimeric states of CRD-2H. Native PAGE analysis confirmed the presence of two oligomeric species and also indicated some dynamic exchange between the two species (figure 20c). Thus, although missing the known determinants contributing to the oligomerisation, the transmembrane region and the two cysteine residues involved in disulfide bond formation, assembly of oligomers was apparent.

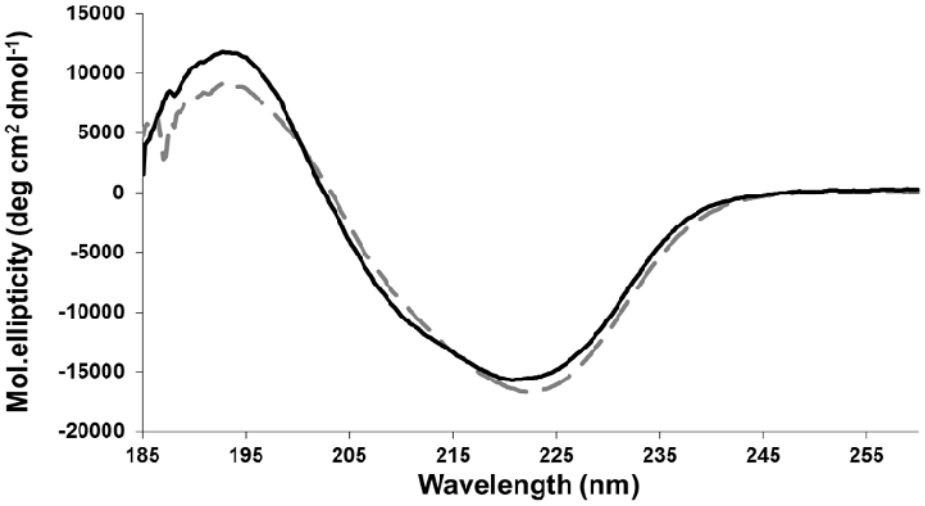


Figure 21. Far-UV CD spectra of the SEC elution fractions of CRD-2H corresponding to the dimeric state (continuous line) and the hexameric state (dotted grey line).

Since an unfolded protein is expected to migrate at a higher molecular weight during SEC and native PAGE we wanted to ensure that the two oligomeric species of CRD-2H were properly folded. To address this issue far-UV CD spectroscopy was performed on the eluted fractions from SEC corresponding to the two oligomeric species. The CD spectra confirmed both oligomeric species to be properly folded, showing contribution predominantly from β-sheet secondary structure, but with evident contribution from α-helical secondary structure (figure 21). Although clearly folded the two oligomeric species show slightly different spectrum profiles, indicating differences in secondary structure composition between the dimeric and hexameric species.

3.4.2 Oligomerisation properties of the isolated stalk domain

In addition, the oligomerisation properties of the isolated stalk domain were investigated by characterising a construct containing the four putative helices and the two membrane proximal cysteins of this domain. The CD spectrum of this construct showed a predominant contribution from α-helical secondary structure, confirming previous predictions (figure 22a). Glutaraldehyde cross-linking under non-reducing conditions in conjunction with SDS-PAGE analysis of this construct revealed the formation of a higher molecular weight species corresponding to a hexameric complex (figure 22c). However, upon SDS-PAGE analysis of non-cross-linked protein, this complex dissociates into dimers and under reducing conditions into monomers (figure 22b). Mass spectrometry analysis further confirmed the presence of a complex consisting of reduction-sensitive disulfide linked dimers by detecting the mass corresponding to the dimer (47699.4 Da) (figure 22d) under non reducing conditions and after reduction with DTT the mass of a monomer (23853.0 Da) (figure 22e).

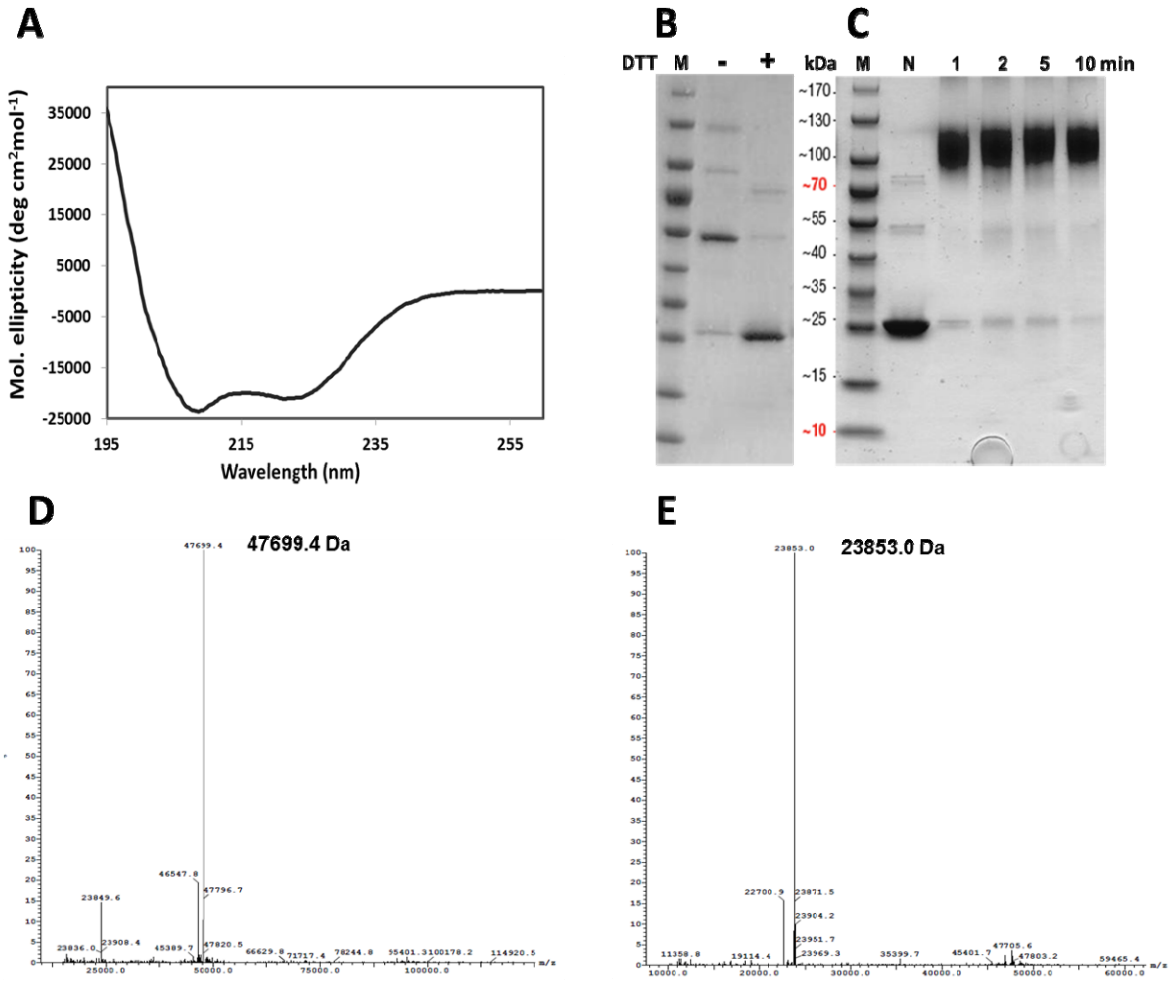


Figure 22. Characterisation of the ERGIC-53 stalk domain (Mw 23853.2). **(A)** Far-UV CD spectra collected under native conditions. SDS-PAGE gels, stained by Coomassie blue with molecular weight of marker indicated. **(B)** lane 1, Marker, lane 2 native stalk domain, and lane 3, after reduction with DTT. **(C)** Glutaraldehyde cross-linking performed under native condition: lane1, marker, lane 2, before addition of glutaraldehyde, followed by lanes 3-5 representing samples collected at different time points after addition of glutaraldehyde. Mass spectrometry analysis of the ERGIC-53 stalk domain, **(D)** before and **(E),** after reduction with DTT (The detected masses are within 3 Da of the expected values).

This data is consistent with previous findings which have shown that one of the oligomeric species of ERGIC-53 exists as SDS-sensitive complex of disulfide-linked dimers. However, the reported complex consisting of a covalently disulfide-linked hexamer shown to exist *in vivo* on studies performed on wild-type ERGIC-53, was not detected in this study. This suggests that the lumenal stalk domain alone is not able to form this complex, or that expression in *E. coli* does not enable the formation of the disulfide-linked hexamer.

In summary, this study demonstrates that the truncated version CRD-2H of ERGIC-53, without the transmembrane region and the two membrane proximal cysteine residues, is able to assemble into hexamers. Hence, this strengthens the evidence for the involvement of the stalk domain in the oligomerisation of ERGIC-53 and highlights the importance of the two helices closest to the CRD. However, it cannot be ruled out that the oligomers formed by this truncated

protein variant could be artefacts and further studies are required to determine if the oligomeric species observed in this study correspond to the native forms.

In addition, the expression of the isolated stalk domain in *E. coli* resulted in a non-covalent complex of three disulfide-linked dimers. The absence of the covalent disulfide-linked hexamer in this study might reflect that formation of this complex requires additional factors or an environment other than that provided by the *E.coli* expression machinery. An interesting theory proposed by Neve *et al.* is the existence of an ER or ERGIC localised disulfide isomerase catalysing re-arrangement of disulfide bonds, and thus acting as a reversible switch between the two disulfide-linked complexes (Neve *et al.* 2005). Although plausible, no evidence for such mechanisms exists thus far. Interesting challenges for future studies will be to clarify the mechanism and role of the different disulfide-linked ERGIC-53 complexes in cycling and cargo transport processes between the ER and the ERGIC. Furthermore, it will be of high interest to reveal if binding of cargo proteins or any post-translational modification, other than disulfide bond formation, influences the oligomeric organisation of ERGIC-53.

3.5 ERGL, a regulator of the ERGIC-53/MCFD2 complex?

ERGL was first identified in prostate tissue and its expression has been found to be restricted to specific tissues including prostate, spleen, salivary gland, cardiac atrium and cells of the central nervous system (Yerushalmi *et al.* 2001). The first 230 amino-terminal residues of the expressed protein show high homology to the CRD of the family of L-type lectins. Particularly, the CRD of ERGIC-53 shows high sequence identity to ERGL (51%) (figure 23a). Interestingly though, ERGL is missing a number of key residues which have been found to be critical for calcium- and sugar-binding in ERGIC-53 and the other animal L-type lectins (figure 23b). Furthermore, ERGL lacks typical transport motifs in the cytoplasmic tail that are found in ERGIC-53, and is predicted to be ER resident. Although ERGL has a different expression pattern compared to the ubiquitously expressed ERGIC-53 and is confined to the ER, the predicted domain organisation is highly similar between these two proteins. While several proteins have been found to interact with ERGIC-53, and it has been shown to form a glycoprotein transport complex with MCFD2, the function of ERGL remains to be revealed.

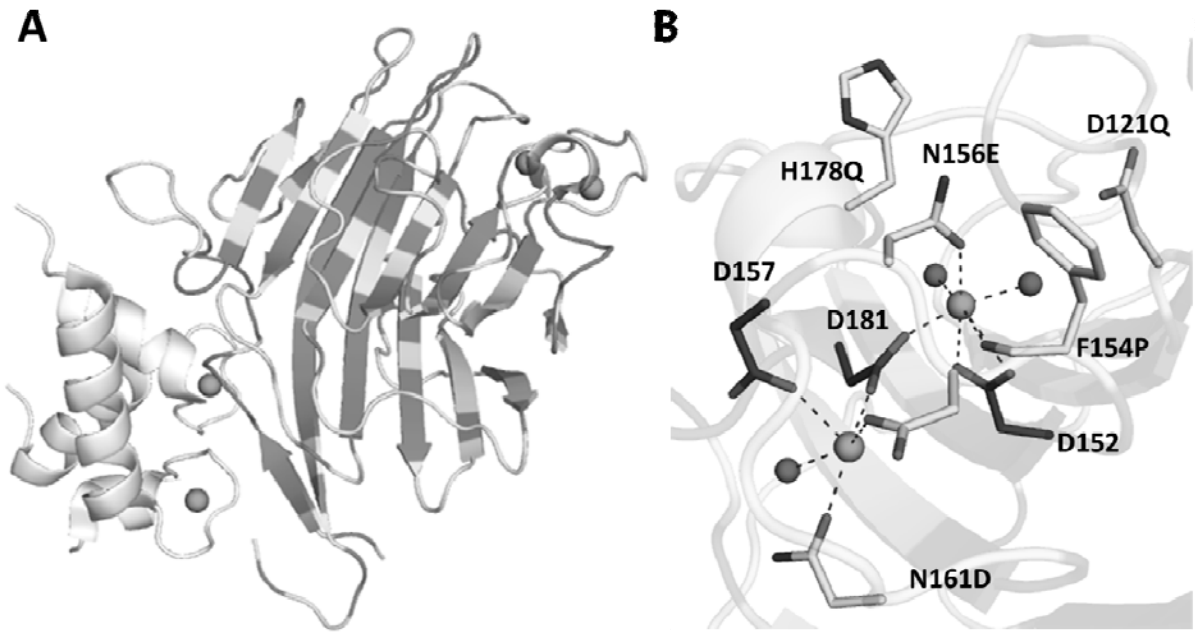


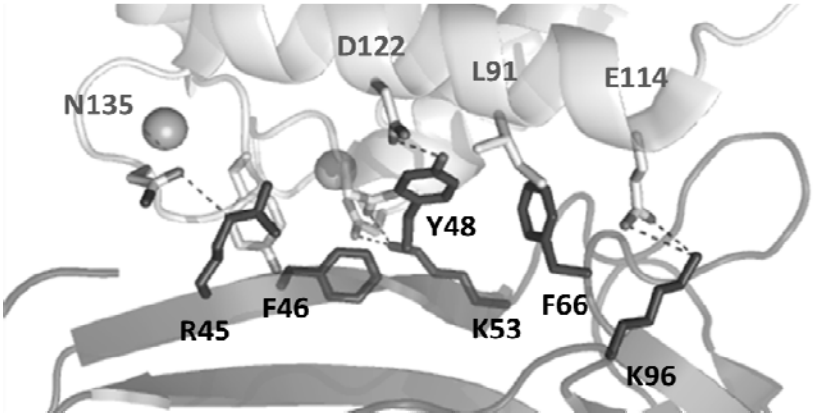
Figure 23. Sequence conservation between the CRDs of ERGIC-53 and ERGL. Conserved amino acids colored dark grey and non-conserved in light grey. **(A)** Structure of the ERGIC-53 CRD in complex with MCFD2. **(B)** A close-up view of calcium and putative carbohydrate binding residue in ERGIC-53 with residues shown as sticks, labelled, and with the corresponding residue found in ERGL denoted for non-conserved residues.

In order to shed light on the role of ERGL in the secretory pathway constructs including full-length human ERGL, the luminal part of it, and the isolated CRD and the stalk domain were generated. Of these constructs, the isolated stalk domain could be expressed and by using CD-spectroscopy this domain was shown to mainly consist of α-helical secondary structure as predicted by bioinformatics. Characterisation of the other constructs was hampered by low yield and low solubility.

3.5.1 ERGL/MCFD2 interaction analysis

The MCFD2-binding site of ERGIC-53, which primarily consists of the loop connecting the β1 and β2 strands, is not conserved within the L-type lectin homologues VIP36 and VIPL. SPR analysis has confirmed that neither VIP36 nor VIPL bind to MCFD2, despite the structural similarities of their CRDs to ERGIC-53 (Kawasaki *et al.* 2008). However, thus far no studies have probed an interaction between MCFD2 and ERGL. Interestingly, the MCFD2-binding site in ERGIC-53 is conserved within the ERGL CRD. We hypothesise that this region of ERGL could bind MCFD2 in a manner similar to that of ERGIC-53 and form a complex. To address this hypothesis we investigated the binding of MCFD2 to this region by NMR spectroscopy and biolayer interferometry using a set of three synthesised peptides of 13 residues each, mimicking this region in ERGIC-53, ERGL and VIP36, respectively (figure 24).

A



B

```
                45        55        65        75        85
ERGIC-53   PHRRFEYKYSFKGPHLVQSDGTVPFWAHAGNAIPSSDQIRVAPSLKSQR
ERGL       PLRRFEYKLSFKGPRLALPGAGIPFWSHHGDAILGLEEVRLTPSMRNRS
VIP36      --EHLKREHSLIKPYQGVGSSSMPLWDFQGSTMLTSQYVRLTPDERSKE
```

Figure 24. (A) The interface between MCFD2 (grey) and ERGIC-53 (dark grey). Hydrogen bonds are shown as dashed lines and interacting residues are shown as sticks and labelled. **(B)** Sequence alignment with light grey shading representing identical residues found in ERGIC-53 and ERGL, and identical residues also found in VIP36 shown with black background and white text. Boxed residues correspond to the synthesised peptides of 13 residues used in this study.

NMR analyses using ^{15}N-labeled MCFD2 show coordinated chemical shift changes in the ^{15}N-HSQC spectra upon addition of the ERGL and ERGIC-53 peptides, whereas the chemical shift perturbation for the VIP36 peptide is much less pronounced (figure 25a). Although these results give a strong indication that MCFD2 interacts with a similar binding mode to, ERGIC-53 and ERGL, further studies are required to support this theory. A major drawback in this study was that spectra had to be collected in different buffer conditions than those of the ^{15}N-HSQC spectrum used as reference in previous experiments, due to aggregation of the peptides. This resulted in limited applicability of our previous chemical shift assignments. In order to proceed with this study, a new set of assignment experiments would therefore be of highest interest.

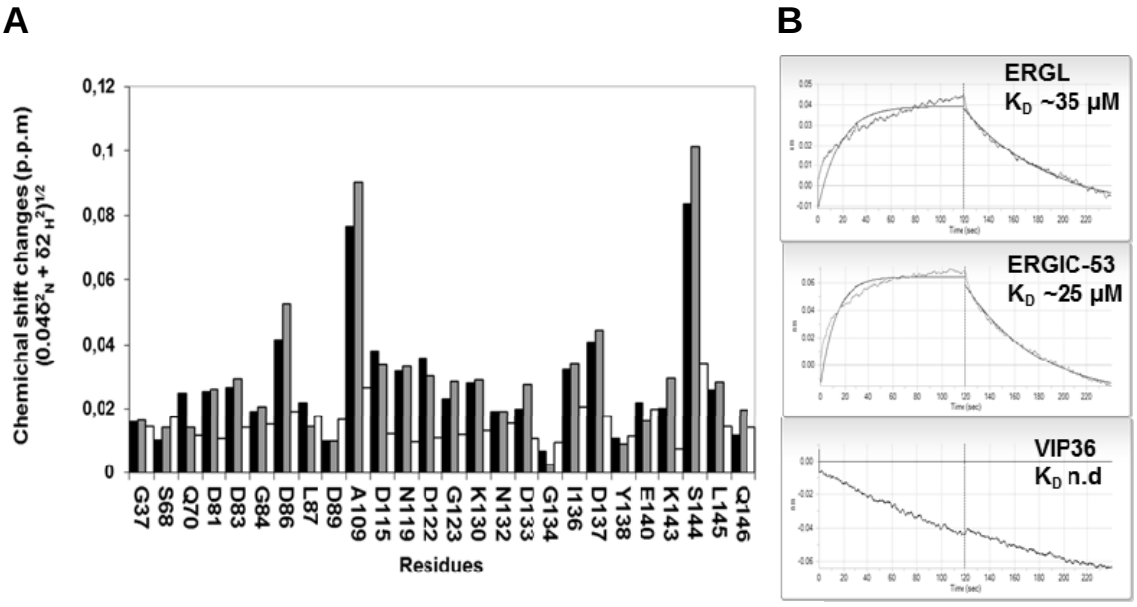


Figure 25. Interaction analysis. **(A)** Chemical shift perturbation data from, ^{15}N-HSQC spectra upon mixing ^{15}N-labelled MCFD2 and the peptides: ERGL-CRD(RRFEYKLSFKGPR) (black), ERGIC-53-CRD(RRFEYKYSFKGPH) (grey) and VIP36-CRD(EHLKREHSLIKPY) (white), respectively. Data were calculated for each of the 26 residues that could be assigned, using the equation $(0.04\delta_N^2+\delta_H^2)^{1/2}$ (ppm), where δ_N and δ_H represent the change in nitrogen and proton chemical shift. **(B)** Biolayer interferometry sensorgrams of binding to MCFD2 for each of the peptides used in the NMR analysis monitored at a concentration of 100 µM. The affinities of the ERGL-CRD and ERGIC-53-CRD peptides to MCFD2 were calculated using a 1:1 binding model.

Real time binding assays between biotinylated MCFD2 captured on a surface of streptavidin and peptides in solution were performed using biolayer interferometry (figure 25b) (OctetRedsystem). The association and dissociation of each peptide to MCFD2 was monitored and from this data K_D values in the low micromolar range could be determined for the ERGL and ERGIC-53 peptides to MCFD2. No interaction could be detected between the VIP36 peptide and MCFD2 confirming previous findings (Kawasaki *et al.* 2008).

In summary, this study strengthens the hypothesis that ERGL forms a complex with MCFD2, although the role of this complex still remains elusive. To date, MCFD2 has been directly linked only to the function as a co-receptor for ERGIC-53 in the transport of the coagulation factors FV and FVIII. However, MCFD2 has been shown to maintain stem cell potential in the adult central nervous system and is recognised as a marker for testicular germ cell tumours (Toda *et al.* 2003, Gashaw *et al.* 2007). Interestingly, the tissue-specific distribution of ERGL is coincident with these yet unresolved processes in which MCFD2 is overexpressed. Although highly speculative, the ER resident protein ERGL could function as regulator of the ERGIC-53/MCFD2 complex by interacting with MCFD2 in the ER and by this means indirectly control secretion of glycoproteins. It will be highly interesting to learn the function of ERGL and the dual function of MCFD2, if there is one, when they are revealed.

4 CONCLUSIONS

Within this thesis work, NMR spectroscopy and X-ray crystallography have been applied to elucidate the organisation and function of the ERGIC-53 /MCFD2 glycoprotein transport receptor complex. The work resulted in the three-dimensional structure of MCFD2 in solution determined by NMR and the crystal structure of MCFD2 in complex with ERGIC-53-CRD. The combination of the two techniques for structural characterisation provided unique insights into the mechanisms of the interaction between these proteins.

NMR studies of MCFD2 revealed that the protein is disordered in the absence of Ca^{2+} ions and adopts a predominantly ordered structure upon binding Ca^{2+} ions. Hence, this work not only resulted in the first three-dimensional structure of MCFD2, but also gave insights into the mechanism by which Ca^{2+} functions as an allosteric regulator on MCFD2 binding to ERGIC-53. Furthermore, from CD and NMR analysis the consequences of all known F5F8D causing missense mutations found in MCFD2 could be explained. The results highlight the importance of intact calcium binding EF-hand motifs for the structural stability of MCFD2 and point toward disruption of the ERGIC-53/MCFD2 interaction as the underlying mechanism for these mutations in causing F5F8D.

The crystal structure of ERGIC-53-CRD/MCFD2-ΔN complex gave a first molecular view of the organisation of a cargo receptor complex in the early secretory pathway. Moreover, the structure revealed that MCFD2 undergoes significant conformational changes upon complex formation, which results in the exposure of its hydrophobic core. Although it has yet to be proven, the structural changes observed in MCFD2 could be part of the mechanism regulating the interaction between the receptor ERGIC-53/MCFD2 complex and the cargo proteins FV and FVIII. Overall, this thesis work has provided new insights into the structural organisation and function of the ERGIC-53/MCFD2 transport complex. From this, a whole new set of questions and challenges for future studies has emerged.

In order to fully understand how the ERGIC-53/MCFD2 complex recognises and transports cargo, a key objective will be to determine how the oligomeric states of ERGIC-53 influence these processes. The preliminary data described in this thesis have strengthened the evidence for that disulfide bond formation, along with interactions between the α-helices within the stalk domain, are involved in the assembly of ERGIC-53 into hexamers. However, the exact organisation of this hexamer still remains elusive and will be of high interest for future studies to reveal. Moreover, preliminary data suggests that ERGL, the closely related ER resident homologue of ERGIC-53, interacts and forms a complex with MCFD2. Although the role of this complex remains poorly understood, one could imagine a scenario where ERGL functions as regulator of the ERGIC-53/MCFD2 complex in the ER by interacting with MCFD2.

In summary, this thesis work has provided new information on the organisation of receptor mediated transport in the secretory pathway. The combination of methods for structural characterisation, such as X-ray crystallography and NMR spectroscopy, applied in this thesis will be a key approach in the future to elucidate the structures and assembly dynamics in the myriad of complexes that exist in the secretory pathway.

5 ACKNOWLEDGEMENTS

This work was carried out at the Division of Molecular Structural Biology of the Department of Medical Biochemistry and Biophysics, Karolinska Institutet. I wish to express my sincere gratitude to the following people for all the help and support:

First of all I would like to thank my main supervisor Professor Ylva Lindqvist for accepting me as a PhD student. It has been truly inspiring to see someone enjoying science so much. Your insightful advices and guidance have been invaluable during my PhD studies and most appreciated. I would also like to thank Professor Gunter Schneider for his admirable dedication to making great science and providing an excellent research environment. Additionally, I would really like to thank you both for running the group so well and for all social arrangement outside the lab. Especially, the dinners at Slätträsk and at your home have been most appreciated.

I am most grateful to my co-supervisor Jodie Guy who was the first to guide me in the world of protein expression and crystallisation, and later for joining me in the bewildering world of NMR. It has been great working with you! I am also most grateful to my co-supervisor Inari Kursula for all your help during my first years in the ERGIC-53 project. It has been truly encouraging to see your devotion to both science and your family. I have learnt much about both science and life from you. Kiitos paljon! I would like to express my gratitude to my old ERGIC-53 co-worker Jean-Marie Bourhis. It was most fun. I really enjoyed working with you as well as our discussions outside work. I hope we can go on and finish off the EAL side project now when this thesis is written.

I would like to express my sincere gratitude to Professor Torleif Härd for all support and contribution to the NMR work presented in this thesis. In addition, for being an excellent and most patient guide in the fascinating, but sometimes confusing world of biomolecular NMR. I would also like to thank PhD Hejer Elmahmoudi and colleagues at Laboratory of Genetics, Immunology and Human Pathologies, Tunis, Tunisia, for a most fruitful collaboration.

I would also like to thank all the students that have contributed to the project - Axel, Isabel, Ella, Javaria, and Lance. It was great getting to know you and you all did a great work. In particular I would like thank Axel for coming back and spending two summers in the lab. Your contribution made a difference for the project.

I would like to thank all the present and past members of the group of Molecular Structure Biology:Robert, a true scientist in heart and soul. I've gained enormously from your lectures and shouting at me what to do and not to do. Köszi "Cha-Cha"! Bernie, the oracle of crystallography, programming and Swedish football players that have, are going to, or might go to play in the Scottish league. Mange, my long time lab colleague and one of the best characters I have ever met - great discussions, always in a good mood and a special thanks for all the liquorice! Ahmad, king of the lab, without you it would stop running. Dominik, now we

start doing real science and I will payback some beer I owe you. Ekatarina, it has been freaking awesome sharing office with you. Maria, it has been great fun working with you and I especially remember our trips to a cold and confusing Uppsala. Doreen, for all your help in teaching, structural biology, and all the superb cakes during the years. Rajesh, for always being helpful, and also for showing all those weird Bollywood dancing moves. Jason, the fastest guy the group has ever seen. One day I will outrun you mate, but first a lager. Ömer, always nice friendly and ready to give a helping hand somewhere. Dominic," le Roi-Soleil" always joyful and in a good mood. Victoria, always helpful fast and organised. All lab new-comers, Katharina, Ming Wei, Eva Maria it has been great starting to get to know you. Of the past members of the lab I would first of all like to thank Eva, for all the years you have been taking care of us. A special thanks for keeping track of all my lost papers and the mysterious "reseräkning". Stort tack och en ännu större kram! Hanna, as a previous member of the "such a nice office". It was great having you as an office colleague and sharing your sunshine view of the world. Daniel, Mr Norrland for being so relaxed and easy going, come-back on the football field soon? Catrin and Stina, for all your help getting me started in the lab in an always helpful and supporting way. Tanja, for help with crystallography and excellent borsjtj. Azmiri, for many interesting lunch discussions and for sharing a common interest in the outcome of "La liga". Jola, one of the most helpful and scientific minded persons I have ever met. Thank you for all your help during your stay in our lab! Leif, for introducing me to bag in box rum and new cloning methods. Mikko, for trying to explain the greatness of polyketides and for sharing a common interest in the great Jari Litmanen. Agatha, for great office company and for teaching me a few, but important polish words. Na zdrowie! Atsushi, for help with crystallography and pokemon nomenclature.

I would also like to acknowledge Åke Rökaeus and the MBB teaching unit, Susanne Björkholm Agneta Nordberg and Susanne Rothstein. You made teaching fun and a great learning experience. I would like to thank all the people at MBB that have helped and support me and especially: Tomas from upstairs, for interesting discussions regarding ditt och datt. Helena, Martin and all helpful SGC/PSF workers. The glorious MBB football team, the dynamic finish duo Jaakko and Juha, Lionel Massa, Joel "magic JJ" Johansson, Yilmaz the running machine, Bernie the bomber, Bruno O Fenomeno , Coach Colin, Asmundur the rock and all other players that have contributed during the years.

Jag skulle också vilja passa på att tacka vänner utanför labbet. Salah, Dali och alla spelare i Inglourious Ballers. John, John igen, Jocke, Jesper och alla andra gamla västerortskaraktärer. Basårsligan, Dan, Prinsen, Håkan, Klingse och Mia.

Slutligen skulle jag vilja tacka min familj. Mor och far, tack för att ni alltid finns där och stöttat mig i mina val. Kico och Alex, de bästa bröderna man kan ha. Stort tack för allt under alla år som gått! Tinna och Mikaela för många trevliga stunder och middagar. Erik, Thea och Kosmos för många minnesvärda upptåg.

Laura tack för allt!

6 REFERENCES

Abdallah, H.E., Gouider, E., Amor, M.B., Jlizi, A., Meddeb, B., &. Elgaaied, A. (2011). Molecular analysis in two Tunisian families with combined factor V and factor VIII deficiency. *Haemophilia*, *16*(5), 801-4

Anelli, T., Ceppi, S., Bergamelli, L., Cortini, M., Masciarelli, S., Valetti, C., & Sitia, R. (2007). Sequential steps and checkpoints in the early exocytic compartment during secretory IgM biogenesis. *EMBO J*, *26*(19), 4177-4188.

Anelli, T., & Sitia, R. (2008). Protein quality control in the early secretory pathway. *EMBO J*, *27*(2), 315-327.

Appenzeller, C., Andersson, H., Kappeler, F., & Hauri, H. P. (1999). The lectin ERGIC-53 is a cargo transport receptor for glycoproteins. *Nat Cell Biol*, *1*(6), 330-334.

Appenzeller-Herzog, C, Nyfeler, B., Burkhard, P., Santamaria, I., Lopez-Otin, C., & Hauri, H.P. (2005). Carbohydrate-and conformation-dependent cargo capture for ER-exit. *Mol Biol Cell*, *16*(3), 1258–1267.

Appenzeller-Herzog, Christian, & Hauri, H.P. (2006). The ER-Golgi intermediate compartment (ERGIC): in search of its identity and function. *J Cell Sci*, *119*(11), 2173-2183.

Appenzeller-Herzog, Christian, Roche, A.-C., Nufer, O., & Hauri, H.P. (2004). pH-induced conversion of the transport lectin ERGIC-53 triggers glycoprotein release. *J Biol Chem*, *279*(13), 12943-12950.

Bannykh, S. I., Rowe, T., & Balch, W. E. (1996). The organization of endoplasmic reticulum export complexes. *J Cell Biol*, *135*(1), 19-35.

Barlowe, C. (2003). Signals for COPII-dependent export from the ER: what's the ticket out? *Trends Cell Biol*, *13*(6), 295-300.

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M. F., et al. (1994). COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell*, *77*(6), 895-907.

Belden, W. J., & Barlowe, C. (2001). Role of Erv29p in collecting soluble secretory proteins into ER-derived transport vesicles. *Science*, *294*(5546), 1528-1531.

Bernasconi, R., & Molinari, M. (2011). ERAD and ERAD tuning: disposal of cargo and of ERAD regulators from the mammalian ER. *Curr Opin Cell Biol*, *23*(2), 176-183.

Bi, X., Corpina, R. A., & Goldberg, J. (2002). Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature*, *419*(6904), 271-277.

Blobel, G. B. D. (1975). TRANSFER OF PROTEINS ACROSS MEMBRANES I. Presence of Proteolytically Processed and Unprocessed Light Chains On Membrane-Bound Ribosomes of Murine Myeloma. *In Vitro*, *67*, 835-851.

Bonifacino, J. S., & Glick, B. S. (2004). The mechanisms of vesicle budding and fusion. *Cell*, *116*(2), 153-166.

Bouckaert, J., Hamelryck, T., Wyns, L., & Loris, R. (1999). Novel structures of plant lectins and their complexes with carbohydrates. *Curr Opin Struct Biol*, *9*(5), 572-577.

Boyd, W. C., & Shapleigh, E. (1954). Specific Precipitating Activity of Plant Agglutinins (Lectins). *Science*, *119*(3091), 419.

Bröcker, C., Engelbrecht-Vandré, S., & Ungermann, C. (2010). Multisubunit tethering complexes and their role in membrane fusion. *Curr Biol*, *20*(21), 943-52.

Budnik, A., & Stephens, D. J. (2009). ER exit sites--localization and control of COPII vesicle formation. *FEBS Lett*, *583*(23), 3796-3803.

Bue, C. A., & Barlowe, C. (2009). Molecular Dissection of Erv26p Identifies Separable Cargo Binding and Coat Protein Sorting Activities. *J Biol Chem*, *284*(36), 24049-24060.

Bue, C. A., Bentivoglio, C. M., & Barlowe, C. (2006). Erv26p directs pro-alkaline phosphatase into endoplasmic reticulum-derived coat protein complex II transport vesicles. *Mol Biol Cell*, *17*(11), 4780-4789.

Bukau, B., Weissman, J., & Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell*, *125*(3), 100.

Caldwell, S. R., Hill, K. J., & Cooper, A. A. (2001). Degradation of endoplasmic reticulum (ER) quality control substrates requires transport between the ER and Golgi. *J Biol Chem*, *276*(26), 23296-23303.
Camire, R. M., & Bos, M. H. A. (2009). The molecular basis of factor V and VIII procofactor activation. *J Thromb Haemost*, *7*(12), 1951-1961.
Caro, L. G., & Palade, G. E. (1964). PROTEIN SYNTHESIS, STORAGE, AND DISCHARGE IN THE PANCREATIC EXOCRINE CELL. AN AUTORADIOGRAPHIC STUDY. *J Cell Biol*, *20*(3), 473-495.
Chatterjee, S., & Mayor, S. (2001). The GPI-anchor and protein sorting. *Cell Mol Life Sci*, *58*(14), 1969-1987.
Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K., & Zerial, M. (1990). Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell*, *62*(2), 317-329.
Cintron, N. S., & Toft, D. (2006). Defining the requirements for Hsp40 and Hsp70 in the Hsp90 chaperone pathway. *J Biol Chem*, *281*(36), 26235-26244.
Cortini, M., & Sitia, R. (2010). ERp44 and ERGIC-53 synergize in coupling efficiency and fidelity of IgM polymerization and secretion. *Traffic*, *11*(5), 651-659.
Dancourt, J., & Barlowe, C. (2010). Protein sorting receptors in the early secretory pathway. *Annu Rev Biochem*, *79*, 777-802.
De Matteis, M. A., & Luini, A. (2011). Mendelian disorders of membrane trafficking. *N Engl J Med*, *365*(10), 927-938.
Denzel, A., Otto, F., Girod, A., Pepperkok, R., Watson, R., Rosewell, I., Bergeron, J. J., et al. (2000). The p24 family member p23 is required for early embryonic development. *Curr Biol*, *10*(1), 55-58.
Dong C., Nichols C.D., Guo J., Huang W., Lambert N.A., & Wu G. (2012). A Triple Arg Motif Mediates $_{\alpha(2)\,(B)}$ -Adrenergic Receptor Interaction with Sec24C/D and Export. *Traffic*, Epub ahead of print
Ellgaard, L., & Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol*, *4*(3), 181-91.
Elmahmoudi, H., Wigren, E., Laatiri, A., Jlizi, A., Elgaaied, A., Gouider, E., & Lindqvist, Y. (2011). Analysis of newly detected mutations in the MCFD2 gene giving rise to combined deficiency of coagulation factors V and VIII. *Haemophilia*, *17*(5), e923-e927.
Fiedler, K., & Simons, K. (1994). A putative novel class of animal lectins in the secretory pathway homologous to leguminous lectins. *Cell*, 3;77(5),625-6.
Fiedler, K., & Simons, K. (1996). Characterization of VIP36, an animal lectin homologous to leguminous lectins. *J Cell Sci*, *109(1)*, 271-276.
Fraldi, A., Zito, E., Annunziata, F., Lombardi, A., Cozzolino, M., Monti, M., Spampanato, C., et al. (2008). Multistep, sequential control of the trafficking and function of the multiple sulfatase deficiency gene product, SUMF1 by PDI, ERGIC-53 and ERp44. *Hum Mol Genet*, *17*(17), 2610-2621.
Gashaw, I., Dushaj, O., Behr, R., Biermann, K., Brehm, R., Rübben, H., Grobholz, R., et al. (2007). Novel germ cell markers characterize testicular seminoma and fetal testis. *Mol Hum Reprod*, *13*(10), 721-727.
Gavel, Y., & Von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng*, *3*(5), 433-442.
Gething, M. J., McCammon, K., & Sambrook, J. (1986). Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. *Cell*, *46*(6), 939-950.
Gething, M. J., & Sambrook, J. (1992). Protein folding in the cell. *Nature*, *355*(6355), 33-45.
Gifford, J. L., Walsh, M. P., & Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. *Biochem J*, *405*(2), 199-221.
Giuliani, F., Grieve, A., & Rabouille, C. (2011). Unconventional secretion: a stress on GRASP. *Curr Opin Cell Biol*, *23*(4), 498-504.
Glick, B. S. (2000). Organization of the Golgi apparatus. *Curr Opin Cell Biol*, *12*(4), 450-456.
Grabarek, Z. (2006). Structural basis for diversity of the EF-hand calcium-binding proteins. *J Mol Biol*, *359*(3), 509-525.

Grinna, L. S., & Robbins, P. W. (1980). Substrate specificities of rat liver microsomal glucosidases which process glycoproteins. *J Biol Chem*, *255*(6), 2255-2258.
Guy, J. E., Wigren, E., Svärd, M., Härd, T., & Lindqvist, Y. (2008). New insights into multiple coagulation factor deficiency from the solution structure of human MCFD2. *J Mol Biol*, *381*(4), 941-955.
Hauri, H. P., Appenzeller, C., Kuhn, F., & Nufer, O. (2000). Lectins and traffic in the secretory pathway. *FEBS Lett*, *476*(1-2), 32-7.
Hauri, H. P., Kappeler, F., Andersson, H., & Appenzeller, C. (2000). ERGIC-53 and traffic in the secretory pathway. *J Cell Sci*, *113 (4)*, 587-596.
Hauri, H. P., Nufer, O., Breuza, L., Tekaya, H. B., & Liang, L. (2002). Lectins and protein traffic early in the secretory pathway. *Biochem Soc Symp*, *82*(69), 73-82.
Helenius, A. (1994). How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol Biol Cell*, *5*(3), 253-265.
Helenius, A., & Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science*, *291*(5512), 2364-2369.
Helenius, A., Trombetta, E. S., Hebert, D. N., & Simons, J. F. (1997). Calnexin, calreticulin and the folding of glycoproteins. *Trends Cell Biol*, *7*(5), 193-200.
Holmes, K. C. (1996). Muscle proteins — their actions and interactions. *Curr Opin Struct Biol*, *6*(6), 781-789.
Hsu, V. W., & Yang, J. S. (2009). Mechanisms of COPI vesicle formation. *FEBS Lett*, *583*(23), 3758-3763.
Itin, C., Roche, A. C., Monsigny, M., & Hauri, H. P. (1996). ERGIC-53 is a functional mannose-selective and calcium-dependent human homologue of leguminous lectins. *Mol Biol Cell*, *7*(3), 483-493.
Ivaskevicius, V., Windyga, J., Baran, B., Bykowska, K., Daugela, L., Watzka, M., Seifried, E., et al. (2008). The first case of combined coagulation factor V and coagulation factor VIII deficiency in Poland due to a novel p.Tyr135Asn missense mutation in the MCFD2 gene. *Blood Coagul Fibrinolysis,* 19(6),531-534.
Jayandharan, G., Spreafico, M., Viswabandya, A., Chandy, M., Srivastava, A., & Peyvandi, F. (2007). Mutations in the MCFD2 gene are predominant among patients with hereditary combined FV and FVIII deficiency (F5F8D) in India. *Haemophilia*, *13*(4), 413-419.
Kamiya, Y., Kamiya, D., Yamamoto, K., Nyfeler, B., Hauri, H. P., & Kato, K. (2008). Molecular basis of sugar recognition by the human L-type lectins ERGIC-53, VIPL, and VIP36. *J Biol Chem*, *283*(4), 1857-1861.
Kane, W. H., & Davie, E. W. (1988). Blood coagulation factors V and VIII: structural and functional similarities and their relationship to hemorrhagic and thrombotic disorders. *Blood*, *71*(3), 539-555.
Kanehara, K., Kawaguchi, S., & Ng, D. T. W. (2007). The EDEM and Yos9p families of lectin-like ERAD factors. *Semin Cell Dev Biol*, *18*(6), 743-750.
Kappeler, F., Klopfenstein, D. R. C., Paccaud, J.P., & Hauri, H.P. (1997). The Recycling of ERGIC-53 in the Early Secretory Pathway. *Biochemistry*, *272*(50), 31801-31808.
Kawasaki, N., Ichikawa, Y., Matsuo, I., Totani, K., Matsumoto, N., Ito, Y., & Yamamoto, K. (2008). The sugar-binding ability of ERGIC-53 is enhanced by its interaction with MCFD2. *Blood*, *111*(4), 1972-1979.
Kondylis, V., Pizette, S., & Rabouille, C. (2009). The early secretory pathway in development: a tale of proteins and mRNAs. *Semin Cell Dev Biol*, *20*(7), 817-827.
Kreis, T. E., & Lodish, H. F. (1986). Oligomerization is essential for transport of vesicular stomatitis viral glycoprotein to the cell surface. *Cell*, *46*(6), 929-937.
Kretsinger, R. H., Nockolds, C. E., Coffee, C. J., & Bradshaw, R. A. (1972). The structure of a calcium-binding protein from carp muscle. *Cold Spring Harb Symp Quant Biol*, *36*, 217-220.
Lahtinen, U., Hellman, U., Wernstedt, C., Saraste, J., & Pettersson, R. F. (1996). Molecular cloning and expression of a 58-kDa cis-Golgi and intermediate compartment protein. *J Biol Chem*, *271*(8), 4031-4037.
Lee, C., & Goldberg, J. (2010). Structure of coatomer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats. *Cell*, *142*(1), 123-132.

Lee, M. C. S., & Miller, E. A. (2007). Molecular mechanisms of COPII vesicle formation. *Semin Cell Dev Biol*, *18*(4), 424-434.
Levine, C. G., Mitra, D., Sharma, A., Smith, C. L., & Hegde, R. S. (2005). The Efficiency of Protein Compartmentalization into the Secretory Pathway. *Mol Biol Cell*, *16*(1), 279 -291.
Lievens, P. M.-J., De Servi, B., Garofalo, S., Lunstrum, G. P., Horton, W. A., & Liboi, E. (2008). Transient dimerization and interaction with ERGIC-53 occur in the fibroblast growth factor receptor 3 early secretory pathway, *Int J Biochem Cell Biol*, *40*(11), 2649-2659.
Loris, R. (2002). Principles of structures of animal and plant lectins. *Biochim Biophys Acta*, *1572*(2-3), 198-208.
Ma, D., Zerangue, N., Lin, Y. F., Collins, A., Yu, M., Jan, Y. N., & Jan, L. Y. (2001). Role of ER export signals in controlling surface potassium channel numbers. *Science*, *291*(5502), 316-319.
Malkus, P., Jiang, F., & Schekman, R. (2002). Concentrative sorting of secretory cargo proteins into COPII-coated vesicles. *J Cell Biol*, *159*(6), 915-921.
Mann, K. G., Jenny, R. J., & Krishnaswamy, S. (1988). Cofactor proteins in the assembly and expression of blood clotting enzyme complexes. *Annu Rev Biochem*, *57*, 915-956.
Mannucci, P. M., Duga, S., & Peyvandi, F. (2004). Recessively inherited coagulation disorders. *Blood*, *104*(5), 1243-1252.
Margittai, E., & Bánhegyi, G. (2010). Oxidative folding in the endoplasmic reticulum: towards a multiple oxidant hypothesis? *FEBS Lett*, *584*(14), 2995-2998.
Martínez-Menárguez, J. A., Geuze, H. J., Slot, J. W., & Klumperman, J. (1999). Vesicular tubular clusters between the ER and Golgi mediate concentration of soluble secretory proteins by exclusion from COPI-coated vesicles. *Cell*, *98*(1), 81-90.
Mellman, I., & Warren, G. (2000). The road taken: past and future foundations of membrane traffic. *Cell*, *100*(1), 99-112.
Miao, H. Z., Sirachainan, N., Palmer, L., Kucab, P., Cunningham, M. A., Kaufman, R. J., & Pipe, S. W. (2004). Bioengineering of coagulation factor VIII for improved secretion. *Blood*, *103*(9), 3412-3419.
Mitrovic, S., Ben-Tekaya, H., Koegler, E., Gruenberg, J., & Hauri, H. P. (2008). The Cargo Receptors Surf4, Endoplasmic Reticulum-Golgi Intermediate Compartment (ERGIC)-53, and p25 Are Required to Maintain the Architecture of ERGIC and Golgi. *Mol Biol Cell*, *19*(5), 1976-1990.
Mizuno, M., & Singer, S. J. (1993). A soluble secretory protein is first concentrated in the endoplasmic reticulum before transfer to the Golgi apparatus. *Proc Natl Acad Sci U S A*, *90*(12), 5732-5736.
Mohorko, E., Glockshuber, R., & Aebi, M. (2011). Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *Journal of Inherited Metabolic Disease*, *34*(4), 869-878.
Morais, V. A., Brito, C., Pijak, D. S., Crystal, A. S., Fortna, R. R., Li, T., Wong, P. C., et al. (2006). N-glycosylation of human nicastrin is required for interaction with the lectins from the secretory pathway calnexin and ERGIC-53. *Biochim Biophys Acta*, *1762*(9), 802-810.
Munro, S., & Pelham, H. R. (1987). A C-terminal signal prevents secretion of luminal ER proteins. *Cell*, *48*(5), 899-907.
Muñiz, M., Nuoffer, C., Hauri, H. P., & Riezman, H. (2000). The Emp24 complex recruits a specific cargo molecule into endoplasmic reticulum-derived vesicles. *J Cell Biol*, *148*(5), 925-930.
Neve, E. P. A., Lahtinen, U., & Pettersson, R. F. (2005). Oligomerization and interacellular localization of the glycoprotein receptor ERGIC-53 is independent of disulfide bonds. *J Mol Biol*, *354*(3), 556-568.
Neve, E. P. A., Svensson, K., Fuxe, J., & Pettersson, R. F. (2003). VIPL, a VIP36-like membrane protein with a putative function in the export of glycoproteins from the endoplasmic reticulum. *Experimental Cell Research*, *288*(1), 70-83.
Nichols, W. C., Seligsohn, U., Zivelin, A., Terry, V. H., Hertel, C. E., Wheatley, M. A., Moussalli, M. J., et al. (1998). Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII. *Cell*, *93*(1), 61-70.
Niki, I., Yokokura, H., Sudo, T., Kato, M., & Hidaka, H. (1996). Ca2+ signaling and intracellular Ca2+ binding proteins. *Journal of Biochemistry*, *120*(4), 685-698.
Nishimura, N., & Balch, W. E. (1997). A di-acidic signal required for selective export from the endoplasmic reticulum. *Science*, *277*(5325), 556-558.

Nishio, M., Kamiya, Y., Mizushima, T., Wakatsuki, S., Sasakawa, H., Yamamoto, K., Uchiyama, S., et al. (2010). Structural basis for the cooperative interplay between the two causative gene products of combined factor V and factor VIII deficiency. *Proc Natl Acad Sci U S A*, *107*(9), 4034-4039.

Nufer, O., Mitrovic, S., & Hauri, H. P. (2003). Profile-based data base scanning for animal L-type lectins and characterization of VIPL, a novel VIP36-like endoplasmic reticulum protein. *J Biol Chem*, *278*(18), 15886-15896.

Nyfeler, B., Kamiya, Y., Boehlen, F., Yamamoto, K., Kato, K., De Moerloose, P., Hauri, H. P., et al. (2008). Deletion of 3 residues from the C-terminus of MCFD2 affects binding to ERGIC-53 and causes combined factor V and factor VIII deficiency. *Blood*. 111(3), 1299-301

Nyfeler, B., Michnick, S. W., & Hauri, H.-P. (2005). Capturing protein interactions in the secretory pathway of living cells. *Proc Natl Acad Sci U S A*, *102*(18), 6350-6355.

Nyfeler, B., Reiterer, V., Wendeler, M. W., Stefan, E., Zhang, B., Michnick, S. W., & Hauri, H.-P. (2008). Identification of ERGIC-53 as an intracellular transport receptor of alpha1-antitrypsin. *J Cell Biol*, *180*(4), 705-712.

Nyfeler, B., Zhang, B., Ginsburg, D., Kaufman, R. J., & Hauri, H. P. (2006). Cargo selectivity of the ERGIC-53/MCFD2 transport receptor complex. *Traffic*, *7*(11), 1473-1481.

Oeri, J., Matter, M., Isenschmid, H., Hauser, F., & Koller, F. (1954). Congenital factor V deficiency (parahemophilia) with true hemophilia in two brothers. *Bibliotheca Paediatrica*, *58*, 575-588.

Otte, S., & Barlowe, C. (2004). Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. *Nat Cell Biol*, *6*(12), 1189-1194.

Paetzel, M., Karla, A., Strynadka, N. C. J., & Dalbey, R. E. (2002). Signal peptidases. *Chem Rev*, *102*(12), 4549-4580.

Paroutis, P., Touret, N., & Grinstein, S. (2004). The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology Bethesda*, *19*(4), 207-15.

Pezzati, R., Bossi, M., Podini, P., Meldolesi, J., & Grohovaz, F. (1997). High-resolution calcium mapping of the endoplasmic reticulum-Golgi-exocytic membrane system. Electron energy loss imaging analysis of quick frozen-freeze dried PC12 cells. *Mol Biol Cell*, *8*(8), 1501-1512.

Pollock, S., Kozlov, G., Pelletier, M.-F., Trempe, J.-F., Jansen, G., Sitnikov, D., Bergeron, J. J. M., et al. (2004). Specific interaction of ERp57 and calnexin determined by NMR spectroscopy and an ER two-hybrid system. *EMBO J*, *23*(5), 1020-1029.

Powers, J., & Barlowe, C. (2002). Erv14p Directs a Transmembrane Secretory Protein into COPII-coated Transport Vesicles. *Mol Biol Cell*, *13*(3), 880-891.

Prins, D., & Michalak, M. (2011). Organellar calcium buffers. *Cold Spring Harb Perspect Biol*, *3*(3).

Reiterer, V., Nyfeler, B., & Hauri, H.-P. (2010). Role of the lectin VIP36 in post-ER quality control of human alpha1-antitrypsin. *Traffic*, *11*(8), 1044-1055.

Russell, C., & Stagg, S. M. (2010). New insights into the structural mechanisms of the COPII coat. *Traffic*, *11*(3), 303-310.

Salama, N. R., Yeung, T., & Schekman, R. W. (1993). The Sec13p complex and reconstitution of vesicle budding from the ER with purified cytosolic proteins *EMBO J*, *12*(11), 4073-4082.

Saraste, J., Palade, G. E., & Farquhar, M. G. (1987). Antibodies to rat pancreas Golgi subfractions: identification of a 58-kD cis-Golgi protein. *J Cell Biol*, *105*(5), 2021-2029.

Sato, K., & Nakano, A. (2002). Emp47p and its close homolog Emp46p have a tyrosine-containing endoplasmic reticulum exit signal and function in glycoprotein secretion in Saccharomyces cerevisiae. *Mol Biol Cell*, *13*(7), 2518-2532.

Sato, K., & Nakano, A. (2003). Oligomerization of a cargo receptor directs protein sorting into COPII-coated transport vesicles. *Mol Biol Cell*, *14*(7), 3055-3063.

Satoh, T., Cowieson, N. P., Hakamata, W., Ideo, H., Fukushima, K., Kurihara, M., Kato, R., et al. (2007). Structural basis for recognition of high mannose type glycoproteins by mammalian transport lectin VIP36. *J Biol Chem*, *282*(38), 28246-28255.

Satoh, T., Sato, K., Kanoh, A., Yamashita, K., Yamada, Y., Igarashi, N., Kato, R., et al. (2006). Structures of the carbohydrate recognition domain of Ca2+-independent cargo receptors Emp46p and Emp47p. *J Biol Chem*, *281*(15), 10410-10419.

Schwarz, F., & Aebi, M. (2011). Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol*, *21*(5), 576-82.

Schweizer, A., Fransen, J. A., Bächi, T., Ginsel, L., & Hauri, H. P. (1988). Identification, by a monoclonal antibody, of a 53-kD protein associated with a tubulo-vesicular compartment at the cis-side of the Golgi apparatus. *J Cell Biol*, *107*(5), 1643-1653.

Sohn, K., Orci, L., Ravazzola, M., Amherdt, M., Bremser, M., Lottspeich, F., Fiedler, K., et al. (1996). A major transmembrane protein of Golgi-derived COPI-coated vesicles involved in coatomer binding. *J Cell Biol*, *135*(5), 1239-1248.

Sousa, M., & Parodi, A. J. (1995). The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *EMBO J*, *14*(17), 4196-4203.

Springer, S., Chen, E., Duden, R., Marzioch, M., Rowley, A., Hamamoto, S., Merchant, S., et al. (2000). The p24 proteins are not essential for vesicular transport in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, *97*(8), 4034-4039.

Strating, J. R. P. M., & Martens, G. J. M. (2009). The p24 family and selective transport processes at the ER-Golgi interface. *Biol Cell*, *101*(9), 495-509.

Tagliavacca, L., Anelli, T., Fagioli, C., Mezghrani, A., Ruffato, E., & Sitia, R. (2003). The making of a professional secretory cell: architectural and functional changes in the ER during B lymphocyte plasma cell differentiation. *Biol Chem*, *384*(9), 1273-1277.

Tisdale, E. J., Plutner, H., Matteson, J., & Balch, W. E. (1997). p53/58 Binds COPI and Is Required for Selective Transport through the Early Secretory Pathway. *J Cell Biol*, *137*(3), 581-593.

Toda, H., Tsuji, M., Nakano, I., Kobuke, K., Hayashi, T., Kasahara, H., Takahashi, J., et al. (2003). Stem cell-derived neural stem/progenitor cell supporting factor is an autocrine/paracrine survival factor for adult neural stem/progenitor cells. *J Biol Chem*, *278*(37), 35491-35500.

Tremblay, L. (1999). Cloning and expression of a specific human alpha 1,2-mannosidase that trims Man9GlcNAc2 to Man8GlcNAc2 isomer B during N-glycan biosynthesis. *Glycobiology*, *9*(10), 1073-1078.

Velloso, L M, Svensson, K., Pettersson, R. F., & Lindqvist, Y. (2003). The crystal structure of the carbohydrate-recognition domain of the glycoprotein sorting receptor p58/ERGIC-53 reveals an unpredicted metal-binding site and conformational changes associated with calcium ion binding. *J Mol Biol*, *334*(5), 845-851.

Velloso, Lucas M, Svensson, K., Schneider, G., Pettersson, R. F., & Lindqvist, Y. (2002). Crystal structure of the carbohydrate recognition domain of p58/ERGIC-53, a protein involved in glycoprotein export from the endoplasmic reticulum. *J Biol Chem*, *277*(18), 15979-15984.

Vollenweider, F., Kappeler, F., Itin, C., & Hauri, H. P. (1998). Mistargeting of the Lectin ERGIC-53 to the Endoplasmic Reticulum of HeLa Cells Impairs the Secretion of a Lysosomal Enzyme. *J Cell Biol*, *142*(2), 377-389.

Wang, X., Matteson, J., An, Y., Moyer, B., Yoo, J.-S., Bannykh, S., Wilson, I. A., et al. (2004). COPII-dependent export of cystic fibrosis transmembrane conductance regulator from the ER uses a di-acidic exit code. *J Cell Biol*, *167*(1), 65-74.

Wendeler, M. W., Paccaud, J. P., & Hauri, H.P. (2007). Role of Sec24 isoforms in selective export of membrane proteins from the endoplasmic reticulum. *EMBO Reports*, *8*(3), 258-264.

Wigren, E., Bourhis, J. M., Kursula, I., Guy, J. E., & Lindqvist, Y. (2010). Crystal structure of the LMAN1-CRD/MCFD2 transport receptor complex provides insight into combined deficiency of factor V and factor VIII. *FEBS Lett*, *584*(5), 878-882.

Wopereis, S., Lefeber, D. J., Morava, E., & Wevers, R. A. (2006). Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin Chem*, *52*(4), 574-600.

Yamada, T., Fujimori, Y., Suzuki, A., Miyawaki, Y., Takagi, A., Murate, T., Sano, M., et al. (2009). A novel missense mutation causing abnormal LMAN1 in a Japanese patient with combined deficiency of factor V and factor VIII. *Am J Hematol*. 84(11), 738-42.

Yap, K. L., Ames, J. B., Swindells, M. B., & Ikura, M. (1999). Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins*, *37*(3), 499-507.

Yerushalmi, N., Keppler-Hafkemeyer, A., Vasmatzis, G., Liu, X. F., Olsson, P., Bera, T. K., Duray, P., et al. (2001). ERGL, a novel gene related to ERGIC-53 that is highly expressed in normal and neoplastic prostate and several other tissues. *Gene*, *265*(1-2), 55-60.

Zanetti, G., Pahuja, K. B., Studer, S., Shim, S., & Schekman, R. (2011). COPII and the regulation of protein sorting in mammals. *Nat Cell Biol*, *14*(1), 20-8.

Zhang, B., Cunningham, M. A., Nichols, W. C., Bernat, J. A., Seligsohn, U., Pipe, S. W., McVey, J. H., et al. (2003). Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex. *Nat Genet*, *34*(2), 220-225.

Zhang, B., Kaufman, R. J., & Ginsburg, D. (2005). LMAN1 and MCFD2 form a cargo receptor complex and interact with coagulation factor VIII in the early secretory pathway. *J Biol Chem*, *280*(27), 25881-6.

Zhang, B., McGee, B., Yamaoka, J. S., Guglielmone, H., Downes, K. A., Minoldo, S., Jarchum, G., et al. (2006). Combined deficiency of factor V and factor VIII is due to mutations in either LMAN1 or MCFD2. *Blood*, *107*(5), 1903-7.

Zhang, B., Spreafico, M., Zheng, C., Yang, A., Platzer, P., Callaghan, M. U., Avci, Z., et al. (2008). Genotype-phenotype correlation in combined deficiency of factor V and factor VIII. *Blood*, *111*(12), 5592-5600.

Zheng, C., Liu, H.-hui, Zhou, J., & Zhang, B. (2010A). EF-hand domains of MCFD2 mediate interactions with both LMAN1 and coagulation factor V or VIII. *Blood*, *115*(5), 1081-1087.

Zheng, C., Liu, H.-H., Yuan, S., Zhou, J., & Zhang, B. (2010B). Molecular basis of LMAN1 in coordinating LMAN1-MCFD2 cargo receptor formation and ER-to-Golgi transport of FV/FVIII. *Blood*, *116*(25), 5698-5706.