

Università degli studi della Tuscia di Viterbo
Dipartimento di Scienze Agrarie e Forestali

PhD course in Plant and Animal Sciences - XXXIII° cycle

Investigation of ERQC and ERAD using the plant as a model organism

s.s.d. BIO/04

PhD student: Andrea Lia
PhD coordinator: Prof. Roberta Bernini

Tutor: Dr Angelo Santino
Co-Tutor: Dr Pietro Roversi

This thesis was written in LaTeX during a world-wide pandemic.

Abstract

In Eukaryotes, about one-third of the genome codes for proteins that are destined for the secretory pathway; they emerge from the ribosomes into the Endoplasmic Reticulum (ER) lumen in an unfolded state. The lumen of the ER provides a highly specialised compartment for the folding of secretory proteins: Endoplasmic Reticulum glycoprotein folding Quality Control (ERQC) and ER-Associated Degradation (ERAD) preside over early stages of cellular glycoprotein secretion and maintain steady glycoproteostasis. ERQC and ERAD are central to health and disease. For example, ERQC-mediated ER retention and ERAD degradation of glycoprotein mutants can originate unfortunate consequences when a glycoprotein bears a mutation inducing a minor folding defect but does not abrogate function ("responsive mutant"): in these cases ERQC/ERAD can cause disease by blocking the secretion of the glycoprotein mutant, even though its residual activity would be beneficial to the organism. Host ERQC and ERAD are also exploited by some viruses to support the folding of viral glycoproteins or degrade anti-viral host glycoproteins: indeed, modulators/inhibitors of ERQC and ERAD exhibit anti-viral activity. The reliance of cancer cells on secreted glycoproteins also suggests that ERQC and ERAD components could constitute potential anticancer targets. Last but not least, ERQC and ERAD also play an important role in the Unfolded Protein Response (UPR) induced by abiotic stress.

A number of *in vitro* and *in cellula* studies in fungi and mammalian cells investigated inhibition and/or modulation of ERQC/ERAD enzymes as anti-cancer, anti-virals or rescuers of secretion in rare disease caused by responsive glycoprotein mutants. In this thesis, we investigate the plant as a simple and cheap alternative model for the study of ERQC and ERAD *in vivo*. In particular, we show that the *Arabidopsis thaliana* (*At*) innate immune response to bacterial peptides - and the availability of plant ERQC/ERAD mutant components - offer an easy means of assaying ERQC/ERAD function and modulation in a whole live organism. While structures of the ERQC usher and checkpoint (respectively ER α-Glucosidase II (GluII) and UDP-Glucose Glycoprotein Glucosyl Transferase (UGGT)) have been determined, no structure of any of the ER Degradation-Enhancing α-Mannosidase-like proteins (EDEMs) is available yet. In the second strand of work described in this thesis, we looked in the *Chaetomium thermophilum* genome for orthologues of the ERAD checkpoint heterodimer EDEM:PDI. We selected two so far uncharacterised genes which we called *Ct*HTM1P and *Ct*PDI1P . We demonstrated mannosidase activity for *Ct*HTM1P in plants and we cloned, expressed and purified from mammalian cells the *Ct*HTM1P :*Ct*PDI1P heterodimer and obtained its first 18 Å Cryo-Electron Microscopy (Cryo-EM) reconstruction.

Acknowledgements

The research activity described in this thesis was carried out and funded within the framework of the Italian Ministry for University and Research (MIUR) under the Project "Sostenibilità dei sistemi Agrari e ForEstali in ambiente Mediterraneo in un contesto di cambiamento globale, SAFE-Med" (Department of Excellence, Law 232/2016). Some of the work was partially supported by research grants awarded to Dr Pietro Roversi of the University of Leicester, England UK: a LISCB Wellcome Trust ISSF award, grant reference 204801/Z/16/Z; and a Wellcome Trust Seed Award in Science, grant reference 214090/Z/18/Z.

The following people helped with one or more experiments: Dr Maria De Benedictis, Mr Charlie J. Hitchman, Dr Lucia Marti, Dr Pietro Roversi, Dr Gábor Tax.

Prof. Richard Strasser kindly sent *A. thaliana* seeds; Prof. Ed Hurt generously shared *C. thermophilum* cDNA.

I thank Pietro Roversi, Angelo Santino, Roberta Bernini and Stefania Masci for their indefatigable supervision and patience; Nicole Zitzmann for having me into her lab as an Erasmus student and showing me a first glimpse of the scientific research world; Matthew Percival, Marta Falcicchio, Gábor Tax, Leonardo Feletto, Alessandro Stirpe, Simon Fung, Aurelia Scarano, Stefania De Domenico, Antonia Gallo, Snežana Vasiljević, Leone D'Amico and Roberta Ibba for helpful discussions and advice; Renato, Lucia, Matteo, Stefania, Mario, Gary and Iliana for all the support.

Contents

1 Introduction and background **8**

1.1 *N*-Glycosylation . . . 9

1.2 Endoplasmic Reticulum Quality Control (ERQC) . . . 11

1.2.1 ER α-Glucosidase I and II and the ERQC lectins . . . 12

1.2.2 UDP-Glucose glycoprotein Glucosyl Transferase (UGGT) . . . 13

1.2.2.1 Wild-type and mutant glycoproteins that are UGGT substrates in the plant. . . . 15

1.3 Endoplasmic Reticulum Associated Degradation (ERAD) . . . 15

1.3.1 ER Degradation Enhancing α Mannosidase-like proteins (EDEMs) . . . 15

1.3.2 The ERAD checkpoint heterodimer: EDEM:PDI . . . 18

1.3.2.1 Elusive disulfide bonds in the ERAD checkpoint heterodimer . . . 19

1.4 ERQC and ERAD modulation . . . 19

1.4.1 Modulation of ERQC/ERAD for rescue of secretion of rare disease associated responsive glycoprotein mutants . . . 19

1.4.2 Modulation of ERQC/ERAD as an anti-viral strategy . . . 20

1.4.3 Modulation of ERQC/ERAD as an anti-cancer strategy . . . 22

1.4.4 Modulation of ERQC/ERAD for a better response to abiotic stress . . . 23

1.5 The case for recombinant protein production, functional and structural characterisation of the ERAD checkpoint mannosidase . . . 24

1.5.1 The choice of an ERAD checkpoint mannosidase from a thermophilic organism 25

1.6 The plant as an *in vivo* model to study ERQC and ERAD . . . 25

1.6.1 Plants carrying responsive mutations in surface receptors as a model system to probe ERQC/ERAD . . . 25

1.6.2 Plant response to pathogens as a model system to probe ERQC/ERAD . . . 27

1.7 Open questions and aims of the thesis . . . 27

2 Materials and methods **30**

2.1 General cloning strategies . . . 30

2.1.1 Polymerase chain reaction (PCR) . . . 30

2.1.2 Gel electrophoresis . . . 31

2.1.3 DNA gel purification . . . 31
2.2 Plant Biology Experiments . . . 31
2.2.1 Plant material, growth conditions and treatments . . . 31
2.2.2 Seed sterilisation . . . 33
2.2.3 Plant growth and cultivation . . . 33
2.2.4 *N*B-DNJ αGlu II inhibition and Elf18/Flg22 elicitation . . . 33
2.2.5 cDNA reverse-synthesis . . . 33
2.2.6 Quantitative Reverse Transcription PCR . . . 34
2.2.7 Statistical Analysis . . . 34
2.2.8 Plant transient transformation . . . 35
2.2.9 Plant stable transformation . . . 36
2.2.10 Confocal Laser Scanning Microscopy . . . 37
2.3 DNA cloning and recombinant protein production . . . 37
2.3.1 List of buffers . . . 37
2.3.2 Searching for the *Ct* orthologues of *Sc*PDI1P and *Sc*HTM1P . . . 37
2.3.3 Cloning of *Ct*HTM1P and *Ct*PDI1P . . . 38
2.3.4 Cloning vector Litmus28i . . . 38
2.3.4.1 Cloning of *Ct*HTM1P full length . . . 39
2.3.4.2 Cloning of full length *Ct*PDI1P . . . 39
2.3.5 Expression vector pHLsec . . . 40
2.3.5.1 Cloning of four *Ct*HTM1P constructs . . . 42
2.3.5.2 *Ct*PDI1P intron excision . . . 43
2.3.6 Gibson assembly . . . 44
2.3.7 *E. coli* transformation . . . 45
2.3.8 Plasmid purification . . . 46
2.4 Protein expression and purification . . . 47
2.4.1 Harvesting of the supernatant and protein extraction from cells . . . 48
2.4.2 Western blot - wet transfer . . . 48
2.4.3 Immobilised metal affinity chromatography (IMAC) . . . 49
2.4.4 Size exclusion Chromatography (SEC) . . . 50
2.5 Protein crystallisation . . . 51
2.6 Mass spectrometry . . . 51
2.7 Cryo-EM data collection and processing . . . 51
2.8 Homology modelling . . . 52

3 Results **53**
3.1 The plant as a model to study ERQC . . . 53
3.1.1 *In planta* ERQC modulation . . . 53
3.1.2 *N*B-DNJ *in planta* toxicity assay . . . 54
3.1.3 ERCQ modulation in *At* mutant plants . . . 55

3.1.4 Bacterial peptides response assay . . . 57
3.1.5 The level of *BiP3* in *rsw3* is higher than in WT and *psl4-2* plants . . . 61
3.2 Structural and functional characterisation of the ERAD checkpoint heterodimer . . . 62
3.2.1 Plant's ERAD . . . 63
3.2.2 *Ct*HTM1P and *Ct*PDI1P are likely thermostable orthologues of known ERAD checkpoint mannosidases . . . 63
3.2.3 *Ct*HTM1P recomplementation of the *mns4-1/mns5-1/bri1-5* triple mutant genetic background . . . 65
3.2.4 Expression of *Ct*HTM1P constructs for recombinant protein production . . . 67
3.2.5 Purification of *Ct*HTM1P$_D$ . . . 70
3.2.5.1 IMAC affinity chromatography . . . 70
3.2.5.2 Size exclusion chromatography . . . 71
3.2.6 Co-expression of *Ct*HTM1P$_D$:*Ct*PDI1P . . . 72
3.2.7 Co-purification of *Ct*HTM1P$_D$:*Ct*PDI1P . . . 72
3.2.7.1 IMAC affinity chromatography . . . 72
3.2.7.2 MALDI-MS/MS protein identification . . . 73
3.2.7.3 Size-exclusion chromatography . . . 74
3.2.8 *Ct*PDI1P structure determination . . . 75
3.2.8.1 Crystallisation of *Ct*PDI1P . . . 76
3.2.9 *Ct*HTM1P$_D$:*Ct*PDI1P complex structure determination . . . 77
3.2.9.1 Crystallisation of *Ct*HTM1P$_D$:*Ct*PDI1P complex . . . 77
3.2.9.2 Cryo-Electron Microscopy of *Ct*HTM1P$_D$:*Ct*PDI1P complex . . . 78

4 Discussion **81**
4.1 The plant as a model to study ERQC . . . 81
4.2 Structural and functional characterisation of the ERAD checkpoint heterodimer . . . 84

5 Conclusions **87**
5.1 Future perspectives . . . 88

A Partial catalytic Cys oxidation of human GAPDH to Cys-sulfonic acid **91**
A.1 Introduction . . . 91
A.2 Experimental Procedures and Results . . . 92
A.2.1 Protein expression . . . 92
A.2.2 Purification of Cys-sulfonated *Hs*GAPDH . . . 93
A.2.2.1 IMAC affinity chromatography . . . 93
A.2.2.2 Size-exclusion chromatography . . . 94
A.2.2.3 Size exclusion Chromatography with Multi Angle Laser Scattering (SEC-MALS) . . . 94
A.2.3 Mass spectrometry protein and Cys-modification identification . . . 95

A.2.4 Cys-sulfonated *Hs*GAPDH structure determination 98
A.2.4.1 Protein crystallisation 98
A.2.4.2 X-ray diffraction collection and processing 98
A.2.4.3 Structure determination and refinement 99
A.3 Discussion 101

B Modulation of ERQC and ERAD: A Broad-Spectrum Spanner in the Works of Cancer Cells? 105
B.1 Introduction 105
B.2 ERQC/ERAD and cancer 105
B.2.1 ERQC and cancer 107
B.2.2 ERAD and cancer 109
B.2.3 Glycoproteins and Anti-cancer Strategies Focusing on ERQC/ERAD Modulation 111
B.3 Conclusions 112

List of activities 113

List of publications 114

Bibliography 115

Chapter 1

Introduction and background

Eukaryotic cells evolved a complex system of endomembranes that allows proteins newly synthesised by a cytoplasmic ribosome to reach a specific sub-cellular or extracellular destination. Every protein that travels along the secretory pathway carries a secretory signal that directs it to traverse a series of sub-cellular compartments in which it is post-translationally modified, checked for its fold quality and eventually targeted to its secretory destination [1].

Cellular compartments that are part of the secretory pathway are in constant communication through specialised vesicles, which continuously travel between their membranes [1]. Such system of endomembranes is highly conserved among eukaryotic organisms. Cargo traffic through these vesicles is tightly organised: a molecule enters the secretory pathway from the endoplasmic reticulum (ER) and it moves towards the Golgi apparatus - for example to lysosomes (vacuoles in a plant cell), the Plasma Membrane (PM) and/or the extracellular space [2]. High specificity of transport is guaranteed by the presence of coat-proteins on the surface of the vesicles: coated vesicles act as carriers that mediate uni- or bi-directional transport between two adjacent membranes in the secretory pathway [3, 4]. Coat-proteins can be of three kinds: SNARE, COPI and COPII. They are usually lost the moment a vesicle merges with the membrane of the target compartment [5, 6].

In Eukaryotes, about one-third of the genome codes for proteins that are destined for the secretory pathway; these polypeptides emerge from ER-membrane associated ribosomes into the ER lumen in an unfolded state [7]. The lumen of the ER provides a highly specialised compartment for the folding of secretory proteins. Their conformational maturation is a complex process determined not only by the amino acid sequence but also by post- and co-translational modifications, and by a variety of ER chaperones and folding enzymes [7]. The ER possesses efficient (but not-fully understood yet) quality control and degradation systems to ensure that transport is limited to properly folded and assembled proteins [8]: Endoplasmic Reticulum glycoprotein folding Quality Control (ERQC) and ER-Associated Degradation (ERAD) preside over early stages of cellular glycoprotein secretion and maintain steady glycoproteostasis. The secretory fate of each protein is surveyed and depends on ERQC/ERAD mediated modifications of polysaccharide chains covalently attached to the folding proteins in the ER (N-linked glycans).

1.1 *N*-Glycosylation

Among the covalent modifications of a newly-synthesised protein, the co-translational addition of *N*-linked oligosaccharides in the form of a 14-saccharides core unit ($Glc_3Man_9GlcNAc_2$) is one of the most common [9]. Without added sugars, secretory proteins misfold/aggregate, cannot proceed to the Golgi and get eventually degraded [7].

The central role of the *N*-glycosylation can be underlined by the fact that 75.8% of all human proteins have at least one *N*-glycosylation consensus sequence (*aka* sequon) NXS/T (where X can be any amino acid but proline) and thus may be glycoproteins, with an average of approximately one sequon per every 118 residues [10]. Cytosolic proteins are not glycosylated in the same way as secretory ones. Amongst *bona fide* glycoproteins, about 90% carry either *N*-linked carbohydrate units alone or both *N*- and *O*-linked ones [11]. Figure 1.1 shows the structure and the composition of the ER high-mannose oligosaccharide $Glc_3Man_9GlcNAc_2$.

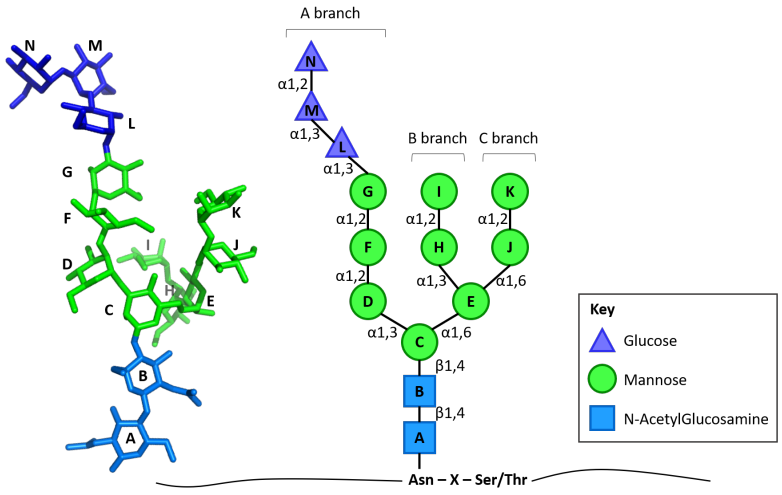


Figure 1.1: The ER high-mannose *N*-glycan. **Left**: one conformation of the $Glc_3Man_9GlcNAc_2$ glycan [12, 13]. **Right**: the *N*-linked glycan represented with the standard shapes and colours for monosaccharides in glycans [14]: light blue squares: N-acetyl glucosamine (GlcNAc); green circles: mannose (Man); blue triangles: glucose (Glc). *N*-glycosylation motif: X is not Pro.

The oligosaccharide precursor $Glc_3Man_9GlcNAc_2$ is kept in the ER lumen covalently linked to a lipidic molecule called dolichol [15]. The entire oligosaccharide chain is assembled on the ER membrane-inserted dolichol prior to transfer to the *N* residue of the *N*-glycosylation sequon of a newly-synthesised protein. The saccharide units contributing to the *N*-linked oligosaccharide are activated in the cytosol by covalent binding to diphospho-nucleotides. The latter molecules donate their sugar to the dolichol precursor in a specific order. During this process, once the $Man_5GlcNAc_2$ has been built on the cytoplasmic side of the ER membrane, the dolichol precursor is flipped to the lumenal side, and four additional Man residues are then sequentially added from a

dolichol-P-Man donor. These four Man additions are catalysed by the sequential action of the ER-membrane associated glycosyl-transferases ALG3, ALG9, ALG12, and again ALG9) to generate $Man_9GlcNAc_2$-PP-Dolichol. Then ALG6, ALG8, and ALG10 catalyse the sequential addition of three Glc residues to form the dolichol precursor bearing $Glc_3Man_9GlcNAc_2$ [16, 17]. After this, the entire oligosaccharide chain is *en bloc* transferred by the enzyme Oligosaccaryl Transferase (OST) to each *N*-glycosylation sequon in the newly-synthesised protein, each transfer happening in a single step. OST is tightly associated with the eukaryotic translocon Sec61 and it has its catalytic site on the lumen side of the ER membrane [18]. Figure 1.2 shows the large complex formed by the ribosome on the cytosolic side of the ER membrane, the translocon and the OST enzyme.

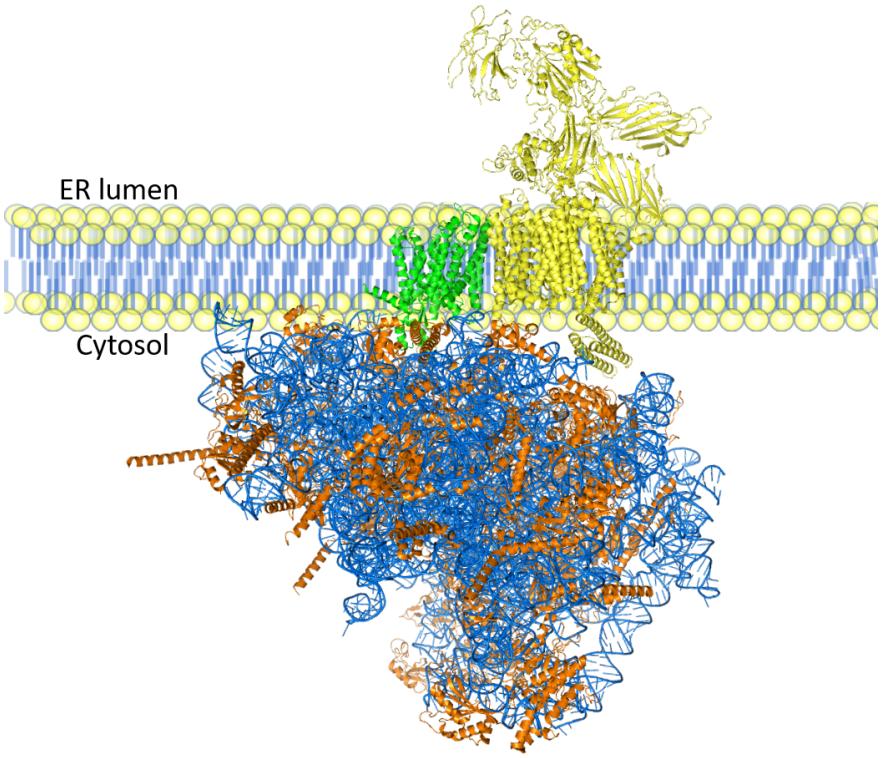


Figure 1.2: *N*-glycosylation in the ER. A model for the structure of the Sec61 translocon (green) in association with the ribosome (blue: ribosomal RNA; orange: ribosomal proteins) and OST-A (yellow) on the ER membrane (the latter not to scale). The model was assembled by overlapping the subunit STT3A from the Cryo-EM structure of human OST-A (PDB entry 6S7O) onto the same subunit in the Cryo-EM Structure of a mammalian OST-A bound to Sec61 and a non-programmed 80S Ribosome (PDB entry 6FTJ). The displayed OST-A is the one from PDB entry 6S7O.

All the different species of *N*-linked glycans on mature glycoproteins are the result of later modifications (mostly taking place in the Golgi) of the oligosaccharide precursor just described. When a glycoprotein is in the ER *N*-linked glycan modifications can be, essentially, of three kinds:

removal/addition of a glucose residue by ERQC enzymes, or removal of a mannose residue by ERAD enzymes.

1.2 Endoplasmic Reticulum Quality Control (ERQC)

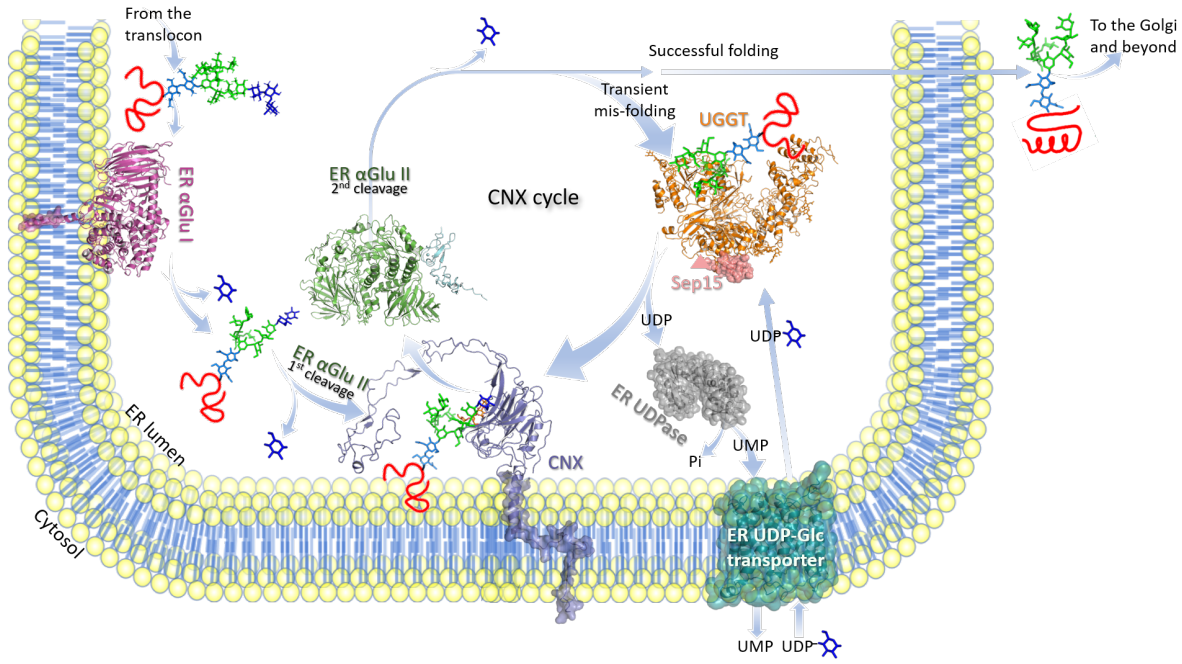


Figure 1.3: Structural view of ERQC. The ER trajectory of the folding stages of a glycoprotein, bound to a glycan of variable composition, is followed through its interactions with ERQC components, either ER-lumenal or ER-membrane associated. Proteins whose structures have been described in the literature for at least one eukaryotic species are in cartoon representation, while homology models of proteins of unknown structure are in semitransparent surface representation. Proteins for which no orthologue of known structure exists are represented as arbitrary 2D shapes. The red squiggly symbol represents the polypeptidic part of a glycoprotein on its way to folding (either incompletely folded, transiently misfolded, or terminally misfolded), while the red helix and two-stranded-β-sheet symbol represents a glycoprotein that has successfully attained its native fold. The colours of the monosaccharides in the glycan are from [14]. The symbolic N-linked glycan on the glycoprotein is not to scale with the proteins, and the proteins themselves are not represented to exact relative scale to one another. The nascent glycoprotein in the upper-left corner has just been synthesised into the ER by a translocon-associated ribosome, see Figure 1.2.

1.2.1 ER α-Glucosidase I and II and the ERQC lectins

Regardless of the folding state of a newly synthesised glycoprotein, ER α-glucosidase I (*aka* ER αGlu I, the product of the *mogs* gene) acts on an *N*-linked $Glc_3Man_9GlcNAc_2$ glycan, removing the outer glucose and enabling further deglucosylation by ER αGlu II [19] (Figure 1.3). ER α Glu I is a singly *N*-glycosylated, type II membrane protein of 837 amino acids in humans [20] (Uniprot entry Q13724). The cytoplasmic sequence is short and is likely responsible for the ER localisation of the protein whilst the lumenal portion is comprised of two domains: an N-terminal domain of unknown function and the C-terminal catalytic domain belonging to the GH63 family of protein folds [21]. All data point to *N*-linked glycans on nascent glycoproteins being de-glucosylated by ER αGlu I almost immediately upon emerging from OST and this led to the speculation that it could be associated with OST and therefore the translocon complex [22]. This notion in turn begs the question of the role of the outer glucose residue in the *N*-linked glycan: if it is to be immediately removed, why is it added in the first place? The outer glucose residue is added to the *N*-linked glycan dolichol precursor by the ALG10 α-1,2-glucosyltransferase. No congenital disorder of glycosylation caused by mutations in the ALG10 gene has so far been identified [23] and not all eukaryotes bear this outer glucose on their *N*-linked glycans [24]. The only apparent role for the outer glucose is promoting efficient *N*-linked glucosylation by OST [25]. One can only speculate that some eukaryotes could dispose of the *alg10*/*mogs* gene pair by evolving OST subunits that do not need the outer Glc residue on the dolichol glycan precursor, but we do not know of published studies looking at the (co-)distribution of *alg10* and *mogs* genes in eukaryotic genomes. What is clear though is that - once ALG10 adds the outer glucose - ER αGlu I is vital for entry of glycoproteins in the ERQC system, as evidenced by the sick phenotypes of fission yeasts Glu I mutants [26], the Type IIb congenital disorder of glycosylation associated with ER αGlu I mutations in humans [27, 28] and an *Arabidopsis thaliana* ER αGluI mutant suffering from embryonic lethality - in turn attributed to a reduction in cellulose deposition [29].

ERQC uses a glucose-on/glucose-off Yin-Yang system [30], shuttling glycoproteins between ER α-Glucosidase II (ER α-Glu II) and UDP-Glucose Glycoprotein Glucosyltransferase (UGGT) to monitor glycoprotein folding status [30, 31].

As the first ER αGlu II-catalysed Glc cleavage controls entry into the CNX cycle, and the second ER Glu II-catalysed cleavage controls exit from it, we refer to this enzyme as the "ERQC usher". ER Glu II recognises a glycoprotein that carries a $Glc_2Man_9GlcNAc_2$ glycan. A first interaction between ER αGlu II and its glycoprotein substrate leads to the hydrolysis of the α-1,3 bond between the two glucose moieties (Figure 1.3, ER αGlu II "first cleavage"). After this cleavage, the resulting mono-glucosylated glycan can associate with the ERQC lectins (Calnexin (CNX) or Calreticulin (CRT)).

Calreticulin and the luminal domain of calnexin have a high-degree of amino acid sequence similarity, and their functions are considered to be largely interchangeable. They can even associate simultaneously with the same glycoprotein [32]. A soluble form of calnexin (residues 1–387) can functionally replace its membrane-bound counterpart; however, when calnexin was expressed as

a soluble protein in L cells, the pattern of associated glycoproteins changed to resemble that of calreticulin; conversely, membrane-anchored calreticulin bound to a similar set of glycoproteins as calnexin; therefore, the different topological environments of calnexin and calreticulin are important in determining their distinct substrate specificities [33].

The ERQC lectins are in association with several chaperones such as ERp57 and ERp29, protein disulfide isomerases the cyclophilin B peptidyl proline isomerase [34, 35, 36, 37] and Erp72, which also displays peptidyl proline isomerase activity when interacting with CRT indirectly via PDIR [38]. Such complexes help a mono-glucosylated glycoprotein to attain its proper fold. CNX and CRT may also promote folding by recruiting folding catalysts to the vicinity of the folding glycoprotein [39, 40]. Interestingly, CNX/CRT can also prevent aggregation of various unfolded proteins *in vitro* [41]. Of course, CNX and CRT interaction with mono-glucosylated substrates prevents their export to the Golgi, so that the ERQC lectins are responsible for ER retention of incompletely folded glycoprotein.

ER αGlu II acts a second time on the same glycoprotein (Figure 1.3, "second cleavage"): this second catalytic cleavage leads to the removal of the inner glucose from the glycan. At this stage, the glycoprotein - now bearing a $Man_9GlcNAc_2$ glycan - is freed from the lectins (CNX/CRT) and their associated chaperones. The glycoprotein bearing no mono-glucosylated glycans can now leave the ER, moving towards the Golgi apparatus, and from there proceed further through the secretory pathway. In the Golgi the *N*-glycans can be heavily modified to become "complex glycans". Modifications that are commonly seen include fucosylation, mannosylation or the addition of galactose or *N*-acetylneuraminic acid [42].

1.2.2 UDP-Glucose glycoprotein Glucosyl Transferase (UGGT)

UGGT (the ERQC misfolding checkpoint enzyme) selectively re-glucosylates misfolded glycoproteins, thus mediating their binding to ER lectins that retain them in the ER. ER α-Glu II then removes the glucose added by UGGT and frees the glycoprotein from the ER lectins. The UGGT/ER α-Glu II pair gives rise to the repeated cycles of re-folding association with ER lectins and associated chaperones known as calnexin cycle [43, 37, 44], Figure 1.3.

How does UGGT sense defects in the fold of its substrates? The UGGT structure comprises a catalytic domain, two β-sandwich domains and four thioredoxin-like domains [45](see Figure 1.4). The latter domains are thought to be responsible for misfold recognition but the molecular mechanisms underpinning this process remain unexplained as yet. The UGGT structure has a good degree of intrinsic mobility and this feature seems to be a key-factor that allows the enzyme to successfully survey the folding of a vast variety of substrates [46]. The way UGGT recruits a misfolded glycoprotein is still unclear but it is tempting to speculate that the enzyme uses a strategy that we can describe as "it takes one to know one" [47]: UGGT may recognise surface hydrophobic residues on misfolded substrates, which is generally a symptom of a folding defect, and this recognition may in turn be mediated by UGGT domains that expose hydrophobic patches themselves.

Upon the discovery of the complex fold and topology of UGGT (see Figure 1.4), the enzyme was

speculated to be sufficiently tricky to fold that it would be subjected to re-glucosylation by itself [45] - a hypothesis recently validated by the first study listing UGGT1 clients in HEK293 cells: UGGT1 is indeed reglucosylated by itself in UGGT2 KO human cells [48]. Whether UGGT1-mediated UGGT1 reglucosylation is simply necessary because of the complexity of the enzyme's fold, or is the hallmark of intrinsinc misfolding in UGGT1 - remains an open question. However, UGGT1 carries de-mannosylated glycans that are a hallmark of ERAD - an observation compatible with UGGT1 possessing itself misfolded regions [49, 50].

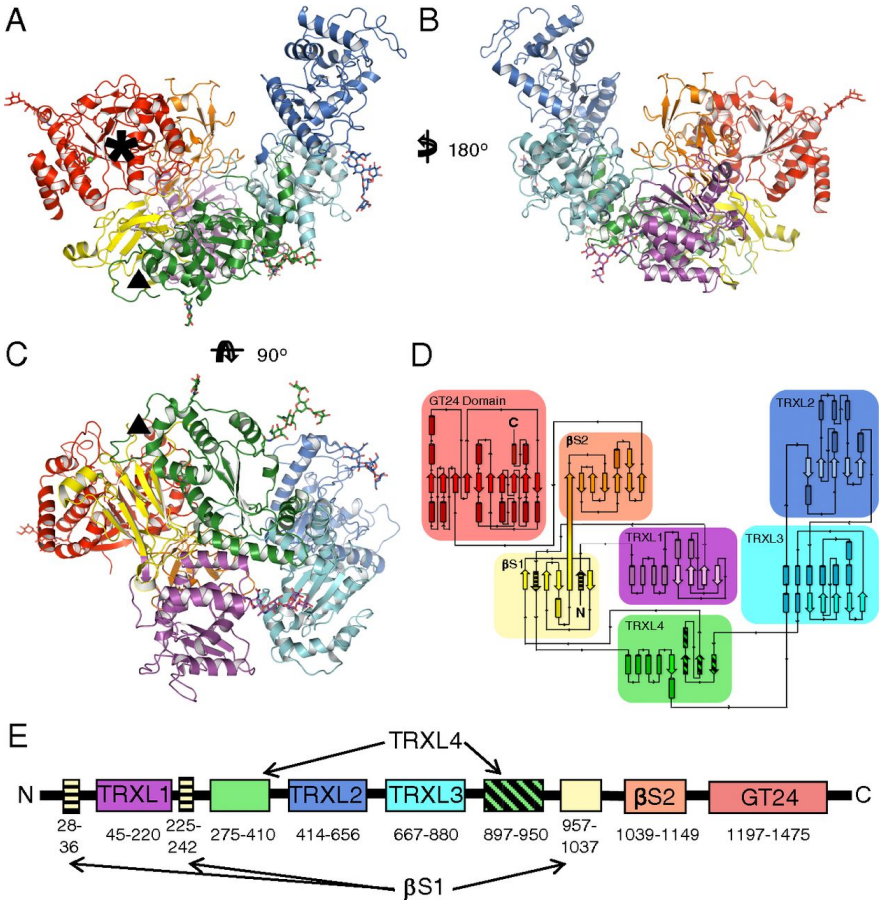


Figure 1.4: Structure and topology of *Ct*UGGT (from [45]). **A**–**C**: Three orthogonal views of *Ct*UGGT crystal structure. **D,E**: The 2- and 1-dimensional topological diagrams of *Ct*UGGT, respectively .

UGGT function is widely conserved across eukaryotes. A few organisms appear to lack a *uggt* gene: some protists that make either very short *N*-linked glycans or no *N*-linked glycans at all (such as *Tetrahymena*, *Giardia*, *Plasmodium* [51]) and - more surprisingly - the yeast *Saccharomyces cerevisiae* [52]. In *Schizosaccharomyces pombe*, UGGT is mostly needed in conditions of ER stress [53]. Nevertheless, *S. cerevisiae* KRE5 - when knocked-out - gives the same cell-wall defects observed in a UGGT-KO *S. pombe* strain and it is likely to be a diverged relative of UGGT [54]. Although most homozygous uggt1$^{-/-}$ mice die in the womb, some survive until birth: mouse embryonic fibrob-

lasts (MEFs) derived from viable uggt1$^{-/-}$ embryos propagate well, do not exhibit any significant morphological change, and do not express UGGT1 protein [55]. The same UGGT1-KO MEF cells show enhanced ER aggregation of misfolded glycoprotein mutants compared to WT cells [16]. Heterozygous UGGT1$^{(+/-)}$ knock-out mouse (possessing ∼50% of the wild-type enzyme levels) appear healthy [56].

1.2.2.1 Wild-type and mutant glycoproteins that are UGGT substrates in the plant.

In *A. thaliana*, the mutant *ebs1-3* carries a mutation in the *At*1g71220 gene (encoding *At*UGGT) [57]. The mutation affects the UGGT DQD motif that is absolutely conserved across all known UGGT sequences because it coordinates UDP-Glc in the active site of the enzyme (see PDB IDs 5H18, 6FSN). In the *At ebs1-3* mutant, lacking UGGT activity, it is possible to rescue the secretion of a defective form of the brassinosteroid receptor (*bri1-9*) that is normally retained in the ER. *At*UGGT seems to be also important for the correct folding of the plant innate immune receptor EFR - suggesting that EFR is a client of UGGT [58, 59, 60]. The localisation of the *bri1-9* gene product and EFR - and the plant phenotypes associated with them in WT and mutated form - are therefore ideal observables for testing UGGT activity in plants.

1.3 Endoplasmic Reticulum Associated Degradation (ERAD)

Given the high energy costs of protein synthesis [61], eukaryotic cells prioritise repair/refolding over protein degradation. However, continuous futile refolding attempts would titrate away chaperones and folding catalysts that are needed to fold newly synthesised proteins continuously entering into the ER, potentially reducing the ER folding capacity, increasing ER accumulation of misfolded and incompletely folded proteins, and disrupting ER proteostasis [17]. Thus, when a glycoprotein, despite several rounds of CNX/CRT-assisted folding, cannot achieve its proper conformation, it is extracted from ERQC by Endoplasmic Reticulum Associated Degradation (ERAD), ultimately leading to its retrotranslocation and degradation [62].

A glycoprotein enters ERAD when acted upon by one of the ER Degradation-Enhancing α-Mannosidase-like proteins (EDEMs): each of these enzymes removes a specific mannose residue from one or more of the glycans of a terminally misfolded glycoprotein. The actual N-glycan ERAD signal is created by the EDEMs enzyme's cleavage of the α1,2Man-α1,6Man linkage of the C-branch (1.1) to expose a hidden α1,6Man residue, subsequently recognised and bound by the highly conserved ERAD lectins, osteosarcoma amplified 9 (OS-9 in mammalian cells and Yos9 in yeast [17]) and XTP-3B [63]. Figure 1.5 illustrates the ERAD machinery.

1.3.1 ER Degradation Enhancing α Mannosidase-like proteins (EDEMs)

In multicellular Eukaryotes, the EDEM family contains three ER proteins: EDEM1, EDEM2 and EDEM3 that are stress-inducible members of the family of the glycosylhydrolase 47 enzymes (GH47), the same family that also includes ERMan I and other Golgi α-mannosidases [64].

EDEMs remove the four α1,2-linked mannose residues ("G", "I", "K" and "F" in Figure 1.1) from the glycan; the resulting glycoprotein carries a $Man_5GlcNAc_2$ glycan. A glycan $Man_{7/5}GlcNAc_2$ that exposes a α1,6-mannose linkage between Man "E" and Man "J" on the C-branch (see Figure 1.1) is recognised by the lectins OS9 and XTP-3B [65, 66], through their Mannose-6-phosphate Receptor Homology (MRH) domain. Via association of the de-mannosylated substrate with the lectin domains of OS9 or XTP-3B, the glycoprotein is delivered to the retrotranslocon complex formed by HRD1/HRD3/DER1/USA1 [67, 68, 69] (Figure 1.5). At this point the de-mannosylated glycoprotein moves from the endoplasmic reticulum to the cytosol, it gets ubiquitinated and earmarked for recruitment to the 26S proteosome, where it is finally degraded. Misfold recognition by EDEMs may also be a case of "it takes one to know one": for example, besides its mannosidase domain, EDEM1 has an intrinsically disordered region (IDR) involved in the interaction with tyrosinase mutants; domain recognition methods identified in the mouse EDEM1 sequence, an extension at the N-terminal (residues from 40 to 119, with highest reaching values between residues 48–93) that shows a high propensity for intrinsic disorder; further analysis showed that this highly disordered region is of low complexity, generally hydrophilic with only few patches of aromatic and hydrophobic motifs [70]. In the absence of the IDR, the interaction between EDEM1 and two misfolded mutants of the protein Tyrosinase was abolished and in HEK293T cells overexpressing EDEM1 and Δ-IDR EDEM1 the degradation of mutants (such as α-1AntiTrypsin Null Hong Kong, hereafter NHK) was considerably enhanced in the presence of EDEM1 but not in Δ-IDR EDEM1 cells [71]. So it is likely that the IDR may confer to EDEM1 the conformational plasticity and/or the molecular determinants needed for the EDEM1-misfolded substrate interactions [70, 71].

The α barrel architecture of the GH47 family is the main region involved in the glycan binding site of EDEM1. Homology modelling using as templates ER class I α1–2 mannosidase from *Saccharomyces cerevisiae* and *Homo sapiens* (PDB code 1DL2 and 1X9D, with template-target identity of 25% and 52% sequence similarity, respectively) has been used to identify the potential aminoacids involved in glycan recognition and processing; the two crystal structures were used to guide molecular docking of a high mannose $Man_9GlcNAc_2$ glycan in the active site of EDEM1; interestingly, all the aminoacids forming the inner region of the cavity are very well conserved in all mannosidases including EDEM1; by contrast, the family of EDEM sequences shows significant sequence variations in loops forming the outer part of the active site cavity, which probably recognises (and endows EDEMs with specificity for) various high-mannose structures [70].

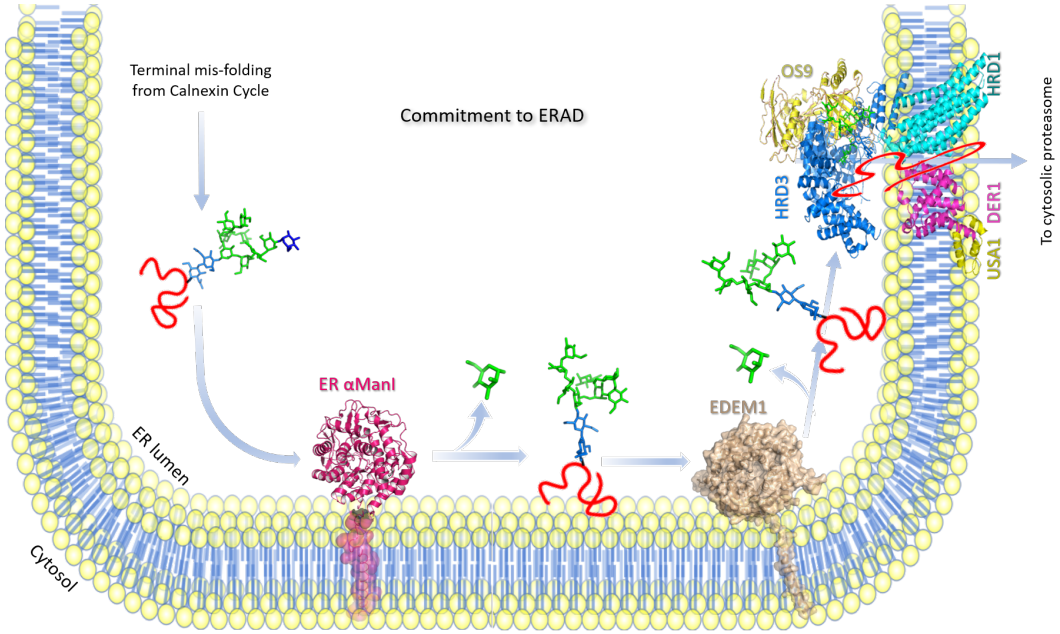


Figure 1.5: Structural view of ERAD. The ER trajectory of the folding stages of a glycoprotein, bound to a glycan of variable composition, is followed through its interactions with ERAD components, either ER-lumenal or ER-membrane associated. Proteins whose structures have been described in the literature for at least one eukaryotic species are in cartoon representation, while homology models of proteins of unknown structure are in semitransparent surface representation. Proteins for which no orthologue of known structure exists are represented as arbitrary 2D shapes. The red squiggly symbol represents the polypeptidic part of a glycoprotein on its way to folding (either incompletely folded, transiently misfolded, or terminally misfolded), while the red helix and two-stranded-β-sheet symbol represents a glycoprotein that has successfully attained its native fold. The symbols of the monosaccharides in the glycan are from [14]: blue squares, N-acetyl glucosamine (GlcNAc); green circles, mannose (Man); blue circles, glucose (Glc). The symbolic N-linked glycan on the glycoprotein is not to scale with the proteins, and the proteins themselves are not represented to exact relative scale to one another. Terminally misfolded glycoproteins are first demannosylated by ER αManI (magenta, PDB ID 1X9D) and then by EDEM1 (wheat), to yield a $Man_{5-7}GlcNAc_2$ glycan. The mannose-6-phosphate receptor homology (MRH) domain of OS-9 (yellow) binds the demannosylated glycan and recruits the terminally misfolded glycoprotein to the ERAD retrotranslocon assembled around the HRD1/HRD3 ubiquitin ligase (magenta, violet, green and cyan, yeast structures, PDB IDs 5V6P and 5V7V, on the right). A cytoplasmic proteasome then degrades the terminally misfolded ubiquitinated glycoprotein (not shown).

EDEM1 can recognise and bind misfolded glycoproteins in a glycan-independent manner - since the putative catalytic residues of its mannosidase-like domain are not required for binding to substrates [72, 71]. The observation that EDEM1 could form auto-regulatory complexes with OS-9, SEL1L, and HRD1 [71] leaves room for the interesting hypothesis that ERAD has a dynamic auto-

regulatory function. EDEM1 overexpression increases slightly the ratio of mannose removal from the C branch over one from the A branch, suggesting that the enzyme is involved in the trimming of the terminal mannose residue in branch C [73]. EDEM1 preferentially trims $Man_8GlcNAc_2$ to $Man_7GlcNAc_2$ and $Man_6GlcNAc_2$ to $Man_5GlcNAc_2$ [74]. EDEM2 on the other hand is required for conversion of $Man_9GlcNAc_2$ to $Man_8GlcNAc_2$, removing only the outermost mannose of the B branch (Mannose "I", Figure 1.1) [75, 76]. This catalyses the first mannose trimming step and thereby initiates ERAD. EDEM3 and EDEM1 are required for conversion of the just formed $Man_8GlcNAc_2$ into $Man_7GlcNAc_2$, $Man_6GlcNAc_2$ and $Man_5GlcNAc_2$ [76]. In addition, in WT cells treated with kifunensine, the cleavage of Mannose "I" was blocked; in contrast, cleavage of Mannose "I" increased in EDEM1- and EDEM3-Knockout (KO) cells, indicating that the latter enzymes are implicated as second-step enzymes in the de-mannosylation of the glycan [75].

1.3.2 The ERAD checkpoint heterodimer: EDEM:PDI

The *Saccharomyces cerevisiae* ER degradation-enhancing mannosidase has been characterised as a heterodimer called HTM1P:PDI1P, comprising the α(1,2)-mannosidase HTM1P subunit and the disulfide isomerase PDI1P subunit; it appears that the HTM1P mannose-hydrolase domain is the N-terminal one, while a large C-terminal domain likely mediates association with PDI1P [77, 78, 79]. *Sc*PDI1P stably associated with *Sc*HTM1P may facilitate ERAD by reducing disulfide bonds in misfolded substrate proteins that are recognised or processed by *Sc*HTM1P.

In mouse, EDEM1 was found to reside in a complex with ERdj5 and BiP [80]. ERdj5 possesses reductase activity and a J domain, which recruits the ER Hsp70 chaperone BiP. It is possible that such complex might confer the ERAD checkpoint the ability to recognise misfolded glycoproteins with mispaired disulfides. The oxidoreductase proteins (PDI1P/ERdj5), through their TRXL domains, may prevent disulfide-linked aggregation and the substrates to fold by maintaining them in reduced states in order to enhance ERAD by increasing their retrotranslocation rates [80]. The proven association of EDEM1 with the ERDJ5 PDI [81, 80] in mouse suggests that ERDJ5 is the most likely mammalian orthologue of yeast PDI1P but no structure of either the ERDJ5:EDEM1 or HTM1P:PDI1P complex exists. Another PDI called TXNDC11 is reported to be a disulfide isomerase involved in ERAD and it was found interacting with EDEM1, EDEM2 [74, 76] and EDEM3 [82]. Complexation with TXNDC11 increases EDEM1 and EDEM2's mannosidase activity [74, 76].

Although *Sc*PDI1P is a soluble protein located in the ER lumen, the structure of *Sc*PDI1P shows several patches of surface-exposed hydrophobic residues. The point mutation Leu313Ala in *Sc*PDI1P prevents complex formation of the oxidoreductase with *Sc*HTM1P, diminishes mannosidase activity, and delays degradation of unfolded glycoproteins *in vivo* [79]. The *Sc*PDI1P:*Sc*HTM1P interface requires several residues in the C-terminal domain of *Sc*HTM1P (Asp607, Glu627, Trp636) and is further facilitated by an inter-molecular disulfide bond involving either Cys579 or Cys644 in the C-terminal domain of *Sc*HTM1P. The Cys65–Cys445 disulfide bond in the MHD, once formed, stabilises the prolonged association of *Sc*PDI1P with *Sc*HTM1P [83].

1.3.2.1 Elusive disulfide bonds in the ERAD checkpoint heterodimer

As mentioned above, a disulfide bond between Cys65 and Cys445 in the mannosidase homology domain (MHD) of *Sc*HTM1P is essential for the ERAD activity, and its formation requires *Sc*HTM1P to stably associate with *Sc*PDI1P in the ER lumen. Indeed, a *Sc*HTM1P mutant lacking the entire C-terminal domain failed to form a disulfide bond between Cys65 and Cys445 as expected [83].

The intermolecular disulfide bond between *Sc*HTM1P and *Sc*PDI1P appears stable at least for 60 min, suggesting that the *Sc*HTM1P:*Sc*PDI1P complex does not merely represent a folding and/or disulfide bond forming intermediate of *Sc*HTM1P; instead, the disulfide bond between *Sc*HTM1P Cys65 and Cys445 appears to stabilise the noncovalent interactions between *Sc*PDI1P and *Sc*HTM1P, thereby stabilising the mixed disulfide between *Sc*PDI1P and *Sc*HTM1P [83]. Although mammalian EDEMs contain conserved Cys65 and Cys445 in the MHD, they lack the C-terminal domain found in *Sc*HTM1P. Besides, mammalian EDEMs possess additional cysteines that are not conserved in *Sc*HTM1P. Therefore the mechanisms of formation of the possible disulfides between subunits of mammalian ERAD checkpoint heterodimers may be considerably different from those observed in *Sc*HTM1P.

On the PDI side, it is not clear either which Cys is involved in the inter-molecular disulfide bridge. *Sc*PDI1P has only six Cysteines, and four of them are part of the two CGHC consensus sequences, which characterise the active sites, forming two disulfide bonds: Cys61-Cys64 and Cys406-Cys409. A third disulfide bond is between the residues Cys90-Cys97 but it is likely to be transient [84]. The observed oxidation states of the PDI catalytic cysteines indicate that yeast *Sc*PDI has on average ∼1.4 free sulphydryl groups per molecule [85]. So either *Sc*PDI1P Cys90 or Cys97 is a good candidate for forming the disulfide bridge to *Sc*HTM1P.

1.4 ERQC and ERAD modulation

Pharmacological modulation of ERQC and ERAD has been relatively understudied, partly because of the risks associated with targeting housekeeping machineries. Yet, were selective and potent ERQC/ERAD inhibitors available, they would likely open novel therapeutic avenues in several areas of medical science. ERQC/ERAD enhancers have potential applications also in biotechnology and agricultural science. In this section, we summarise the potential impact of ERQC/ERAD modulators.

1.4.1 Modulation of ERQC/ERAD for rescue of secretion of rare disease associated responsive glycoprotein mutants

The wonderfully efficient ERQC and ERAD protein-folding and degradation machineries in the ER of eukaryotic cells ensure that only correctly folded glycoproteins can exit the ER, proceed to the Golgi, and from there continue along the secretory pathway toward their cellular or extracellular destinations [86]. The stringency of the ERQC/ERAD systems is of great advantage to healthy cells. They allow time for complex glycoproteins to fold in the ER and prevent premature secretion

of incompletely folded species. In the background of a misfold-inducing missense mutation in a secreted glycoprotein gene, the resulting misfolded glycoprotein is either retained in the ER by ERQC or degraded by ERAD machinery [87]. ERQC-mediated ER retention and ERAD degradation of glycoprotein mutants bear particularly unfortunate consequences when the mutation induces a minor folding defect but does not abrogate the function of the glycoprotein ("responsive mutant"). In these cases, ERQC/ERAD cause disease by blocking the secretion of the glycoprotein mutant, even though its residual activity would be beneficial to the organism [31].

Studies in *Arabidopsis* have confirmed that exogenous Kifunensine application blocks the degradation of at least two ER-retained mutant variants of BRI1, *bri1-5* and *bri1-9*, and a misfolded plant innate immunity receptor (EFR), providing additional support for a role of ERQC and ERAD checkpoint inhibitors to rescue secretion of misfolded glycoproteins carrying activity [88, 89, 90]. Of course, such molecules could be of use in many a rare disease, effectively representing broad-spectrum rescuers of secretion. As the therapeutic scope is limited to patients carrying responsive mutations, such application are an example of personalised/precision medicine.

1.4.2 Modulation of ERQC/ERAD as an anti-viral strategy

Most antiviral drugs target viral proteins critical to the viral life cycle. For example, the Hepatitis C receptor SR-BI seems to fold under UGGT1 control [91]. In recent years, increasing levels of resistance to direct-acting antivirals have become a major public health concern, highlighting the urgent need for the development of alternative treatments [92]. One appealing strategy to avoid antiviral drug resistance is the targeting of virus–host interactions [93]. Drugs that act by blocking specific host functions required by many different viruses have the additional potential to treat a wide range of viral infections and coinfections (broad-spectrum antivirals) [94].

The biosynthesis of enveloped viruses depends heavily on the host cell ERQC. This dependency exceeds the dependency of host glycoproteins, offering a window for the targeting of ERQC for the development of broad-spectrum antivirals [95]. Inhibitors of the ERQC machinery, used by most enveloped viruses for the correct folding of their surface glycoproteins [96], represent one class of promising host-targeting, broad-spectrum antivirals. Inhibition of the ERQC machinery, including α-GluII inhibition, causes viral glycoproteins to misfold and reduces virion secretion and/or infectivity [93]. The centrality of α-GluII to viral glycoprotein folding and secretion makes it an appealing target for host-targeting antivirals, with clinical relevance for broad-spectrum antiviral therapy. Some ER α-Glu iminosugar inhibitors are known to possess antiviral activity: castanospermine; DNJ; and the alkylated derivatives of DNJ, *N*-butyl-1-deoxynojirimycin (*N*B-DNJ, also known as miglustat) and M*ON*-DNJ (also known as UV-4) [97]. M*ON*-DNJ is the best-known inhibitor of α-GluII *in vivo* [98], is active against dengue and influenza [99, 100]. All these molecules interact with the enzyme similarly to glucose [95], Figure 1.6.

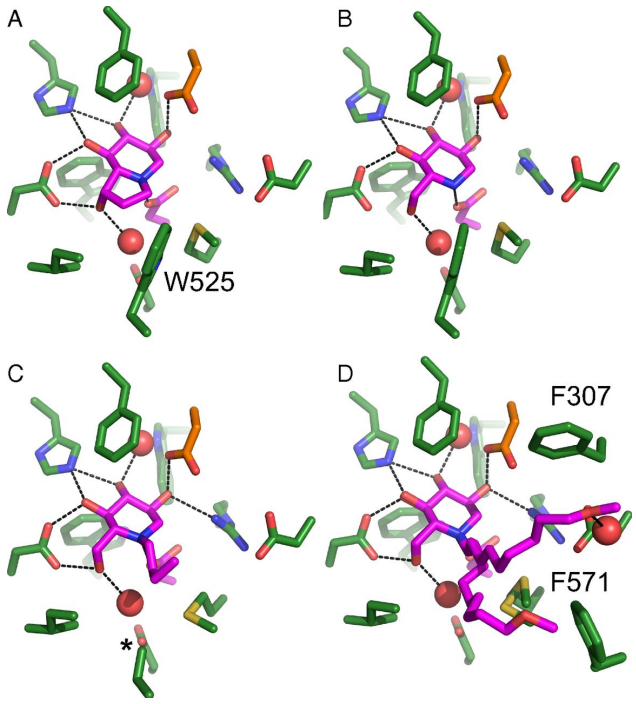


Figure 1.6: Crystal structures of *Mm* ER α-Glucosidase II in complex with iminosugar inhibitors (from [95]). Structures of complexes of *Mm* ER α-GluII (in green) with castanospermine **(A)**; DNJ **(B)**; *N*B-DNJ **(C)**; and M*ON*-DNJ **(D)** (in magenta). Figure taken from [95].

Recently, genome-scale RNAi and CRISPR-Cas9 screens have led to identification of many crucial host factors for flavivirus replication - active for example against dengue virus (DENV) and Zika virus (ZIKV) - including several key regulators of the HRD1-mediated ERAD pathway (see Figure 1.5). To find a molecule that can have broad-spectrum anti-flaviviral activity, a number of small molecules were tested as inhibitor of HRD1 complex-mediated dislocation; the molecule CP26 was observed to target HRD1 complexes and to inhibit ERAD of a misfolded glycoprotein like the Null Hong Kong mutant (NHK); *in cellula*, the presence of CP26 resulted in inhibition of the HRD1 complex-mediated dislocation and hampered DENV and ZIKV infection, yet increased the level of ER stress and UPR caused by the accumulation of misfolded NHK in the ER lumen [101]. On the other hand, ERAD proteins such as EDEM1 and EDEM3 were found to enhance the ubiquitylation of hepatitis C virus envelope (HCV E) proteins via direct physical interaction; knocking down of EDEM1 and EDEM3 increased the half-life of the HCV E2 protein, as well as virus production, whereas exogenous expression of these enzymes reduced the production of infectious virus particles; when HCV-infected cells were treated with kifunensine, a potent inhibitor of the ERAD pathway, the half-life of HCV E2 increased and so did virus production; kifunensine inhibited EDEM1 and EDEM3, thus blocking the ubiquitylation of HCV E2 protein [102].

Despite these results, the role of the ERAD machinery in the production of viruses is still largely unknown and it seems that modulation of ERAD affects only a subset of viruses. An example of this is represented by the life cycle of Japanese encephalitis virus (JEV), which is another member

of the *Flaviviridae* family; in contrast to HCV, kifunensine treatment had little effect on JEV production in infected cells or on the steady-state level of viral E glycoprotein: neither EDEMs nor ER ManI were found to interact with JEV E in cells, indicating no significant role of the ERAD pathway in the JEV life cycle [102].

The different sensitivity of DENV, ZIKV, HCV and JAV to ERAD inihibitors could be explained by different level of complexity in the fold of the respective viral glycoproteins. Indeed, an easy-to-fold viral glycoprotein does not necessarily interact with the ERAD machinery and sometimes not even with the calnexin cycle.

1.4.3 Modulation of ERQC/ERAD as an anti-cancer strategy

Plasticity, an intrinsic characteristic of healthy cells in biological contexts as varied as embryonal development [103], tissue development and repair [104], adaptation to injury [105], and wound healing [106], is also central to cancer initiation, progression, and metastasis. The proteins establishing and maintaining cancer plasticity are good anticancer drug targets in the fight against cancer initiation, progression, and therapy resistance itself [107]. Plasticity of cancer cells relies heavily on glycoproteins that traverse the secretory pathway, such as cell surface receptors and signalling molecules released in the extracellular medium [108, 109]. These secreted glycoproteins often respond to and steer changes in the surroundings of a cancer cell and contribute to tumour immunity [110], tumour growth, and cancer cell division, adhesion, and metastasis.

The reliance of cancer cells on secreted glycoproteins begs the question as to whether the ERQC and/or ERAD systems (together with the parallel misfolding-associated protein secretion system, MAPS [111]) could constitute potential anticancer targets. It is conceivable that ERQC/ERAD would make attractive targets for the treatment of cell malignancies [112], in that the fitness of the cancer cells, particularly those bearing a high secretory burden such as multiple myeloma cells [113], is critically dependent on the functional integrity of the ER, which in turn relies on ERQC/ERAD as ER stress-attenuating mechanisms.

The therapeutic value of pharmacological chaperones (small molecules specifically stabilising a misfolded glycoprotein as it traverses the ER) is already well established in a number of congenital glycoprotein misfolding endocrine and metabolic disorders [114], further supporting the idea that therapeutic modulation of ER glycoprotein folding and degradation systems could also be successfully applied to cancer treatment, at least in cases where ERQC-assisted glycoprotein folding and ERAD play a major role.

Importantly, while pharmacological chaperones are designed to bind individual misfolded glycoproteins, any drug targeting a specific ERQC/ERAD component would affect folding of all glycoproteins that are dependent on it for their folding/degradation. Given the unique and central role of ERQC/ERAD in the fate of hundreds of secreted glycoproteins - and remembering that plasticity of different cancers depends on different subsets of secreted glycoproteins - ERQC/ERAD modulating drugs may have the potential to represent broad-spectrum anticancer agents.

The modulation of ERQC/ERAD as an anti-cancer strategy is further discussed in Appendix

B of this thesis.

1.4.4 Modulation of ERQC/ERAD for a better response to abiotic stress

A cell carefully maintains the balance between protein folding demand and folding capacity. When that balance is disturbed by adverse environmental conditions - such as heat stress, salt stress, ageing and other abiotic stresses - unfolded and misfolded proteins accumulate in the ER, causing "ER stress", a condition that can trigger Unfolded Protein Response (UPR). UPR is a homeostatic mechanism to alleviate ER stress through transcriptional and translational events that reduce the production of secreted and membrane proteins and increase the synthesis of chaperones, foldases and other components involved in ERQC and ERAD systems. In this sense, ERQC and ERAD ultimately represent cell-sparing activities, and if they fail, then programmed cell death (PCD) may ensue [37]. For example, in the brain of *Drosophila melanogaster* ERAD activity decreases with ageing, and this leads to accumulation of misfolded protein in the ER lumen and consequent ER stress; the overexpression of EDEMs suppresses age-dependent neurodegeneration and behavioural decline and extends the lifespan without inducing toxicity and without affecting the UPR gene expression network [115]. Consistent with those results, UPR is closely related to salt, heat and drought stress in plants [116]. For example, overexpression of BiP proteins and/or CRT in *Nicotiana tabacum* leaves leads to better tolerance to drought [117, 118].

Protein folding is already challenging for the cell in ordinary circumstances but it becomes more difficult when environmental stressors can denature proteins and cause even more non-native folding to occur. For example, heat afflicts the folding of glycoproteins, as exemplified by heat shock-induced unfolding: when sensing excessive heat, the cell increases the number of molecular chaperones (Heat Shock Proteins - Hsps) to face the negative effects on proteins [119]. It is therefore not surprising that ERQC and ERAD machineries also play an important role in the response to heat, since they represent one of the main means of alleviating the stress level in the ER, acting on misfolded glycoproteins and promoting their interactions with the lectins CNX/CRT and the associated chaperones and/or relieving the ER of misfolded proteins by shunting them to degradation (ERAD). Moreover, it is important to remember that ERQC and ERAD components are ER glycoproteins themselves, and they are subjected to the misfold-inducing abiotic stress just like any other glycoproteins in the ER.

All in all, empowering ERQC and ERAD through overexpression of their key enzymes can be a strategy to achieve a better tolerance to abiotic stress. More specifically, we speculate that an interesting strategy to enhance and ensure the functioning of ERQC and ERAD could be the complementation of a cell or organism (*e.g.* a plant) with versions of ERQC/ERAD enzymes that are better prone to resist specific stresses. In particular, the stress caused by heat shock could be relieved using ERQC/ERAD enzymes orthologues that are thermo-resistant and therefore would ensure a more stable quality control and/or degradation system at higher-than-physiological ambient temperatures.

1.5 The case for recombinant protein production, functional and structural characterisation of the ERAD checkpoint mannosidase

There are many open questions about the early stages of commitment of misfolded glycoproteins to ERAD; in particular, it is not yet clear how EDEMs can accomplish their function in the extraction of a misfolded glycoprotein from the calnexin cycle and the precise role of EDEMs in the quality control machinery is unknown. For example, a misfolded glycoprotein can be intercepted by EDEMs when in association with calnexin: upon immunoprecipitation of cell lysates with antibodies against CNX and flu hemagglutinin (HA), pulse-labelled EDEMs-HA was co-precipitated by an anti-CNX antibody [120]. This observation cannot be explained by EDEM simply being a client of the calnexin cycle - because EDEM remained in the CNX complex even after 1 hour of chase - and because such association was not abrogated by treatment with castanospermine, a known inhibitor of ER glucosidases [120]. In the same set of experiments, the CRT:EDEMs association was observed for a chimera of CRT with the CNX transmembrane region, but not for CRT, suggesting that the CNX transmembrane domain be needed for CNX:EDEMs interaction; indeed, removal of the CNX transmembrane domain abrogated the CNX:EDEMs interaction [120]. The Intrinsic Disordered Region (IDR) of EDEM1 does not seem to be involved in such interaction: when cells were transfected with both EDEM1 and EDEM1-ΔIDR, it was observed through pulse-chase and immunoprecipitation with anti-CNX antibodies that all EDEM proteins were co-immunoprecipitated with CNX, indicating that indeed the IDR is not required for this interaction [70].

The ERAD pathway is a one-way street - since the removal of mannose residues from the glycan is not reversible; if EDEMs demannosylate a misfolded glycoprotein while it is interacting with the ER lectins and before it is recognised by αGluII, further access to the calnexin cycle is not possible at all for that molecule. Variations in the amounts of intra-cellular EDEMs do not affect the time of association of misfolded glycoproteins with CNX; and the overall maturation of glycoproteins with normal folding and the release from CNX of ERAD candidates was observed to be delayed only when kifunensine was used as inhibitor [121].

These observations leave room for questions about a putative competition for glycoprotein substrates between EDEMs and αGluII. We decided to express and purify recombinant ERAD checkpoint proteins in order to carry out their functional and structural characterisation, and generate hypotheses about their biochemical and biological roles, to be tested by further experiments. Availability of recombinant ERAD checkpoint proteins will also enable *in vitro* assays and pave the way to the determination of their structures. The structures in turn can form the basis for structure-based lead design for potent and selective modulation of early stages of ERAD (as discussed above in Section 1.4).

1.5.1 The choice of an ERAD checkpoint mannosidase from a thermophilic organism

As we have discussed in Section 1.3.2, the yeast ERAD checkpoint has been characterised as a heterodimer called *Sc*HTM1P:*Sc*PDI1P, comprising the α(1,2)-mannosidase HTM1P subunit (Uniprot entry MNL1_YEAST, described at `https://www.uniprot.org/uniprot/P38888`) and the disulfide isomerase PDI1P subunit (Uniprot entry PDI_YEAST, described at `https://www.uniprot.org/uniprot/P17967`). The heterodimeric complex has a mass in excess of 250 kDa and preferentially de-mannosylates misfolded glycoproteins [83, 79, 78, 122]. It is likely that the PDI TRXL domains confer to the mannosidase enzyme the ability to recognise misfolded glycoproteins and/or enhance its activity on mispaired disulfide-containing substrates [74].

Towards the elucidation of the structural basis for EDEM recognition of misfolded substrates, we set out to discover, clone and purify the ERAD checkpoint heterodimer, in order to study it by X-ray crystallography and/or Cryo-EM.

For recombinant protein production of the ERAD checkpoint proteins and their structural studies, we selected *Chaetomium thermophilum* (*Ct*, a thermophilic fungus that survives at temperatures of up to 60 °C) as the organism source, in order to maximise protein stability. Indeed, many recent studies of Eukaryotic proteins have profited from targeting orthologues from *Ct* [123, 124]. In particular, the first Cryo-EM and crystal structures of a UGGT were obtained for *Ct*UGGT [45]) and a crystal structure of ER αGlucosidase II was obtained for *Ct*GluII [125].

1.6 The plant as an *in vivo* model to study ERQC and ERAD

The high conservation of sequence and function of ERQC and ERAD components across Eukaryotes makes it possible to use fungi and plants for basic studies of ERQC and ERAD *in cellula* and *in vivo* [26], respectively, without the ethical and economical complications of mammalian animal models, and the limitations of *in cellula* studies. Any phenotype mediated by glycoproteins whose folding is controlled by ERQC, or whose degradation is controlled by ERAD, enables the study of ERQC and ERAD function *in vivo*.

1.6.1 Plants carrying responsive mutations in surface receptors as a model system to probe ERQC/ERAD

The roles of the *Arabidopsis thaliana* ERAD mannosidases MNS4 and MNS5 during ERAD of misfolded glycoproteins have been probed with plants carrying responsive mutations in surface glycoproteins. Specifically, the severe growth defect of the *Arabidopsis bri1-5* mutant is caused by a structurally defective but still functional BRI1-5 receptor [88] that is targeted to ERAD. Strasser et al. demonstrated that overexpression of MNS4 or MNS5 promotes BRI1-5 degradation, leading to an enhanced *bri1-5* dwarf phenotype. Moreover, *in planta* ERAD of BRI1 variants is impaired in the absence of MNS4 and MNS5, leading to a less severe *Arabidopsis bri1-5* dwarf phenotype [126].

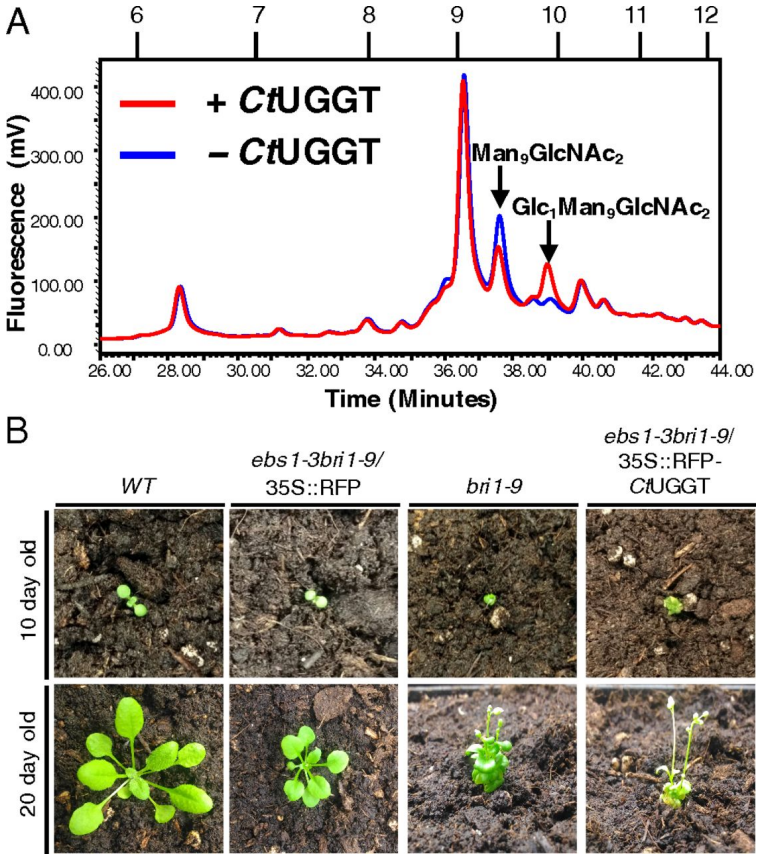


Figure 1.7: *Ct*UGGT is active *in vitro* at 37 °C and *in vivo* at 25 °C. **(A)** *in vitro* CtUGGT-mediated reglucosylation of urea-misfolded bovine thyroglobulin at 37 °C. The HPLC elution profile of the *N*-linked glycans obtained from urea-misfolded bovine thyroglobulin is in blue. In red, the glycans from the same protein after treatment with *Ct*UGGT at 37 °C. The amount of $Man_9GlcNAc_2$ is reduced and the amount of $Glc_1Man_9GlcNAc_2$ increases, proving that the enzyme is active. **(B)** *Ct*UGGT rescues the bri1-9 phenotype. Representative images of double mutant plants *ebs1-3/bri1-9* stably expressing the red-fluorescent protein *Ct*UGGT fusion 35S::RFP–*Ct*UGGT are compared with WT, *ebs1-3/bri1-9*, and *bri1-9* plants at 10 and 20 days after germination. The typical dwarf phenotype of the *bri1-9* mutant was restored in the *ebs1-3/bri1-9* genetic background (which carries an inactive *At*UGGT gene and therefore displays a phenotype similar to WT) by the functional *Ct*UGGT protein. Figure taken from [45].

In a paper published soon before the start of the work described in this thesis [45], we set out to test *Ct*UGGT activity *in vitro* and *in vivo*. *Ct*UGGT is active *in vitro* at 37 °C in reglucosylating urea-denatured bovine thyroglobulin, a well-known UGGT substrate [127, 128] and likely an ensemble of misfolded molecules (Figure 1.7A). To check the *in vivo* activity of the protein, we chose genetic recomplementation of *Ct*UGGT in the well-studied *A. thaliana* double mutant *At ebs1-3 bri1-9* plant, the first and only example of a living organism whose growth phenotype depends on UGGT activity [57]. The double mutant *At ebs1-3/bri1-9* plant is derived from the *At bri1-9*

strain, which carries a brassinosteroid hormone receptor gene *bri1* point mutation that impairs the fold of the BRI1 receptor but does not abrogate completely its activity (mutation *bri1-9*) [129]. The *bri1-9* point mutation leads to UGGT-mediated ER retention of the BRI1-9 receptor and consequently a dwarf phenotype. In the *At ebs1-3/bri1-9* strain, which carries an inactive *At*UGGT gene (mutation *ebs1-3*) as well as the *bri1-9* point mutation, a phenotype similar to wild type is observed: in the absence of UGGT, the misfolded and yet active BRI1-9 receptor is not retained in the ER; it is secreted and can signal for plant growth [57]. Upon genetic transformation of the *At ebs1-3/bri1-9* double mutant plants with the RFP–*Ct*UGGT fusion protein, the enzyme correctly localises to the ER and these plants revert to the *bri1-9* dwarf phenotype (Figure 1.7B), confirming that *Ct*UGGT is active in reglucosylating the misfolded BRI1-9 receptor *in vivo* at 25 °C [45].

1.6.2 Plant response to pathogens as a model system to probe ERQC/ERAD

Plants are sessile organisms that are under constant attack from microbes. The plant immune system is composed of surveillance systems that detect many microbial elicitors; in so doing, the plant immune system allows plants to switch from growth and development into defence mode, eventually enabling rejection of most potentially harmful microbes. Elicitors sensed by plants are typically essential structures for pathogen survival and are conserved among pathogens. The conserved microbe-specific molecules, referred to as microbe- or pathogen-associated molecular patterns (MAMPs or PAMPs), are recognised by the plant innate immune system through pattern recognition receptors (PRRs) [130]. Among the best studied bacterial MAMPs there is EF-Tu (Elongation Factor - Thermo unstable), which in *Arabidopsis thaliana* (*At*) is recognised by the EF-Tu receptor (EFR), a pattern-recognition receptor [131]. Plant cells recognise and bind EF-Tu, preventing genetic transformation and protein synthesis in pathogens such as *Agrobacterium tumefaciens* [131]. *At*Glu II point mutations and deletions impair the plant response to bacterial peptide elf18, derived from EF-Tu, likely because the EF-Tu receptor (EFR) or another protein in the elf18 signalling pathway requires calnexin cycle assisted folding [131]. Two more genetic studies showed that UGGT — like Glu II — is also important for the correct function of the plant innate immune response to elf18 [58, 59].

1.7 Open questions and aims of the thesis

ERQC and ERAD control early stages of glycoprotein folding and prevent progress of misfolded glycoproteins from the ER to the Golgi apparatus and down the secretory pathway. The established models for both systems centre around *N*-glycan modification as a proxy signal monitoring glycoprotein folding status and mediating ER retention: re-glucosylation/ER retention in ERQC vs. de-mannosylation/degradation in ERAD. The major difference is that ERQC is a glucose-on/glucose-off Yin-Yang cycle, while ERAD seems to be a one-way street of no return. Both ERQC and ERAD have evolved checkpoint enzymes that detect misfolded glycoproteins, and chemically modify *N*-glycans on them, thus enabling binding of the modified glycoprotein to ER lectins but

the molecular determinants of misfold recognition on the part of the ERQC and ERAD checkpoints (UGGT and EDEMs, respectively) remain elusive to date. In the ERQC/ERAD fields this and many open questions remain unanswered, such as:

1. how does a misfolded glycoprotein become an ERAD substrate instead of lingering forever in the calnexin cycle?
2. how and when does the switch between ERQC ("a new hope" for a glycoprotein to fold) and ERAD ("no hope, only degradation") happen?
3. why are EDEMs involved in the formation of heterodimers with PDIs (*e.g.* in the case of *Sc*PDI1P and TXNDC11, discussed above in Section 1.3.2)?
4. what is the contribution of the PDI to the mannosidase activity?
5. can we make a plant more resistant to abiotic stress, like heat-shock and drought, empowering the ERQC and ERAD machineries?
6. can we study ERQC/ERAD in *Arabidopsis thaliana*, using:
   - the responsive mutants *bri1-9* and *bri1-5* [129, 45, 126];
   - the plant response to microbe-associated molecular patterns (MAMPs) mediated by EFR [59]?

Starting from these broader questions we chose to work on the the following ones:

1. can we modulate ERQC *in planta* using *N*B-DNJ?
2. is the *A. thaliana* immune response (MAMP) useful to establish an *in vivo* assay of ERQC activity?
3. can we study the activity of EDEMs through the mutant *bri1-5*?
4. what is the structure of the EDEM:PDI ERAD checkpoint heterodimer?

Thus, to better characterise the plant ERQC and to design a fast and easy-to-use *in vivo* ERQC inhibition/modulation assay, we used *Arabidopsis thaliana* as a model organism and we studied the effect of *N*B-DNJ on the plant ERQC measuring the plant fresh-weight and its capability to produce a well-folded and active EFR receptor and thus to respond to bacterial elicitation. We compared the resulting phenotypes with the ones observed in *A. thaliana rsw3* and *pls5-1*, mutants with a reduced αER GluII activity, and we further characterised the effect of *N*B-DNJ by qRT-PCR analysis.

Aiming to produce a plant that is more resistant to abiotic stress, we started recomplementing the *mns4-1/mns5-1/bri1-5* triple mutant with the *Chaetomium thermophilum* thermostable EDEMs orthologue *Ct*HTM1P in order to evaluate the mannosidase activity of *Ct*HTM1P .

Finally, we expressed and purified the recombinant ERAD checkpoint mannosidase *Ct*HTM1P - together with its putative partner *Ct*PDI1P - for structural studies of the heterodimer, using crystallography and Cryo-EM.

Chapter 2

Materials and methods

2.1 General cloning strategies

The concentration of every DNA sample was measured using the NanoDrop instrument, absorbance at 260 nm. The primers used are reported in Table 2.1. In every figure showing DNA bands in agarose gel the marker used was GeneRuler™1kb DNA ladder (ThermoFisher SM0311), Figure 2.1.

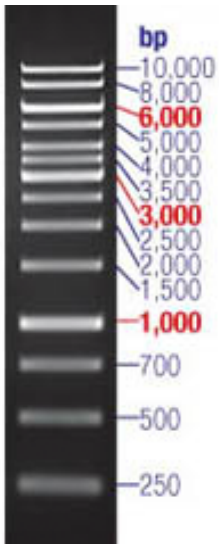


Figure 2.1: GeneRuler™1kb DNA ladder.

2.1.1 Polymerase chain reaction (PCR)

PCR mix always contained:

- Q5 High-Fidelity 1X Master Mix (New England Biosciences M0494) 25 μL;
- 2.5 μL of 10 μM Forward Primer stock to 0.5 μM final concentration;
- 2.5 μL of 10 μM Forward Primer stock to 0.5 μM final concentration;

- Template DNA 1ng;
- Nuclease-Free Water to 50 μL.

The PCR amplification was carried out in steps as follows:

1. Initial denaturation: T=98 °C, t=30";
2. Denaturation: T=98 °C, t=10";
3. Annealing: T= the annealing temperature of each pair of primers is indicated in Table 2.1, t=30";
4. Elongation: T=72 °C, t= chosen considering an average Q5 Hot Start polymerase rate of 1kb/30sec;
   Then back to step 2 for 33 cycles;
5. Final extension: T=72 °C, t=2 minutes.

The single colony-PCR was carried out as follows: Single colonies were picked from the plates and dipped in the same solutions for PCR amplification as described above, using the same primers and PCR protocol utilised for the amplification of the different inserts.

2.1.2 Gel electrophoresis

The quality of every reaction of amplification and/or restriction enzyme digestion was investigated through DNA gel-electrophoresis: 1% or 2% agarose gels were run at 100V for 1 hour in TAE buffer with the SYBR™Safe DNA Gel Stain (Thermo Fisher Scientific).

2.1.3 DNA gel purification

The bands corresponding to the amplified inserts and/or the restriction enzyme-linearised vectors were purified from the agarose gels using the GeneJet™gel extraction kit (Thermo Fisher Scientific).

2.2 Plant Biology Experiments

2.2.1 Plant material, growth conditions and treatments

In this thesis the *Arabidopsis thaliana* ecotype Col-0 was used and referred to as "wild-type (WT)". *At* Col-0, *At* pls5-1, *At* rsw3 and *At* pls4-2 seeds were obtained from The *Arabidopsis* Information Resource (TAIR).

Primers	Forward primer	Reverse primer	$T_{annealing}$
*Ct*HTM1P _Litmus28i	CATCGCGCGCACCGT ACGTCTCGAGATGAG CTGCAGATTTCTGGC	GGATCCAGATATCCT GCAGGAATTCTCCA TCAACTCCGCCGTC	63.8°C
*Ct*PDI1P _Litmus28i	CATCGCGCGCACCGT ACGTCTCGAGATGC ACAAGGCTCAGAGG	GGATCCAGATATCC TGCAGGAATTCTCAC AGCTCATCATGGGC	65.9°C
pHLsec_*Ct*PDI1P exon1	GATGGGTTGCGTAG CTGAAACCGGTTCC GACGTCATCCAGCTC	GGGGCGAAGAACTC GGCAAGGACCAGGTC	67.9°C
*Ct*PDI1P exon2_pHLsec	CTTGCCGAGTTCTT CGCCCCGTGGTGCG	GTGATGGTGGTGCT TGGTACCGGCAGTCT CCTTACCCTCCTCAG	72°C
*Ct*HTM1P _pHLsec A	GATGGGTTGCGTAGC TGAAACCGGTGATAG AGTTAGAGCCCTGCG	GTGATGGTGGTGC TTGGTACCTCCAT CAACTCCGCCGTC	65°C
pHLsec_N-termGFP B1	GATGGGTTGCGTAG CTGAAACCGGTATGG TGAGCAAGGGCGAG	TGCAGCTCATGGGG CCCTGGAACAGCAC	69°C
*Ct*HTM1P _pHLsec B2	CCAGGGCCCCATGAG CTGCAGATTTCTGGC	GTGATGGTGGTGC TTGGTACCTCCAT CAACTCCGCCGTC	66.4°C
pHLsec_*Ct*HTM1P C1	GATGGGTTGCGTAGC TGAAACCGGTATGAG CTGCAGATTTCTGGC	TGCTCACCATTCCA TCAACTCCGCCGTC	66.4°C
C-termGFP_pHLsec C2	GTGATGGTGGTGC TTGGTACCGGGGC CCTGGAACAGCAC	GTGATGGTGGTGC TTGGTACCGGGGC CCTGGAACAGCAC	69°C
pHLsec_N-termGFP D1	GATGGGTTGCGTAG CTGAAACCGGTATGG TGAGCAAGGGCGAG	TAACTCTATCGGGG CCCTGGAACAGCAC	69°C
*Ct*HTM1P _pHLsec D2	CCAGGGCCCCGATAG AGTTAGAGCCCTGCG	GTGATGGTGGTGC TTGGTACCTCCAT CAACTCCGCCGTC	65°C
*Ct*HTM1P _pDONR221	GGGGACAAGTTTGTACA AAAAAGCAGGCTTTATGA GCTGCAGATTTCTGGCC	GGGGACCACTTTGTACA AGAAAGCTGGGTCTCCA TCAACTCCGCCGTCGCC	69°C
*Ct*HTM1P _pK7FWG2	GGCAGCCTGG TGTCTCACTGG	CCAGGTTCTGTCC GCTGAGCATA	69°C
*Ct*HTM1P _eGFP	GGCAGAGATGAG TACGTGGTCATCA	CTGAAACGGCC ACAAGTTCAGC	68°C

Table 2.1: List of primers used for PCR cloning. Each sequence is indicated following the 5' → 3' direction. The annealing temperatures of each couple of primers were calculated using the New England Biolabs T_m calculator tool (`https://tmcalculator.neb.com`).

2.2.2 Seed sterilisation

To surface-sterilise seeds of *Arabidopsis thaliana* they were mixed in an Eppendorf® tube with a 1:5 solution of H_2O:NaClO (*bleach*) and incubated for 10 minutes at room temperature. Later, the seeds were spun down by centrifugation at 3,000 g for 30 seconds. The bleach was removed washing the seeds 5 times with dH_2O. The sterilised seeds were kept in a fridge at 4 °C for 3 days for vernalisation.

2.2.3 Plant growth and cultivation

To allow them to grow, the seeds of *Arabidopsis thaliana* were either sown in 12-well plates containing 2 mL of one-half strength Murashige and Skoog medium [132] (1/2 MS medium, described below) or placed in a Petri dish (Corning™ BB6401) containing solid 1/2 MS medium supplemented with 1% (w/v) of plant agar (Duchefa Biochemie P1001). **1/2 MS medium**: Murashige and Skoog medium (Duchefa Biochemie M0221) 1.1 gram per litre, 0.5% (w/v) sucrose, pH 5.7.

To grow the seeds of *Arabidopsis thaliana* on soil we used a mixture of peat and vermiculite in a proportion of 5:1.

Seedlings were grown at 22 °C and 70% relative humidity under a 16-h-light/8-h-dark cycle (approximately 120 μmol m^{-2} s^{-1}).

2.2.4 *N*B-DNJ αGlu II inhibition and Elf18/Flg22 elicitation

For the growth inhibition assay, the seeds were surface-sterilised and vernalised for 4 days prior to being sown in 12-well plates containing five seeds per 2 mL of 1/2 MS medium. *N*B-DNJ treatment was carried out on three-day-old seedlings germinated and grown in 1/2 MS medium (see Section 2.2.3), together with 100 nM elf18 (ac-SKEKFERTKPHVNVGTIG) or flg22 (QRLSTGSRINSAKDDAAGLQIA) peptide elicitors. The effect of treatment on seedling growth was analysed 13 days post-sowing by photography and recording the fresh weight of plants.

For gene transcription analysis, the seeds were surfaced, sterilised and sown in 12-well plates (∼10 seeds per well) containing 1/2 MS medium (2 mL per well). After 9 days, the medium was replaced with fresh medium. On the tenth day, elf18/flg22 elicitors were added to the medium, for 30 minutes. Afterwards, the seedlings were frozen in liquid nitrogen and used for RNA extraction.

2.2.5 cDNA reverse-synthesis

Gene transcription analysis was carried out from the seedlings frozen in liquid nitrogen and reduced to a fine powder using mortar and pestel. Total RNA was extracted from homogenised plant tissues of three independent replicates, each composed of 20 seedlings, using the NORGEN-Total RNA Purification Kit (Cat. 17200 Norgene Biotek Corp.) according to the manufacturer's protocol. RNA samples were treated with RNase-free DNase I (ThermoFisher scientific EN0521) to eliminate any potential DNA contamination. First-strand cDNA was synthesised from 1 μg of total RNA by

using Maxima™ H Minus cDNA Synthesis Master Mix (ThermoFisher scientific M1661) according to the manufacturer's instructions.

2.2.6 Quantitative Reverse Transcription PCR

Quantitative reverse transcription-PCR (qRT-PCR *aka* real-time PCR) was used to analyse the expression levels of the *PHI1* (AT1G35140), *RET-OX* (AT1G26380) genes and *BiP3* (AT1G09080); the constitutively expressed *UBIQUITIN5* gene (AT3G62250) served as an internal reference to normalise gene expression [133]. Analysis was performed on a StepOne™ Real-Time PCR System (Applied Biosystem, Warrington, United Kingdom) in a reaction volume of 20 μL containing 2× iTaq Universal SYBR Green Supermix (Bio-Rad 1708880) and 0.5 μM of each primer. For the amplification, the following PCR settings were used:

1. an initial denaturation step at 95 °C for 2 minutes;
2. 95 °C for 15 seconds;
3. 60 °C for 1 minute;
4. Back to Step 2 40×.

The specificity of the PCR amplification was confirmed by dissociation curve analysis [134]. The relative quantification of gene expression was established using the comparative $2^{-\Delta\Delta Ct}$ method [133] and the transcription levels of each gene relative to UBIQUITIN5 were determined using a modification of the Pfaffl method [133] as described in reference [135] and expressed in arbitrary units.

The primers used for the qRT-PCR are listed below in Table 2.2

Primers	Forward primer	Reverse primer
UBQ5	GGAATCGACGCTTCATCTCG	ATGAAAGTCCCAGCTCCACA
PHI1	TTGGTTTAGACGGGATGGTG	ACTCCAGTACAAGCCGATCC
RET-OX	AGGTTCTCGAACCCTAACAACA	GCACAGACGACACGTAAGAAAG
BiP3	CGAAACGTCTGATTGGAAGAA	GGCTTCCCATCTTTGTTCAC

Table 2.2: List of primers used for gene expression analysis through qRT-PCR amplification. Each sequence is indicated following the 5' → 3' direction.

The PCR efficiency of each oligonucleotide pair was calculated from linear regressions of the standard curves. Real-time PCR derived data were relatively quantified taking into account the divergent efficiencies.

2.2.7 Statistical Analysis

For the qRT-PCR gene transcription analysis three replicates were performed for each experiment in addition to a no-template control included for each primer pair. Statistical analysis was performed

using Student's t-test [136]. The average fresh-weight of the *Arabidopsis thaliana* plants was taken from five plants per treatment. The fresh-weight was compared and statistically analysed using ANOVA with Scheffè's *post hoc* test [137]. Statistical analysis and histograms-graphs in Figures 3.1, 3.5, 3.6, 3.7, 3.8, and 3.9 were obtained using `GraphPad Prism` version 8.0.0 for Windows (GraphPad Software, San Diego, California USA, `www.graphpad.com`).

2.2.8 Plant transient transformation

The full-length coding sequence of *Ct*HTM1P was amplified by PCR using the Accuprime Pfx DNA Polymerase (Invitrogen), cloned into pDONR™221 vector (Invitrogen) and sequenced. DNA primers used for cloning are "*Ct*HTM1P _pDONR™221" in Table 2.1.

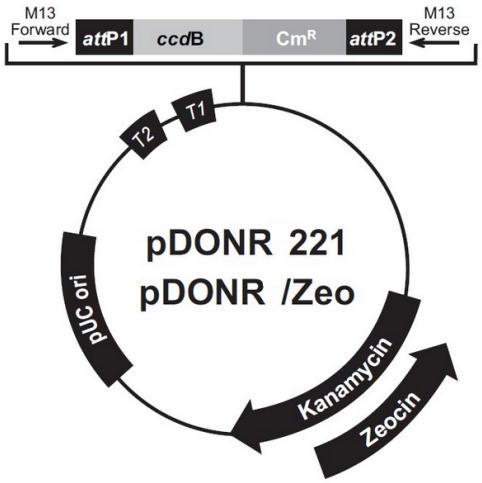


Figure 2.2: pDONR™221 map. The pDONR™221 vector has a pUC origin for high plasmid yields and universal M13 sequencing sites for ease of use.

*Ct*HTM1P was fused at the C-terminal with the enhanced GFP (eGFP), using the pK7FWG2 binary vector and the Gateway® Technology according to the manufacturer's instructions.

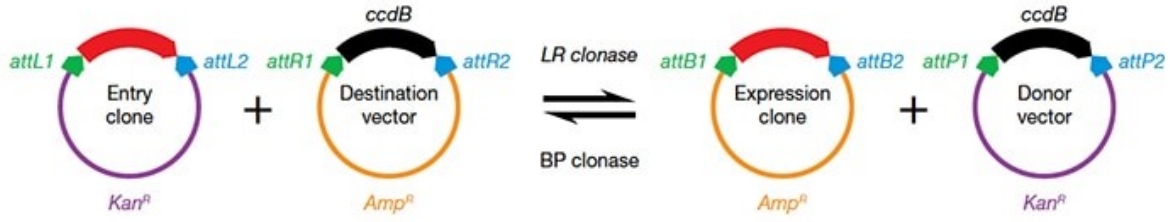


Figure 2.3: Cloning of a binary vector using Gateway® Technology. Figure taken from `lifetechnologies.com/gateway`.

Binary vector, containing the construct, was amplified in *Escherichia coli* DH5α before transformation of *Agrobacterium tumefaciens* (strain GV3101). The transformation of *Agrobacterium tume-*

faciens was checked through colony-PCR, using the primers described as *Ct*HTM1P _pK7FWG2 and *Ct*HTM1P _eGFP in Table 2.1: the first couple amplifies the portion of the gene between bases 1363-2263 while the second couple amplifies a portion between the base number 2383 of the insert and the sequence of the eGFP, Figure 2.4.

Then, *Agrobacterium tumefaciens* was inoculated in 5ml of LB supplemented with rifampicin (Duchefa R0146), gentamycin (Duchefa G0124) and spectinomycin (Duchefa S0188). The culture was incubated at 28 °C in a shaking incubator until it reached a stationary grow phase. The culture was harvested and centrifuged at 4000g for 15 minutes at room temperature. The pellet was washed twice with 10 mL of infiltration buffer (50mM MES pH 5.6; 2mM Na 3 PO 4 ; 0.5% (w/v) glucose, 100 μM acetosyringone), and then resuspended in the same buffer to a concentration of 0.5 OD_{280}. The transient transformation was performed injecting the culture in 4-5 weeks old tobacco leaf epidermal cells (*Nicotiana tabacum* plants, cv Petit Havana). Plants were grown at 24 °C, 14h light, 10h dark and were used for *Agrobacterium tumefaciens* (strain GV3101)-mediated transient expression. The bacterial optical density (OD_{600}) used for plant transformation was 0.5.

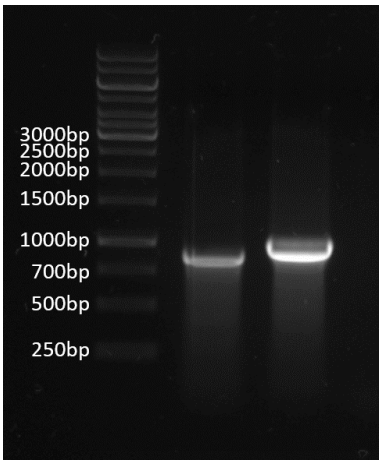


Figure 2.4: Colony-PCR of *Agrobacterium tumefaciens* colonies transformed with *Ct*HTM1P _pK7FWG2 vector. The success of the transformation with the *Ct*HTM1P _pK7FWG2 vector was investigated with two different sets of primers through colony-PCR. The PCR-amplified bands show DNA segments of the expected size: the band on the left is 900 bp and the band on the right is 986 bp.

2.2.9 Plant stable transformation

Arabidopsis plants were transformed using the floral dip method [138]:

1. Healthy *Arabidopsis* plants were grown until they were flowering;
2. *Agrobacterium tumefaciens* (strain GV3101) carrying gene *Ct*HTM1P on the pK7FWG2 binary vector was prepared;

3. *Agrobacterium tumefaciens* was spun down and resuspended to OD_{600}=0.8 in 500 mL of 5% Sucrose, 10 mM $MgCl_2$, 200 μM acetosyringone;

4. Before dipping, Silwet L-77 was added to a concentration of 0.05% (500 μL/L);

5. *Arabidopsis thaliana* was dipped in *Agrobacterium* solution for 1 hour, with gentle agitation;

6. The dipped plant was kept under a cover for 16 to 24 hours to maintain high humidity;

7. After 5 days, the dipping procedure was repeated;

8. The cover was removed 2-3 days later;

9. Dry seeds were harvested and the transformants were selected on 1/2 MS medium with phosphinothricin (aka Basta, Duchefa P0159), final concentration 20 μg/ml.

2.2.10 Confocal Laser Scanning Microscopy

An inverted laser scanning confocal microscope (LSM Pascal; Carl Zeiss) was used for confocal analyses. Imaging of *Ct*HTM1P-GFP was performed using 488 nm excitation of an Argon ion laser, 25 mW. GFP was detected with a 505–530 nm filter set. A 488/543/633 beam splitter was used for acquisition. Imaging was performed using 40x Zeiss plan-neofluar oil, 1.3 NA, DIC.

2.3 DNA cloning and recombinant protein production

2.3.1 List of buffers

All 0.22μm filtering of solutions was performed with Millipore Express PLUS 0.22μm PES membranes (Millipore).

PBS: Phosphate buffered saline - NaCl 137 mM, KCl 2.7 mM, Na_2HPO_4 10 mM, KH_2PO_4 1.8 mM; pH 7.4. Obtained by dissolving Phosphate Buffered Saline tablets (Sigma Aldrich P4417).

IMACBW: Na_2HPO_4 50 mM pH 7.5, NaCl 300 mM.

IMACE: Na_2HPO_4 50 mM pH 7.5, NaCl 200 mM, 500 mM imidazole (Honeywell Fluka 56750).

SEC buffer: HEPES (Sigma H3375) 20 mM pH 7.5, NaCl 100 mM.

WB - Transfer buffer: Tris 25 mM pH 8.8, Glycine 200 mM, MetOH 20%.

WB - Blocking buffer: TrisHCl 20mM pH 7.6, NaCl 140 mM, Tween20 1%.

2.3.2 Searching for the *Ct* orthologues of *Sc*PDI1P and *Sc*HTM1P

In order to discover thermostable ERAD checkpoint proteins, the *Chaetomium thermophilum* genome (at `https://c-thermophilum.bork.embl.de/`) was searched with *Sc*HTM1P (Uniprot P38888) and *Sc*PDI1P (Uniprot P17967) sequences.

2.3.3 Cloning of *Ct*HTM1P and *Ct*PDI1P

The *Chaetomium thermophilum* ERAD α-1,2-mannosidase *Ct*HTM1P (CTHT_0058730, Uniprot G0SCX7, G0SCX7_CHATD) DNA was purchased from GeneArt/LifeTechnologies, nucleotides 1-3276 (^{1}ATGAGCTGCAGATT(...)ACGGCGGAGTTGATGGA^{3276}) encoding *Ct*HTM1P amino-acids 1-1092, with codon optimisation for expression in mammalian cells; the gene was inserted after the NdeI site of vector pMK-RQ (kanR), plasmid *Ct*HTM1P -pMK-RQ.

The DNA inserts mentioned from this point onwards were PCR-amplified for insertion in Litmus28i and pHLsec vectors. PCR primers were designed with a 5' 15-25 bp overlap with the target DNA vector or fragment.

2.3.4 Cloning vector Litmus28i

The Litmus28i vector (New England Biolabs®, Figure 2.5) was linearised using the restriction enzyme EcoRI at the 5' and XhoI at the 3' end. The template DNA used for the double digestion was Litmus28i_GBAN370S, a plasmid encoding for Non-lysosomal glucosylceramidase (Uniprot P04062). The Figure 2.6 shows the linear EcoRI/XhoI Litmus28i vector. The DNA bands in the yellow circle correspond to the size of linear Litmus28i vector, the lower bands correspond to the size of the restriction enzymes freed GBAN370S (1608bp).

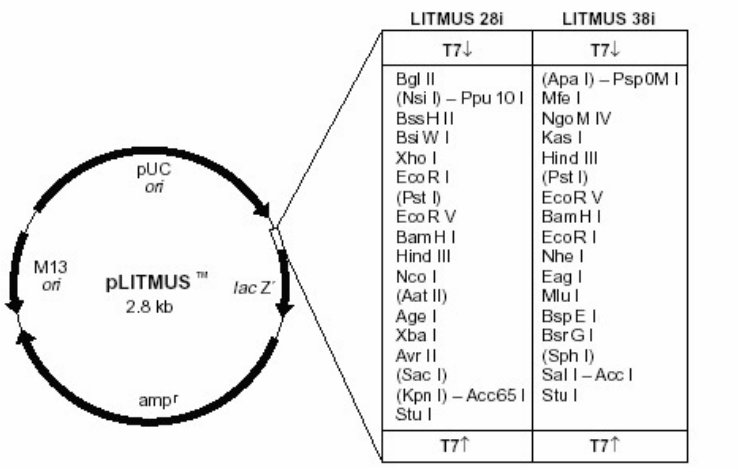


Figure 2.5: Litmus28i map. Litmus28i is a multi-purpose cloning/*in vitro* transcription phagemid vector. The molecule is a small, double-stranded circle 2,823 base pairs in length [139].

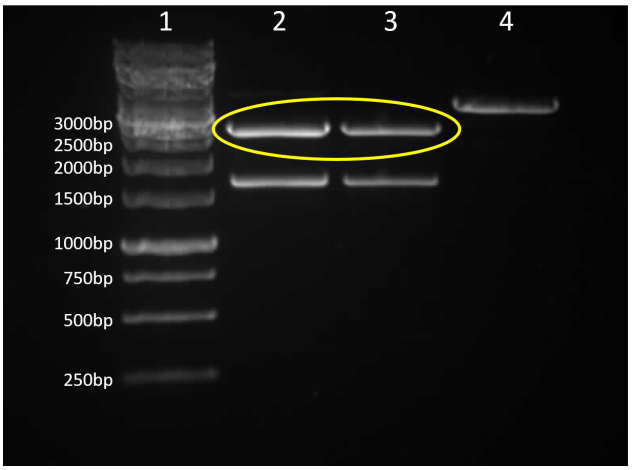


Figure 2.6: Linearisation of Litmus28i. **1**: Molecular Mass marker; **2 and 3**: Litmus28i_GBAN370S digested with the restriction enzymes EcoRI and XhoI; **4**: Litmus28i_GBAN370S digested with the restriction enzyme EcoRI.

2.3.4.1 Cloning of *Ct*HTM1P full length

The full gene for *Chaetomium thermophilum* HTM1P (Uniprot entry G0SCX7, `https://www.uniprot.org/uniprot/G0SCX7`) was purchased from GeneArt/LifeTechnologies, further details in Section 2.3.3. The *Ct*HTM1P full length gene, containing the EcoRI and XhoI sites for the insertion in linearised Litmus28i vector [139], was carried out by polymerase chain reaction (PCR). The template DNA used for the reaction was *Ct*HTM1P -pMK-RQ, using the pair of primers labelled as "*Ct*HTM1P _Litmus28i" as described in 2.1.

DNA template used was 1μg of linearised Litmus28i vector mixed with 2μL of 10X FastDigest™green buffer and 0.5μL FastDigest™enzyme (Thermo Fisher Scientific), adding deionised water up to a total volume of 20μL. The mix was incubated for 1 hour at 37 °C. The reaction was analysed through gel electrophoresis as described in 2.1.2. The DNA corresponding to the linearised vector was purified from the agarose gel as described in 2.1.3.

The amplified DNA was cloned in Litmus28i vector using ligation independent cloning (aka Gibson assembly), described below in Section 2.3.6.

2.3.4.2 Cloning of full length *Ct*PDI1P

The gene of the *Chaetomium thermophilum* PDI1P (Uniprot entry G0SGS2, `https://www.uniprot.org/uniprot/G0SGS2`) was amplified from *Chaetomium thermophilum* cDNA (a gift of Ed Hurt, Biochemistry Centre, Heidelberg University, Heidelberg, Germany).

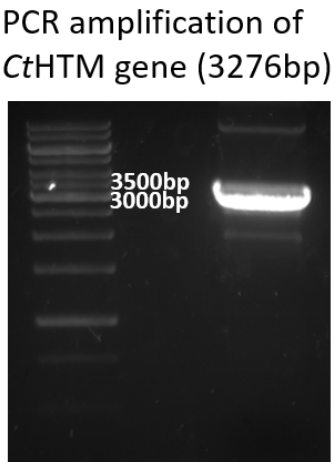


Figure 2.7: *Ct*HTM1P full length amplification. The gene was amplified by PCR creating the EcoRI/XhoI overlapping ends for the insertion in the linearised Litmus28i vector.

The amplified DNA was cloned in Litmus28i vector using ligation independent cloning (aka Gibson assembly), described below in Section 2.3.6.

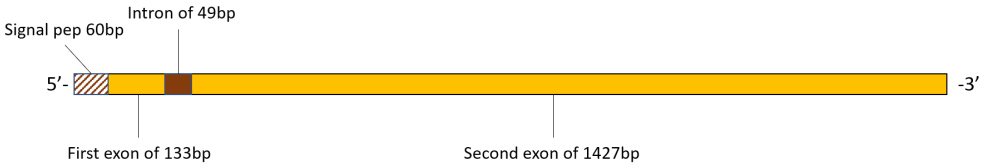


Figure 2.8: Schematic sequence of *Ct*PDI1P gene. The *Ct*PDI1P gene is constituted by a 5' short signal peptide of 60bp (white with brown lines), the first exon of 133bp (yellow), an intron of 49bp (brown) and the second exon of 1427bp (yellow).

2.3.5 Expression vector pHLsec

The vector termed pHLsec is a vector for secreted constructs and it is based on the pLEXm backbone, the multiple cloning site (MCS) however was reduced when several new features were introduced, such as: a Kozak sequence; a secretion signal sequence; a C-terminal 6xHis tag. Constructs can be cloned into this vector, preserving and making use of the features listed above, by using the AgeI and KpnI restriction sites [140].

The pHLsec vector (Figure 2.10) was linearised using AgeI at the 5' and KpnI at the 3' end. DNA template used: pHLsec vector encoding a pEGFP fusion of the DNA encoding human Solute carrier family 35 member B1 (SLC35B1 aka UGTrel1, Uniprot P78383, S35B1_HUMAN) plasmid pEGFP_SLC35B1-pHLsec, courtesy of Snežana Vasiljević, Department of Biochemistry, Oxford University. 1μg of the template pEGFP_SLC35B1-pHLsec DNA was mixed with 2μL of 10X FastDigest™green buffer and 0.5μL FastDigest™enzyme (Thermo Fisher Scientific), adding deionised water up to a total volume of 20μL. The mix was incubated for 1 hour at 37 °C. Figure 2.9 shows

the linear AgeI/KpnI pHLsec vector. The DNA bands in the yellow circle correspond to the size of the linear pHLsec vector, the lower bands correspond to the size of the restriction enzymes freed pEGFP_SLC35B1 (1704bp).

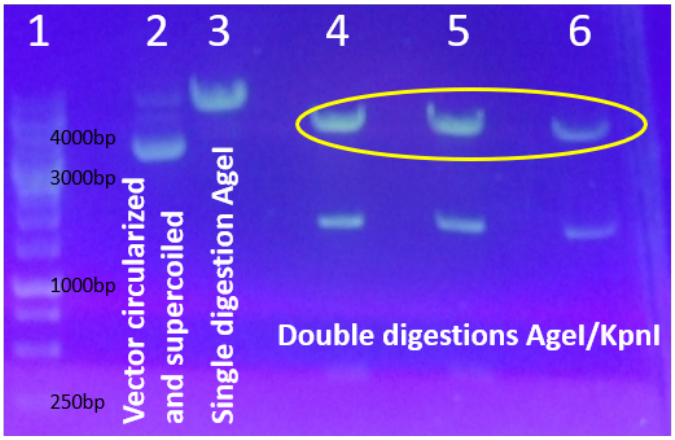


Figure 2.9: Linearisation of pHLsec. **1**: Molecular Mass marker; **2**: circular and supercoiled pEGFP_SLC35B1-pHLsec; **3**: pEGFP_SLC35B1-pHLsec digested with the restriction enzyme AgeI; **4,5,6**: pEGFP_SLC35B1-pHLsec digested with the restriction enzymes AgeI and KpnI.

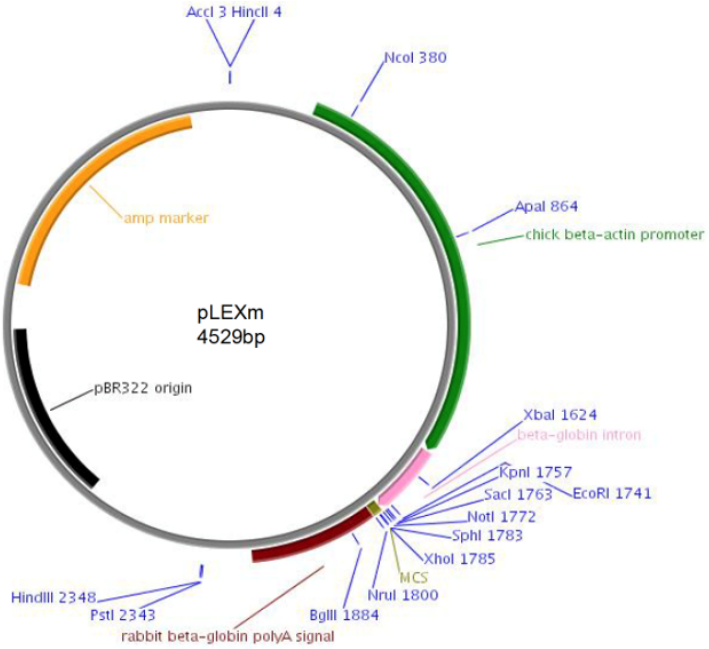


Figure 2.10: pHLsec map. Figure taken from [140]

2.3.5.1 Cloning of four *Ct*HTM1P constructs

Different portions of the sequence of *Ct*HTM1P were amplified by PCR, using *Ct*HTM1P _Litmus28i as a template for the reaction. The DNA sequence encoding for GFP protein was amplified using the plasmid pHLsec:GFP-SLC35B1 (Uniprot ID P78383), kindly donated by Dr Snežana Vasiljević (Biochemistry Department, University of Oxford), as a template for the reaction.

The inserts are described below:

A 5' pHLsec overlapping end-*Ct*HTM1P ^{193}AGAGTTAGAGC(...)GTTGATGGA^{3276}-3' pHLsec overlapping end. Length of the insert is 3084 bp;

B1 5' pHLsec overlapping end-GFP-3' *Ct*HTM1P overlapping end. Length of the insert is 738 bp;

B2 5' GPF overlapping end-*Ct*HTM1P ^{1}ATGAGCTGCA(...)TTGATGGA^{3276}-3' pHLsec overlapping end. Length of the insert is 3276 bp;

C1 5' pHLsec overlapping end-*Ct*HTM1P ^{1}ATGAGCTGCA(...)TTGATGGA^{3276}-3' GFP overlapping end. Length of the insert is 3276 bp;

C2 5' *Ct*HTM1P overlapping end-GFP-3' pHLsec overlapping end. Length of the insert is 738 bp;

D1 5' pHLsec overlapping end-GFP-3' *Ct*HTM1P overlapping end. Length of the insert is 738 bp;

D2 5' GPF overlapping end-*Ct*HTM1P ^{193}AGAGTTAGAG(...)TTGATGGA^{3276}-3' pHLsec overlapping end. Length of the insert is 3084 bp.

Figure 2.11 shows the PCR amplification of the inserts listed above.

Such inserts were used for insertion between the AgeI and KpnI sites of the pHLsec vector. Insert "**A**" was mixed with the AgeI/KpnI linearised $^{\text{AgeI/KpnI}}$pHLsec DNA in 3:1 molar ratio: 0.06 pmol of insert DNA and 0.02 pmol of AgeI/KpnI linearised pHLsec vector.

The pairs of DNA inserts "**B1**"-"**B2**", "**C1**"-"**C2**" and "**D1**"-"**D2**" were mixed with the AgeI/KpnI linearised $^{\text{AgeI/KpnI}}$pHLsec DNA in 3:1 molar ratio: 0.06 pmol of each fragment forming a couple and 0.02 pmol of AgeI/KpnI linearised pHLsec DNA.

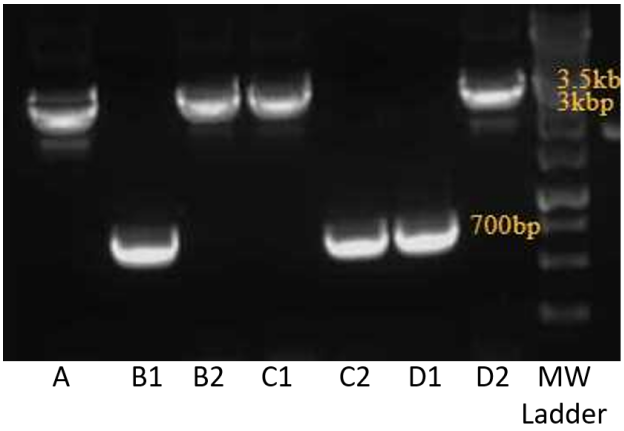


Figure 2.11: PCR amplification of GFP and *Ct*HTM1P gene sequence. **A**: Amplification of *Ct*HTM1P soluble domain sequence with overlapping ends for insertion in pHLsec. **B1**: Amplification of GFP sequence with the addition of overlapping ends, 5' overlaps pHLsec sequence and 3' overlaps *Ct*HTM1P sequence . **B2**: Amplification of *Ct*HTM1P sequence with the addition of overlapping ends, 5' overlaps GFP sequence and 3' overlaps pHLsec sequence. **C1**: Amplification of *Ct*HTM1P sequence with the addition of overlapping ends, 5' overlaps pHLsec sequence and 3' overlaps GFP sequence. **C2**: Amplification of GFP sequence with the addition of overlapping ends, 5' overlaps *Ct*HTM1P sequence and 3' overlaps pHLsec sequence. **D1**: Amplification of GFP sequence with the addition of overlapping ends, 5' overlaps pHLsec sequence and 3' overlaps *Ct*HTM1P sequence. **D2**: Amplification of *Ct*HTM1P soluble domain sequence with the addition of overlapping ends, 5' overlaps GFP sequence and 3' overlaps pHLsec sequence. 1% agarose in 50mL TAE buffer, electrophoresis conditions: 100V for 1 hour.

The amplified DNA was cloned in pHLsec vector using ligation independent cloning (aka Gibson assembly), described below in Section 2.3.6.

2.3.5.2 *Ct*PDI1P intron excision

In order to remove the intron from the *Ct*PDI1P gene sequence the two exons, flanking the intron, were amplified separately using *Ct*PDI1P _Litmus28i as a template. The PCR reaction mix and the setting used are described in 2.1.1.
The length of the first exon is 133 bp, ^{61}TCCGACGTC(...)TTGCCGAGT^{133}. The length of the second is 1412 bp, ^{182}TCTTCGCC(...)AGGAGACTGCC^{1594}. The portions of the gene sequence encoding the N-terminal signal peptide, the C-terminal retrieval signal (HDEL) and the stop codon were not amplified. The primers used are reported in Table 2.1, primer XFWD, XREV, YFWD and YREV.

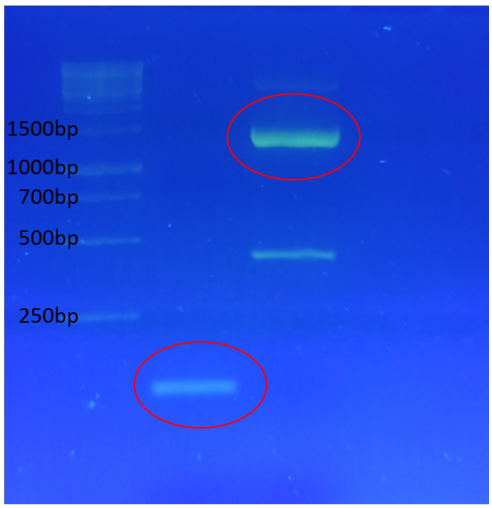


Figure 2.12: PCR amplification of *Ct*PDI1P two exons. **Left**: molecular-mass size marker; **middle**: amplification of the first exon (133bp); **right**: amplification of the second exon (1427bp). Agarose 2% in TAE buffer (see Section 2.1.2). Electrophoresis conditions: 100V for 1 hour.

The two *Ct*PDI1P fragments ^{61}TCCGACGTC(...)TTGCCGAGT^{133} and ^{182}TCTTCGCC(...)AGGAGACTGCC^{1594}, were cloned in the pHLsec vector using ligation independent cloning (aka Gibson assembly), described below in Section 2.3.6.

2.3.6 Gibson assembly

New England Biolabs® (NEB) recommends a total of 0.02–0.5 pmols of DNA fragments when 1 or 2 fragments are being assembled into a vector and a 2-3:1 molar ratio of each insert:vector [141]. To calculate the number of pmols of each fragment for optimal assembly, based on fragment length and weight, the following formula was used:

$$pmols = \frac{(\text{weight in ng}) \times 1,000}{(\text{base pairs} \times 650 \text{ Daltons})}$$

The amplified insert DNA was mixed with the linearised vector DNA (Litmus28i or pHLsec) in 3:1 molar ratio: 0.06 pmol of insert DNA and 0.02 pmol of linearised vector DNA. To this DNA, 10 μL of GIBSON ASSEMBLY MasterMix(2X) (New England Biolabs® NEB E2611) were added and the total volume made 20 μL with H_2O. The mix was incubated at 50 °C for 1 hour. 2μL of the Gibson assembly mix was used to transform *E. coli* DH5α as described below in Section 2.3.7.

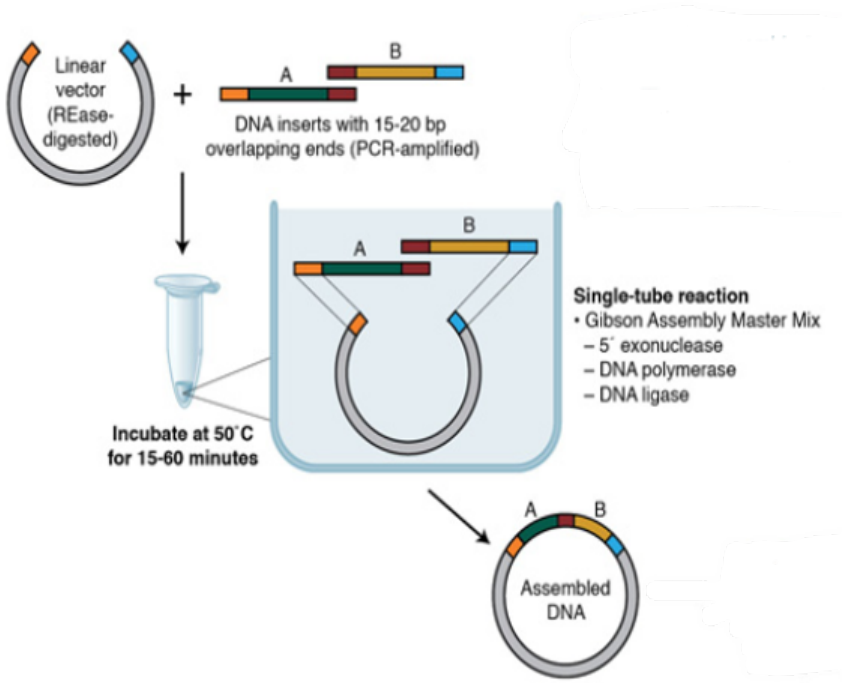


Figure 2.13: Gibson assembly strategy. Gibson Assembly employs three enzymatic activities in a single-tube reaction: 5' exonuclease, the 3' extension activity of a DNA polymerase and DNA ligase activity. The 5' exonuclease activity chews back the 5' end sequences and exposes the complementary sequence for annealing. The polymerase activity then fills in the gaps on the annealed regions. A DNA ligase then seals the nick and covalently links the DNA fragments together. Figure taken from [141].

2.3.7 *E. coli* transformation

A volume of 25μL of frozen glycerol stock of competent DH5α *E. coli* (high efficiency) cells (NEB C2987I) were thawed on ice. 2μL of the Gibson assembly mix (as described in 2.3.6) was added to it, mixed, and left on ice for 2 minutes. The cells were then heat-shocked in a water bath at 42°C for 30 seconds and put back on ice for 2 minutes. A volume of 950 μL of Super Optimal broth with Catabolite repression (SOC) media (New England Bioscience) was added and the sample left in a shaking incubator at room temperature for 1 hour. A volume of 100 μL of cells in SOC media were plated on pre-heated LB agar + antibiotic (carbenicillin 100 μg/mL) plates and incubated overnight at 37°C.

To confirm the transformation of *E. coli* DH5α the presence of the *Ct*PDI1P _pHLsec plasmid was investigated through colony-PCR using the method described in 2.1.1. The result of the colony-PCR on two different colonies used as a template of the PCR reaction is shown in Figure 2.14.

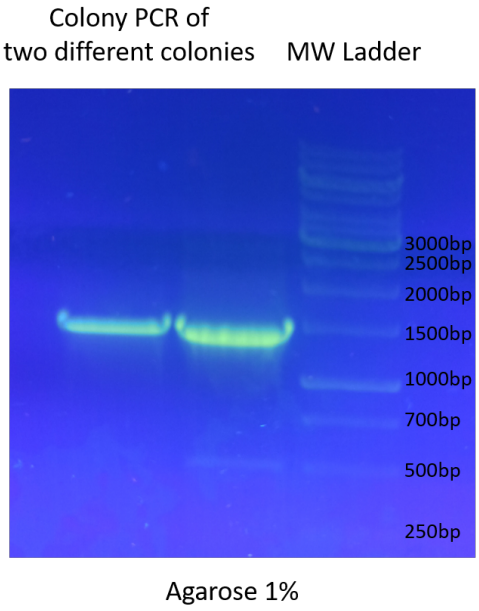


Figure 2.14: Colony-PCR of *E. coli* colonies transformed with *Ct*PDI1P _pHLsec vector. The success of the transformation with the *Ct*PDI1P _pHLsec vector was investigated in two different colonies through colony-PCR. The PCR-amplified bands show DNA segments of the expected size (1560bp).

The *Ct*PDI1P _pHLsec DNA vector maxiprep was performed as described below in Section 2.3.8.

2.3.8 Plasmid purification

The plasmid DNAs were miniprepped using the the GeneJet Plasmid Miniprep Kit (ThermoFisher Scientific). The plasmid DNAs were maxiprepped using the following steps:

1. a single colony of competent DH5α *E. coli* cells transformed with the plasmid was inoculated in 15 mL of LB with carbenicillin 100 μg/mL and grown overnight (O/N) at 37°C, then inoculated into 750 mL of the same media in a 2 L baffled flask and grown O/N at 37°C shaking at 220 rpm.

2. the cells were pelleted with a centrifuge spin at 4,000g for 10 mins at 4°C and the pellet resuspended in 12 mL 10 mM Tris/Cl pH 8, 10 mM EDTA. A volume of 4 mL of the same buffer + 50 mg lysozyme was added. The sample was split into 2x 50 mL tubes. The cells were incubated at room temperature (RT) for 5 minutes and then 15 mL of 0.2 M NaOH 1% SDS was added to each tube to complete cell lysis, incubating at RT for 5 mins. A volume of 11 ml of 3 M KAcO pH 4.8 was then added to each tube. The solutions were incubated at RT for 5 mins and then spun at 4,000g for 10 minutes at 4°C to pellet cellular debris.

3. the supernatant (S/N) was filtered through Miracloth (Merck Millipore): two Miracloth squares were cut and each placed on top of a clean 50 mL tube, after which the S/N was

poured gently onto the Miracloth square and collected in the tube below. A volume of isopropanol equal to 60% of the solution was then added to precipitate the plasmid DNA. and the solution incubated at RT for 5 minutes. The plasmid DNA was then pelleted by spinning at 4,000g for 10 minutes at 4°C. The two DNA plasmid pellets were resuspended by pipetting in a total of 15 mL of 10 mM Tris/Cl pH 8, 10 mM EDTA. An equal volume of 5 M LiCl (kept at –20°C) was added and incubated on ice for 5 minutes to precipitate contaminating RNA and proteins. The solution was spun at 4,000g for 10 minutes at 4°C, and the pellet discarded.

4. a volume of isopropanol equal to 60% of the solution was then added and incubate at RT for 5 minutes to re-precipitate the purified plasmid DNA. The solution was spun at 4,000g for 10 minutes at 4°C, the S/N discarded and the pellet resuspended by pipetting in 10 mL of 10 mM Tris/Cl pH 8, 1 mM EDTA. To get rid of the remaining RNA, a volume of 50 μL of 10 mg/ml heat-treated RNase A was added, and the solution incubated at RT for 15 minutes. A volume equal to 25% of the sample volume wad added of a solution 30% w/v PEG 6000, 2.5 M NaCl: this re-precipitates the plasmid DNA (the solution went cloudy). The solution was incubated on ice for 30 minutes and then spun at 4,000g for 10 minutes at 4°C.

5. the pellet was re-suspended by pipetting in 10 mL of 10 mM Tris/Cl pH 8, 1 mM EDTA. The PEG was then extracted by the addition of 2 mL chloroform and a quick vortex before spinning at 4,000g for 2 minutes at 4°C. The aqueous layer (top) was retained by pipetting and transferred to a clean 50 mL tube. The pure plasmid DNA was then precipitated by the addition of a 10% volume of 5 M NaCl (or 3 M NaAcO) and 3 volumes of ethanol. The solution was spun at 4,000 g for 10 minutes at 4°C.

6. the DNA plasmid pellet was re-suspended/washed with 20 mL 100% ethanol and spun again at 4,000g for 2 minutes at 4°C. The S/N was discarded and the pellet left to dry overnight in the tube upside down on the bench at RT. The DNA plasmid was re-suspended by pipetting in 2 mL of 10 mM Tris/Cl pH 8, 0.1 mM EDTA (filtered) and left overnight for the plasmid DNA to resuspend fully. The pure plasmid DNA was filtered-sterilised with a 0.22 μm filter and stored at -20°C.

2.4 Protein expression and purification

Human epithelial kidney (HEK293F) cells (ThermoFisher Scientific) were transfected using the method described in [142]:

1. HEK293F cells were split in GIBCO FreeStyle 293 Media (ThermoFisher Scientific 12338018) at 0.5 x 10^6 cells/mL;

2. cells were kept in an orbital shaker incubator at 37°C, 120 rpm and 5% CO_2. Twenty-four hours later, or when they reached the 1x10^6 cells/mL, >90% cell viability was confirmed by trypan blue exclusion [143];

3. HEK293F at a concentration of 10^6 cells/mL suspended in GIBCO FreeStyle 293 Media (ThermoFisher Scientific 12338018), were transfected with the plasmid: 1 μg of DNA per mL of culture. The DNA plasmid was initially dissolved in 10mL of PBS per 100 mL of culture (see the recipe in 2.3.1) and vortexed vigorously for 3 seconds;

4. 4 μL per mL of culture of a filter-sterilised solution of 0.5 mg/mL Polyethylenimine (PEI, Sigma catalogue 408727) was added to the PBS/DNA solution and vortexed vigorously for 3 seconds. The mixture was incubated at room temperature for 20 minutes;

5. the DNA/PEI mixture was added to the cell culture. The cell culture was poured in Erlenmeyer flasks with 0.2 μm vent caps (Corning), and incubated in an orbital shaker incubator at 37°C, shaking at 120 rpm, under a 5% CO_2 atmosphere. The volume of the chosen flask was 3 times the volume of the culture.

2.4.1 Harvesting of the supernatant and protein extraction from cells

The HEK293F cell culture was harvested 4 days post-transfection by centrifuging at 4,000 g for 5 minutes, the supernatant was made 1× PBS by adding an appropriate 10× PBS stock solution volume (see the recipe in 2.3.1), and the pH was adjusted to 7.4 by adding a few drops of 2 M NaOH, before bottle-top vacuum filtration through a 0.2 μm filter. The cell pellet was lysed following the steps below:

1. the cell pellet was resuspended into ∼40 ml of pre-chilled lysis buffer (100 mM potassium acetate, 50 mM Tris pH 7.5, 5% glycerol, 0.3% Triton X-100, protease inhibitors - COmplete EDTA free) per litre of culture, pipetting several times;

2. the cells were sonicated for 4 cycles (20 sec on, 20 sec off);

3. the lysate was centrifuged at 20,000 x g for 35 min at 4°C and the supernatant was retained;

4. the lysate was filtered through a 0.2 μm filter.

2.4.2 Western blot - wet transfer

Western blot wet-transfer was performed in order to detect the presence of the expressed proteins. The transfer of the proteins from an SDS-PAGE gel to a PVDF membrane was performed following the step below:

1. after running an SDS-Pade, the unstained gel was used for the transfer;

2. 1x transfer buffer was prepared following the recipe in 2.3.1;

3. a square of filter paper (larger than the acrylamide gel) was soaked in transfer buffer;
4. the PVDF membrane was first activated in pure MeOH for 30", then it was equilibrated in transfer buffer;
5. a sandwich was assembled as following: Sponge - 2 pieces of filter paper (soaked) - Gel (washed with transfer buffer) - Membrane (activated and equilibrated) - 2 pieces of filter paper (soaked) - Sponge;
6. a transfer-tank was filled with 600mL of transfer buffer, a freezer-block was used to avoid overheating;
7. the power supplier was set to run at 20V overnight.

The blotting of the PVDF membrane and the detection of the antibody signal followed as described below:

1. the membrane was incubated with blocking buffer (see the recipe in 2.3.1) + 5% lyophilised milk for 1 hour;
2. the membrane was washed 3 times with blocking buffer, 5 minutes for each wash;
3. the membrane was then incubated with Anti-GFP antibody (Mouse monoclonal, abcam, [9F9.F9] (ab1218)) for 1 hour;
4. point 2 was repeated;
5. the membrane was incubated with secondary antibody (IRDye® 800CW Goat anti-Mouse, LI-COR 926-32210) for 1 hour;
6. to prepare the membrane for the detection of proteins, it was washed twice with blocking buffer, 5 minutes for each wash;
7. the last wash was applied using a version of the blocking buffer without Tween20 for 5 minutes;
8. the Odyssey® CLx Imaging System (LI-COR Biosciences) was used to visualise the result using the software Image Studio (LI-COR Biosciences).

2.4.3 Immobilised metal affinity chromatography (IMAC)

The filtered supernatant and/or lysate was then passed onto a 1 mL HisTrap HP Ni IMAC column (GE Healthcare 17-5248-01) equilibrated against binding buffer IMACBW (see the recipe in 2.3.1), using a peristaltic pump at room temperature, at a flow rate of approximately 3 mL/min. The column was washed with 5 column volumes (CV) buffer IMACBW, fitted to an ÄKTA purifier FPLC machine in a 4 °C cabinet and further washed with 6 mL of buffer IMACBW. The proteins were eluted with a 15 CV gradient from 0% to 100% of elution buffer IMACE (see the recipe in 2.3.1).

2.4.4 Size exclusion Chromatography (SEC)

The protein-containing fractions from IMAC were pooled and concentrated using a polyethersulfone (PES) membrane, 10 kDa MM cutoff centrifugal ultrafiltration device (Thermo Fisher Scientific 88513). This sample was then filtered through a 0.2 μm filter and applied to a HiLoad Superdex 200 16/60 size exclusion chromatography (SEC) column (GE Healthcare 28-9893-35) equilibrated against SEC buffer (see the recipe in 2.3.1) running at 0.4 mL/min collecting 1 mL fractions; SDS-PAGE (Invitrogen) of SEC fractions was used to assess purity (Thermo Fisher NuPAGE Bis-Tris 4-12% gel), run at 200 V for 30' in NuPAGE™ MES or MOPS buffer (respectively NP0002 and NP000102), stained in Coomassie Blue (Sigma ISB1L) for 1 h and destained with water. The fractions containing a single peak in the chromatogram were pooled and concentrated using a PES membrane and protein concentration measured by loading 1.5 μL of sample on a NanoDrop 1000 spectrophotometer (Thermo Scientific). The absorption was measured setting the NanoDrop to a wavelength of λ=280 nm, at such wavelength the aromatic amino acids tryptophan (Trp) and tyrosine (Tyr) exhibit strong light absorption so the absorption of proteins and peptides at 280 nm is proportional to the content of these amino acids. To calculate the concentration of a protein sample the Lambert-Beer's law was used [144]:

$$A_{280} = \epsilon_{280} c L \tag{2.1}$$

where A_{280} is the absorbance of the protein solution at a wavelength of λ=280 nm, ϵ_{280} is the protein molar absorptivity in $M^{-1}cm^{-1}$ or mL/mg cm^{-1}, c is the concentration of the protein in M or mg/mL, and L is the length of the spectrometer cell in cm. The equation becomes:

$$c = \frac{A_{280}}{\epsilon_{280} L} \tag{2.2}$$

The ϵ_{280} of the studied proteins are listed in Table 2.3 below:

Protein	Molecular Mass	ϵ_{280}
*Ct*PDI1P	55,481 kDa	0.83
*Ct*HTM1P $_{A}$	113,102 kDa	1.34
*Ct*HTM1P $_{B}$	152,109 kDa	1.32
*Ct*HTM1P $_{C}$	152,109 kDa	1.32
*Ct*HTM1P$_{D}$	142,005 kDa	1.22
*Hs*GAPDH	36,053 kDa	0.79

Table 2.3: The calculated Molecular Mass and the ϵ_{280} of the different proteins studied in this thesis. Extinction coefficients are in units of mL/mg cm^{-1}, at λ=280 nm measured in water and computed using `http://www.protparam.net/index.html`.

2.5 Protein crystallisation

Vapour diffusion crystallisation 200 nL sitting drops were set up by mixing the protein in mother liquor:protein ratios 1:1 (drop 1) and 1:2 (drop 2) with the 96-conditions MORPHEUS crystallisation screen (Molecular Dimensions, [145, 146]), Pact Prémiere, JCSG+ and Stura Macrosol, using a Mosquito liquid handler crystallisation robot (SPT Labtech). Two drops per condition were created: a drop mixing 100 nL of protein solution with 100 nL of mother liquor (drop 1); and one mixing 130 nL of protein solution with 70 nL of mother liquor (drop 2). Each mother liquor well of the screen contains 40 μL of the crystallisation condition. The sitting drops were left equilibrating against the mother liquor in an incubator at 21 °C.

2.6 Mass spectrometry

Each protein band was excised with a sterile blade from an SDS-PAGE gel and sent to the FAO Metabolomics and Proteomics Laboratory at York University (`https://www.york.ac.uk/biology/technology-facility/proteomics/pr-services/protein-id/` for trypsin digestion followed by MS analysis.

For post-translational modification analysis, LC-MS was used. Peptides were eluted from a C18 column over a gradient before infusion into the MS, separating the peptides and enhancing the signal.

For protein identification, MALDI-MS and MS/MS were performed using a Bruker ultraflex III MALDI-TOF/TOF, which produces higher-confidence identifications by including tandem spectral data in Mascot searches [147].

2.7 Cryo-EM data collection and processing

Cryo-EM data collection was carried out by Dr Christos Savva on the FEI Titan Krios G3 microscope on March 25 2021 at the Midlands Regional Cryo-Electron Microscope Facility at the University of Leicester, England, UK, with the following settings: acceleration Voltage: 300 kV; spherical aberration coefficient (Cs): 2.7mm; nominal magnification: 81,000X; imaging mode: EFTEM; calibrated physical pixel size: 1.086 Å; second condenser aperture:50 μm; objective aperture: 100 μm; spot size: 4; illumination area: 1.8 μm; defocus range: from -2.6 μm to -1.7 μm in 0.3 μm intervals; detector: K3 (Counting or Super-resolution bin 2); energy filter slit Width: 20 eV; dose rate on detector: 15e$^-$/pixel/sec (measured over typical specimen area); dose rate on specimen: 16.5 $^-$/pixel/sec (measured over a hole); exposure time: 3 sec/micrograph; images/Hole: 1; AFIS Number of fractions: 43.

Cryo-EM data processing was carried out by Mr. Charlie J. Hitchman. Particle picking was carried out with Topaz [148]. A number of 52,000 particles were picked from 170 micrographs. 2D classification [149] was carried out in multiple rounds, each time removing particles that gave rise to obviously poor-quality classes. The final number of particles taken forward for the initial model

was 13,955. 3D initial modelling [149] produced 4 models from these particles and best 2D classes. A single model was taken forwards, selected based on its visual quality in Chimera, and its class distributions. Three rounds of 3D classification with 3D auto-refine in `Relion` [149] produced a 18.1 Å model from a single 3D class obtained from 4,447 particles.

2.8 Homology modelling

The HHPred server [150] was used to align the protein sequences with the ones of orthologues of known structure and create homology models with MODELLER [151]. The transmembrane helix of *Mm*MOGS (mouse GCS1, UniProt Q80UM7, MOGS_MOUSE, residues 42–62) was homology modelled on PDB entry 1HH0, residues A20-A40. The C-terminal part of human calnexin (UniProt P27824, CALX_HUMAN, residues 461–484) was homology modelled using PDB ID 6A69, residues B223–B246. The C-terminal part of human Sep15 (UniProt O60613, SEP15_HUMAN, residues 46–134) was homology modelled using PDB ID 2A4H, residues A11-A99. The human ER UDPase (UniProt O75356, ENTP5_HUMAN, residues 22–404) was homology modelled using PDB ID 5U7W, residues A4–A412. The human UDP-Glc transporter (putatively identified as UniProt P78383, S35B1_HUMAN, residues 9–321; although a recently published paper reports ATP/ADP antiporter activity [152]) was homology modelled on the basis of PDB ID 5OGE, residues C16–C333. The ER lumenal domain of human EDEM1 (UniProt Q92611, EDEM1_HUMAN, residues 126–587) was homology modelled on the basis of PDB ID 1X9D A84-A535; its N-terminal transmembrane part, residues 1–34, was homology modelled on the basis of PDB ID 5MRW, residues E57–E90. The HRD1/HRD3 complex was modelled by docking the crystal structures (PDB IDs 5V6P and 5V7V) in the Cryo-EM map for the complex (Electron Microscopy Data Bank ID EMD-8638 [153]), using Chimera [154].

Chapter 3

Results

3.1 The plant as a model to study ERQC

Eukaryotic glycoproteins in the cellular secretory pathway require the ERQC machinery in order to fold correctly [155]. Small molecule modulators that target the ERQC α-glucosidases hamper glycoprotein entry into and/or exit from the ER glycoprotein folding cycle [156]. Such inhibitors have broad-spectrum antiviral activity against a number of enveloped viruses [157] and have the potential to rescue the secretion of misfolded but active glycoproteins in rare diseases [45]. The strong conservation of the ERQC machinery across eukaryotes makes plants an attractive system for low-cost, easy and fast screening of candidate ERQC modulators, much in the same way as plants sensitive to the action of human chemotherapeutics have been used for preliminary screening of anti-cancer drugs efficacy [158]. Published experiments exploited the *Arabidopsis thaliana* immune response to bacterial peptides [159] as a means of assaying iminosugar activity in the ER *in vivo*.

3.1.1 *In planta* ERQC modulation

The best-known class of ERQC modulators are iminosugars, glucose analogues that have the potential to act as broad-spectrum antivirals. Iminosugars target the host's ERQC and negatively affect the folding of a number of enveloped viral glycoproteins [160, 161, 162, 163, 164]. *In vitro* screens have been proposed to offer several advantages over phenotype-based *in vivo* methods as a starting point for drug repositioning [165] but modulators of the activity of ER α-glucosidases need to cross both the plasma- and ER-membranes. This challenge makes cheap and easy *in vivo* screening a more appealing strategy towards the development of compounds that possess both activity and the ability to reach their cellular targets in the ER lumen.

In order to develop an *in vivo* assay for ERQC inhibition by candidate molecules, it is necessary to relate their activity to an observable phenotype that depends on ERQC. *Arabidopsis thaliana* plants (Columbia-0 ecotype, *aka* Col-0) recognise characteristic bacterial structures, such as the N-terminus of bacterial Elongation Factor - Thermal unstable (EF-Tu) and flagellin [166, 131]. The N-acetylated peptide elf18 comprises the first 18 amino acids of EF-Tu, while the flg22 peptide

spans 22 conserved amino-acids of bacterial flagellin. The response to elf18 is mediated by the EF-Tu receptor (EFR) [167] and the response to flg22 is mediated by the Flagellin Sensitive 2 (FLS2) receptor [168].

Both EFR and FLS2 are secreted glycoproteins that reach the plant cell membrane after attaining their native folding in the ER. Accumulation and signalling of EFR (but not of FLS2) are impaired in the plant by mutations in the genes coding for calreticulin3 and UGGT [59], both of which are ERQC components. Similarly, plants carrying mutations in the genes coding for the two subunits of ER α-Glucosidase II (α-GluII, the central ERQC enzyme) are insensitive to elf18 but responsive to flg22 [159].

3.1.2 *N*B-DNJ *in planta* toxicity assay

The iminosugar *N*B-DNJ [95, 169] is well-tolerated in humans, with concentrations as high as 1 mg/mL (around 5 mM) displaying no cytotoxicity [170]. To test the toxicity of *N*B-DNJ in *Arabidopsis thaliana*, 15-day-old plants were treated with increasing concentrations of inhibitor, on the basis of a previously developed growth inhibition assay [168].

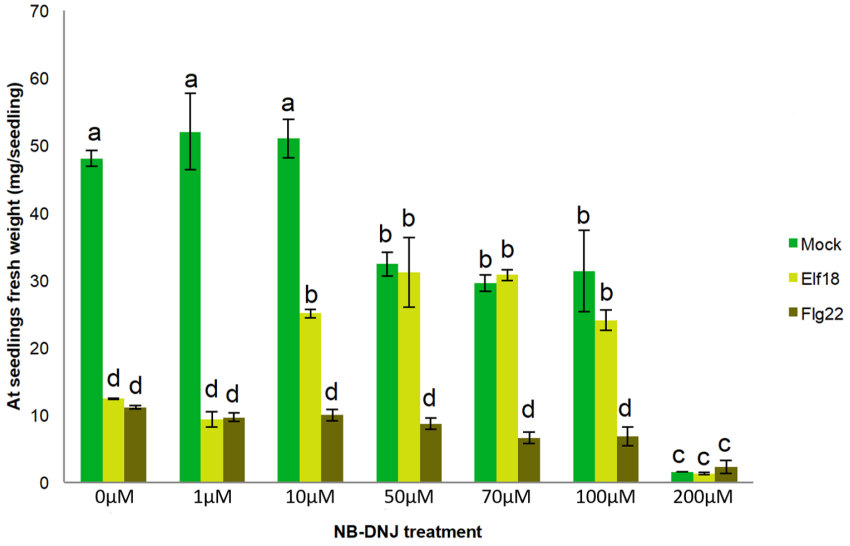


Figure 3.1: *N*B-DNJ toxicity assay. *Arabidopsis thaliana* (*At*) seedlings, Col-0 ecotype, were grown in liquid 1/2 MS medium containing different concentrations of *N*B-DNJ (0 μM, 1 μM, 10 μM, 50 μM, 70 μM, 100 μM, 200 μM) in combination with 100 nM of elf18 or flg22 (see colours in the legend). Values represent the mean of at least five independent experiments ($\pm$ s.e.; n = 5) with similar results in each experiment. Each letter (a, b, c and d) indicates a group of samples whose values are statistically equivalent to the ones in the same group, at the ($p < 0.01$) level determined by ANOVA with Scheffè's *post hoc* test.

In the mock treatment experiments in Figure 3.1, the impact of *N*B-DNJ treatment on plant

growth was measured. Arrest of wild-type seedling growth at *N*B-DNJ concentrations higher than 200 μM indicated toxicity. At lower *N*B-DNJ concentrations (1 and 10 μM), there was no impact on seedling growth. Supplementation of *N*B-DNJ at the intermediate concentrations of 50 μM, 70 μM and 100 μM resulted in reduced plant growth (about 30–40% of wild-type), with clear differences between elf18- and flg22-treated plants at the same concentrations. The concentration of 70 μM *N*B-DNJ was chosen for further experiments, Figure 3.1.

3.1.3 ERCQ modulation in *At* mutant plants

A number of mutants that are useful to study ERQC *in planta* have been described in the literature, but for most of them the innate immune response to bacterial elicitors has been measured only by monitoring plant weight and some physio-pathological parameters related to plant immune response (see Table 3.1). We decided to compare the effect of the *N*B-DNJ treatment on a WT plant with a plant that carries mutations on the genes encoding for the α or β subunits of α-GluII.

Mutant name	Kind of mutation	Affected gene encoded for
rsw3	S599F	α-subunit of *At*α-GluII
psl5-1	S517F	α-subunit of *At*α-GluII
psl4-2	T-DNA insertion	β-subunit of *At*α-GluII
uggt1-1	T-DNA insertion	*At*UGGT

Table 3.1: *Arabidopsis thaliana* α-GluII and UGGT mutants. List of the mutants used in this work. The mutations in *rsw3* and *psl5-1* affect the α-subunit of *At*α-GluII. The mutation in *psl4-2* affects the β-subunit of *At*α-GluII. The mutation in *uggt1-1* affects *At*UGGT. The mutants *rsw3*, *psl5-1* and *psl4-2* are further described in [159] while the mutant *uggt1-1* is further described in [171].

Two missense mutants are reported in *Arabidopsis* for the ER α–Glu II α catalytic subunit: the *At*Glu II S517F also known as *psl5-1* mutant and the *At*ER α–Glu II S599F also known as *rsw3* mutant (see Table 3.1). Both these mutants grow normally under standard conditions but show impaired response to elf18 treatment [159]. Both *psl5-1* and *rsw3* are missense mutations on the sequence of the α-subunit of *At*α-GluII, raising the question as to any residual ER α–Glu II activity.

To check if we could inhibit the putative residual *At*ER α-Glucosidase II activity in these mutants, we grew *At psl5-1* mutant in the presence of 70 μM *N*B-DNJ. Under these experimental conditions, the growth of germinated embryos is completely impaired - and treatment with 70 μM *N*B-DNJ is lethal to the *At psl5-1* mutant, Figure 3.2. These results suggest that *At psl5-1* may possess residual activity, the latter only completely abrogated by the iminosugar treatment.

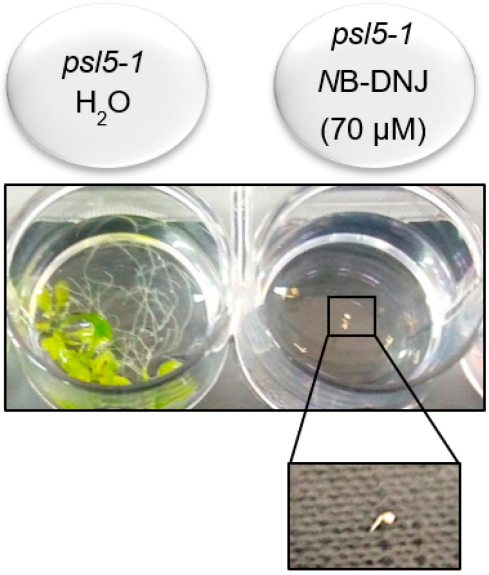


Figure 3.2: *At psl5-1* and *N*B-DNJ. *Arabidopsis thaliana psl5-1* seedlings show impaired growth in presence of 70 μM *N*B-DNJ.

Alternatively, the extra toxic effects due to glycolipid metabolism inhibition by *N*B-DNJ may explain the observed toxicity [157]. As the site of the *At*Glu II *rsw3* S599F amino acid substitution is located further away from the catalytic residues (*At*Glu II Asp512 and Asp588) than the site of the S517F mutations in *psl5-1*, it is expected that the *rsw3* mutant may also carry residual activity, as suspected also by Soussillane et al. [172].

To check if this is the case, we repeated the *N*B-DNJ iminosugar treatment on the *At*ER α–Glu II *rsw3* mutant. *N*B-DNJ was used in different concentrations (10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 300 nM, 100 nM, 50nM) on three-day-old seedlings, concentrations of *N*B-DNJ lower 500 nM were not toxic to *rsw3* mutants: embryo growth was not significantly impaired by 300 nM *N*B-DNJ (Figure 3.3). At *N*B-DNJ iminosugar dose higher than or equal to 500 nM, the growth of *rsw3* embryos was completely abolished, Figure 3.3.

When the same treatment was applied to the *psl4-2* mutant plant, which has a T-DNA insertion within the sequence encoding for the β-subunit of *At*α-GluII, the growth of germinated embryos was not impaired and treatment with 70 μM *N*B-DNJ was not lethal. The *psl4-2* mutant showed much better tolerance to *N*B-DNJ, as shown in Figure 3.4. Concentrations of *N*B-DNJ lower than 50 μM, had almost no impact on seedling growth of *psl4-2* mutant plants. The *psl4-2* mutant showed a phenotype closer to Col-0 rather than *psl5-1*. The β subunit of α-GluII carries the retrieval signal sequence "HDEL" at its C-terminal end and thus mediates ER-retention of the α-GluII α subunit. The observation of no severe phenotype, in contrast to what previously observed when treating *psl5-1* with a concentration of *N*B-DNJ higher than 500nM, made us wonder about the residual α-GluII activity in absence of its retrieval signal. We like to call this residual α-GluII activity "residual activity in transit" (see Figure 4.1 and its accompanying text in the Chapter 4).

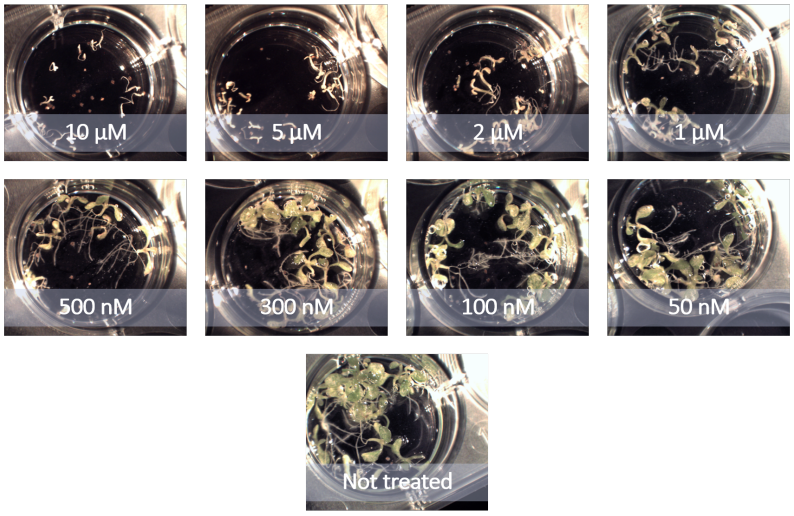


Figure 3.3: Effects of *N*B-DNJ-treatment on the *Arabidopsis thaliana rsw3* mutant. 15 days-old *rsw3* plants grown with different concentration of *N*B-DNJ. Each well contains about 10 seeds. Above 300 nM, *N*B-DNJ impairs plant embryo growth. Results derive from three independent experiments (± s.e. n = 10).



Figure 3.4: Effects of *N*B-DNJ-treatment on the *Arabidopsis thaliana psl4-2* mutant. 15 days-old *psl4-2* plants grown with different concentration of *N*B-DNJ. Each well contains about 10 seeds. Results derive from three independent experiments (± s.e. n = 10).

3.1.4 Bacterial peptides response assay

The plant response to elf18 can be used to monitor the folding of the EFR receptor, which in turn depends on ER α-Glucosidase II activity, with the flg22 treatment providing a negative control. To better investigate the ability of the two mutants to respond to bacterial peptides, we preliminary evaluated the response of WT plants (Col-0 ecotype) to elf18 or flg22 elicitation (as described in Section 2.2.4). As expected, both treatments reduced significantly plant growth. The *rsw3* mutant was insensitive to elf18 and highly sensitive to flg22, see Figure 3.5, in keeping with what is reported in the literature [159]. When the same treatment was applied on the *psl4-2* mutant, a reduced sensitivity to elf18 was observed, with a phenotype that was intermediate between the

WT and the *rsw3* ones, while the same *psl4-2* mutant still displayed high sensitivity to flg22 with a loss of weight comparable to the one observed for the *rsw3* mutant.

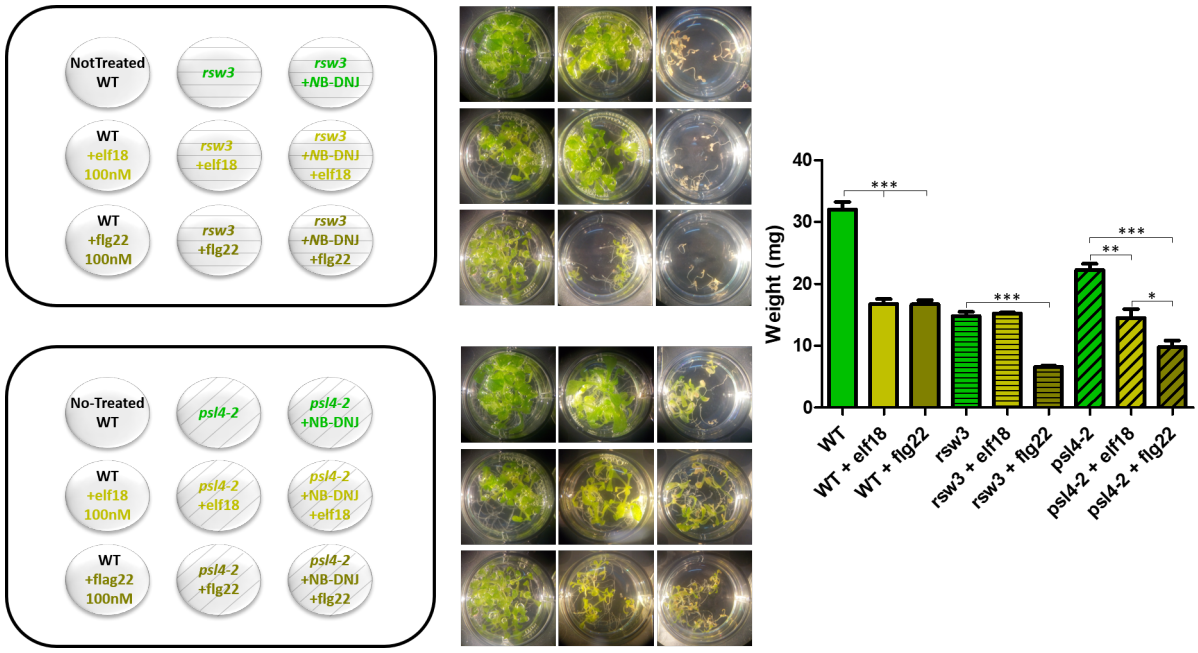


Figure 3.5: Effects of elicitors-treatment on *rsw3* and *psl4-2* mutants. 13-day-old seedlings control or treated with 100 nM elf18 or flg22 elicitors with and without 70 μM *N*B-DNJ. **Left** Schemes of the distribution of the treatment on the plate. The nine wells are depicted, each with the plant genotype and the amounts of *N*B-DNJ and/or elf18/flg22 used. **Centre** Images of the respective plants after the treatments. Each well contains about 10 seedlings. **Right** Mean of the plants fresh-weight (expressed in mg) of at least three independent experiments ($\pm$ s.e. n = 10), statistical analysis determined by ANOVA with Bonferroni's test (* = $p < 0.5$, ** = $p < 0.01$, *** = $p < 0.001$). WT: wild type.

Next, we investigated the expression level of two genes typically induced by elf18 and flg22 elicitation in the *At*ER α-Glu II *rsw3*, *psl5-1*, *psl4-2* and *uggt1-1* mutant plants (see Table 3.1). *Phosphate-induced 1* (*PHI-1*) and *RET-OX* are well known early elicitor-induced genes [173]. 10-day-old Col-0, *rsw3* and *psl4-2* seedlings were treated with elf18 (or flg22 used as a control) for 30 min, as described in Section 2.2.4. Low expression levels of both *PHI-1* and *RET-OX* were recorded in the *rsw3* mutant seedlings in response to elf18 but not to flg22, while Col-0 seedlings showed a normal defence response characterised by higher expression levels of defence genes with both elicitors. The *psl4-2* mutant confirmed to have an intermediate response to the elicitor elf18: the expression level of *PHI-1* and *RET-OX* indicated a less severe reaction to the presence of the elicitor elf18 compared to the WT plant, but not to the same extent as the *rsw3* mutant. Figure 3.6.

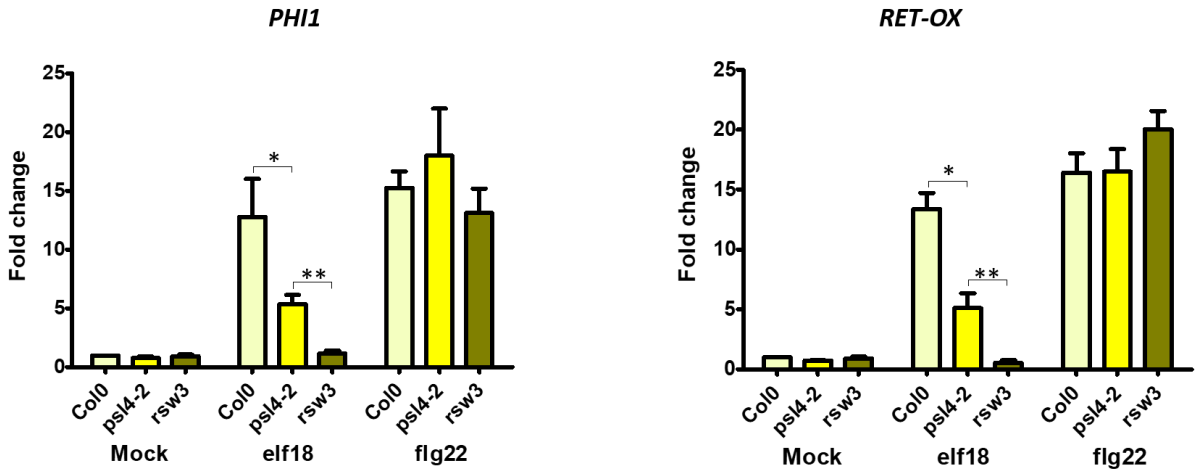


Figure 3.6: Expression of defence genes in the *At rsw3* and *psl4-2* plants. 10-day-old seedlings were treated with water or elf18 and flg22 elicitors. Expression of phosphate-induced 1 (*PHI-1*) and reticulin-oxidase homologue (*RET-OX*) genes were determined by quantitative reverse transcription-PCR (qRT-PCR) 30 min after treatment. Transcript levels are shown as the mean of at least three independent experiments (± s.e. n = 20 in each experiment) normalised to ubiquitin 5 (*UBQ5*) used as a reference. In both panels, asterisks indicate statistically significant differences according to Student's *t*-test (* = $p < 0.5$, ** = $p < 0.01$, *** = $p < 0.001$).

At the calnexin cycle checkpoint end, both insertional UGGT mutants and point mutants were previously reported: in the first category fall *uggt1-1* and *uggt1-2*, (carrying T-DNA insertions in the 11th and 16th exons, respectively [171]) and the *uggt-4* mutant, characterised by a T-DNA insertion in the 24th intron [58]. With regard to point mutants that showed impaired immune response to EFR, Saijo et al. [59] reported three point mutants (*psl2-1*: *At*UGGT D1497N, *psl2-2*: *At*UGGT E306K and *psl2-4*: *At*UGGT W1443STOP), together with a double mutant (*psl2-3*: *At*UGGT R1409K/W1443STOP) and a large deletion mutant (*psl2-5*: *At*UGGT Δ304-1613). Furthermore, Li et al. [58] reported another point mutant of UGGT, which was named uggt-3: *At*UGGT W1523STOP. Concerning the insertional mutants *uggt1-1* and *uggt1-2*, Blanco-Herrera et al. [171] reported some residual re-glucosylation activity for *uggt1-1* in comparison with *uggt1-2*, used in the same study. Both these mutants were reported to be more susceptible to pathogen attack, but their response to bacterial elicitors has not been investigated so far. However, the *uggt1-1* mutant is characterised by a T-DNA insertion in the 11th exon, at a position corresponding to amino acid *At*UGGT H499, and thus much earlier than the *uggt1-2* mutant (where T-DNA was mapped on the 24th intron) and the *uggt-4* mutant (where T-DNA was mapped on the 16th exon) for which efl18 insensitivity was reported [58].

We chose to investigate the *A. thaliana* T-DNA UGGT insertion mutant (*uggt1-1*) using EFR signalling. To confirm the dependency of the elf18 response on *At*ER α–Glu II we decided to inves-

tigate the effects of *N*B-DNJ treatment on the *uggt1-1* mutant response to elf18 elicitation, using the plant response to flg22 as a positive control. Results shown in Figure 3.7 indicated that, unlike what was observed with the Glu II mutant *rsw3*, *N*B-DNJ is not lethal in the background of the *uggt1-1* mutant: treatment with the iminosugar caused a growth reduction similar to that observed in Col-0 WT plants. Elicitation experiments showed that this mutant was highly susceptible to flg22. Elf18 elicitation impacted on *uggt1-1* plant growth much less than flg22, causing a lower plant weight reduction, which was not statistically significant when compared to untreated plants. Only co-treatment of *At uggt1-1* with 70 μM *N*B-DNJ caused complete insensitivity to elf18, likely because of Glu II inhibition, Figure 3.7.

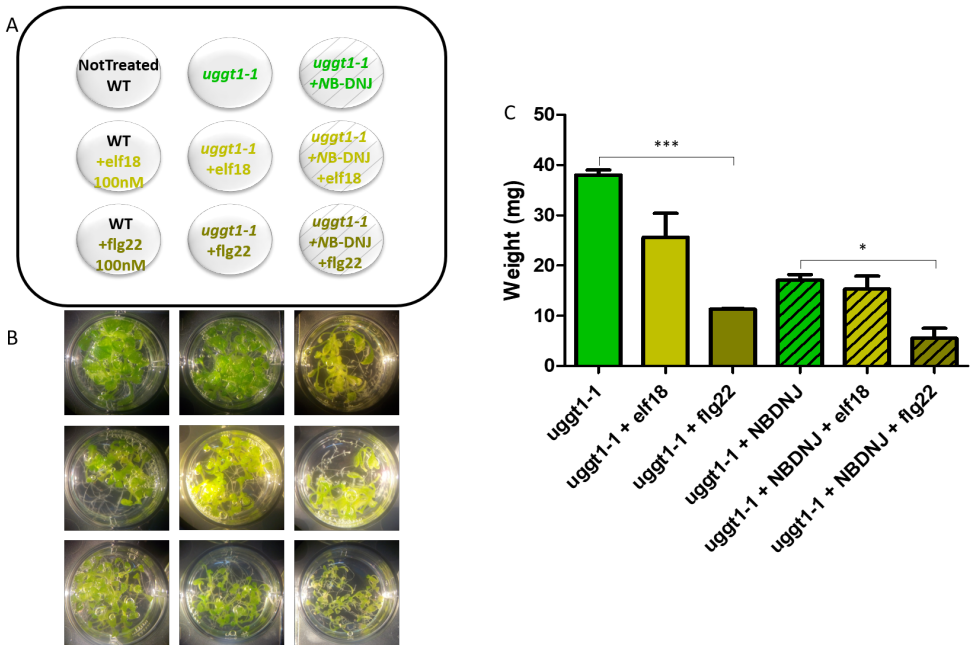


Figure 3.7: Effects of elicitor treatments on the *At uggt1-1* mutant. 13-day-old seedlings control or treated with 100 nM elf18 or flg22 elicitors with and without 70 μM *N*B-DNJ. **(A)** Scheme of the treatment distribution on the plate. The nine wells are depicted, each with the plant genotype and the amounts of *N*B-DNJ and/or elf18/flg22 used. **(B)** Images of the plants after the treatments. Each well contains about 10 seedlings. **(C)** Mean of the plants fresh-weight (expressed in mg) of at least three independent experiments ($\pm$ s.e. n = 10), statistical analysis determined by ANOVA with Bonferroni's test (* = $p < 0.5$, ** = $p < 0.01$, *** = $p < 0.001$).

Finally, we verified the expression of innate immune defence genes in the *uggt1-1* mutant upon elf18 or flg22 elicitation. Under our experimental conditions, we found that both elicitors triggered a significant increase in the expression levels of both *PHI-1* and *RET-OX* and only the concomitant application of *N*B-DNJ blocked the EFR signalling pathway, Figure 3.8. *N*B-DNJ treatment induces resistance to elf18 in the same manner in WT and *uggt1-1* plants: indeed, there is no statistical difference between *PHI-1* and *RET-OX* transcript levels in WT and *uggt1-1* plants after

treatment with elf18.

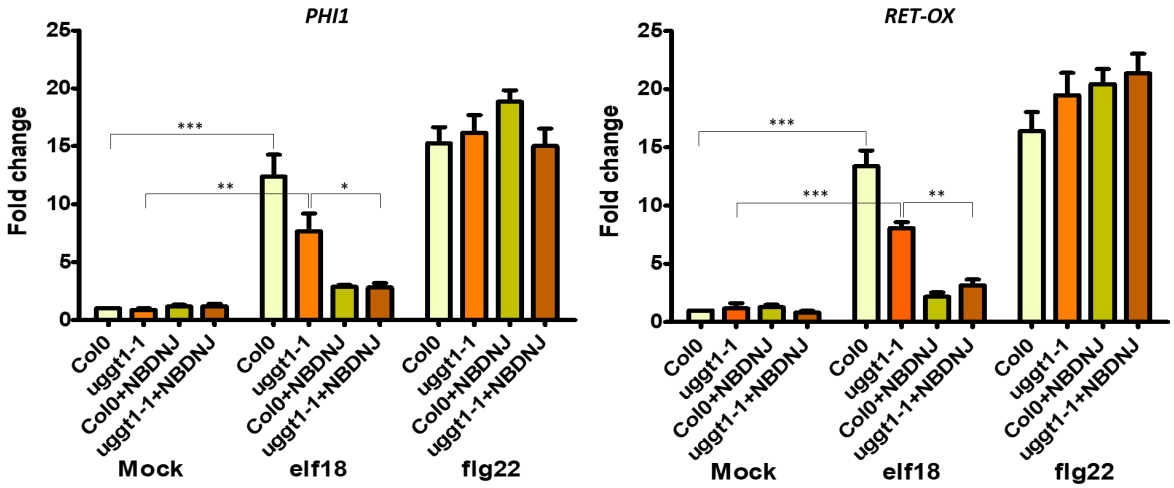


Figure 3.8: Expression of defence genes in the *At uggt1-1* plant. 10-day-old seedlings were treated with water or elf18 and flg22 elicitors, with and without 70 μM *N*B-DNJ. Expression of *PHI1* and *RET-OX* genes were determined by qRT-PCR 30 min after treatment. Transcript levels are shown as the mean of at least three independent experiments ($\pm$ s.e. n = 20 in each experiment) normalised to *UBQ5* (*ubiquitin 5*) used as a reference. In both panels, asterisks indicate statistically significant differences among mock and treatments within the same genotype (same color) according to Student's *t*-test (* = $p < 0.5$, ** = $p < 0.01$, *** = $p < 0.001$).

3.1.5 The level of *BiP3* in *rsw3* is higher than in WT and *psl4-2* plants

To evaluate whether the two mutants, *psl4-2* and the *rsw3*, were intrinsically and/or differently stressed by the respective mutation, the expression level of *BiP3* was investigated. The gene *BiP3* encodes for a major chaperone in the ER lumen and plays an important role in the unfolded protein response [174]. Figure 3.9 shows the expression level of *BiP3* in Col-0, *psl4-2* and the *rsw3* plants. The plants were not treated with elicitors nor with *N*B-DNJ, the seeds were sterilised as described in Section 2.2.2 and grown in 1/2 MS liquid medium for 10 days. The RNA was purified and the cDNA synthesised as described in Section 2.2.5, the qRT-PCR analysis was performed as described in Section 2.2.6. The expression level of *BiP3* was significantly higher in *rsw3* than in Col-0 and *psl4-2* plants.

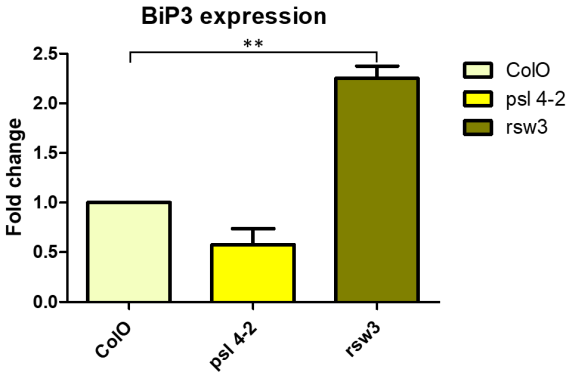


Figure 3.9: Expression of *BiP3* in the *At rsw3* and *psl4-2* plants. Expression of *BiP3* in 10-day-old seedlings was determined by quantitative reverse transcription-PCR (qRT-PCR). Transcript levels are shown as the mean of at least three independent experiments ($\pm$ s.e. n = 20 in each experiment) normalised to ubiquitin 5 (*UBQ5*) used as a reference. Asterisks indicate statistically significant differences among Col-0 and the mutants according to Student's *t*-test (* = $p < 0.5$, ** = $p < 0.01$, *** = $p < 0.001$).

The results in this section show that the iminosugar *N*B-DNJ, which reached phase I clinical trials as an antiviral drug, can inhibit *A. thaliana* ER α–Glu II and treatment of plants with *N*B-DNJ influences the immune response to the bacterial EF-Tu peptide elf18 but not to the bacterial flagellin-derived flg22 peptide. Additionally, the treatment of the ER α–Glu II *rsw3* mutant with *N*B-DNJ at concentrations higher than 500 nM (see Figure 3.3) resulted to be lethal to *A. thaliana* embryos while, on the other hand, the treatment of a *psl4-2* mutant plant with a 100-fold higher concentration of *N*B-DNJ showed much less severe a phenotype (Figure 3.4). The differences between *rsw3* and *psl4-2* are further proven by the observation that the level of *BiP3* transcription is higher in *rsw3* than in WT and *psl4-2*, indicating a higher level of ER stress.

3.2 Structural and functional characterisation of the ERAD checkpoint heterodimer

At the same time, our lab was also interested in better characterising the ERAD system studying the structure of EDEMs, the ERAD glycoprotein misfolding checkpoints, whose structure is still unknown and in using the plant to design an *in vivo* EDEMs activity assay. ERAD has a critical role in the maintenance of the integrity of the ER proteome by constantly monitoring the folding of polypeptides under normal and stress conditions. Glycoprotein folding in the ER is not always successful: in mammals, 30% of all newly synthesised proteins are estimated to be incorrectly folded [175]. Moreover, genetic mutations, errors in transcription and translation, toxic compounds and cellular stresses such as defects in cellular redox regulation due to hypoxia, oxidants and reducing agents that interact with disulfide bonds in the ER lumen, glucose starvation and abnormalities in

calcium regulation, lead to a significant increase in incorrectly folded glycoproteins [176]. Unfolded glycoprotein accumulation induces stress in the ER, which triggers an evolutionarily conserved series of signal-transduction events, which constitutes the unfolded protein response (UPR). These signalling events aim to ameliorate the accumulation of unfolded proteins in the ER; however, when these events are severe or protracted they can induce cell death [176]. ER stress has been associated with a wide range of diseases, including neurodegeneration, stroke, bipolar disorder, cardiac disease, cancer, diabetes, muscle degeneration and others [176, 115] for which modulation/enhancement of EDEM enzymes could be a therapeutic strategy. For example in [115] it was shown that the overexpression of EDEM1 in *Drosophila melanogaster* led to the suppression of age-associated physiological decline. We set out to use the plant to better characterise the activity of the EDEM enzymes and to use the *bri1-5* mutant to design an *in vivo* activity assay. Moreover, since the structures of the EDEMs are still unknown we also started work towards the structural determination of *Ct*HTM1P - the *Chaetomium thermophilum* orthologue of the EDEM enzymes.

3.2.1 Plant's ERAD

Arabidopsis thaliana orthologues of HRD1, HRD3/SEL1L, and the Man binding lectin YOS9 have been identified [177, 178, 179]. The *Arabidopsis* GH47 family is composed of five members, termed MNS1 to MNS5. MNS1 and MNS2 are Golgi-α-mannosidases and display a redundant function during N-glycan trimming in *Arabidopsis* [180]. MNS3 fulfills the biosynthetic role of ER-α-mannosidase in plants, acting upstream of MNS1/MNS2 in the N-glycan processing pathway. Thus, MNS1 to MNS3 together are responsible for N-glycan processing from $Man_9GlcNAc_2$ to $Man_5GlcNAc_2$, which is a prerequisite for complex N-glycan formation in the Golgi apparatus. MNS4 and MNS5 are ER-resident proteins required for ERAD of the misfolded BRI1 variants *BRI1-5* and *BRI1-9* [126]. Professor Richard Strasser (University of Natural Resources and Life Sciences, Vienna) kindly donated to us seeds of *Arabidopsis thaliana mns4-1/mns5-1* double and *mns4-1/mns5-1/bri1-5* triple mutants, generated from the combination of *mns4-1* (SALK_119093), which transcript contains a deletion of the N-terminal targeting region and part of the mannosidase homology domain, *mns5-1* (GT5_84786), which contains a Ds activation-tagging T-DNA in the first exon of the *MNS5* gene and *bri1-5* which carries the mutation Cys69Tyr.

3.2.2 *Ct*HTM1P and *Ct*PDI1P are likely thermostable orthologues of known ERAD checkpoint mannosidases

Our search of the *Ct* genome (at `https://c-thermophilum.bork.embl.de/`) for orthologues of *Sc*HTM1P and *Sc*PDI1P yielded *Ct*HTM1P (CTHT_0058730, Uniprot entry G0SCX7, `https://www.uniprot.org/uniprot/G0SCX7`) and *Ct*PDI1P (CTHT_0067360, Uniprot entry G0SGS2, `https://www.uniprot.org/uniprot/G0SGS2`), respectively, as the best hits. Sequence alignments were performed with BLAST® (at `https://blast.ncbi.nlm.nih.gov`): the *Ct*HTM1P and *Sc*HTM1P sequences showed a sequence identity of 37.35%, best aligned at the N-terminal of the *Ct*HTM1P (the catalytic domain). *Ct*PDI1P and *Sc*PDI1P showed a sequence identity of 40.39%.

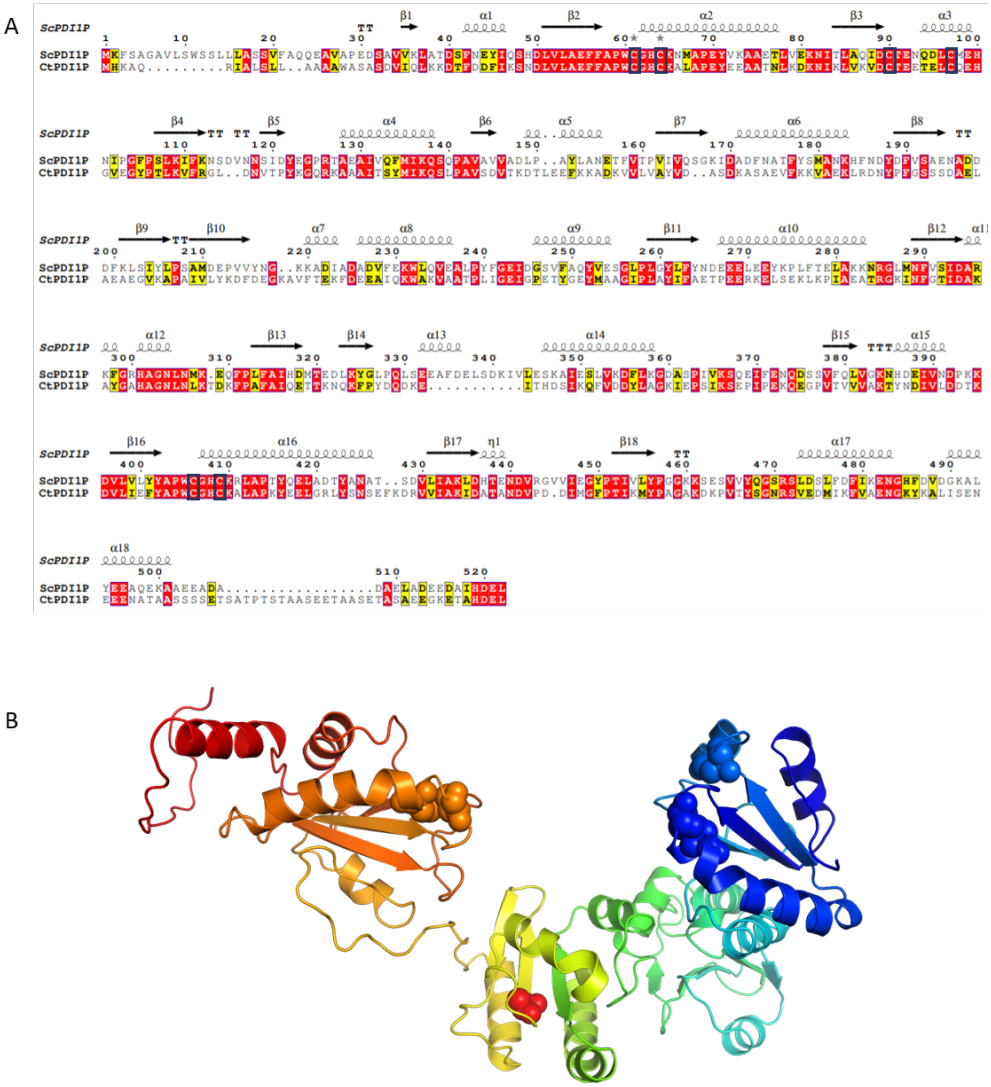


Figure 3.10: Homology modelling of *Ct*PDI1P based on a *Sc*PDI structure. **A**: sequence alignment of *Ct*PDI1P and its closest sequence homologue, *Sc*PDI (Uniprot entry `https://www.uniprot.org/uniprot/P17967`). Red: conserved residues. Black font on yellow background: Risler Similarity Score [181] greater than 0.7. The conserved Cysteines forming disulfide bridges in *Sc*PDI and likely involved in disulfide bridges in *Ct*PDI1P too (*Sc*PDI1P Cys61-Cys64, Cys90-Cys97 and Cys406-Cys409 corresponding respectively to *Ct*PDI1P Cys50-Cys53, Cys79-Cys86 and Cys385-Cys388) are boxed in blue. The secondary structure of the *Sc*PDI1P is indicated above the sequence and it is taken from the PDB ID 2B5E. Figure generated in `ESPript` [182]. **B**: homology model for *Ct*PDI1P (residues 21-501) built with `Modeller` run from the `HHPred` server at `https://toolkit.tuebingen.mpg.de`, using PDB ID 2B5E as a template. The model is in cartoon representation, painted blue to red from N- to C-terminus. *Ct*PDI1P Cys50-Cys53: dark blue spheres; *Ct*PDI1P Cys79-Cys86: lighter blue spheres; *Ct*PDI1P Cys385-Cys388: orange spheres; red spheres: *Ct*PDI1P Ala303, the site corresponding to the *Sc*PDI1P mutation L313P [79]. Figure generated in `PyMOL` [183].

The homology model of *Ct*HTM1P used in this thesis is based on the structure at the PDB ID 1X9D, the human ER α1,2-Mannosidase I. This provides a model of the N-terminal catalytic domain which has high-degree sequence conservation within the Glycosyl Hydrolase family 47 (GH47), but it was not possible to build a model for the C-terminal domain of *Ct*HTM1P whose sequence shows variability across Eukaryotes. The best sequence homology between *Ct*PDI1P and structures in the PDB is the one with PDB ID 2B5E [84], the structure of a *Saccharomyces cerevisiae* Protein Disulfide Isomerase. In particular, based on the sequence alignment performed with `HHPred` at `https://toolkit.tuebingen.mpg.de`, the following disulfide bridges are likely to be conserved: Cys50-Cys53 and Cys79-Cys86 in the first active Thioredoxin (TRX) domain; and Cys385-Cys388 in the second active TRX domain (see sequence conservation in Figure 3.10A). Figure 3.10B illustrates the homology model for *Ct*PDI1P built with `Modeller` run from the `HHPred` server at `https://toolkit.tuebingen.mpg.de`, using PDB ID 2B5E as a template.

A disulfide bond between *Sc*HTM1P and *Sc*PDI1P involves either the fifth (Cys579) or the sixth Cys (Cys644) residue of the *Sc*HTM1P mannosidase [83] These Cys residues are conserved in the fungal ERAD mannosidase C-terminal domain of *Vanderwaltozyma polyspora*, *Candida glabrata*, *Ashbya gossyppi*, *Kluyveromyces lactis* and *Pichia stipitis* [83] but not in *Chaetomium thermophilum*. Additionally, the ERAD mannosidase C-terminal domain is not conserved in mammalian EDEMs: the C-terminal domain is only found in fungal orthologues of *Sc*HTM1P and does not show any homology to known protein domains [83].

Mutation *Sc*PDI1P L313P weakens the *Sc*HTM1P :*Sc*PDI1P interaction [79]. Within the sequence of *Ct*PDI1P , the residue corresponding to *Sc*PDI1P Leu313 is *Ct*PDI1P Ala303, coloured in red in Figure 3.10B.

3.2.3 *Ct*HTM1P recomplementation of the *mns4-1/mns5-1/bri1-5* triple mutant genetic background

Arabidopsis thaliana MNS4 and MNS5 are the plant's orthologues of *Ct*HTM1P . Our *in vivo* *Ct*HTM1P activity assay is based on *Ct*HTM1P transfection in *Arabidopsis thaliana mns4-1 mns5-1 bri1-5* triple mutant plant. The *At bri1-5* mutant plant harbours a misfolding mutation in the brassinosteroid receptor BRI1 [179]: the mutated, misfolded BRI1 receptor is re-glucosylated by the plant UGGT and remains trapped in the ER and is eventually committed to ERAD. As a result, the *At bri1-5* mutant displays a dwarf phenotype.

In the triple mutant plant *Arabidopsis thaliana mns4-1/mns5-1/bri1-5*, inactivation of the *At*MNS4 and MNS5 genes, in the *bri1-5* background, rescues secretion of the misfolded BRI1 receptor and restores a phenotype similar to WT [126]. For our *in vivo* *Ct*HTM1P activity assay, we have re-complemented *Arabidopsis thaliana mns4-1/mns5-1/bri1-5* triple mutant plant with *Ct*HTM1P , and monitored plant growth. The hypothesis was that if *Ct*HTM1P were active in recognising and demannosylating BRI1-5 (i.e. acting as an EDEM), the plant would revert to a dwarf phenotype.

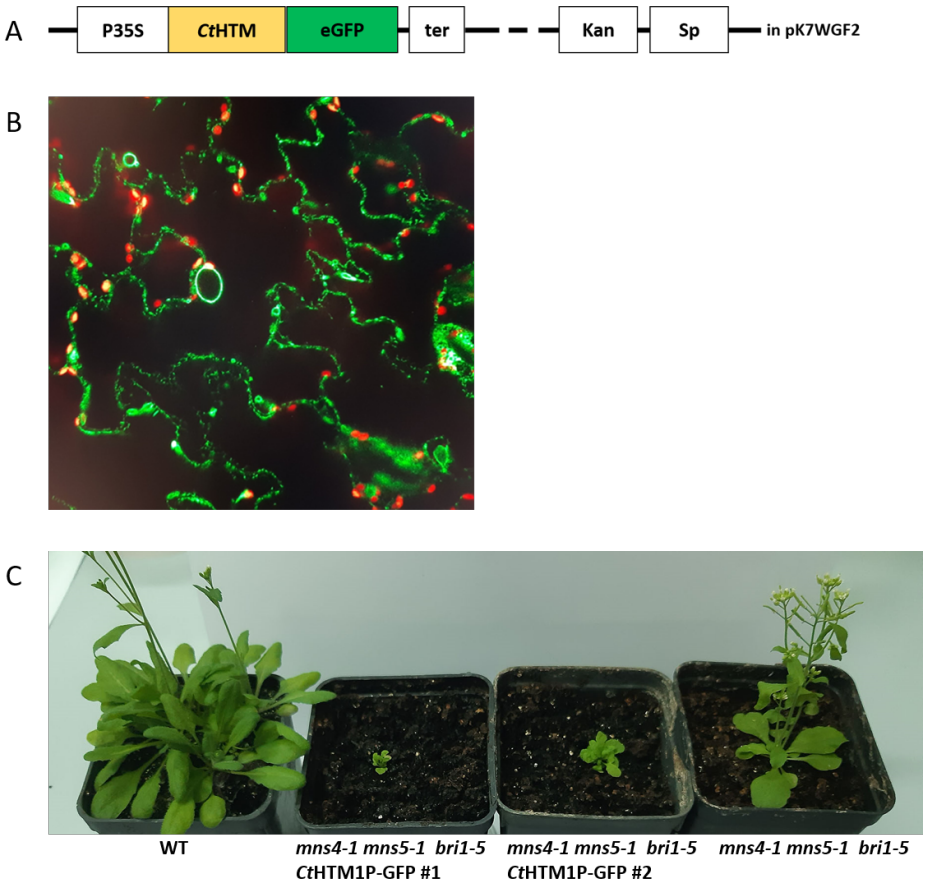


Figure 3.11: Stable transformation of *Arabidopsis thaliana* with *Ct*HTM1P :GFP. **A**: Scheme of the construct *Ct*HTM1P :GFP_pK7WGF2 used the plant transformation. **B**: confocal microscopy image of the *Ct*HTM1P localised in the ER membrane. Imaging of *Ct*HTM1P -GFP was performed as described in Section 2.2.10. **C**: Stable transformation of the *mns4-1/mns5-1/bri1-5* triple mutant (on the right) with *Ct*HTM1P -GFP (plants in the middle), compared to the WT *Arabidopsis thaliana* (on the left). The plants indicated as *Ct*HTM1P -GFP#1 and *Ct*HTM1P -GFP#2 are just two different batches of plants with identical genetic background and treated in the same way.

The gene *Ct*HTM1P was first cloned into the pDONR221 vector (Invitrogen) using the primers listed in Table 2.1, *Ct*HTM1P was then cloned using the gateway cloning system into the destination vector pK7WGF2 and fused at the C-terminal with enhanced GFP (eGFP) as described in Section 2.2.8, Figure 3.11A.

Agrobacterium tumefaciens was transformed with the construct and it was used first to transiently transform *Nicotiana tabacum* leaves (Section 2.2.8) and later for stable transformation of *Arabidopsis thaliana mns4-1/mns5-1/bri1-5* triple mutant (Section 2.2.9), Figure 3.11B.

The results in Figure 3.11B, indicate that *Ct*HTM1P localises to the ER as expected, and the ones in Figure 3.11C suggest *Ct*HTM1P is active in recognising the misfolded receptor *bri1-5* and consequently removing mannose residues from the glycans on *bri1-5*, dispatching the receptor to

ERAD and its final degradation. Our results indicate that the expression of *Ct*HTM1P gene in the triple mutant background restores the *bri1-5* dwarf phenotype in the background of the double plant EDEM mutation, likely by re-endowing the plant with ERAD mannosidase activity.

3.2.4 Expression of *Ct*HTM1P constructs for recombinant protein production

We first attempted to express *Ct*HTM1P by transfecting HEK293F cells with the *Ct*HTM1P $_{50-1092}$-pHLsec plasmid as described in Section 2.4, but the experiment yielded negligible amounts of *Ct*HTM1P , and resulted in the purification of a contaminant instead: n-Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), as described in Appendix A.

In an attempt to optimise the expression level of *Ct*HTM1P , four different *Ct*HTM1P constructs were then designed, as shown in Figure 3.12, and cloned in the expression vector pHLsec as described in Section 2.3.5.1.

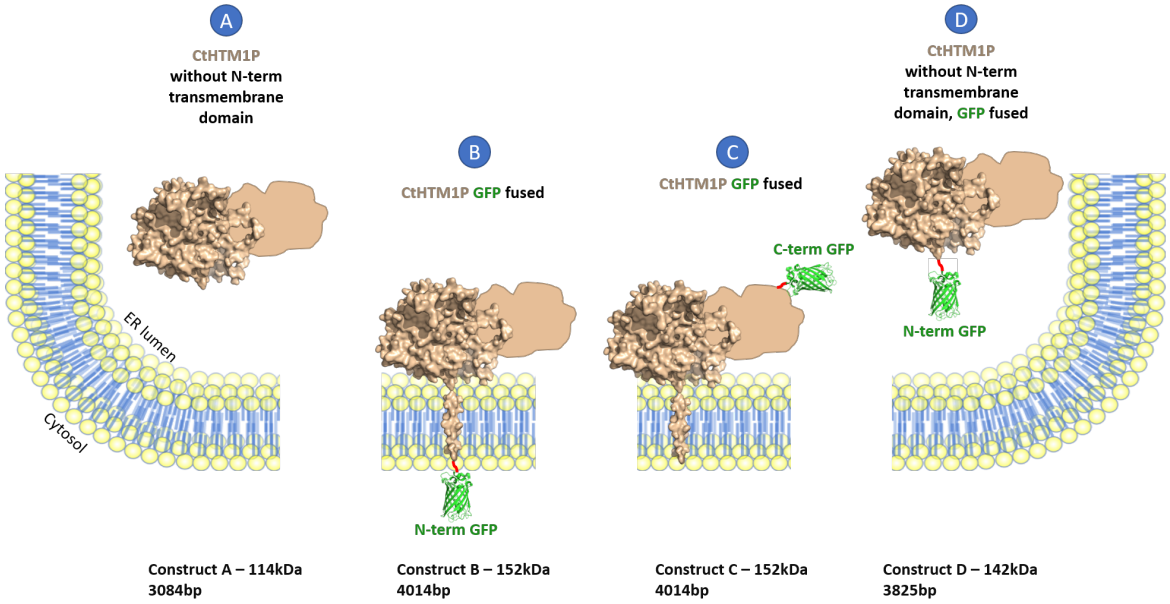


Figure 3.12: Scheme of the four different *Ct*HTM1P constructs. **A**: *Ct*HTM1P soluble domain only, the N-terminal sequence containing the transmembrane domain was removed, the protein sequence starts from residue 64 onwards. **B**: *Ct*HTM1P full length with GFP-fusion at the N-terminal. The GFP is on the cytosolic side of the ER membrane. **C**: *Ct*HTM1P full length with GFP-fusion at the C-terminal. The GFP is on the lumenal side of the ER membrane. **D**: Construct in A GFP-fused at the N-terminal domain. In brown: homology model of the Mannosidase domain of *Ct*HTM1P$_{\mathrm{D}}$ (Uniprot G0SCX7, residues 67-618) built using PDB 1X9D residues 245-698 as a template; the *Ct*HTM1P$_{\mathrm{D}}$ N-terminal transmembrane part, residues 28-53, was homology modelled on the basis of PDB ID 6Z0I, residues 174-199; the brown shape represent the C-terminal domain of the *Ct*HTM1P$_{\mathrm{D}}$ for which it was not possible to build a homology model (the orientation between the homology model and the shape symbolising the C-terminal domain is also arbitrary); in red: peptide linker (LEVLFQGP); in green: GFP protein (PDB ID 5B61).

HEK293F cells were then transfected with the four plasmids (A, B, C and D) as described in 2.4. The four different HEK293F cell cultures were harvested 4 days post-transfection by centrifuging at 4,000 g for 5 minutes. The supernatant was made 1× PBS by adding an appropriate 10× PBS stock solution volume and the pH was adjusted to 7.4 by adding a few drops of 2 M NaOH, before bottle-top vacuum filtration through a 0.2 μm filter, while the pellet was lysed and treated as described in 2.4.1. Samples from both the culture supernatant and the lysed cells of the cultures transfected with the four plasmids were Western blotted, in order to detect the presence and/or the localisation of the expressed recombinant proteins. The Western blot using an Anti-GFP antibody was performed following the protocol described in Section 2.4.2 and the result is shown in Figure 3.13.

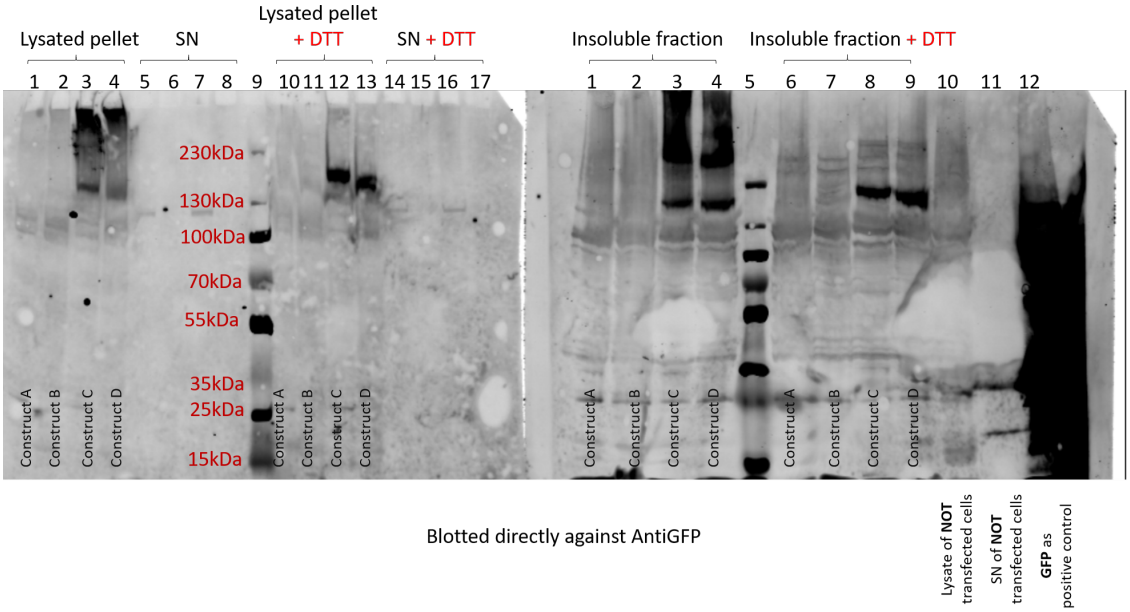


Figure 3.13: Western blot analysis using Anti-GFP antibody. **Left**; lanes 1, 2, 3 and 4: pellet from the lysed cells from four different cultures transfected separately with constructs A, B, C and D. Lanes 5, 6, 7 and 8: supernatant of the four different culture. Lane 9: Molecular Mass marker. Lanes 10, 11, 12, 13: pellet from the lysed cell of the four different cultures treated with DTT. Lanes: 14, 15, 16, 17: supernatant of the four different cultures treated with DTT. **Right**; lanes 1, 2, 3, 4: insoluble fractions from the four different cultures transfected separately with constructs A, B, C and D. Lane 5: Molecular Mass marker. Lanes 6, 7, 8 ,9: insoluble fractions from the four different cultures treated with DTT. Lane 10: lysate of not-trasfected cells (as negative control). Lane 11: supernatant of not-trasfected cells (as negative control). Lane 12: GFP protein (as positive control).

The Western blot showed bands of the expected mass, indicating that *Ct*HTM1P construct C and construct D were expressed in the HEK293F cells. Both proteins seemed to be present only in the lysate sample, and not in the culture supernatant where we expected to find them, as the

pHLsec vector is designed for secreted expression (see vector map in Figure 2.10) and our constructs lack ER-retrieval sequences.

Moreover, aggregation/oligomerisation was apparent in absence of reducing agent. Indeed, the addition of 1× DTT to the sample disrupted such oligomerisation while the SDS did not: this points towards disulfide-bridge mediated interactions with other copies of the same protein or with endogenous HEK293F cells' partner protein(s).

To assess the concentration of reducing agent needed to disrupt the aggregation, *Ct*HTM1P$_{\mathrm{C}}$ and *Ct*HTM1P$_{\mathrm{D}}$ samples were treated with a gradient of the reducing agent Tris (2-Carboxyethyl) phosphine (TCEP, Sigma-Aldrich 51805-45-9). The result is shown in Figure 3.14.

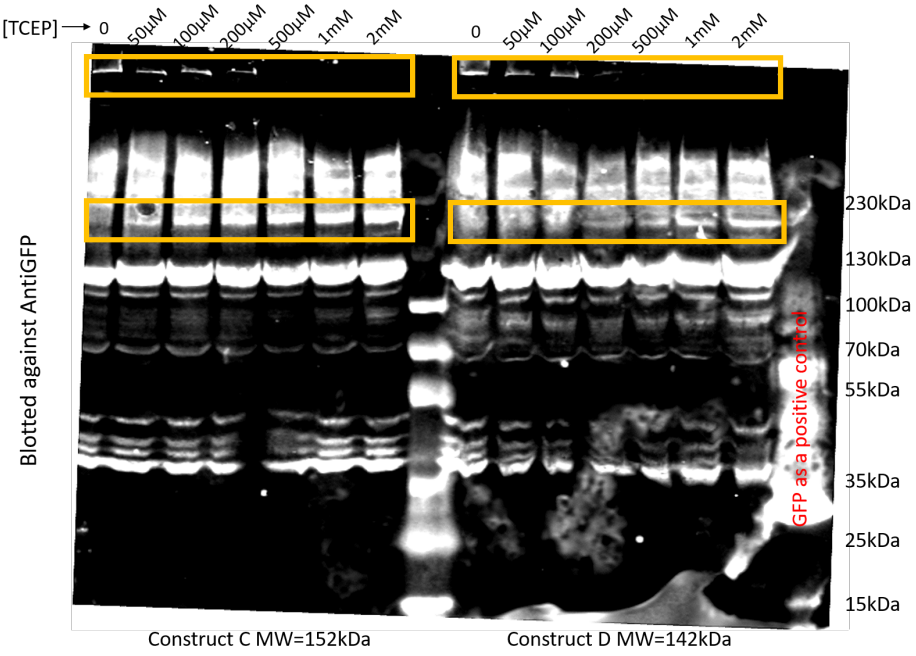


Figure 3.14: Western blot analysis of a gradient of reducing agent (TCEP). Protein expressed after cell transfection with constructs C (**Left**) and D (**Right**) were treated with an increasing concentration of TCEP. Each lane has a different concentration of TCEP: 0, 50 μM, 10 μM, 200 μM, 500 μM, 1 mM and 2 mM. GFP protein was used as a positive control. The upper yellow rectangle shows the disappearance of the highest molecular mass band as the concentration of reducing agent increases. The lower yellow rectangle shows the increase in the concentration of ≈140kDa protein band as the highest molecular mass band disappears.

The Western blot was performed using the anti-GFP antibody as described in Section 2.4.2; the samples treated with a concentration of 500 μM TCEP did not show any visible band corresponding to aggregation/oligomerisation, and while the upper disulphide-bonded species disappeared, the ≈140kDa protein band became stronger, confirming that recombinant *Ct*HTM1P$_{\mathrm{D}}$ undergoes intermolecular oligomerisation/aggregation mediated by SS bonds in the conditions used for protein production/purification.

3.2.5 Purification of *Ct*HTM1P$_D$

We chose to work with *Ct*HTM1P$_D$ first because it lacks its N-terminal transmembrane domain, so that its purification does not require extraction from the ER membrane (unlike the one of *Ct*HTM1P$_C$).

3.2.5.1 IMAC affinity chromatography

Our WB analysis in Figure 3.13 showed the presence of *Ct*HTM1P$_D$ in the lysate, thus the HEK293F cell culture, transfected with *Ct*HTM1P$_D$, was harvested 4 days post-transfection and the cells were processed as described in Section 2.4.1. The filtered lysate was then passed onto a 1 mL HisTrap HP Ni IMAC column and the IMAC affinity chromatography followed the protocol described in Section 2.4.3. The result is shown in Figure 3.15.

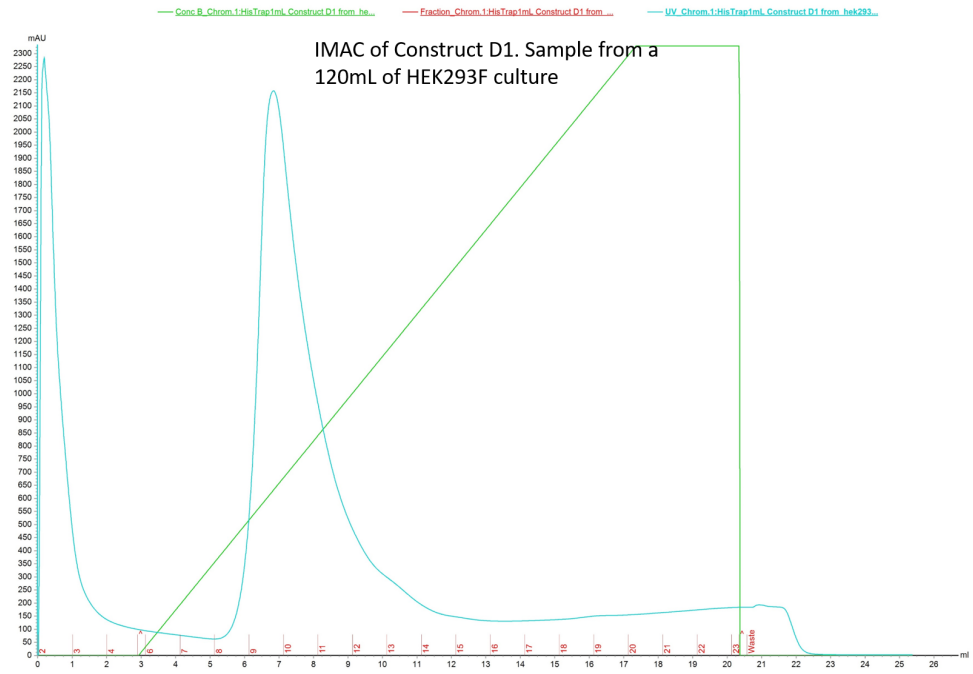


Figure 3.15: IMAC HisTrap elution. A 1mL HisTrap Ni^{++} column was loaded with the cell lysate of HEK293F cells transfected with the construct D (see Figure 3.12). The elution of proteins bound to the column was determined by an increasing gradient of buffer IMACE (see the recipe of the buffer in Section 2.3.1). The elution of the bound protein started at 5mL and was collected in fractions from 8 to 15.

A single peak of UV light absorption at 280 nm was observed indicating the presence of proteins in our sample, eluted when the gradient of imidazole (contained in buffer IMACE) was around 15-20%. This is usually the amount of imidazole needed to detached a 6×His-tagged protein from the Ni^{++} IMAC column. We proceeded directly with a further step of purification using a size exclusion column to evaluate the mass of the proteins contained in the peak-corresponding fractions and to

purify our protein away from the contaminants that could have aspecifically interacted with the column.

3.2.5.2 Size exclusion chromatography

The protein-containing fractions were pooled together and concentrated before being applied to a HiLoad Superdex 200 16/60 size exclusion chromatography (SEC) column as described in Section 2.4.4. The result is shown in Figure 3.16. The SEC elution profile showed different peaks. To identify fractions containing *Ct*HTM1P$_D$, a Western blot was performed using an anti-GFP antibody (the method is described in Section 2.4.2). Figure 3.16 shows the result of a Western blot protein detection, which was performed as described in Section 2.4.2.

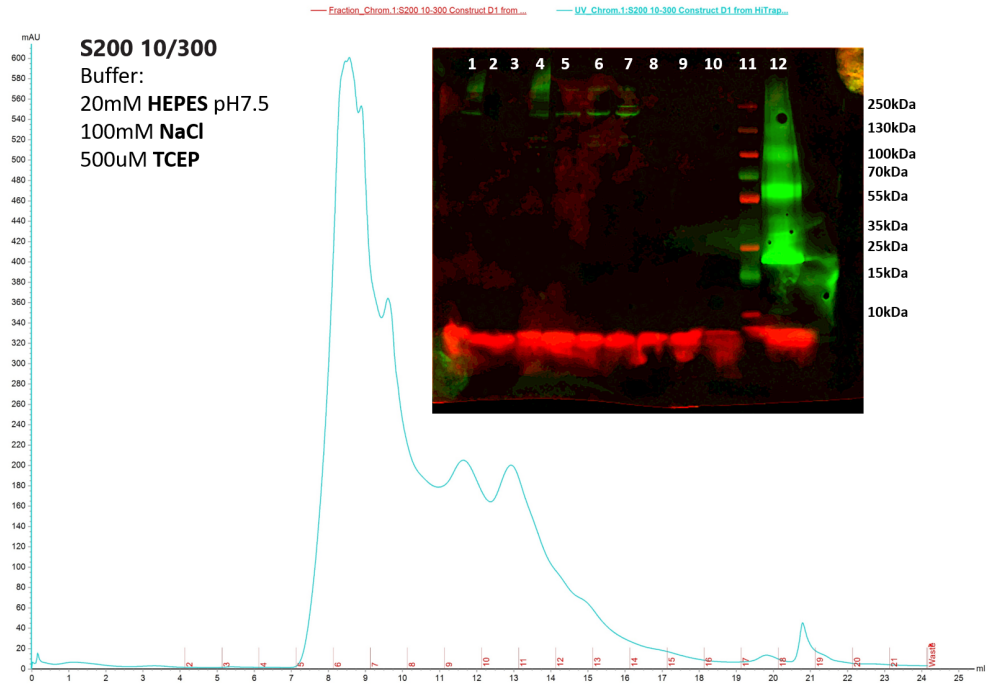


Figure 3.16: Size exclusion chromatography and Western blot analysis. **SEC chromatogram**: Fractions from 8 to 15, eluted from IMAC (3.15), were pooled together and concentrated to 500μL. The concentrated sample was loaded onto an S200 10/300 Gel Filtration column. The entire elution volume (1CV) was collected in fractions of 1.5 mL and the elution peaks were analysed through Western blot using an anti-GFP antibody. **Western blot**: sample is lane 1 - insoluble fraction (pellet after lysis); lane 2 - supernatant of HEK293F cell culture; lane 3 - HisTrap flow-through; lane 4 - pooled fractions eluted from IMAC; lane 5 - SEC fraction 6; lane 6 - SEC fraction 7; lane 7 - SEC fraction 9; lane 8 - SEC fraction 10; lane 9 - pellet of not transfected cells (as negative control); lane 10 - empty lane; lane 11 - molecular mass marker; lane 12 - GFP protein (as positive control).

The green bands showed in the inset are proteins recognised by the anti-GFP antibody, indi-

cating the expression of our *Ct*HTM1P$_D$ construct. The protein *Ct*HTM1P$_D$ was identified in the fractions 6, 7 and 9 corresponding to the first three protein peaks. The detected protein bands were of the expected molecular size but unfortunately, the expression level of *Ct*HTM1P$_D$ in HEK293F cells resulted to be very low and the amount of the purified protein was not enough to be used for structural investigations.

3.2.6 Co-expression of *Ct*HTM1P$_D$: *Ct*PDI1P

As discussed in Section 2.3.2 we identified *Ct*PDI1P as the orthologue of *Sc*PDI1P, which was reported to form the heterodimer *Sc*HTM1P:*Sc*PDI1P. In an attempt to increase the expression level of the protein *Ct*HTM1P$_D$ we decided to co-express in HEK293F cells both components of the *Chaetomium thermophilum* putative heterodimer *Ct*HTM1P$_D$: *Ct*PDI1P . Transient co-transfection of HEK293F cell lines with *Ct*HTM1P$_D$: *Ct*PDI1P constructs was carried out in collaboration with Dr Gábor Tax as follows.

HEK293F cells were cultured in FreeStyle^{TM} 293 Expression Medium. The initial concentration of the culture was 0.5x10^6 cells/mL in 50 mL GIBCO FreeStyle 293 Media. Cells were kept in an orbital shaker incubator at 37 °C, shaking at 120 rpm and in 5% CO_2. Twenty-four hours later or when they reached the 1x10^6 cells/mL concentration, the cells were transferred to Batch 1: 200 mL volume in a 500 mL flask; Batch 2: 600 mL volume (three 200 mL aliquots, each in a 500 mL flask). The flasks were transfected with 0.5 μg of each plasmid DNA per million cells, following the protocol described in 2.4.

3.2.7 Co-purification of *Ct*HTM1P$_D$: *Ct*PDI1P

The *Ct*HTM1P$_D$: *Ct*PDI1P complex was purified in collaboration with Dr Pietro Roversi by IMAC affinity chromatography followed by SEC, as described in what follows.

3.2.7.1 IMAC affinity chromatography

Batches 1 and 2: (in parentheses the values for Batch 2) The 200 mL (600 mL) of HEK293F cells were harvested 72 hours (96 hours) post-transfection. The cells were harvested as described in Section 2.4.1 and the supernatant was made 1× PBS by the addition of 20 mL (60 mL) of 10× PBS and filtered through a 0.22 μm filter. It was loaded on a 5 mL HisTrap affinity column (equilibrated in PBS) at a speed of approximately 1.6 mL/min. The column was washed with 8 column volumes (CV) of wash/binding buffer (PBS, 10 mM Imidazole) and the protein eluted as described in Section 2.4.3, collecting 1.0 mL (2.0 mL) fractions. Protein-containing fractions were loaded on a 4-12% SDS-PAGE gel. Each lane was loaded with 15 μL sample. Non-reducing lanes: 6 μL 4× dye. Reducing lanes: 4 μL 4× dye and 2 μL of 10× DTT. The gel was run at 200 mV for 35' in NuPAGE^{TM} MES SDS Running Buffer (Thermo Fisher Scientific). Molecular Mass markers: Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder, 10 to 250 kDa. Figure 3.17 illustrates the SDS-PAGE gel of the elution profile of Batch 1.

The protein-containing fractions were analysed through an SDS-PAGE with and without reducing agent. The heterodimer *Sc*HTM1P:*Sc*PDI1P is held together by hydrophobic interactions and it is further stabilised by an intermolecular disulfide bridge [83], a different distribution of protein bands in presence of reducing agent may indicate that also the heterodimer $Ct\text{HTM1P}_\text{D}:Ct\text{PDI1P}$ is mediated by a disulfide bridge. According to the result shown in Figure 3.17 this does not seem to be the case.

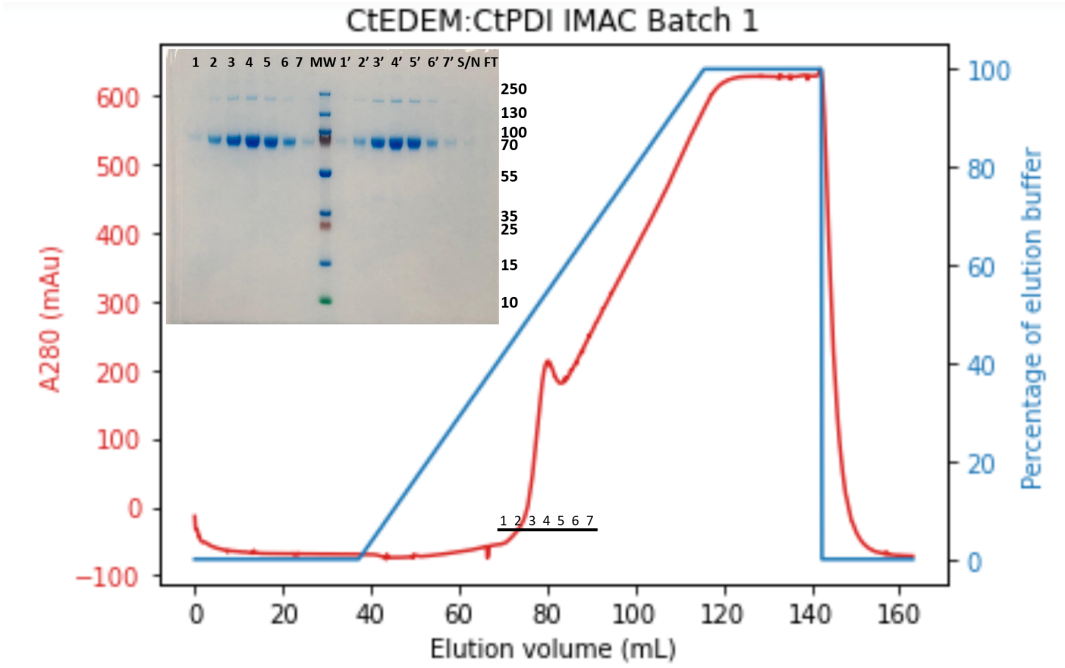


Figure 3.17: Elution profile of the IMAC purification of the first batch of $Ct\text{HTM1P}_\text{D}:Ct\text{PDI1P}$. The blue line describes the gradient of elution buffer (IMACE). The red line describes the UV absorbance at λ=280 nm. The elution buffer absorbs at λ=280 nm, likely because of metal impurities forming a ferrocene complex with the imidazole. Eluted protein, contained in fractions from 1 to 6, is visible as a peak on top of the linear imidazole gradient. In the upper left panel is shown the SDS-PAGE 4-12% Bis-Tris gel of the fractions 1-6, indicated in the chromatogram. **Left**: Lane 1 to 7 show samples from eluted fraction 1 to 7, respectively; **Centre**: MM marker; **Right**: lanes labelled with a prime contained samples from eluted fraction 1 to 7 treated with reducing agent. Gel run at 200V for 55' in MOPS buffer.

3.2.7.2 MALDI-MS/MS protein identification

Two bands were excised from the SDS-PAGE gel in Figure 3.17, the one above that looks of ∼142 kDa - which is the expected size for $Ct\text{HTM1P}_\text{D}$ - and the one below that looks of ∼70 kDa, and were sent to the FAO Metabolomics and Proteomics Laboratory at York University for trypsin digestion and MS analysis. For protein identification, MALDI-MS and MS/MS were performed using a Bruker ultraflex III MALDI-TOF/TOF, which produces higher-confidence identifications

by including tandem spectral data in `Mascot` searches [147]. As a general rule MALDI tends towards identifying the most dominant proteins in any given sample. The ID of the proteins via a two unique peptides matched by each `Mascot` search (see Table 3.2).

Protein	Observed MM	Calc. MM	Expect	Peptide
*Ct*HTM1P$_D$	1090.6160	1089.6183	7.4×10^{-3}	^{186}K.VQVFETVIR.G^{196}
*Ct*HTM1P$_D$	1286.7172	1285.6878	3.3×10^{-4}	^{771}K.ITLQLVEEGER.E^{783}
*Ct*PDI1P	2594.1448	2593.1312	4.1×10^{-8}	^{76}K.VDCTEETELCQEHGVEGYPTLK.V^{99}
*Ct*PDI1P	2355.1311	2354.1026	6.5×10^{-6}	^{170}K.LRDNYPFGSSSDAELAEAEGVK.A^{193}

Table 3.2: MALDI-MS/MS *Ct*PDI1P and *Ct*HTM1P$_D$ protein identification after trypsin digestion (Batch 1). List of peptides. The Expect value indicates the probability that the observed match between MS/MS spectra and peptide sequence would be found by chance. Confident matches typically have Expect values less than 0.1. The number at both sides of the sequences indicate the position of the fragments in the corresponding protein sequence.

The results indicated in Table 3.2 show that amongst the peptides identified after trypsin digestion, two fragments belong to the *Ct*HTM1P sequence and two other ones to *Ct*PDI1P , thus confirming that the expressed proteins are the correct ones.

3.2.7.3 Size-exclusion chromatography

Once confirmed that the proteins eluted from IMAC contained *Ct*HTM1P and/or *Ct*PDI1P , the protein-containing fractions were pooled, concentrated and loaded on a size-exclusion chromatography column.

Batch 1: nine protein-containing fractions were pooled, diluted with 5 mL of SEC buffer (120 mM NaCl, 50 mM HEPES pH 7.2) and concentrated to 5 mL using a Vivaspin 20 mL centrifugal concentrator (MMCO 30 kDa), spinning for 10' at 5,000 × g. The sample was filtered through a 0.22 μm filter and loaded onto a 5 mL loop - connected to an ÄKTA purifier FPLC machine. An S200 16/60 SEC column was equilibrated in SEC buffer described above, and the experiment ran at 1 mL/min flow rate, collecting 1.8 mL fractions.

Batch 2: three protein-containing fractions were pooled (6 mL total volume). The sample was prepared as for Batch 1. An S200 16/60 SEC column was equilibrated in 150 mM NaCl, 20 mM MES pH 7.0, 1 mM $CaCl_2$, and the experiment run at 1 mL/min flow rate, collecting 1.8 mL fractions. Figure 3.18.

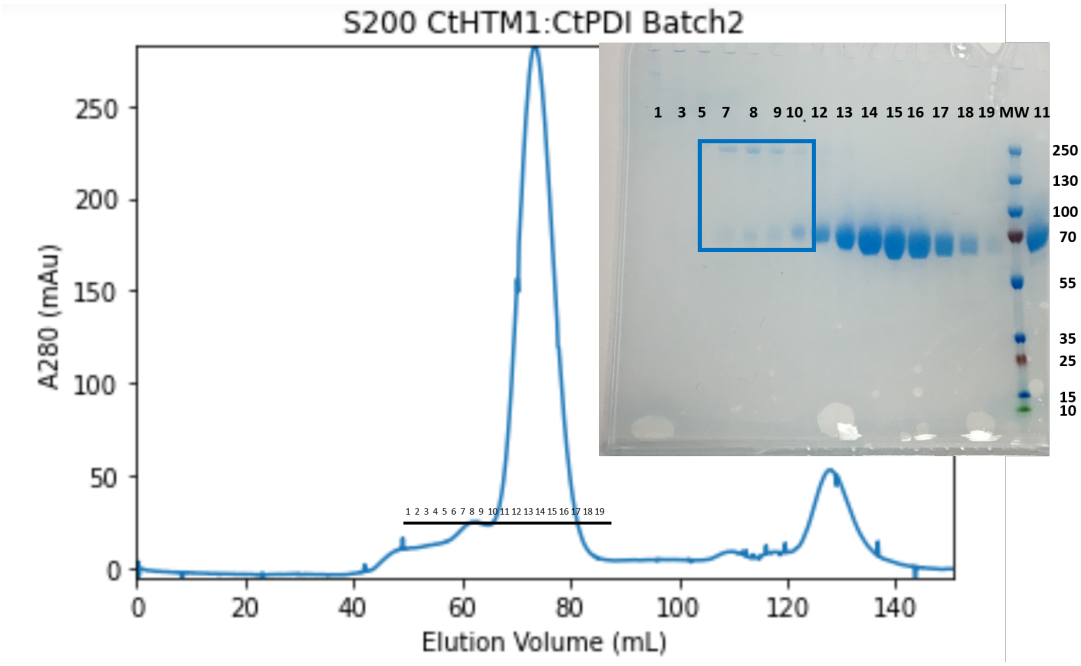


Figure 3.18: Elution profile of the S200 16/60 SEC purification of Batch 2 of CtHTM1P$_{\mathrm{D}}$:CtPDI1P. The blue line describes the UV absorbance at λ=280 nm. Eluted proteins, contained in fractions from 1 to 19, are visible as a peak. In the upper right panel is shown the SDS-PAGE 4-12% Bis-Tris gel of some of the fractions indicated in the chromatogram. The boxed fractions contain the CtHTM1P$_{\mathrm{D}}$:CtPDI1P heterodimer. Gel run at 200V for 55' in MOPS buffer.

Figure 3.18 shows the results of SEC carried out on Batch 2. One main peak was detected and the protein-containing fractions were loaded on a 4-12% Bis-Tris SDS-PAGE gel. Results shown in the right panel indicated that the heterodimer CtHTM1P$_{\mathrm{D}}$:CtPDI1P, which has an expected mass around 196 kDa (CtHTM1P$_{\mathrm{D}}$ is 142 kDa while CtPDI1P is 55 kDa, these masses do not take into account the mass of N-linked glycans on the two proteins), co-eluted in fractions from 7 to 10 and that in later fractions only CtPDI1P was detectable on an SDS-PAGE gel.

3.2.8 *Ct*PDI1P structure determination

The CtPDI1P containing fractions of the S200 elutions from Batch 1 and Batch 2 were separately pooled (Batch 1: 10.8 mL; Batch 2: 16 mL) and concentrated, each batch of protein in a separate 30,000 kDa MMCO Vivaspin concentrator, spinning at 5,000 g for 25'. At the end of this concentration: Batch 1: volume 60 μL, OD$_{280}$=0.877; Batch 2: volume 2.5 mL, OD$_{280}$=1.218. 120 μL of the latter were stashed aside, and the remaining sample concentrated for further 5' to volume 500 μL, OD$_{280}$=3.082. Figure 3.20 shows the SDS-PAGE gel of the Batch 1 sample at OD$_{280}$=0.877 (lanes 3, 3') and the Batch 2 samples at OD$_{280}$=1.218 (lanes 4, 4') and OD$_{280}$=3.082 (lanes 5, 5').

3.2.8.1 Crystallisation of *Ct*PDI1P

The Batch 2 *Ct*PDI1P sample was devoted to crystallisation trials. Pooled samples at OD_{280}=1.218 and OD_{280}=3.082 were concentrated in a 500 μL spin concentrator (MMCO=30,000 Dalton). After centrifugation, the sample was filtered through a 0.22 μm filter. The final volume was 160 μL and the absorbance of the sample was OD_{280}=12.84 - corresponding to [*Ct*PDI1P]=298 μM (calculated from ϵ_{280}=43,030 ($M^{-1}cm^{-1}$, as computed at `https://www.scripps.edu/cgi-bin/cdputnam/protcalc3`). With a molecular mass of 55,481 Dalton this corresponds to [*Ct*PDI1P]=5.37 mg/mL.

A Mosquito liquid handler robot was used to set up four 96-wells vapour diffusion sitting drop crystallisation experiments using the four commercial screens: Morpheus, Pact Prémiere, JCSG+ and Stura Macrosol, as described in Section 2.5.

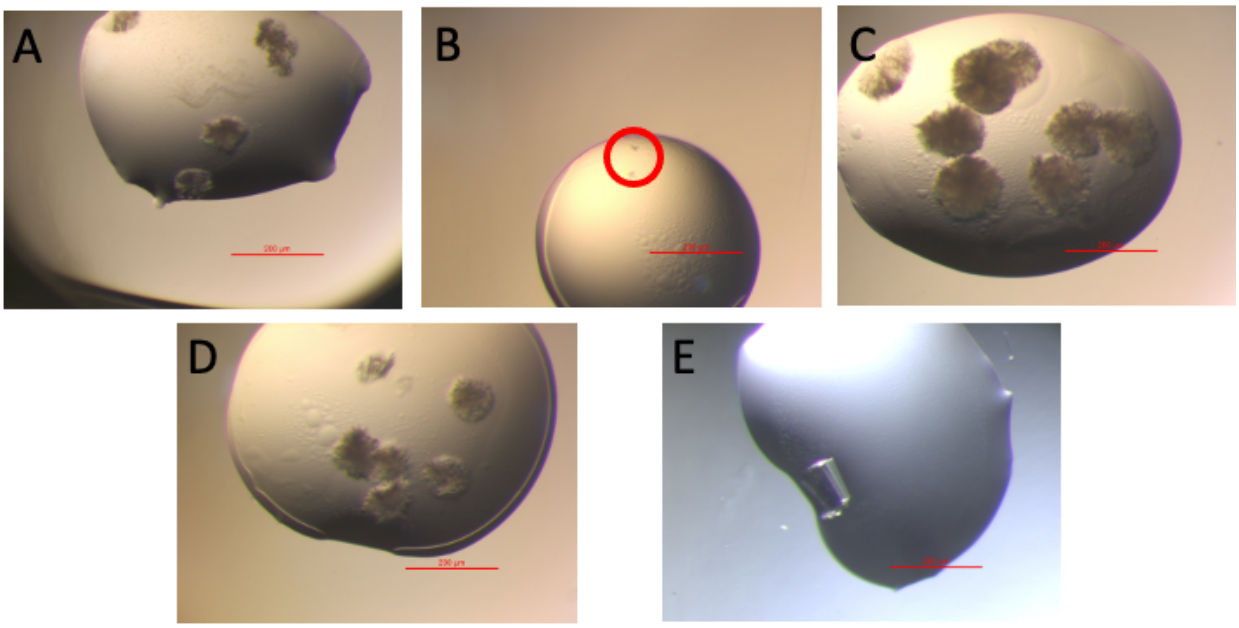


Figure 3.19: Crystals in *Ct*HTM1P_D :*Ct*PDI1P and *Ct*PDI1P crystallisation plates. All red bars correspond to 200 μm. **A**: *Ct*HTM1P_D :*Ct*PDI1P , Plate 1, Morpheus Screen condition H9 (upper drop); **B**: *Ct*PDI1P Plate 4, Morpheus Screen condition C1 (upper drop). The crystals are barely visible at this magnification and are circled in red; **C**: *Ct*PDI1P Plate 4, Morpheus Screen condition H9 (upper drop); **D** *Ct*PDI1P Plate 4, Morpheus Screen condition H9 (lower drop); **E**: *Ct*PDI1P Plate 5, Pact Premiere Screen condition A4 (lower drop).

The first crystallisation screens (from *Ct*PDI1P at 5.37 mg/mL) produced a number of samples that look crystalline but are not suitable for X-ray diffraction (see Figure 3.19B-D). The second attempt at crystallisation - using *Ct*PDI1P at 1.04 mg/mL - produced the crystal in Figure 3.19E) which is of excellent quality and will be used for X-ray diffraction.

3.2.9 *Ct*HTM1P$_D$:*Ct*PDI1P complex structure determination

The fractions of the S200 elution from Batch 2 containing *Ct*HTM1P$_D$:*Ct*PDI1P complex were pooled and concentrated in a 30,000 kDa MMCO Vivaspin concentrator, spinning at 5,000 g for 25'. At the end of this concentration: Batch 1: volume 40 μL, OD$_{280}$=0.017; Batch 2: volume 500 μL, OD$_{280}$=0.265. 60 μL of the latter were stashed aside, and the remaining sample concentrated for further 5' to volume 150 μL, OD$_{280}$=0.741. Figure 3.20 shows the SDS-PAGE gel of the Batch 2 samples at OD$_{280}$=0.401 (lanes 1,1') and OD$_{280}$=0.741 (lanes 2,2').

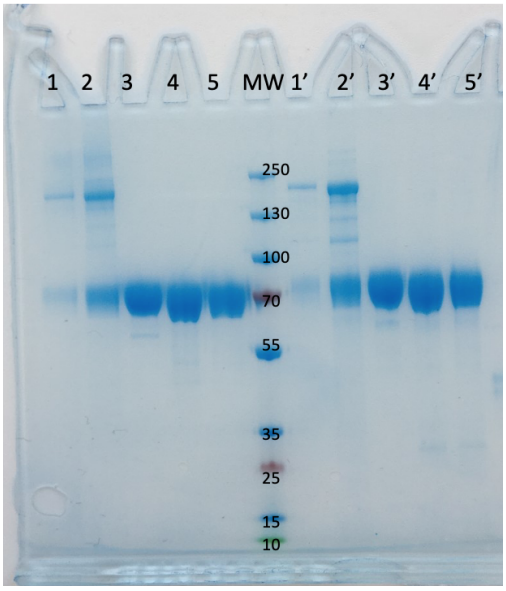


Figure 3.20: SDS-PAGE 4-12% Bis-Tris gel of the pooled and concentrated S200 16/60 SEC fractions of the *Ct*HTM1P$_D$:*Ct*PDI1P heterodimer and the *Ct*PDI1P. Gel run at 200 V for 55' in MOPS buffer. Lanes with primed numbers contain reducing agent. Lanes 1,1': 2 μL of *Ct*HTM1P$_D$:*Ct*PDI1P complex Batch 2, OD$_{280}$=0.265; lanes 2,2': 4 μL of *Ct*HTM1P$_D$:*Ct*PDI1P complex Batch 2, OD$_{280}$=0.741; lanes 3,3': 4 μL of *Ct*PDI1P Batch 1, OD$_{280}$=0.877; lanes 4,4': 3 μL of *Ct*PDI1P Batch 2, OD$_{280}$=1.218; lanes 5,5': 1 μL of *Ct*PDI1P Batch 2, OD$_{280}$=3.082. Lane 6: MM ladder with molecular mass in kDalton.

3.2.9.1 Crystallisation of *Ct*HTM1P$_D$:*Ct*PDI1P complex

The Batch 2 *Ct*HTM1P$_D$:*Ct*PDI1P complex sample at OD$_{280}$=0.741 was devoted to crystallisation trials. The sample was concentrated in a 500 μL spin concentrator (MMCO=30,000 Dalton) by centrifugation, and filtered through a 0.22 μm filter. The final volume was 40 μL and the absorbance of the sample was OD$_{280}$=1.03 - corresponding to [*Ct*HTM1P$_D$:*Ct*PDI1P]=5 μM (calculated from ϵ_{280}=210,620 (M^{-1}cm^{-1}, as computed at `https://www.scripps.edu/cgi-bin/cdputnam/protcalc3`). With MM=197,468 Dalton this corresponds to [*Ct*HTM1P$_D$:*Ct*PDI1P]=1.0 mg/mL. To this sample, 1 μL of a 20 mM stock solution of kifunensine was added, making it 500 μM kifunensine.

A Mosquito liquid handler robot was used to set up a 96-wells vapour diffusion sitting drop crystallisation experiment using the Morpheus commercial screen as described in Section 2.5. The first crystallisation screens (from *Ct*HTM1P$_D$:*Ct*PDI1P at OD$_{280}$=0.741) produced samples that look crystalline but are not suitable for X-ray diffraction (see Figure 3.19A).

3.2.9.2 Cryo-Electron Microscopy of *Ct*HTM1P$_D$:*Ct*PDI1P complex

The Batch 2 *Ct*HTM1P$_D$:*Ct*PDI1P complex sample at OD$_{280}$=0.265 was devoted to Electron Microscopy.

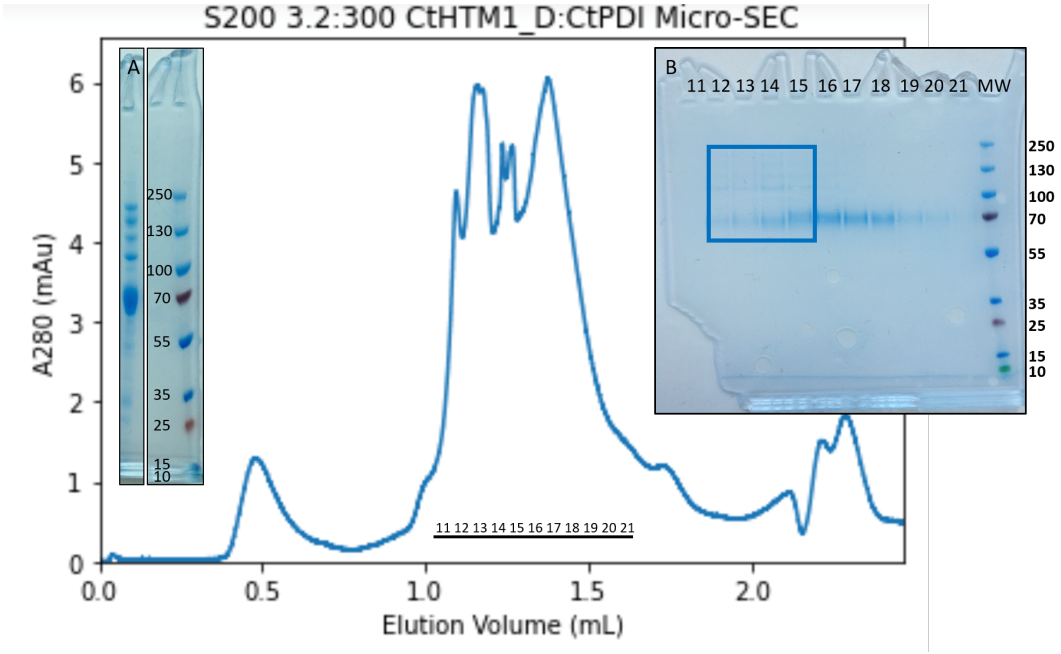


Figure 3.21: Elution profile of the S200 3.2:300 Micro-SEC purification of the pooled Batches 1 and 2 of *Ct*HTM1P$_D$:*Ct*PDI1P . The blue line describes the UV absorbance at λ=280 nm. The numbers on the black bar indicate the elution fraction numbers. **Inset A**: a 4-12% SDS-PAGE lane of the Batch 2 *Ct*HTM1P$_D$:*Ct*PDI1P complex sample at OD$_{280}$=0.741 after 10 days at 4 °C (gel run in MOPS buffer at 200 V for 55'). Proteolysis of *Ct*HTM1P$_D$ is visible as three additional bands at MM $\approx$ 135, 120 and 115 kDa. **Inset B**: a 4-12% SDS-PAGE gel of elution fractions 11-21 of the S200 3.2:300 Micro-SEC purification of the pooled Batches 1 and 2 of *Ct*HTM1P$_D$:*Ct*PDI1P (gel run in MOPS buffer at 200 V for 55'). Each sample was prepared with 15 μL of S200 3.2:300 Micro-SEC fraction plus 5 μL of non-denaturing 4x SDS-PAGE loading dye, and denatured for 1' at 90 °C prior to loading on the gel. Boxed fractions 12 to 15 appear to contain the 70 kDa *Ct*PDI1P and a number of proteolytic fragments derived from *Ct*HTM1P$_D$ (the lightest of which likely corresponds to the loss of GFP, $\approx$ 115 kDa, see Inset A). Fraction 13 was used for Cryo-EM on a graphene-oxide grid.

The grids for Cryo-EM screening were prepared by Dr Christos Savva of the Midlands Cryo-EM

Facility at Leicester University. Four samples of 10 μL each of *Ct*HTM1P$_{\mathrm{D}}$:*Ct*PDI1P complex in the SEC buffer (150 mM NaCl, 20 mM MES pH 7.0, 1 mM $CaCl_2$) at OD_{280}=0.265, 0.132, 0.062 and 0.031 (corresponding to 1.26, 0.63, 0.31 and 0.165 μM) were prepared by serial dilution. A 3 μL volume of each sample was applied to glow-discharged Quantifoil R1.2/1.3 gold mesh (Quantifoil Micro Tools GmbH, Germany) graphene oxide (Sigma- Aldrich)-coated holey carbon grids. After incubation for 1 min, excess volume was blotted with filter paper and sample-loading repeated 3X times prior to plunge freezing using a Vitrobot (FEI).

The Cryo-EM micrographs obtained from the grid made from the most concentrated sample showed protein oligomers/aggregates. It was hypothesised that oligomerisation/aggregation would be a result of the concentration step. The remaining *Ct*HTM1P$_{\mathrm{D}}$:*Ct*PDI1P samples from Batches 1 and 2 were pooled and the 95 μL were concentrated on a 500 μL spin concentrator down to 40 μL, filtered through a 0.22 μm filter. This sample has an OD_{280} of 0.88. These 40 μL were injected onto a Cytiva Superdex 200 Increase 3.2/300 size-exclusion chromatography column, equilibrated in 150 mM NaCl, 20 mM MES pH 7.0, 1 mM $CaCl_2$, collecting 50 μL fractions. Fraction 13 of the elution profile (see Figure 3.21) was used to prepare two graphene oxide Cryo-EM grids.

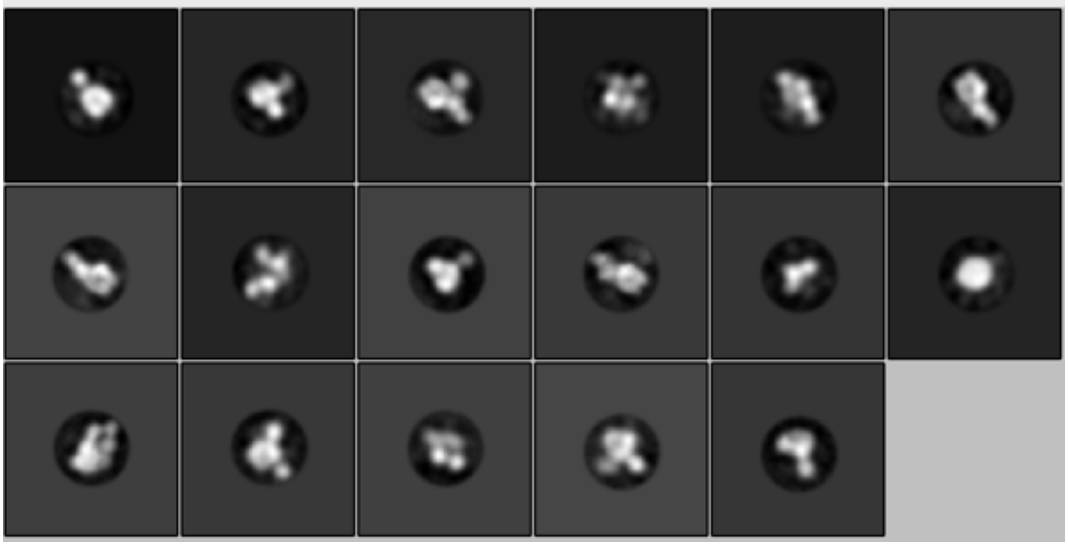


Figure 3.22: 2D classification of *Ct*HTM1P$_{\mathrm{D}}$:*Ct*PDI1P particles imaged by Cryo-EM. Every class illustrated in the panel comes from an average of many particles that showed the same orientation (bar a rotation around the direction of the electron beam). The 2D classes come from a set of 4,447 particles selected amongst the initial 52,000 picked in 170 micrographs.

Cryo-EM (cryo-electron microscopy) uses an electron beam transmitted through a biological sample to generate projection images. The projection images of a sample can be used to reconstruct the 3D structure when many views are available [184]; 2D classification is useful for investigating particle quality before moving into the 3D reconstruction stage, and to qualitatively explore the distribution of views within the dataset.

The Cryo-EM data were acquired by Dr Christos Savva of the Midlands Regional Cryo-Electron

Microscope Facility at the University of Leicester, England, UK - as described in section 2.7. Figure 3.22 shows some of the 2D classes for 4,447 particles selected amongst the initial 52,000, coming from 170 micrographs. The 18 Å auto-refined 3D map shown in Figure 3.23 was obtained by Mr. Charlie Hitchman following the data processing protocol described in section 2.7. The homology models of the catalytic domain of *Ct*HTM1P and of *Ct*PDI1P were placed manually to fit the map using Chimera [154].

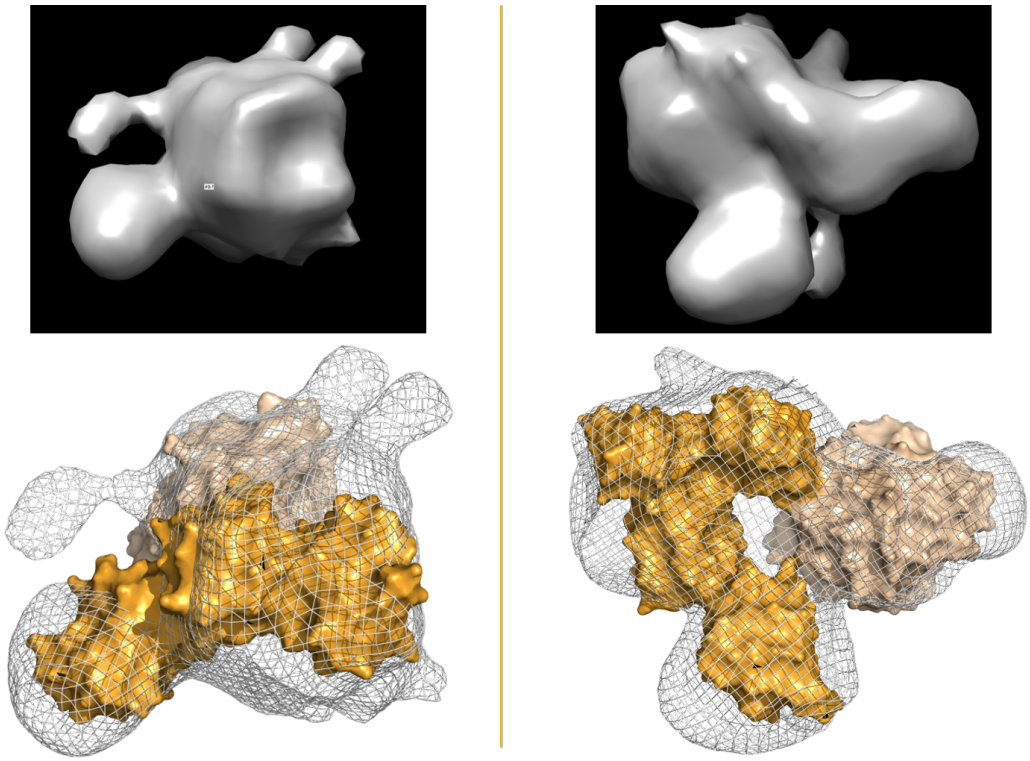


Figure 3.23: The two insets above show the 18 Å 3D model after autorefine (thanks to Dr Pietro Roversi, Dr Christos Savva and Mr. Charlie Hitchman). Below, homology models of the catalytic domain of *Ct*HTM1P (wheat-coloured surface) and *Ct*PDI1P (dark-orange-coloured surface) manually docked into the 18 Å Cryo-EM map. The Cryo-EM data collection and processing are described in section 2.7 and the homology modelling is described in section 2.3.2.

Chapter 4

Discussion

4.1 The plant as a model to study ERQC

High sequence and functional conservation of the two main components of the ERQC, ER α–Glu II, and UGGT, enables the use of plant ER α–Glu II and UGGT mutants for *in vivo* ERQC studies. For example, in a piece of work published prior to the work described in this thesis [45], we were able to demonstrate the activity of the UGGT from the thermophilic filamentous fungus *C. thermophilum* by recomplementing its gene into the *A. thaliana* with genetic background *ebs1-3/bri1-9*.

Iminosugars, currently the best known broad-spectrum antivirals, show cross-reactivity with intestinal glucosidases and enzymes of the glycosphingolipid pathway [93], creating a need to expand the search for new selective ERQC modulators. The development of low-cost platforms for the screening of molecules with potential pharmaceutical effects before testing them in animal models is only just emerging [185]. In this context, plants have been already proposed for preliminary screening of anti-cancer compounds [186].

In the work described in this thesis, we show that the iminosugar *N*B-DNJ, which reached phase I clinical trials as an antiviral drug, can inhibit *A. thaliana* ER α–Glu II. Our results confirm the potential of *A. thaliana* as a low-cost, easy-to-use organism to study ERQC function. In this context, leucine-rich repeat receptor kinases (LRR_RKs) are very useful ERQC client glycoproteins. Even though EFR, FLS2 and the brassinosteroid receptor BRI1 belong to the same sub-family (LRR-RK, sub-family XII), EFR-mediated signalling depends on ERQC, while the BRI1- and FLS2-mediated responses do not [58], enabling good control experiments. For example, in Section 3.1.1, we show that the treatment of plants with the known ER α-glucosidase iminosugar inhibitor *N*B-DNJ influences the immune response to the bacterial EF-Tu peptide elf18 but not to the bacterial flagellin-derived flg22 peptide. A possible explanation of this different behaviour is that EFR (or one of its downstream effectors) evolved only recently and depends on the ERQC machinery more than FLS2 and BRI1 [58]. Alternatively, macroscopic phenotypes - such as loss of plant weight or decreased plant size - are not subtle enough to probe the complex roles that ERQC components play in the folding of the many glycoproteins likely involved in plant responses to elicitors. Differences in protein structure, and/or numbers of *N*-linked glycans and/or the relative positions of the *N*-

glycosylation site(s) with respect to the sites of misfold can all affect the degree to which either the EFR/FLS2/BRI1 receptors, or other glycoproteins involved in their signalling cascades, depend on Glu II and UGGT.

A second set of results, described in Section 3.1.3, confirm those from Lu et al. [159], who showed for the first time that EFR-dependent ligand binding activity was impaired in *psl5-1* and *rsw3 A. thaliana* mutants. While a T-DNA insertion mutant was reported for the β-subunit of ER α–Glu II [159], no similar mutant is available for the α-subunit: homozygous insertional mutants disrupting the fold of the α catalytic subunit of ER α–Glu II cause embryonic death [187], pointing to a likely essential role of the enzyme for seedling development.

Interestingly, treatment of the ER α–Glu II *rsw3* mutant with *N*B-DNJ at concentrations higher than 500 nM (see Figure 3.3) was lethal to *A. thaliana* embryos, likely because of complete inhibition of *At*Glu II, and/or extra toxic effects due to *N*B-DNJ glycolipid metabolism inhibition [157]. On the other hand, the treatment of a *psl4-2* mutant plant with a 100-fold higher concentration of *N*B-DNJ showed a much less severe phenotype (Figure 3.4). It must be underlined that iminosugars like *N*B-DNJ are analogues of glucose and they competitively inhibit glucosidases in general and therefore *N*B-DNJ is not a specific inhibitor for ER α-Glu II. Indeed, it was observed that in the ER *N*B-DNJ can inhibit both ER α-Glu II and α-Glu I, so the extra toxicity observed with *N*B-DNJ in the *rsw3* background could be explained by inhibition of ER α–Glu II and ER α–Glu I by *N*B-DNJ. However, concentrations of *N*B-DNJ higher than 100 μM are required to achieve Glu I inhibition *in cellula* [93]. To check for Glu I inhibition, tri-glucosylated glycans on glycoproteins would need quantifying with and without *N*B-DNJ treatment in the *rsw3* mutant.

Taken together, our results confirm that ER α–Glu II is essential for seedling development and EFR signalling, and that *At*Glu II *rsw3* may retain some residual activity. Moreover, our data suggest that the catalytic domain of ER α–Glu II in absence of the β subunit (and the latter protein's ER retrieval signal) may also ba able to function in the ER. The enzymatic activity might be carried out by the α subunit while transiting in the ER lumen, before leaving it due to the lack of association with the β subunit (as illustrated in Figure 4.1). This model is consistent with observations reported in [95]: the single mutant α-Glu II$^{\mathrm{R840E}}$ abrogates the α/β interaction *in vitro* and corresponds to the plant *At*α-Glu II$^{\mathrm{R787E}}$ mutant showing reduced ER localization *in planta*. Such mutant displays an enzymatic activity in digesting the substrate 4-methylumbelliferyl α-D-glucopyranoside (4-MUG) was estimated to be around 10% compared to WT, *in vitro*. The work described in this thesis did not establish if the mutant *psl4-2* has a similar degree of residual activity. If so, the 10% of *At*α-Glu II$^{\mathrm{R787E}}$ activity would correspond to what we call "activity in transit". The main difference between the two sets of experiments is constituted by the fact that the results in this thesis (Section 3.1.3) come from *in vivo* models where the GluII α subunit is continuously translated into the ER lumen and that the mutation in *At*α-Glu II$^{\mathrm{R787E}}$ (even more the ones in *At*α-Glu II$^{\mathrm{R784E\text{-}R787E}}$), impair the interaction of the α subunit with the β subunit while the mutant *psl4-2* produces a native α subunit that lacks the β subunit completely. In any case, the results in [95] describe the secretion of the α subunit from the ER when the interaction

with its β subunit - an observation compatible with our "activity in transit" model.

As an alternative to such model, it can be hypothesised that the residual α subunit activity observed be granted by its association with another protein-partner that retains it in the ER lumen. The latter hypothesis is compatible with the recent observation that both ER α-Glucosidase II subunits associate with TXNDC11, the protein disulfide isomerase disulfide-bound to the EDEM2 ERAD checkpoint [76].

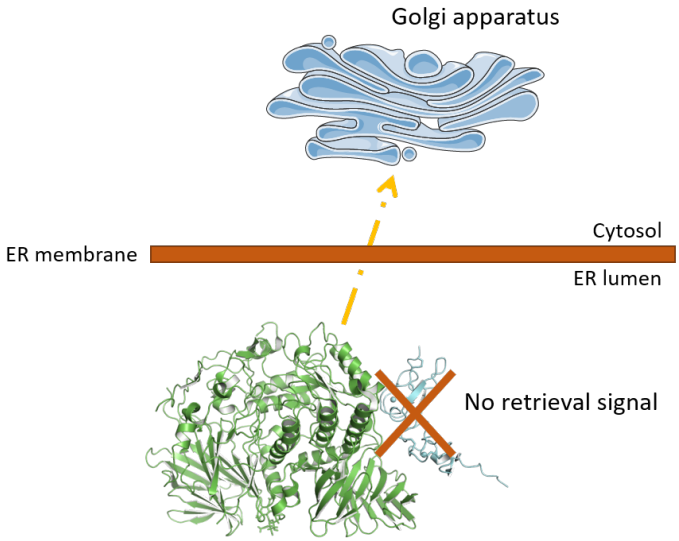


Figure 4.1: Our model of ER α–Glu II activity "in transit". In the mutant *psl4-2*, ER GluII enzymatic activity might be due to the α subunit transiently present in the ER before the lack of the ER retrieval signal granted by the β subunit causes the α subunit to leave for the Golgi. The ER α–Glu II α subunit is in green, while the β subunit is in cyan.

In the *At uggt1-1* mutant, EFR signalling as measured by plant size is not significantly impaired, as reported in previous works for the *psl2-1*, *uggt-3*, and *uggt-4* UGGT mutants [59]. Yet, under our experimental conditions, both elicitors triggered a significant increase in the expression levels of both *PHI-1* and *RET-OX*, and only the concomitant application of *N*B-DNJ blocked the EFR signalling pathway completely, see Figure 3.8. These results indicate that EFR-mediated *PHI-1* and *RET-OX* upregulation depends on glycoproteins that can fold successfully thanks to mono-glucosylated glycans afforded by ER α–Glu II alone; in other words, UGGT activity is not needed for some of the EFR response. We note that both the WT and the *uggt1-1* mutant survive 70 μM *N*B-DNJ treatment, which suggests that this dose of iminosugar only affords partial inhibition of *At*α–Glu II. Indeed, the observation that insertional mutants have been reported for *At*UGGT but not for ER α–Glu II indicate that UGGT activity modulation might have a less drastic impact on the glycoprotein folding machinery and ER homeostasis than ER α–Glu II deletion, boding well for future studies of UGGT modulation for the rescue of the secretion of misfolded but functional glycoprotein mutants in congenital protein folding diseases, or for antiviral and anti-cancer

strategies [47].

In a paper published prior to the work described in this thesis [45], we were able to recomplement an *A. thaliana* insertional UGGT mutant in the genetic background *ebs1-3/bri1-9* using a functional UGGT from the thermophilic filamentous fungus *C. thermophilum*. That paper demonstrated that *Ct*UGGT was active in the *ebs1-3/bri1-9* plant, so much so that *Ct*UGGT retained in the ER the mutated *BRI1-9* receptor, causing the plant to revert to the *bri1-9* dwarf phenotype. Similarly, in the work described in this thesis (see Section 3.2.3) we show that the enzyme *Ct*HTM1P, which is the *Chaetomium thermophilum* orthologue of the EDEMs ERAD checkpoint enzymes, does restore ERAD activity to the triple mutant plant *mns4-1/mns5-1/bri1-5*. The *mns4-1/mns5-1/bri1-5* plant shows a WT-like phenotype because mutations of the MNS4 and MNS5 ERAD mannosidases abrogate ERAD degradation of the mutant receptor *BRI1-5*: instead of being degraded, *BRI1-5* manages to escape the ER and reach the plasma membrane - where it signals and enables plant growth. Our recomplementation of the *mns4-1/mns5-1/bri1-5* triple mutant with *Ct*HTM1P causes the plant to revert to the dwarf phenotype similar to the *bri1-5* plant described in [126]. This result suggests that *Ct*HTM1P can dispatch to ERAD the mutant receptor *BRI1-5*, thus restoring functional ERAD in absence of the plant ERAD mannosidases. This experiment establishes the plant as a good model system for studying the fungal ERAD mannosidase(s) and inhibition/modulation of ERAD using the triple mutant and recomplemented planta as control experiments.

4.2 Structural and functional characterisation of the ERAD checkpoint heterodimer

No published work has so far reported recombinant protein production of an ERAD checkpoint heterodimer in quantities large enough to attempt *in vitro* assays and structural determination. What we do know is that the heterodimer *Sc*HTM1P:*Sc*PDI1P is held together by hydrophobic interactions and it is further stabilised by an intermolecular disulfide bridge [83]. For our recombinant protein production effort, and the functional and structural studies that will use recombinant protein, we selected from the genome of *C. thermophilum* the most similar genes to *Sc*HTM1P and *Sc*PDI1P, in order to profit from the structural stability of proteins from this thermophilic eukaryote.

The mannosidases *Sc*HTM1P and *Ct*HTM1P do not show sequence conservation in the C-terminal region where *Sc*HTM1P has the cysteine residues suspected to be involved in the disulfide bridge with its PDI partner, and neither of the hydrophobic residues known to be involved in the *Sc*HTM1P:*Sc*PDI1P interface are conserved in *Ct*HTM1P. Setting out to attempt recombinant protein production of the *Ct*HTM1P:*Ct*PDI1P ERAD checkpoint heterodimer, we hypothesised that the interaction *Ct*HTM1P:*Ct*PDI1P may still be mediated by different hydrophobic interactions and/or by a disulfide bridge between Cys residues that differ from the *S. cerevisiae* ones.

The results described in Sections 3.2.4 and 3.2.5 suggest that - when HEK293F mammalian cells

are transfected with a plasmid encoding *Ct*HTM1P for secreted expression alone - *Ct*HTM1P is not secreted and appears to aggregate via intermolecular disulfide bridges (see Figure 3.13). Such aggregates are completely disrupted by the addition of 500 μM of TCEP reducing agent to the protein sample (see Figure 3.14). Interestingly, when *Ct*HTM1P and *Ct*PDI1P are co-expressed, we find both proteins in the culture supernatant, and the *Ct*HTM1P expression level is much higher.

Importantly, we also confirmed that *Ct*HTM1P and *Ct*PDI1P also form a heterodimer (see the early fractions of the size-exclusion elution profile in Figure 3.2.7.3), supporting the hypothesis that these proteins are the orthologues of the *S. cerevisiae Sc*HTM1P:*Sc*PDI1P and of the human EDEMs:TXNDC11 [83, 74].

Taken together, our observations are compatible with a model in which - in absence of excess PDI - the overexpressed *Ct*HTM1P may expose a free Cys (normally buried in the mannosidase:PDI interface) which in turns may lead to intermolecular disulfides, aggregation, and lack of secretion. Alternatively, it is possible that some of *Ct*HTM1P , when expressed alone, be kept in the ER lumen by the interaction with the human orthologue of the PDI1P (endogenous to the HEK293F cells) whose ER retrieval signal would prevent secretion of any associated *Ct*HTM1P .

On the structural front, the crystals we obtained from the *Ct*HTM1P$_{\mathrm{D}}$:*Ct*PDI1P crystallisation screen (see Figure 3.19A-D) are not suitable for X-ray diffraction and further optimisation of the crystallisation condition will be needed. On the other hand, the *Ct*PDI1P crystallisation trial produced a good crystal (see Figure 3.19E) that will be analysed by X-ray diffraction in the near future.

The first Cryo-EM micrographs obtained from *Ct*HTM1P$_{\mathrm{D}}$:*Ct*PDI1P grids showed protein oligomers/aggregates, likely formed in the last step of concentration. The protein gels and grids also reveal the presence of a lighter *Ct*HTM1P$_{\mathrm{D}}$ species, probably due to proteolysis and loss of GFP. Further purification steps will likely be needed to have a sample suitable for the determination of the structure of the heterodimer.

Despite the preliminary nature of the result and the low resolution of the map so far (18 Å), the auto-refined 3D map showed in Figure 3.23 has about the right volume to fit our homology models of the *Ct*HTM1P$_{\mathrm{D}}$ mannosidase domain and *Ct*PDI1P . Some of the 2D classes show a structure similar to an "apple petiole", for example see the first 2D class in the top-left corner in Figure 3.22: we hypothesise that the small protruding part could be the N-terminal *Ct*HTM1P$_{\mathrm{D}}$ GFP tag. Since the GFP-tag is flexible and its varying orientation causes heterogeneity in the micrographs and adds noise that affects the 2D class averaging, an optimisation of the sample for the Cryo-EM grids can be also achieved by removing the GFP-tag with the 3C protease that cuts at the linker between the GFP-tag and *Ct*HTM1P$_{\mathrm{D}}$.

Amongst the four different construct of *Ct*HTM1P , described in Figure 3.12, co-purification with *Ct*PDI1P was attempted only for *Ct*HTM1P$_{\mathrm{D}}$. We did not have time to check the co-expression of construct *Ct*HTM1P $_{\mathrm{A}}$ with *Ct*PDI1P due to the time constraints imposed by the COVID-19 pandemic. Construct *Ct*HTM1P $_{\mathrm{B}}$ did not express (Figure 3.13), likely because its fold-

ing is further complicated by the presence of the cytoplasmic N-terminal GFP fusion. Construct *Ct*HTM1P$_{\mathrm{C}}$ did express and it was detected on a Western blot using an Anti-GFP antibody (Figure 3.13) but we preferred to work with *Ct*HTM1P$_{\mathrm{D}}$, initially, because the latter construct does not carry the N-terminal trans-membrane domain and thus it does not require additional solubilisation and membrane-extraction steps. We plan to purify construct *Ct*HTM1P$_{\mathrm{C}}$ using Styrene Maleic Anhydride Lipid Particles LipoDisq® technology [188].

Chapter 5

Conclusions

The work described in this thesis supports the idea that *Arabidopsis thaliana* is a good model to study ERQC and ERAD machineries. We used the plant immune system and a set of available mutants such as *bri1-9* and *bri1-5*, which are associated to specific phenotypes, to probe the plant as a valid model for *in vivo* ERQC and ERAD activity assays. In particular, we were able to demonstrate that plants can be used to investigate the role and the activity of ERQC and ERAD key enzymes such as ER α-Glucosidase II and EDEMs, respectively.

ER Glucosidase II is comprised of two subunits. The catalytic α-subunit is necessary for the life of the cell and the organism. Homozygous deletion of the gene encoding for the α-subunit is lethal to the plant [187] and the same phenotype is apparent when the mutants *rsw3* and *psl5-1* grow in presence of the inhibitor *N*B-DNJ, see Figure 3.2. This indicates that, in *rsw3* and *psl5-1*, a mutation in the catalytic domain is not enough to abrogate completely ER α-Glucosidase II activity. When the enzyme loses the retrieval signal HDEL granted by the β-subunit, ER α-Glucosidase II is able to accomplish its activity in the ER, see Section 3.1.3. The WT plant treated with *N*B-DNJ behaves in a similar way to *rsw3* and *psl5-1* mutants: the iminosugar inhibits ER α-Glu II, confirming that the plant is a good system to study inhibitors so far tested in animals and animal cells.

In a previous study, we used the mutant plant *bri1-9* as a means to demonstrate the high degree of conservation of ER α-Glucosidase II across Eukaryotes [45]. In a similar way, here we used the *bri1-5* mutant to confirm the functional relevance of a fungal EDEM enzyme. The mutant *bri1-5* is dwarf because it carries a mutation within the sequence encoding a misfolded BRI1 receptor which is targeted to ERAD. The *msn4-1/msn5-1/bri1-5* triple mutant plant shows a phenotype similar to the WT because the absence of MSN4 and MNS5 activity prevents ERAD to act on *BRI1-5*. We show that when this *msn4-1/msn5-1/bri1-5* triple mutant plant is complemented with the *Chaetomium thermophilum* orthologue *Ct*HTM1P the plant grows dwarf again: *Ct*HTM1P is therefore active in the ER lumen in recognising and targeting *BRI1-5* to degradation. This is the first proof of functional activity of *Ct*HTM1P at room temperature *in vivo*.

In this thesis, we managed to obtain an *Arabidopsis thaliana* plant that expresses the *Chaetomium thermophilum* EDEMs orthologue (*Ct*HTM1P). Experiments to assess if such *Arabidopsis thaliana*

plant is more resistant to heat stress than a WT plant have not been performed yet, but such result would eventually pave the way for the production of plants, that are of agronomic/economic interest, resistant to the heat stress caused by the global warming.

Taken together our results indicate the plant as a good model organism to develop *in vivo* assays to shed light upon fine details of ERQC and ERAD function, and to test the modulation/inhibition effect of new molecules targeting the ERQC and/or ERAD machineries in a way that is easy, fast, cheap and free from ethical issues.

In the second strand of work described in this thesis (see Section 3.2.6) we looked in the *Ct* genome for orthologues of the ERAD checkpoint heterodimer EDEM:PDI. We selected two so far uncharacterised genes which we called *Ct*HTM1P and *Ct*PDI1P . We managed to express and purify the so far uncharacterised *Ct*HTM1P and *Ct*PDI1P proteins. We showed that not only the two proteins travel together down the size-exclusion column but the mannosidase is only secreted and expressed in good amounts only when co-expressed with the PDI. These data suggest that the proteins form a *Ct*HTM1P :*Ct*PDI1P heterodimer, supporting our initial hypothesis that they constitute the *Ct* orthologues of *Sc*HTM1 and *Sc*PDI1P. This work represents to our knowledge the first successful production of recombinant ERAD checkpoint heterodimer in quantities large enough for *in vitro* and structural studies. We also obtained the first 18 Å Cryo-Electron Microscopy (Cryo-EM) reconstruction of the *Ct*HTM1P :*Ct*PDI1P complex.

5.1 Future perspectives

ERQC and ERAD checkpoint enzymes (UGGT and EDEMs play an important role in the recognition and degradation of rare-disease associated glycoprotein responsive mutants. Indeed, retention of glycoprotein mutants causes a broad spectrum of rare diseases and although each of these life-threatening disorders affects less than 5 out of 10,000 individuals, there are approximately 2,000 known glycoprotein-related congenital rare diseases in the medical literature. The size of the pharmaceutical market for each rare disease is small, and this has discouraged investment in this area. The inhibition of the ERQC or ERAD checkpoints may lead to the rescue of secretion of responsive mutants - like is the case for the *BRI1-9* and *BRI1-5* plant receptor mutants discussed in this thesis. Furthermore, the reliance of cancer cells on secreted glycoproteins begs the question as to whether the ERQC/ERAD checkpoint enzymes could constitute potential anticancer targets (this is further discussed in Appendix B). Moreover, ERQC is also used by many viruses that use host's calnexin cycle to produce their viral glycoproteins: drugs targeting the ERQC usher - like *N*B-DNJ - are already studied as antivirals (*e.g.* anti-Dengue and Zika).

Like any strategy aimed at inhibition of basic cell housekeeping machineries, molecules developed to interfere with ERQC/ERAD checkpoints have the potential to be toxic. In particular, ERQC/ERAD checkpoint inhibition could lead to increased levels of prematurely secreted misfolded glycoproteins, a scenario we like to call "the opening of the ER Pandora's box" [47]. Encouragingly, unlike what observed for ER αGlucosidase II in [187] and in Section 3.1.3, the homozygous

UGGT$^{(-/-)}$ plant suffers under stress conditions only [171], and UGGT1$^{(-/+)}$ heterozygous mice (which possess ∼50% of the wild-type enzyme levels) show no obvious phenotype [56]. The secretome of a UGGT$^{(-/-)}$ plant can be studied as a limiting case of therapeutic UGGT1 activity modulation, and compared with the secretome of a UGGT1$^{(-/+)}$ heterozygous organism to inform on unwanted side effects of UGGT1 inhibition/modulation (respectively) at the cellular level.

Nevertheless, by loosening UGGT/EDEM misfold recognition over-zealousness, drugs modulating (rather than inhibiting) either ERQC or ERAD checkpoint enzymes could rescue secretion of a slightly defective and yet active rare-disease associated glycoprotein, allowing the mutant to carry its residual activity to its target destination. This scenario could be shared in more than one rare disease. To this purpose, the plant model offered by the mutants *BRI1-5* and *BRI1-9* described in Sections 3.1.1 and 3.1.3, can be used to test UGGT and/or EDEM inhibitor candidates, before moving to human cells or animal models.

UGGT/EDEM modulators could also prevent viruses from successfully hijacking the host calnexin cycle, and again the plant can be employed studying the plant-virus infection in the presence of such drugs.

As explored in detail in Appendix B, it is conceivable that UGGT/EDEM modulators be attractive targets for the treatment of cancer cells, particularly those bearing a high secretory burden, whose fitness is critically dependent on the functional integrity of the ER, which in turn relies on ERQC/ERAD misfold recognition as ER stress-attenuating mechanisms. For example, it was recently proven that the receptor IGF-1R - a target in cancer biology as it is important for cell growth - is substrate of UGGT1 and UGGT1/2$^{-/-}$ cells have a reduced efficiency in the accumulation of mature IGF-1R [48].

In the future, such modulation could be easily tested *in vivo* using the plant and assays similar to the ones described in this thesis, before expensive and time-consuming clinical trials.

Whenever ERQC/ERAD misfold recognition activity needs to be upregulated, for example to endow organisms with stress resistance, a mirror strategy would be to potentiate ERQC and ERAD checkpoint function. An example would be the production of a plant that is resistant to abiotic stresses, like global warming-related raised temperatures and/or environmental pollutants, which are among the factors that contribute to misfolded protein accumulation in the ER of a plant. The accumulation of misfolded protein in turns triggers the unfolded protein response (UPR), which restores glycoproteostasis. If the presence of misfolded proteins persists, the cells go towards apoptosis instead. It was shown that in *Drosophila melanogaster* the overexpression of EDEMs led to the suppression of age-associated physiological decline [115]: we speculate that the same principle can be applied to plants of agronomic interest, thanks to genetic potentiation of their ERAD machinery. Such plant would be more tolerant towards the accumulation of misfolded proteins caused by abiotic stresses. For example, one experiment would use qRT-PCR analysis to quantify and compare the expression level of genes involved in stress response (such as BiP genes) between a WT plant and a plant with a thermophilic version of EDEMs (see the plant described in Section 3.2.3) when subjected to heat/drought stress.

The overexpression and purification of *Ct*HTM1P recombinant protein paves the way for *in vitro* functional characterisation of the ERAD checkpoint. For example, its demannosylation kinetics could be studied with an ERAD substrate *in vitro*. The characterisation could involve *in vitro* experiments where *Ct*HTM1P and αGluII (and/or UGGT) act on the same *N*-linked $Glc_1Man_9GlcNAc_2$ (and/or $Man_9GlcNAc_2$) substrate. Enzyme activity could be monitored by analysing the glycans of such misfolded glycoprotein substrates - removed from the protein using enzymes like Endoglycosidase H (EndoH) and/or Glycopeptidase F (PNGaseF), then labelled and quantified through HPLC. Such experiments could reveal the mechanisms behind the interaction between ERQC and ERAD, explaining how and when a misfolded glycoprotein stops to be helped by the calnexin cycle and undertakes the pathway to its degradation. Another possible experiment consists in the expression of *Ct*HTM1P in presence of the inhibitor kifunensine: this *Ct*HTM1P would carry *N*-linked $Man_9GlcNAc_2$ glycans. The incubation *in vitro* of *Ct*HTM1P alone and in absence of kifunensine, followed by an HPLC glycan analysis, could be useful to assess the case we called "it takes one to know one": do EDEMs dispatch themselves to ERAD?

The first structure of the ERAD checkpoint heterodimer *Ct*HTM1P:*Ct*PDI1P (determined either by X-ray crystallography and/or Cryo-EM) once determined will reveal the molecular basis for their interaction and generate a range of hypotheses to be tested with point mutants *in vitro*, *in cellula* and *in vivo*. Cross-linking MS and/or Hydrogen–Deuterium eXchange (HDX) can be also be used, together with surface sequence conservation analysis, to investigate which portion of the *Ct*HTM1P and *Ct*PDI1P are involved in the formation of the heterodimer, and in particular to understand which hydrophobic and Cys residues are involved in the interaction. Availability of crystals of the heterodimer will enable structure-based lead discovery of molecules that are inhibitors/modulators of specific ERAD checkpoints.

All in all, the experiments described in Chapter 3.2 led us very close to this goal. Further optimisations of our Cryo-EM grids and data will hopefully afford a structure, before complexation of the heterodimer with a substrate glycoprotein is attempted, ultimately revealing the molecular basis for misfolded glycoprotein recognition by the ERAD checkpoint.

Appendix A

Partial catalytic Cys oxidation of human GAPDH to Cys-sulfonic acid

A.1 Introduction

Mammalian HEK293F cells are routinely used in conjunction with secreted protein expression vectors for recombinant protein production [189]. They owe their popularity to ease of handling, robust growth rate, excellent transfectability, high capacity for recombinant protein expression, low-cost media requirements and low levels of secreted contaminants [140]. Of course, in absence of secretion of the desired recombinant protein at hand, contaminants are the only proteins present in recombinant protein expression systems [190, 191]. In the course of this thesis' research effort aimed at the purification of recombinant *Chaetomium thermophilum* ERAD mannosidase *Ct*HTM1P from the supernatant of HEK293F cells, I purified, crystallised and determined crystal structures of human n-Glyceraldehyde-3-phosphate dehydrogenase, partially sulfonated at its catalytic Cys residue (*Hs*GAPDH, EC 1.2.1.12) [192]).

GAPDH is essential for glycolysis and gluconeogenesis; it catalyses the NAD+-dependent oxidative phosphorylation of n-glyceraldehyde-3-phosphate to 1,3-diphospho-n-glycerate (and its reverse reaction) [193]. The GAPDH-catalysed forward reaction occurs at an important transition point in glycolysis between the enzymatic steps that consume and generate ATP. Several studies indicate that the enzyme has pleiotropic functions independent of its canonical role in glycolysis [194, 195, 196]. In addition to the somatic cells *Hs*GAPDH isoform, the human genome encodes a testis-specific isoform *Hs*GAPDHS (G3PT_HUMAN (Uniprot O14556)), which is expressed only in the post-meiotic period of spermatogenesis. In multiple mammalian species the sperm isozyme (GAPDHS) shares about 70% amino acid identity with the somatic isozyme, and possesses an additional 72-residues N-terminal extension. The latter enables association of GAPDHS to the sperm flagellum fibrous sheath, so that the enzyme – together with other glycolytic enzymes - provides a localised source of ATP that is essential for sperm motility [197].

GAPDH was one of the first enzymes to be crystallised [198, 199] and one of the first enzymes

whose structure was determined by X-ray crystallography [200, 201]. A number of studies have reported the GAPDH catalytic Cys carrying oxidised post-translational modifications. In particular, cysteine sulfenic acid (CSX, see A in Figure A.1) has been observed in the structure of rabbit muscle GAPDH [202] (of course, by X-ray diffraction alone, this modification may be difficult to distinguish from S-hydroxy-cysteine (CSO, see C in Figure A.1). Cysteine-S-sulfinic acid (CSD, see B in Figure A.1) has been observed in structures of rat GAPDHS and *E. coli* GAPDH [203]. Cysteine-S-sulfonic acid (CSU, see D in Figure A.1) modifies the catalytic Cys in a structure of *Streptococcus pneumoniae* GAPDH [204].

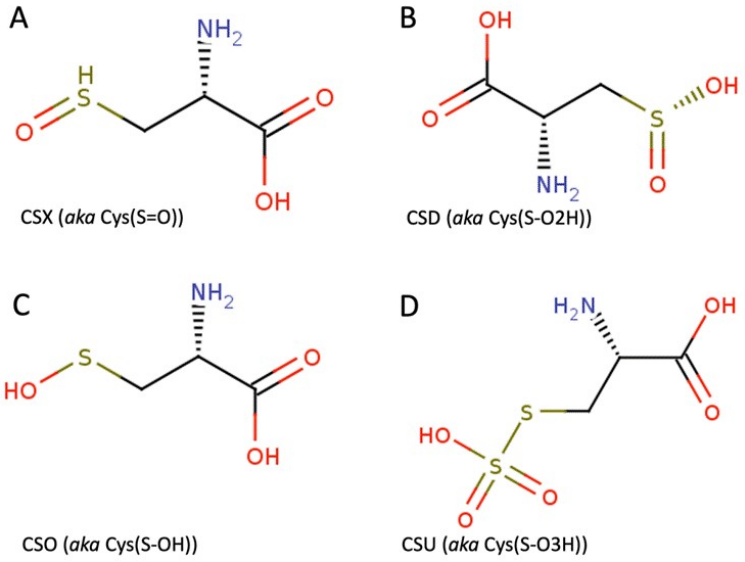


Figure A.1: Cys oxidised modifications and GAPDH PDB entries that carry them at the catalytic Cys152. **A**: S-Oxy-Cysteine *aka* Cysteine sulfenic acid (CSX, PDB ID 1J0X); **B**: S-Cysteinesulfinic acid *aka* S-sulfinocysteine aka 3-Sulfino-L-alanine (CSD, PDB IDs 2VYN and 2VYV); **C**: S-hydroxy-cysteine (CSO, not described in any GAPDH structures); D: Cysteine-S-sulfonic acid (CSU, PDB ID 5M6D and our work: 6YND, 6YNE, 6YNF, 6YNH)

A.2 Experimental Procedures and Results

A.2.1 Protein expression

450 mL of Human epithelial kidney (HEK293F) cells (ThermoFisher Scientific) at a concentration of 10^6 cells/mL suspended in GIBCO FreeStyle 293 Media (ThermoFisher Scientific 12338018), supplemented with 5 μM kifunensine (Cayman Chemical Company 109944-15-2), were transfected with the *Ct*HTM1P $_{50-1092}$-pHLsec plasmid as described in Section 2.4.

A.2.2 Purification of Cys-sulfonated *Hs*GAPDH

The HEK293F cells' supernatant was harvested 4 days post-transfection by centrifuging at 4,000 x g for 5 minutes, it was made 1x PBS by adding an appropriate 10xPBS stock solution volume and the pH was adjusted to 7.4 by adding a few drops of 2 M NaOH, before bottle-top vacuum filtration through a 0.2 μm filter

A.2.2.1 IMAC affinity chromatography

The filtered supernatant was then passed onto a 1 mL HisTrap HP Ni IMAC column equilibrated against binding buffer IMACBW (see 2.3.1) using a peristaltic pump at room temperature, at a flow rate of approximately 3 mL/min. The column was washed with 5 column volumes (cV) buffer IMACBW, fitted to an ÄKTA purifier FPLC machine in a 4 °C cabinet and further washed with 6 mL of elution buffer IMACE (see 2.3.1) supplemented with 1 mM TCEP. Figure A.2 shows the elution of proteins that didn't bind to the Ni in the HisTrap column. The chromatogram shows a λ=280 nm early peak due to proteins passing through the columns as washing buffer IMACBW flows. No peak corresponding to proteins elution is seen in the elution step, flowing an increasing gradient of buffer IMACE as indicated in Section 2.4.3.

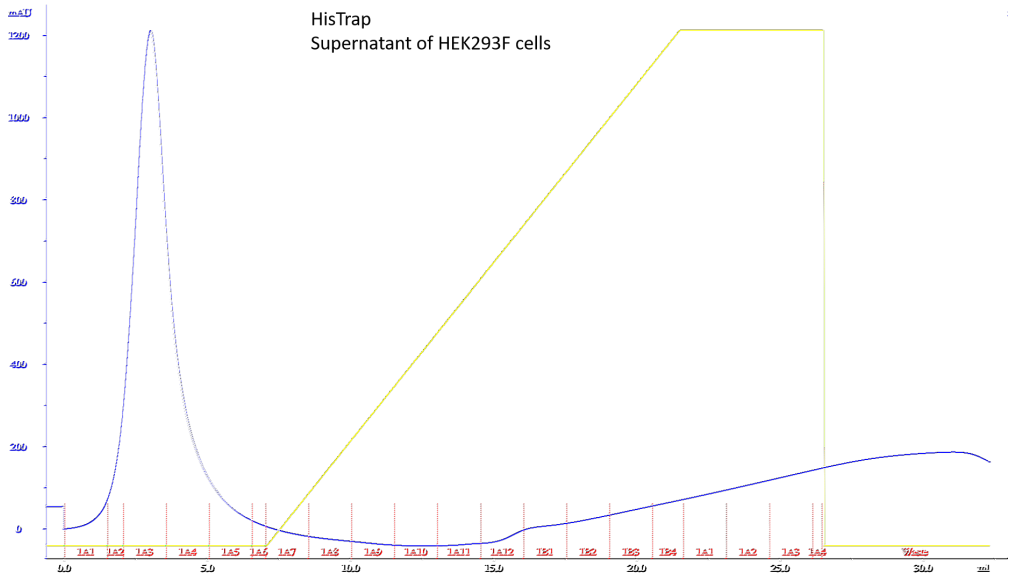


Figure A.2: Elution profile of the IMAC purification of the first batch of *Ct*HTM1P $_{50-1092}$. **Blue line**: absorption at λ=280 nm. **Yellow line**: gradient of elution buffer IMACE. Eluted protein is visible as a peak in fractions from 0 to 5 mL. The elution buffer absorbs at λ=280 nm, likely because of metal impurities forming a ferrocene complex with the imidazole.

The construct *Ct*HTM1P $_{50-1092}$-pHLsec used for the HEK293F transfection was built to express the protein *Ct*HTM1P $_{50-1092}$ carrying a C-terminal 6×His-Tag. The early elution peak in Figure A.2 indicated the expression of contaminants rather than the expression of *Ct*HTM1P $_{50-1092}$

6×His-Tagged protein.

A.2.2.2 Size-exclusion chromatography

To further investigate about the nature of the proteins eluted from the IMAC affinity chromatography described above the three wash 1.5 mL fractions (A2-A3-A5) were pooled together and concentrated down to a volume of 2 mL using a polyethersulfone (PES) membrane, 10 kDa MM cutoff centrifugal ultrafiltration device (Thermo Fisher Scientific 88513). This sample was then filtered through a 0.2 µm filter and applied to a HiLoad Superdex 200 16/60 size exclusion chromatography (SEC) column (GE Healthcare 28-9893-35) equilibrated against SEC buffer (see the recipe in 2.3.1).

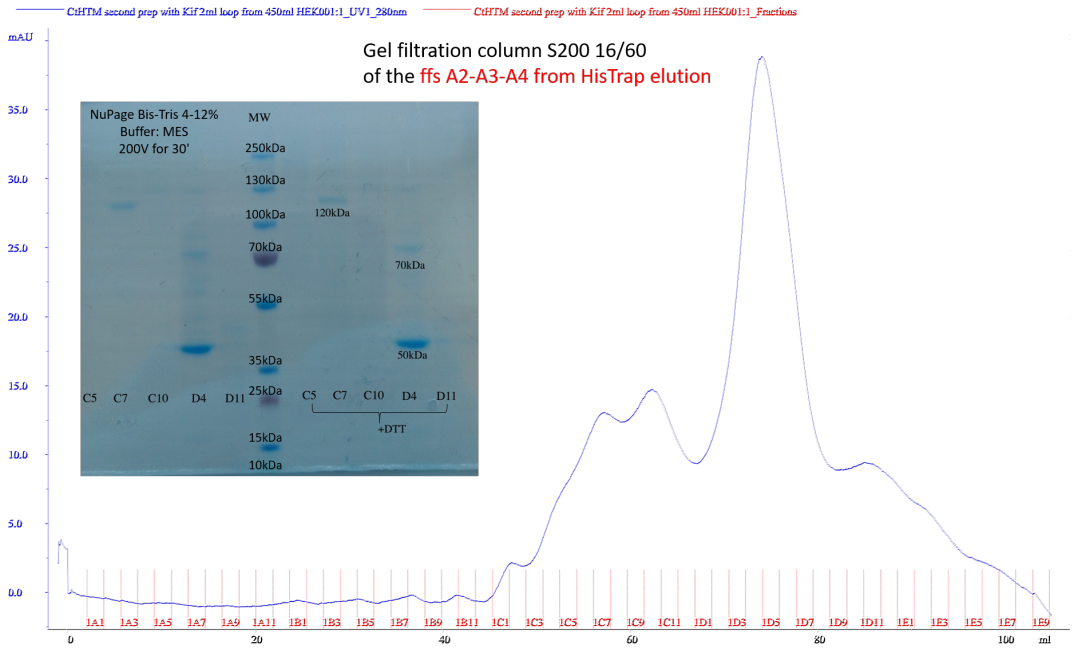


Figure A.3: Size exclusion chromatography of the contaminants from IMAC. The elution profile of the S200 16/60 column on which 2 mL of the concentrated IMAC wash fractions were run. Inset: the SDS-PAGE Bis-Tris 4-12% gel run in MES buffer for 30' at 200 V, stained in Simply Blue™ SafeStain (Thermo Fisher) for 1 h and destained with water. Lanes 1-5: non reducing dye. Lanes 7-11: reducing dye.

A.2.2.3 Size exclusion Chromatography with Multi Angle Laser Scattering (SEC-MALS)

A volume of 30 μL of the concentrated SEC sample was diluted to 500 μL with filter-sterilised and degassed buffer HEPES 20 mM pH 7.5, NaCl 100 mM, TCEP 1mM and injected onto a Superdex S200 10/300 column equilibrated in the same buffer. The mass of the sample was detected on elution with an 18-angle MALS light scattering detector (DawnR HELEOS-R II©) coupled

with a differential Refractive Index detector (OptilabR© T-rEX) (Wyatt Technology). Figure A.4 illustrates the results of the run. Fractions C3-C4-C5 (the former two corresponding to a tetramer of apparent mass 144 kDa) were pooled and concentrated to a volume of 45μL - the absorbance was $OD_2$80=2.46 equivalent to a concentration of 3.0 mg/mL (calculated from ϵ_{280}=0.7963 $(mg/mL)^{-1}cm^{-1}$, as computed at `https://www.scripps.edu/cgi-bin/cdputnam/protcalc3`).

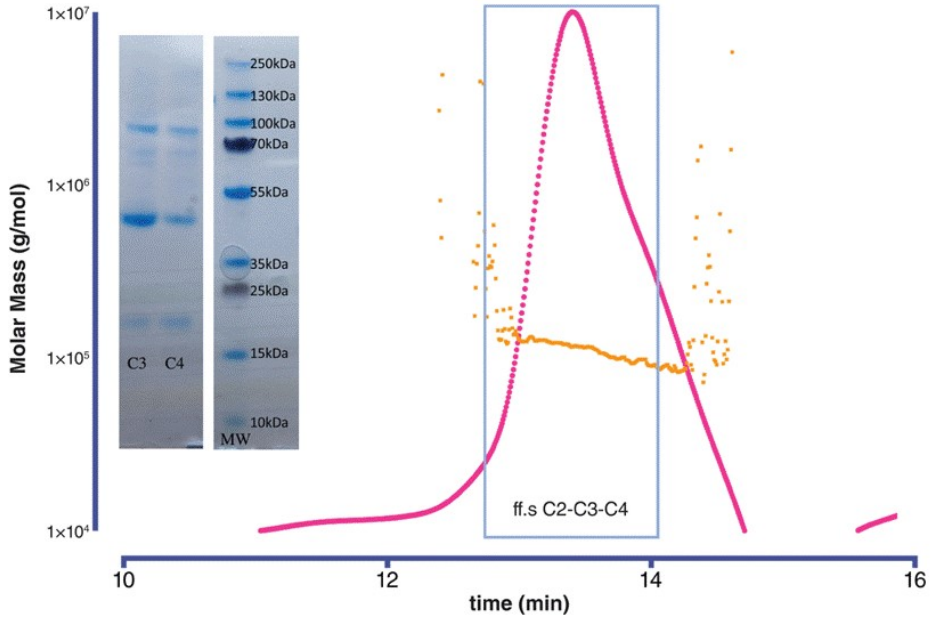


Figure A.4: Size exclusion chromatography multiangle light scattering (SEC-MALS) elution profile of HsGAPDH. Elution profile of the HsGAPDH sample run on the SEC S200 10/300 column interfaced with the 18-angle MALS light scattering detector (Dawn HELEOS II ®) coupled with a differential refractive index detector (Optilab ® T-rEX) (Wyatt Technology). The fractions C3-C4 eluting between 13 and 14 mL (boxed) correspond to a HsGAPDH tetramer of apparent mass 144 KDa. Inset: the SDS-PAGE NuPAGE Bis-Tris 4-12% gel (Thermo Fisher NP0322PK2), run at 200 V for 30' in MES buffer, stained in Simply Blue™SafeStain (Thermo Fisher LC6060) for 1 h and destained with water.

A.2.3 Mass spectrometry protein and Cys-modification identification

The protein bands were excised from the SDS-page gel and sent for mass spectrometry analysis to "FAO Metabolomics & Proteomics Lab", University of York (UK). Protein was digested in-gel with the addition of 20 ng sequencing grade trypsin (Promega V5111). No reduction or alkylation was performed during digestion to help preserve the native state of oxidation. Resulting peptides were analysed over a 1 h liquid chromatography–mass spectrometry acquisition with elution from a 50 cm EasyNano C18 column as detailed in [205] onto a Thermo Orbitrap Fusion Tribrid mass spectrometer. Spectra were acquired in data dependent acquisition mode with sequential fragmentation of all selected peptide precursors using both higher-energy collisional dissociation and

electrontransfer dissociation (ETD). All spectra were acquired in the Orbitrap mass analyser with internal calibration from the ETD reagent. Product ion spectra were searched against the expected sequence of *Hs*GAPDH (Uniprot P04406) appended to a custom in-house database. All pieces of software were run with default parameters, with the exception that propionamide (+71.037114) and dehydro (-1.007825 Da) were added as manually defined variable modifications. The analysis was carried out using Mascot version 2.6.1 (Matrix Science) [147], PEAKS Studio X+,10.5 (Bioinformatics Solutions Inc.) [206] and Byonic™ 3.0 (Protein Metrics Inc) [207] search engines. The same data analysis was also carried out with freeware, using a combination of MSConvert version 3.0 [208], SearchGui version 3.3.18 [209] and PeptideShaker version 1.16.45 [210]. The resulting peptide assignments are equivalent to the ones obtained with the commercial pieces of software, see Figures A.5 and A.6.

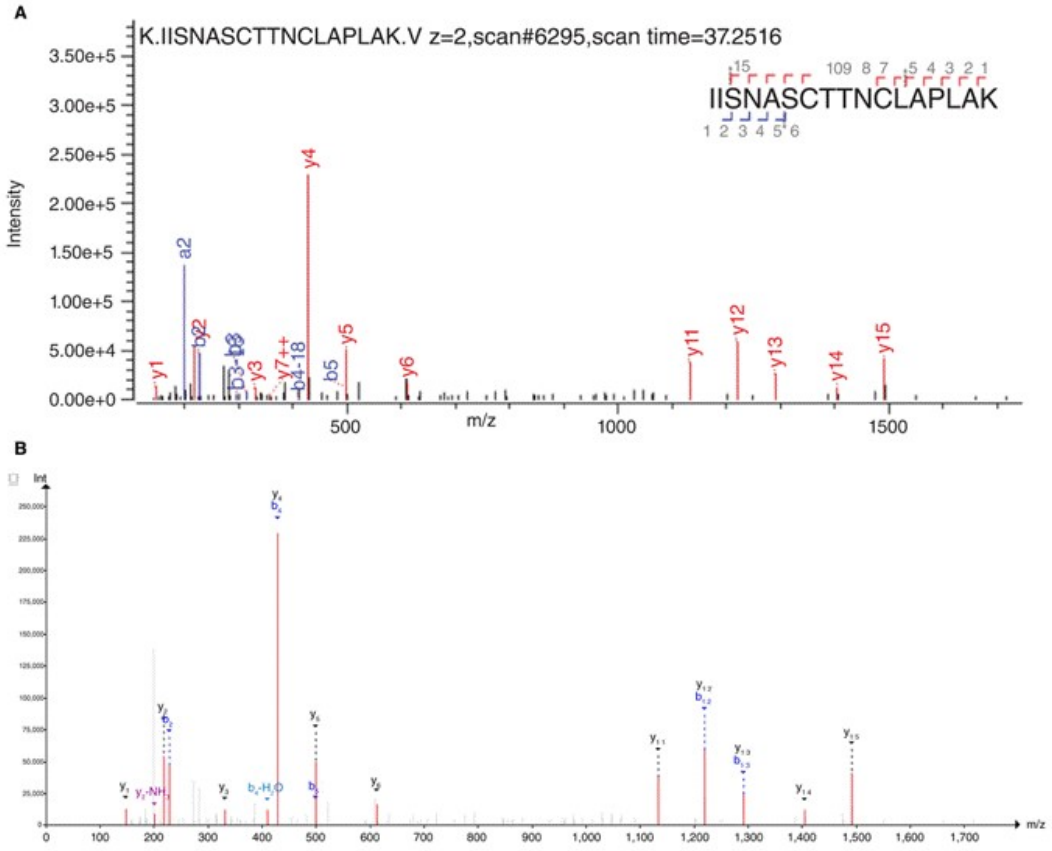


Figure A.5: Product ion spectrum of GAPDH-derived tryptic peptide ^{146}IISNASCTTNCLAPLAK^{162} lacking Cys modification. Product ions are annotated with theoretical b- and y- ions (fragment ions that appear to extend from the amino- or carboxy-terminus of a peptide, respectively) from the peptide assignment IISNASCTTNCLAPLAK. **A**: data analysed using Mascot [147], PEAKS [206] and Byonic™[207] search engines, with all default parameters. **B**: data analysed using MSConvert [208], SearchGui [209] and PeptideShaker [210].

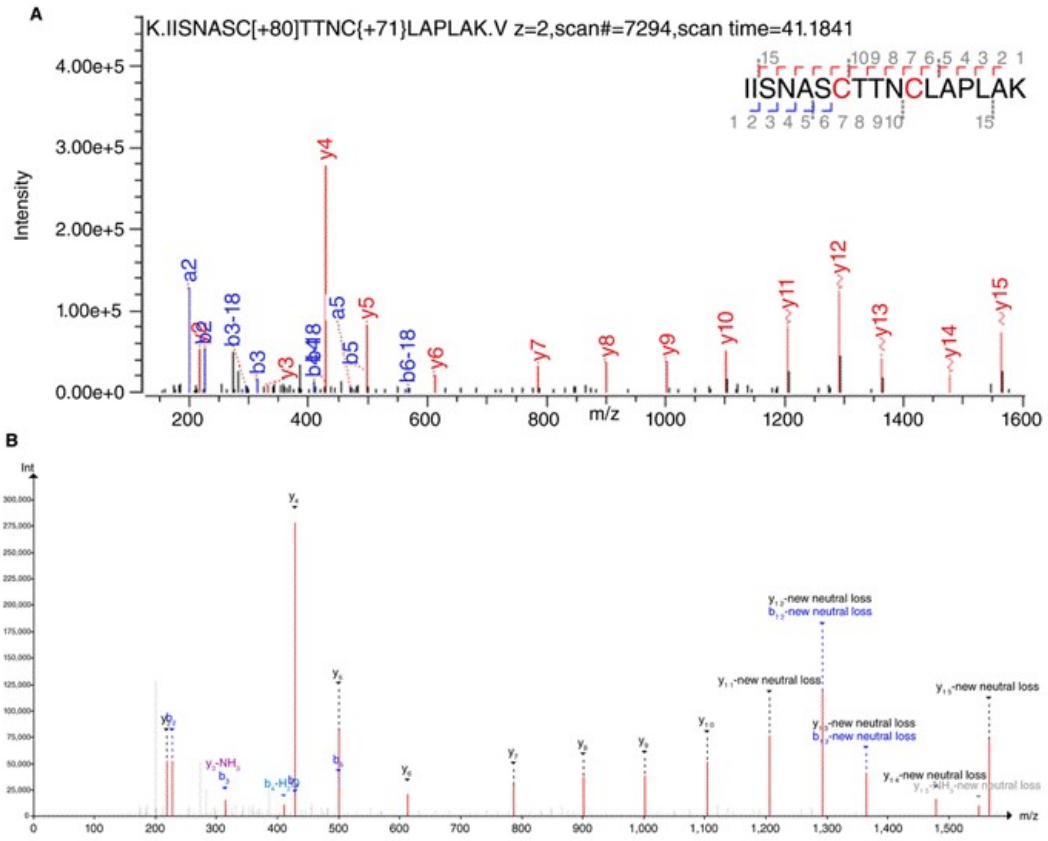


Figure A.6: Product ion spectrum of GAPDH-derived tryptic peptide ^{146}IISNASCTTNCLAPLAK^{162}. A precursor equating to a mass of 1869.8623 Da was selected and fragmented by HCD. Resulting ions are annotated with theoretical b- and y- ions from the peptide assignment IISNASC(+79.9568 Da)TTNC(+71.036 Da)LAPLAK using the Byonic™search engine. Annotation of acrylamide adduct modification (+71.036 Da) at Cys156 is implicitly localised from the m/z difference between the product ions y6 and y7. Cys-S-sulfonic acid modification at Cys152 is implied from the difference in mass between the precursor selected and the resulting y11-15 m/z values observed as neutral mass products labelled with "∼". The measured neutral mass difference is 79.9568 Da, equating to a mass error of 0.018 mDa for S-sulfonic acid modification. **A**: data analysed using Mascot [147], PEAKS [206] and Byonic™[207] search engines, with all default parameters. **B**: data analysed using MSConvert [208], SearchGui [209] and PeptideShaker [210].

Searching specified a precursor tolerance of 3 ppm and a fragment ion tolerance of 0.02 Da. Initial searches included variable modification of: S-Oxy Cysteine (A in Figure A.1), Cysteine S-sulphinic acid (B in Figure A.1), and Cysteine S-sulphonic acid (D in Figure A.1). Searches were then expanded in PEAKS [206] to include 313 of the most frequently observed proteomic modifications and in Byonic™ [207] to include disulfide bonding and wildcard mass addition to Cys. Peptide spectra matches were filtered to p-values (or equivalents) of <0.05. Individual spectra were

manually inspected for site localisation or ambiguity before reporting. Global protein coverage of over 90% was achieved. Individual post-translational modification (PTM) modified spectra were manually inspected for site localisation specificity and potential PTM ambiguity before reporting.

In figure A.3 *Hs*GAPDH is visible as band between the 35 and 55 kDa MM markers in lanes for fraction D4 (lanes 4 and 10)

A.2.4 Cys-sulfonated *Hs*GAPDH structure determination

A.2.4.1 Protein crystallisation

Vapour diffusion crystallisation 200 nL sitting drops were set up by mixing *Hs*GAPDH in mother liquor:protein ratios 1:1 (drop 1) and 1:2 (drop 2) with the 96-conditions MORPHEUS crystallisation screen (Molecular Dimensions, [145, 146]), using a Mosquito crystallisation robot (SPT Labtech) and the drops were left equilibrating at 18 °C. **$P2_1$ Forms A and B**: 40 μL of the the protein from the S200 SEC run were used to set up for crystallisation at a concentration of 18.8 mg/mL. **$P2_1$ Forms A**: these crystals grew in drop 2 equilibrated against condition H9 of the MORPHEUS crystallisation screen (Molecular Dimensions, [145, 146]): 0.1 M Amino acids solution (DL-Glutamic acid, DL-Alanine, Glycine, DL-Lysine, DL-Serine)Serine); 0.1 M Buffer System 3 (Tris (base), bicine pH 8.5); 30% v/v Precipitant Mix 1 (40% v/v PEG 500MME, 20% w/v PEG 20000). **$P2_1$ Forms B**: these crystals grew in drop 2 equilibrated against condition F9 of the MORPHEUS crystallisation screen [145, 146]: 0.12 M monosaccharides solution (D-Glucose, D-Mannose, D-Galactose, L-Fucose, D-Xylose, N-Acetyl-D-Glucosamine); 0.1 M Buffer System 3 (Tris (base), bicine pH 8.5) 30% v/v Precipitant Mix 1 (40% v/v PEG 500MME, 20% w/v PEG 20000). **$P2_1$ Forms C and D**: the protein from the SEC-MALS run was set up for crystallisation at a concentration of 3.0 mg/mL. Both form C and from D grew in drop 1 equilibrated against the F1 condition of the MORPHEUS crystallisation screen [145, 146]: 0.12 M monosaccharides solution (D-Glucose, D-Mannose, D-Galactose, L-Fucose, D-Xylose, N-Acetyl-D-Glucosamine); 0.1 M Buffer System 3 (Tris (base), bicine pH 6.5) 30% v/v Precipitant Mix 1 (40% v/v PEG 500MME, 20% w/v PEG 20000).

A.2.4.2 X-ray diffraction collection and processing

All *Hs*GAPDH crystals were cryo-cooled by plunging them into liquid nitrogen. X-ray diffraction were collected at beamline I03 of the Diamond Light Source in Harwell, England, UK, using wavelength λ=0.97622 Å, and a beam size 80x20 μm. Other data collection parameters are listed in Table A.1. Diffraction data were processed with the `autoPROC` suite of programs [211].

	Form A	**Form B**	**Form C**	**Form D**
PDB ID	6YND	6YNE	6YNF	6YNH
Det. dist., d_{max} (mm, Å)	198.35, 1.5	288.19, 2.0	253.03, 1.8	356.69, 2.4
Photon flux (photons/s)	8.84×10^{11}	8.88×10^{11}	8.85×10^{11}	3.88×10^{12}
Transmission	25%	25%	25%	100%
Number of images	3,600	3,600	3,600	3,600
Oscillation range (°)	0.10	0.10	0.1	0.25
Exposure time (s)	0.05	0.05	0.03	0.013
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell edges: a,b,c (Å)	81.88, 124.45, 141.99	81.79, 124.65, 79.64	87.14, 111.43, 135.94	87.02, 111.30, 69.74
Cell angle β (°)	99.38	117.04	96.02	98.33
Resolution Range (Å)	93.04-1.52 (1.71-1.52)	72.85-1.85 (2.05-1.85)	135.18-2.39 (2.74-2.39)	86.10-2.62 (2.89-2.62)
R_{merge}	0.08 (1.07)	0.226 (1.488)	0.27 (1.28)	0.40 (2.47)
R_{meas}	0.09 (1.16)	0.245 (1.618)	0.29 (1.38)	0.41 (2.54)
Observations	2,029,459 (100,354)	568,580 (26,433)	418,478 (20,849)	496,488 (24,336)
Unique observations	295,270 (14,763)	81,465 (4,074)	59,243 (2,962)	28,207 (1,409)
Average $I/\sigma(I)$	11.7 (1.7)	6.5 (1.5)	6.5 (1.6)	9.0 (1.5)
Completeness	69.4 (12.2)	67.6 (12.8)	58.2 (8.7)	71.3 (14.0)
Multiplicity	6.9 (6.8)	7.0 (6.5)	7.1 (7.0)	17.6 (17.3)
$CC_{1/2}$	0.997 (0.608)	0.991 (0.456)	0.986 (0.591)	0.990 (0.464)

Table A.1: Human GAPDH crystal X-ray diffraction data collection parameters and data processing statistics. Values in parentheses refer to the highest resolution shell.

A.2.4.3 Structure determination and refinement

Structure determination started from the experimental structure factor amplitudes. **$P2_1$ Form A**: initial phases were computed by molecular replacement, searching for eight copies of the *Hs*GAPDH monomer in PDB ID 1U8F in space group $P2_1$ using the program CCP4-Molrep [212]. The eight copies of the *Hs*GAPDH monomer are arranged in two tetramers in the asymmetric unit. Initial automated water addition and positional and individual B-factor refinement were carried out in autoBUSTER [213, 214]. Automated non-crystallographic restraints were used throughout [215], including water molecules (assigned to each chain using CCP4-Sortwater). At each catalytic Cys152 site, a 0.5:0.5 occupancy ratio mixture of Cys and Cys S-sulfonic acid was initially modelled in Fo-Fc unbiased residual density (Figure A.7). At each Cys152 site, occupancies for Cys and Cys S-sulfonic acid were refined and constrained so that they sum up to 1.000±0.005. Crystal form A was deposited in the Protein Data Bank with PDB ID code 6YND. Refinement statistics are reported in Table A.2. **$P2_1$ Forms B,C and D**: initial phases were computed by molecular replacement, searching with the $P2_1$ Form A tetramer (PDB ID 6YND) in space group $P2_1$, placing one tetramer per asymmetric unit in $P2_1$ Forms B and D, and two tetramers per asymmetric unit in $P2_1$ Form C, using the program CCP4-Molrep [212]. The same refinement protocol and restraints were used as described for $P2_1$ Form A, but for the addition of external secondary structure restraints [215] to the highest resolution $P2_1$ Form A structure. Crystal forms B, C and D were deposited in the Protein Data Bank with PDB ID codes 6YNE, 6YNF and 6YNH, respectively. Refinement

statistics are reported in Table A.2.

	Form A	**Form B**	**Form C**	**Form D**
PDB ID	6YND	6YNE	6YNF	6YNH
Space group (Z)	$P2_1(16)$	$P2_1(8)$	$P2_1(16)$	$P2_1(8)$
Resolution range	140.10-1.52(1.63-1.52)	72.85-1.85 (1.98-1.85)	135.18-2.39 (2.62-2.39)	86.10-2.62 (2.72-2.62)
Reflex.s working set	280,585 (5,242)	77,398 (1,536)	56,303 (1,124)	25,751 (545)
Reflex.s free set	14,686 (332)	4,059 (94)	2,940 (61)	1,459 (20)
R,R_{free}	0.182,0.199 (0.209,0.225)	0.190,0.207 (0.218,0.220)	0.179,0.219 (0.222,0.347)	0.176,0.215 (0.231,0.359)
$Rmsd_{bonds}$ (Å)	0.008	0.008	0.009	0.008
$Rmsd_{angles}$ (°)	1.02	1.05	1.04	1.03
Ramachandran fav.	97.4% (2,697/2,768)	97.5% (1,304/1,338)	97.0% (2,586/2,667)	97.5% (1,300/1,333)
Ramachandran allow.	99.7% (2,760/2,768)	99.7% (1,334/1,338)	99.5% (2,653/2,667)	99.7% (1,329/1,333)
Tetramers	(A,C,E,H) and (B,D,F,G)	(B,D,F,G)	(A,C,E,H) and (B,D,F,G)	(B,D,F,G)
Occ. Cys-SO_3H	A:0.13 B:0.15	N/A	A:0.5 B:0.41	B:0.41
	C:0.09 D:0.18	N/A	C:0.46 D:0.33	D:0.42
	E:0.13 F:0.15	N/A	E:0.38 F:0.43	F:0.40
	G:0.15 H:0.08	N/A	G:0.42 H:0.72	G:0.35
Prot.(Wat.) Atoms	20,428 (1,639)	10,108 (569)	20,152 (726)	10,164 (132)
$<B>_{prot}$ ($<B>_{wat}$) (Å^2)	27.26 (33.12)	29.40 (36.17)	39.74 (22.93)	45.38 (29.13)

Table A.2: Human GAPDH crystal structures refinement statistics. Values in parentheses refer to the highest resolution shell, unless otherwise specified.

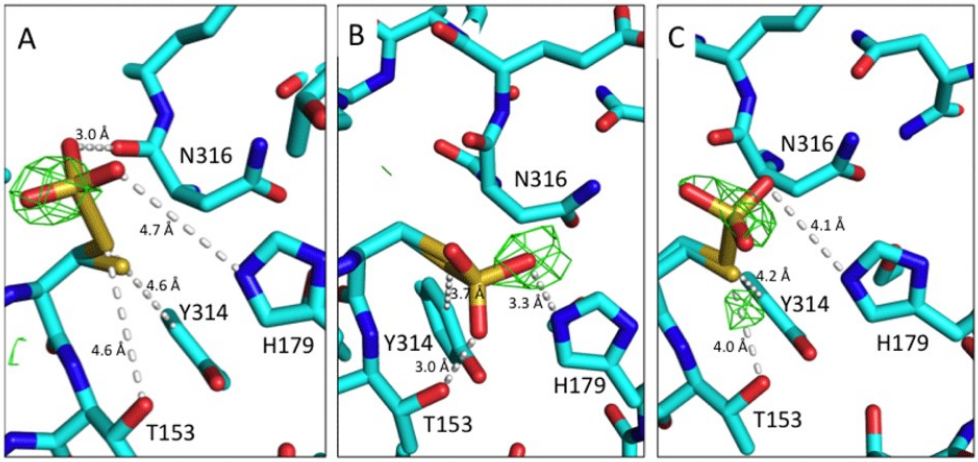


Figure A.7: Fo-Fc difference crystal electron density map around *Hs*GAPDH Cys152 residues. Representative *Hs*GAPDH monomers in crystal form C (PDB ID 6YNF) are shown: A: monomer B; B: monomer E; C: monomer G. The 2.5 σ contour of the unbiased Fo-Fc map is depicted as a green mesh (the map was calculated with phases from the model before the Cys-S-sulfonic acid modification was built). Residue *Hs*GAPDH 152 was modelled and refined as a 0.28:0.72 superposition of Cys:Cys-S-sulfonic acid. C atoms: cyan; N atoms: blue; O atoms: red; S atoms: yellow. Selected non-bonding interactions between the modified Cys and neighbouring residues are represented as white dashed lines with distances in Å. The Figure was made in PyMOL. An alternative Open Source piece of software is USF Chimera [154].

A.3 Discussion

Since the first crystal structures of the lobster GAPDH enzyme [216, 217, 218] several crystal structures of prokaryotic and eukaryotic GAPDHs have been determined. At the time of this writing, the PDB contains more than 142 GAPDH entries, 11 of which are of human GAPDH isoforms, see Table A.3. All human GAPDH crystal structures described so far contain tetramers, sitting in crystal sites of three different symmetries: point group 1 (four copies of the monomer all slightly different from each other); point group 2 (dimer of dimers); and point group 222 (four identical copies in a tetramer of exact 222 symmetry).

Isoform (Uniprot)	**PDB ID**	**Polymorph (Z)**	**Resolution (Å)**	**Reference**
G3P_HUMAN (P04406)	6YND	$P2_1$ Form A (16)	1.54	This work
	6YNE	$P2_1$ Form B (8)	1.85	This work
	6YNF	$P2_1$ Form C (16)	2.39	This work
	6YNH	$P2_1$ Form D (8)	2.62	This work
	1U8F, 1ZNQ	$P2_12_12_1$ Form A (16)	1.75, 2.50	[219, 220]
	3GPD	C2 Form A (8)	3.50	[221]
	4WNC	$P2_1$ Form E (16)	1.99	[222, 223]
	4WNI	$P2_12_12_1$ Form B (16)	2.30	[222, 223]
	6IQ6	$P2_1$ Form F (16)	2.29	[224]
	6ADE	I222 (24)	3.15	N/A
G3PT_HUMAN (O14556)	3H9E, 3PFW, 5C7O	C2 Form B (8)	1.72, 2.15, 1.73	[197, 225]
	5C7O	$P3_121$ (12)	1.86	[197]

Table A.3: Human GAPDH crystal structures.

*Hs*GAPDH crystals of form A and form B grew from the *Hs*GAPDH protein sample purified from the supernatant of HEK293F cells by two chromatography steps (IMAC+SEC); crystals of forms C and form D grew from the same *Hs*GAPDH protein sample after an extra SEC-MALS chromatography step, see Figure A.4. Our *Hs*GAPDH crystal structures all contain tetramers of point group 1 (two tetramers per asymmetric unit in forms A and C, and one tetramer per asymmetric units in forms B and D). Overall, the four novel crystal forms contain 24 crystallographically independent observations of the *Hs*GAPDH molecule, adding to the 29 ones already present in the PDB at the time of this writing, see Table A.3. The overall $\text{rmsd}_{C\alpha}$ is 0.170 Å across these 24 crystallographically independent copies of *Hs*GAPDH.

The protein sample from which crystal forms C and D were grown was analysed by mass spectrometry. In addition to peptides corresponding to unmodified Cys152 (see Figure A.5), we detected ions derived from fragmentation of the Cys152-containing peptide ^{146}IISNAS**C**TTN**C**LAPLAK^{162}, which suggests Cys152 S-sulfonic acid modification (Figure A.6). Annotation is complicated by the presence of a further modification at Cys156 with a mass addition of +71.036 Da, as directly observed in the fragmentation spectrum by the mass spacing between the y-6 and y-7 ions (Figure A.6). The +71.036 Da mass addition at Cys156 is best annotated as acrylamide adduct (pro-

pionamide) imparted during PAGE separation [226] rather than being a native modification at this residue. The measured mass of the peptide was 1869.8623 Da and the theoretical mass for the unmodified peptide is 1718.8695 Da. Including the annotated acrylamide adduct, this leaves 79.9568 Da unaccounted for, which is an extremely close match to S-sulfonic acid modification of Cys (+79.9568 Da). A potential ambiguity for the assignment is the similarity in mass of phosphorylation (+79.966331); however, phosphorylation would equate to mass error of 9.5 mDa as opposed to just 0.018 mDa for S-sulfonic acid. Unlike in the case of the acrylamide adduct, the +79.9568 Da mass addition cannot be observed in any of the individual b- or y-ions in the fragmentation spectrum (Figure A.6), implying complete neutral loss of the moiety between precursor selection and fragmentation. Although neutral loss is frequently observed with Ser and Thr phosphorylation, it is rarely seen with 100% efficiency, further suggesting sulfation of Cys152 as the most credible assignment. Relative quantification of modified *versus* unmodified peptide forms is not possible based on the mass spectrometry data alone, due to unknown but likely significant differences in relative response resulting from the changes in the physico-chemical properties of the Cys152-containing ions upon Cys-S-sulfation.

Peaks suggesting Cys152 S-sulfonic acid modification are also visible in the initial crystal electron density Fo-Fc difference maps of crystal forms A, C and D [189], contoured at 2.5σ level or higher, in close proximity to the Cys152 γ sulfur atom of most *Hs*GAPDH chains in the asymmetric unit (see Figure A.7). The orientation of the Cys-S-sulfonic acid moiety in the active site is not unique, but it varies from chain to chain across the three crystal forms. Figure A.8 illustrates the modification in the eight copies of *Hs*GAPDH in crystal form C. The Cys152 S-sulfonic acid chain makes only loose non-bonding contacts to residues in the active site (see also Figure A.7). Modelling of the Cys152 residue as a superposition of Cys and Cys-S-sulfonic acid with variable occupancies in crystal forms A, C and D enables estimation of the extent to which the crystals contain the Cys152 modification. The average occupancy for the Cys152-S-sulfonic acid modification over these 20 crystallographically independent copies of *Hs*GAPDH is 0.31±0.17 (see Table A.2).The *Hs*GAPDH monomer with the largest occupancy for the modification ($Occ_{Cys-S-SO3}$=0.72) is monomer H in crystal form C. This predominantly Cys-S-sulfonated HsGAPDH monomer superposes with an $rmsd_{C\alpha}$=0.423 Å (over 334 C_α atoms) with the *Streptococcus pneumoniae* GAPDH monomer carrying the same modification [204], see B in Figure A.8. When comparing the modified structure to the one of *Hs*GAPDH with unmodified Cys152 (PDB ID 1U8F) no significant local changes to active site side chains are observed either ($rmsd_{C\alpha}$=0.127 over 333 C_α atoms), see C in Figure A.8. None of the *Hs*GAPDH molecules in the crystals contain NAD^+ nor any bound ligand in the active site.

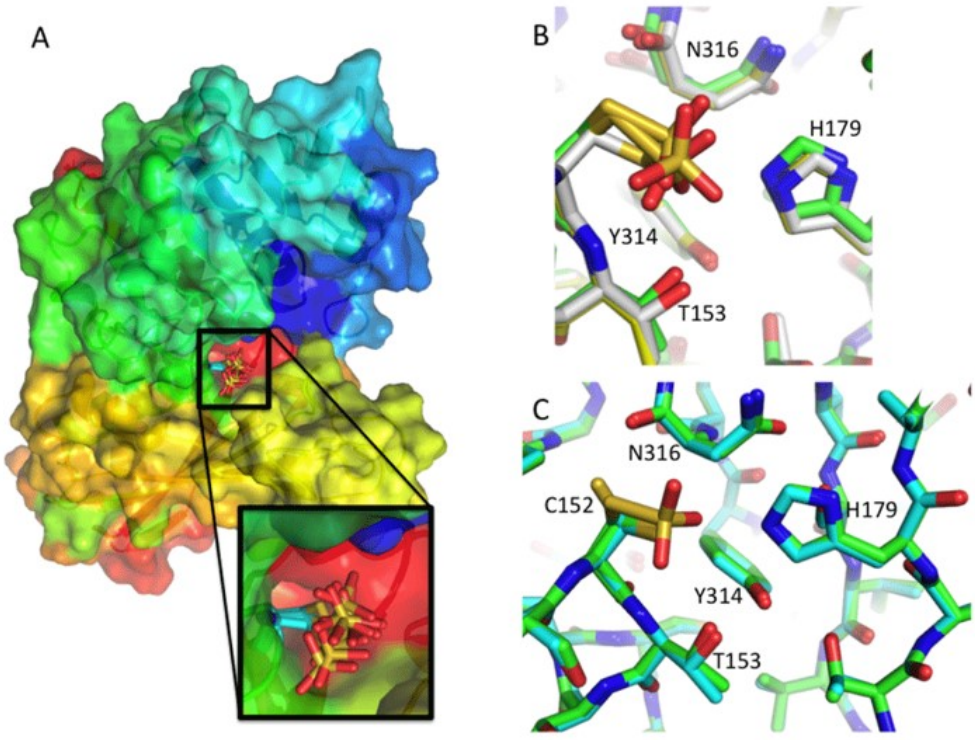


Figure A.8: Cys152-S-sulfonic acid modification of *Hs*GAPDH. N atoms: blue; O atoms: red; S atoms: yellow. **A**: the Cys152-S-sulfonic acid modification in the 8 copies of *Hs*GAPDH in crystal form C. The *Hs*GAPDH monomer is in surface representation and coloured blue to red from N- to C-terminus. The inset shows the Cys152 S-sulfonic acid chains adopting slightly different conformations in different molecules in the asymmetric unit. **B**: superposition of the active site of Cys152-S-sulfonic acid modified *Hs*GAPDH (chain H in crystal form C, green C atoms) with two copies of Cys152-S-sulfonic acid modified *Streptococcus pneumoniae* GAPDH (*Sp*GAPDH, PDB ID 5M6D, chain "A", white C atoms; chain "B", yellow C atoms). **C**: superposition of the active site of Cys152-S-sulfonic acid modified *Hs*GAPDH (chain H in crystal form C, green C atoms) with native *Hs*GAPDH (chain "O" of PDB ID 1U8F, cyan C atoms). The side chain of the catalytic Cys152 of native *Hs*GAPDH in PDB ID 1U8F was modelled and refined as a superposition two alternate conformations. The Figure was made in PyMOL. An alternative Open Source piece of software is USF Chimera [154].

GAPDH is mainly a cytoplasmic enzyme, but it has been reported to moonlight as a cell surface and/or extracellular enzyme in healthy human cells [227, 228],in human cells subjected to stresses such as micronutrient starvation, hypoxia, infection, and cancer [228, 229, 230], and to act as a extracellular effector in fungi and bacteria [231, 232, 233, 234, 235, 236, 237].

In our study, *Hs*GAPDH was purified from the supernatant of HEK293F cells treated with the mannosidase inhibitors kifunensine and kept for 4 days at 37 °C after transfection with a DNA plasmid for recombinant expression of an endoplasmic reticulum glycoprotein. Although

kifunensine is toxic to cells only in concentrations higher than the one used (5μM), during the 4 days between transfection and harvesting of the supernatant the experimental conditions are likely harsher than the basal levels of oxidative stress in the HEK293F cell culture media [238].

Stable oxidation of the GAPDH catalytic cysteine has been observed in a number of studies, for example to Cys sulfenic acid mediated by thiolate oxidation [239]: the modification is reversible but it confers the modified cysteine electrophilic properties similar to the ones of sulfenyl halides [240]. For example, GAPDH catalytic cysteine modification to sulfenic acid converts the enzyme from dehydrogenase to an acyl phosphatase [241, 242, 243]. In the presence of excess oxidant, further oxidation of sulfenic acid to sulfinic acid or sulfonic acid is possible. However, in contrast to GAPDH catalytic Cys oxidation to sulfenic acid, which is reversible, oxidation to sulfonic acid leads to irreversible enzyme inactivation and Cys S-sulfonation is generally considered an irreversible form associated with protein misfolding, degradation, and pathology [240]. In particular, S-sulfonated GAPDH has been reported to translocate to subcellular domains where it does not normally occur, where it may stimulate a "gain of function" that could provoke apoptosis [227].

It is unclear whether the Cys S-sulfonic acid modification we observe in *Hs*GAPDH purified from the supernatant of HEK293F cells represents a physiological response of the cells to the recombinant secreted protein expression system conditions, or whether it is simply the outcome of a protracted oxidative environment on the pool of extracellular enzyme instead. GAPDH allostery is mediated by intra-dimer contacts (the active site of one dimer subunit is in close proximity to residues in the neighbouring subunit). The observation that *Hs*GAPDH tetramers pack in the crystals in ordered fashion even in presence of partial Cys-S-sulfonic acid modification, together with the overall structural similarity of the Cys S-sulfonated and native monomeric enzymes, suggests that the secreted *Hs*GAPDH is constantly inactivated by catalytic Cys S-sulfonation in the expression system's oxidising extracellular environment. Of course, it is also possible that the relative amounts of oxidised vs. reduced GAPDH could vary during protein purification (after harvesting the cells' supernatant) due to changes in the ionic strength and pH; all buffers used were degassed (minimising O_2 content) but contain 1 mM Tris (2-Carboxyethyl) phosphine (TCEP), a reducing agent that prevents intermolecular disulfide bond-mediated oligomerisation of GAPDH. The crystallisation conditions do not nominally contain redox active ingredients but crystal forms A,B grow at pH 8.5, while forms C,D grow at pH 6.5. MS characterisation of GAPDH samples purified at sequential stages during the expression/ purification procedure would elucidate the time-dependence of the enzyme catalytic Cys residue redox composition in the chosen experimental conditions.

Appendix B

Modulation of ERQC and ERAD: A Broad-Spectrum Spanner in the Works of Cancer Cells?

B.1 Introduction

In this Appendix, published as a review in [47], we explore the evidence suggesting that the ability of cancer cells to create and spread tumours around the body, to resist current therapies, and to recur post-treatment hinges vitally on ERQC/ERAD. We review our current understanding of how ERQC/ERAD preserves ER glycoproteostasis and discuss how we may harness the molecular detail so far established on these systems in order to develop new broad-spectrum anticancer therapeutics.

B.2 ERQC/ERAD and cancer

Glycoproteins traversing the secretory pathway of eukaryotic cells reach their cellular or extracellular destinations after folding in the ER [244]. To deal with the constant challenge of protein misfolding in the ER, eukaryotic cells have evolved the ERQC system, centred around the calnexin cycle [245]. Collectively, ERQC components (see Figure 1.3) identify, retain in the ER, and aid folding of misfolded glycoproteins on the way down the secretory pathway. ERQC surveys glycoprotein folding, prevents premature glycoprotein secretion, and is integrated with the adaptive stress response [112]. ERQC proteins either reside in the ER lumen or are inserted in/associated with the ER membrane. A second ER-resident machinery called endoplasmic reticulum-associated degradation (ERAD, see Figure 1.5) comprises proteins that commit terminally misfolded glycoproteins to demannosylation, retrotranslocation to the cytoplasm, and ubiquitination, ultimately targeting them to cytoplasmic proteasomes. Both ERQC and ERAD support cells in their effort to fine tune the rate of glycoprotein synthesis and entry into the ER to match the ER folding capacity (glycoprotein homeostasis or glycoproteostasis) [246].

Malignant cells are deprived of nutrients and their protein synthesis is dysregulated, so that they are especially prone to ER stress. The latter results from protein misfolding within the ER, and it has profound effects on cancer cells' proliferation and survival [247]. It is therefore not surprising that ERQC and ERAD play a key role in cancer biology. Yet, the complexity of ER glycoproteostasis, coupled with the galaxy of cancer cell phenotypes, makes it nontrivial to predict if the activity of a specific ERQC/ERAD component helps or hinders establishment and progression of a specific type of cancer. Indeed, ER quality control and degradation systems have been suggested to represent a double-edged sword that may aid progression as well as prevention of cancer cell growth in a context-dependent manner [248].

Table B.1 lists a number of ERQC/ERAD components and association of their expression levels with cancer patient survival in the Human Protein Atlas (HPA) [249, 250], as evidenced by Kaplan–Meier survival plots [251] derived from cancer tissue images. Quite a few of these ERQC/ERAD components have been identified as unfavourable prognostic markers in cancer studies. We also list the frequency of somatic mutations detected in the same genes, as reported by the Catalogue of Somatic Mutations in Cancer (COSMIC), the world's largest source of manually curated somatic mutation information relating to human cancers [252]. Other useful resources are the database of therapeutic vulnerability of cancer [253], lists of oncogenes [254], and the tumour suppressor gene TUSON ranking [255] (`https://bioinfo.uth.edu/TSGene/` in the cancer cell metabolism genes database [256], but in the interest of simplicity, we did not compile values from these online sources in Table B.1.

In the following subsections, we briefly review some of the published evidence of direct cancer association for a selected subset of ERQC/ERAD components, before examining the second-order involvement of ERQC/ERAD with cancer, through their regulation of folding and degradation of specific cancer-associated secreted glycoproteins.

Protein/gene and UniProt entry/name	**Prognosis upon overexpression (Human Protein Atlas, p<0.001)**	**Frequency (from COSMIC) [252]**
GCS1/MOGS, Q13724/ MOGS_HUMAN	Unfavourable prognosis in renal, liver and colorectal cancers.	191/47211 (0.4%)
ER α Glu II α subunit/GANAB, Q14697/GANAB_HUMAN	Unfavourable prognosis in liver and urothelial cancers.	254/47211 (0.5%)
ER α Glu II β subunit/PRKCSH, P14314/GLU2B_HUMAN	Unfavourable prognosis in renal cancer.	191/47211 (0.4%)
UGGT1/UGGT1, Q9NYU2/UGGG1_HUMAN	Unfavourable prognosis in renal cancer.	333/47297 (0.7%)
UGGT2/UGGT2, Q9NYU1/UGGG2_HUMAN	Unfavourable prognosis in lung and liver cancers	406/47212 (0.8%)
Sep15/Sep15, O60613/SEP15_HUMAN	Unfavourable prognosis in liver and head and neck cancers but favourable prognosis in colorectal cancer.	17/47187 (0.04%))
Calnexin/CANX, P27824/CALX_HUMAN	Favourable prognosis in colorectal cancer but unfavourable in thyroid cancer.	151/47211 (0.3%)
Calreticulin/CALR, P27797/CALR_HUMAN	Favourable prognosis in ovarian but unfavourable in renal cancer.	4344/81169 (5.3%)
ER UDPase, O75356, ENTP5_HUMAN	Favourable prognosis in renal cancer.	110/47209 (0.2%)
ER αMan I, Q9UKM7, MA1B1_HUMAN	Unfavourable prognosis in liver cancer.	178/47354 (0.4%)
EDEM1, Q92611, EDEM1_HUMAN	N/A	141/47255 (0.3%)
EDEM2, Q9BV94, EDEM2_HUMAN	Unfavourable prognosis in renal cancer.	162/35626 (0.4%)
EDEM3, Q9BZQ6, EDEM3_HUMAN	Unfavourable prognosis in renal cancer.	231/35629 (0.6%)
ERDJ5, Q8IXB1, DJC10_HUMAN	Favourable prognosis in endometrial cancer but unfavourable in renal and thyroid cancers.	205/47347 (0.4%)
HRD1, Q86TM6, SYVN1_HUMAN	Favourable prognosis in head and neck cancer.	143/47298 (0.3%)
OS-9, Q13438, OS9_HUMAN	N/A	188/47797 (0.4%)

Table B.1: ERQC/ERAD components and association of their expression levels with cancer patient survival in the Human Protein Atlas (HPA) [249, 250]

B.2.1 ERQC and cancer

Genetic defects in MOGS, the gene encoding α-Glucosidase I (GCS1), cause rare congenital disorder of glycosylation type IIb (CDG-IIb) and confer decreased susceptibility to infections due to viruses whose life cycle depends on the host cell's calnexin cycle [257]. The Human Protein Atlas (HPA) [249, 250] reports unfavourable prognoses in human renal, liver, and colorectal cancers overexpressing the MOGS gene (see Table B.1).

ER α-Glucosidase II (ER αGlu II, in green and cyan in Figure 1.3) acts as an usher, mediating both entry and exit of a glycoprotein into the cycle [7]. Overexpression of ER αGlu II β subunit (ER αGlu II β) in different tumour tissues has been reported [31, 258]. More recently, it has been suggested that activation of ERQC through ER αGlu II can help tumour cells to escape from autophagy and apoptosis [259]. A study of molecular chaperones regulating the invasion phenotype of head and neck cancer (HNC) established that loss of the tumour suppression function of the ER αGlu II α subunit contributed to aggressive cancers [260].

Calnexin (CNX, ER-membrane inserted, in violet in Figure 1.3) and calreticulin (CRT, ER lumenal and soluble) are ERQC lectins specific for mono-glucosylated glycans ($Glc_1Man_9GlcNAc_2$). They recruit monoglucosylated glycoproteins to oxido-reductases, isomerases, and foldases, effectively constituting the refolding end of the calnexin cycle. In one lung cancer study, low levels of CNX contributed to poor prognosis: in a cell-culture model, targeted depletion of calnexin reduced cancer proliferation, invasion, and migration [49]. CNX expression positively correlates with metastasis of breast cancer to the brain [50]. CNX was also significantly upregulated in oral squamous cell carcinoma, and its levels correlated with poor prognosis in patients affected by this tumour [70].

UGGT (UDP-glucose glycoprotein glucosyltransferase) is the ERQC checkpoint, detecting misfolded glycoproteins and reglucosylating them in order to enable further rounds of association with CNX/CRT, beyond the initial one(s) afforded by the OST-transferred N-glycan(s) after the initial ER αGlu II cleavages [31]. In higher vertebrates, there are two UGGT isoforms, UGGT1 and UGGT2. The Human Protein Atlas (HPA) [249, 250] reports unfavourable prognoses in renal cancers and lung and liver cancers overexpressing UGGT1 or UGGT2, respectively. A majority of cancers are reported to overexpress the UGGT1 gene (see Figure B.1(a)), and a few cancer types report a significant rate of mutations in the same gene (see Figure B.1(b))—although without functional data, it is difficult to assess if they are likely to impair or enhance protein function. No studies have directly tested the role of UGGT in cancer plasticity, but a recent paper reports that the IGF-1R receptor - overexpressed in many cancers is trafficked to the plasma membrane at a decreased rate in UGGT1 and UGGT2 KO cells [48]. Therefore, a potential inhibitor of UGGT1 and UGGT2 may function as a treatment against these cancers.

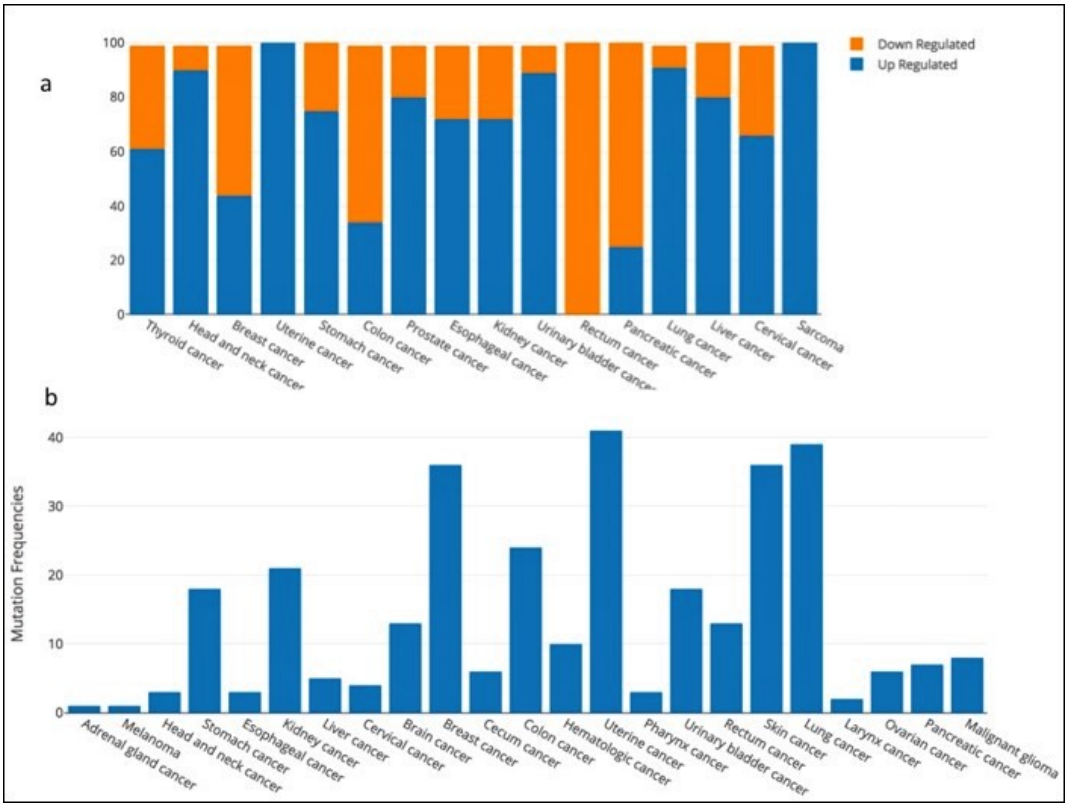


Figure B.1: (a); Percentage of upregulation and downregulation of the UGGT1 gene in a number of cancers from RNA-seq experiments on bulk cancer tissues. (b); Mutation frequencies in various cancer types in the same gene (from the Q9NYU2 entries in BioMuta and BioXpress at the OncoMX server [261])

Sep15 (aka Selenoprotein F, Selenof, in pink in Figure 1.3) is a 15 kDa protein which in humans (but not in fruit fly, mosquito, zebrafish or rat) contains a selenocysteine residue [262]. Selenium has been implicated in cancer prevention [263], but the mechanism and possible involvement of selenoproteins in this process are not well understood. Based on the fact that abnormal glycoprotein folding and secretion were observed in conjunction with Sep15 deficiency, it has been proposed that it may have an important role in the ER maturation of N-glycosylated proteins [264], in particular M-immunoglobulins [265]. Sep15 mitigates oxidative stress and apoptosis [266]. Its C-terminal domain (residues 46–134) folds as a thioredoxin-like domain [267]; the N-terminal domain (residues 1–45), whose fold is not easily predictable from sequence, likely mediates Sep15 nanomolar association with UGGT1 [268]. Indeed, Sep15 enhances UGGT1-mediated reglucosylation of IL-8 and crambin containing mispaired disulfides [259, 269], suggesting that the Sep15 redox potential may have evolved to selectively reduce/isomerise disulfides in nonnative over native environments. A number of studies point to a role of Sep15 in cancer aetiology. The Sep15 coding gene is located in a highly mutated region of chromosome 1, and several mutations and deletions of Sep15 coding gene are involved in cancer progression and tumorigenesis [262]. The expression levels of Sep15 were

investigated in various cancer models: downregulation of the protein was found in hepatocarcinomas and colorectal, gastric, and prostate cancers [262, 263, 270, 271]. On the other hand, decreased expression of Sep15 reduces proliferation and growth of liver and colon cancer cell lines, pointing to a role of Sep15 in tumour progression [270, 272, 273]. Single-nucleotide polymorphisms in the Sep15 gene have been studied in conjunction with differential levels of selenocysteine insertion and susceptibility to lung and breast cancer [274, 275], highlighting the need for a stratified medicine approach in the development of Sep15 modulators as anticancer therapeutics.

Once UGGT has transferred a Glc molecule from UDP-Glc to a misfolded glycoprotein glycan, a molecule of UDP is produced, which would inhibit UGGT [276]. As is the case for other nucleoside diphosphates produced by sugar transferases [277], an ER-specific UDPase (NTPD5, UniProt O75356, ENTP5HUMAN, in grey in Figure 1.3) hydrolyses the ER UDP pool to UMP [276, 278]. NTPD5 may mediate some of the cancer-related phenotypes associated with AKT1 activation: NTPD5 is upregulated in cell lines and primary human tumor samples with active AKT and, together with cytidine monophosphate kinase-1 and adenylate kinase-1, is part of an ATP hydrolysis cycle that converts ATP to AMP, resulting in the cancer-associated compensatory increase in aerobic glycolysis known as the Warburg effect [279]. Many studies have correlated dysregulation of the expression of the ER UDPase with a range of cancers, explaining why the enzyme has been proposed as a potential target for anticancer therapy [280, 281, 282].

B.2.2 ERAD and cancer

High growth rate, impaired ATP generation, hypoxia, hypoglycemia, and specific mutations perturb cancer cells' ER homeostasis [283, 284] and may also induce UPR [285]. This in turn can lead to cell death. ERAD unwittingly (but effectively) helps cancer cells by conferring them tolerance to glycoproteotoxic stress. Indeed, survival under chronic ER stress is a feature of aggressive cancers [286], and tumour cells attempt survival by hijacking ERAD [287]. For these reasons, terminal ERAD components inhibitors have been proposed as targets to specifically impair the survival of cancer cells [247, 288]. Blocking ERAD can also trigger cellular apoptosis [289].

The ERAD components acting early in the pathway are the endoplasmic reticulum degradation-enhancing mannosidases (EDEM), committing misfolded glycoproteins to degradation. To date, no EDEM-specific inhibitors are known, and the effects of EDEM inhibition/deletion on cancer cells have not been investigated, although the generic α-mannosidase inhibitor kifunensine [290] increased adhesion of breast cancer cells to endothelial cells [291] and 1-deoxymannojirimycin (another broad-spectrum mannosidase inhibitor) induced cellular ER stress in a human hepatocarcinoma cell line [292].

A conserved motif within the 3′UTR of ER αMan I is a target of miR-125b, a microRNA frequently downregulated in numerous types of cancers, including hepatocellular carcinoma (HCC), with the expression of ER ManI significantly elevated in HCC, as measured by immunohistochemistry in a liver disease spectrum tissue microarray [293].

ER ManIA aka mannosyl-oligosaccharide 1,2-α-mannosidase IA (ER Man IA, UniProt P33908,

MA1A1HUMAN)— originally annotated as resident in the Golgi—has been shown to co-localise with ER αMan I in quality control vesicles (QCVs) and is also implicated in targeting to ERAD [294]. In cancer, the enzyme levels showed impact on degradation of cell-surface glycoprotein involved in cell-cell adhesion and metastasis: reduced ER ManIA expression or mannosidase inhibition lead to a significantly increased adhesion of breast cancer cells to endothelial cells [291]. Conversely, ER ManIA was downregulated in metastatic hepatocellular cancer (HCC) cell lines and orthotopic xenograft tumours, in comparison with nonmetastatic HCC controls [295].

The Human Protein Atlas (HPA) [249, 250] reports unfavourable prognoses in renal cancers overexpressing EDEM2 and EDEM3. A somatic variant of EDEM1 (N198I), which loses one of its five *N*-linked glycans, was found to confer a selective advantage to hepatocellular carcinoma cells [296].

In mouse, EDEM1 has been found in association with the ERDJ5 protein disulfide isomerase (PDI) [80, 297], and the same interaction was observed between human EDEM1 and the human ERAD PDIs ERDJ5 [298] and TXNDC11 [74]. High levels of ERDJ5 and TXNDC11 are unfavourable prognostic markers in renal and thyroid cancers and glioma, respectively [249, 250]. Knockdown of ERDJ5 by RNA interference in neuroectodermal tumour cells increases the apoptotic response to fenretinide [299]. These data make the case for selective ERDJ5 and/or TXNDC11 modulators as novel chemotherapeutic targets. On the other hand, high levels of TXNDC11 or ERDJ5 were a favourable prognostic marker in endometrial cancer [249, 250], and overexpression of ERDJ5 sensitises neuroblastoma cells to ER stress-induced apoptosis [300], so it is clear that inhibiting this PDI may not work against some cancers. Similar results were observed for TXNDC11, whose elevated levels of expression correlated with suppression of tumour-promoting genes [301].

Once demannosylated by EDEMs, misfolded glycoproteins in the ER lumen and membrane are recruited by the osteosarcoma 9 (OS-9) and XTP3B ERAD lectins [302] which direct them to the ER membrane-bound complexes assembled around E3 ubiquitin ligases [303, 304]. Both OS-9 and XTP3B are localised in the ER lumen [302]. OS-9 and XTP3B specifically recognise Man α(1,6)-Man α(1,6)-Man residues on the processed C-arm of the *N*-linked glycan (Figure 1.1) [66]. XTP3B also inhibits the degradation of nonglycosylated proteins [63]. Yet again, different studies report opposite roles of the ERAD lectins in different cancers. For example, OS-9 is highly upregulated in osteosarcoma [305] and XTP3B was found to be critical for metastatic properties of human lung cancer cell lines [306], while a long noncoding RNA suppresses pancreatic ductal adenocarcinoma (PDAC) cell invasion by increasing both mRNA and protein levels for OS-9 [307].

The ERAD E3 ubiquitin-protein ligases accept ubiquitin specifically from an ER-associated E2 ligase and transfer it to glycoprotein substrates that need degradation [308]. After ubiquitination, the p97 (aka VCP) ATPase helps feeding substrates to a cytosolic proteasome [309]. Most of these ERAD ubiquitin-protein ligases are poorly characterised, and only few targets for each of them, for example, for the ERAD E3 enzymes HRD1 and MARCH6 [310, 311], have been identified. HRD1 protects cells from ER stress-induced apoptosis [312], and its upregulation promotes cell migration and invasion in colon cancer [313]. Another ERAD E3 enzyme, AMFR (aka gp78),

mediates tumour invasion and metastasis functioning as a receptor for the GPI/autocrine motility factor [314]. Modulation of components of HRD1 partners [315] has been proposed as a novel point of intervention for cancer therapies, although there is published evidence that HRD1 suppresses the growth and metastasis of breast cancer cells [316] and the decrease of HRD1 expression contributed to tamoxifen resistance in breast cancer [317] (the latter by promoting the degradation of S100A8, a divalent metal ion binding protein involved in the chemistry-drug resistance in many tumours).

B.2.3 Glycoproteins and Anti-cancer Strategies Focusing on ERQC/ERAD Modulation

Cancer cells survive by adjusting to ER stress, and a number of studies in the literature have pointed out that components of the ERQC/ERAD machineries may constitute anticancer therapeutic targets [299, 300]. The centrality of ERQC/ERAD to glycoproteostasis would potentially endow such compounds with broad-spectrum activity, but depending on their glycoprotein secretory burden, different cancers will vary in their sensitivity to strategies that interfere with ER stress and glycoprotein folding and degradation [112]. Importantly, as is the case for any drug that interferes with basic cellular pathways, ERQC/ERAD modulators are potentially toxic to healthy cells as well. For these reasons, the most promising use for ERQC/ERAD modulators will likely be in combination with existing chemotherapeutics. For example, inhibition of homeostatic ER stress responses enhances apoptosis induced by oxidative stress-inducing drugs acting through the ER stress pathway [299].

Any attempt to develop ERQC/ERAD modulators as anticancer therapeutics would want to aim at ER stress-mediated selective killing of malignant cells without imposing significant damage to surrounding healthy cells. To be selective in aid of anticancer therapy, any ERQC/ERAD inhibitor of this kind needs to exploit different folding requirements of specific glycoproteins in cancerous vs. healthy cells. Amongst the many glycoprotein-dependent strategies used specifically by cancer cells are the expression of tumour-specific glycoprotein isoforms (with patterns of alternative splicing of mRNAs differing between tumour and normal tissues from which they are derived [318]); tumour-specific glycoprotein conformations [319]; upregulation of membrane-embedded drug transporters mediating chemotherapic multidrug resistance [320]; and expression of surface adhesion glycoproteins involved in tissue penetration and/or metastasis in leukemic cells [321] and solid malignancies [322]. Cancer cells also rely extensively on receptor tyrosine kinases (RTK): these glycoproteins are important in squamous cell carcinomas, breast/pancreas/prostate adenocarcinomas, and malignant gliomas. Indeed, nanomolar concentration of tunicamycin, a well-known inhibitor of N-glycosylation, reduces protein levels of at least four RTKs involved in tumour cell proliferation and survival [252, 254, 253].

Glycoproteins are also central to cancer immunotherapy [323, 324]: therapeutic anticancer antibodies, their cell-surface receptors, most of their epitopes [323], and complement components [325] are all glycoproteins. Many glycoproteins also underpin cancers' lack of response to immunotherapy response [114]. Drugs altering glycoprotein secretion/degradation will alter a patient's glycose-

cretome, including the surface antigens targeted by immunotherapy monoclonal antibodies, Fcγ receptors (FcγRs), and components of the complement system. Indeed, recent evidence has implicated polymorphisms of FcγR in the efficacy of monoclonal antibody- (mAb-) mediated therapy [326]. As the molecular basis for the opposite effects between inhibitory vs. activating FcγR resides in different intracellular phosphotyrosyl-based motifs [327], the folding/degradation requirements of different FcγRs may differ. Unfortunately, we have only partially uncovered the roles played by ERQC/ERAD during anticancer mAb therapy and, in particular, the folding and stability of cancer-specific surface glycoprotein epitopes, FcγRs and complement components [110]: the hypothesis that drugs that selectively impair glycoprotein folding and degradation may aid cancer immunotherapy remains to be tested.

B.3 Conclusions

A large number of published studies have highlighted the dependency of a number of cancers on specific ERQC/ERAD components, but the lack of specific inhibitors of the components in both pathways has hampered proper characterisation of the roles played by ERQC/ERAD in cancer biology. Even if such specific inhibitors were available, in order to make a convincing case for ERQC/ERAD as valid anticancer targets, several aspects of ERQC/ERAD biology in healthy and cancer cells need to be better elucidated.

Important open questions involve the degrees of redundancy and interplay between ERAD and ERQC checkpoints (again, UGGTs and EDEMs) in deciding the fate of a specific misfolded glycoprotein. Whether there is a general mechanism by which the dilemma ER retention vs. secretion is solved or whether different individual glycoproteins are taken care by the ERQC and/or the ERAD branch to different extents during their lifetime in the ER still remains to be elucidated. The extents to which specific cancers tip the EQRC/ERAD balance for glycoproteins that are crucial to their survival will be of course one of the next big questions to answer, ultimately helping in each case to make choices between ERQC vs. ERAD modulation for the most effective anticancer prescription of this kind.

Last but not least, when it comes to toxicity, although evidence of ER-retention and/or ER associated degradation exists for a few cancer-associated glycoproteins, we do not know which EQRC/ERAD clients would risk premature and unwanted secretion in healthy cells (a scenario we dubbed the "ER Pandora's box") upon administration of an ERQC/ERAD modulator. Thus, the relative toxicity of such drugs to healthy vs. cancer cells is difficult to predict. Targeting ERQC/ERAD may well prove a broad-spectrum spanner in the plasticity works of cancer cells, but—as it often happens with cancer biology—winning this battle will require a better understanding of the roles that these machineries play in cells at various stages of the cell cycle (in healthy cells as well as in cancer tissues). Only then may ERQC/ERAD inhibitors reach the clinic, adding to the expanding arsenal of anticancer therapeutics.

List of activities

- 17th-19th May 2018, Padova (Italy): Galileo Festival;
- 18th-21st September 2018, Roma (Italy): Poster presentation to XV° FISV congress, "*Arabidopsis thaliana* enables low-cost, easy and fast *in vivo* screening of ERQC modulators";
- September 2018, Lecce (Italy): Poster presentation at the "Notte dei ricercatori 2018", "*Arabidopsis thaliana* come modello per lo studio di ERQC: dalla pianta all'uomo";
- 3rd May 2019, Leicester (UK): Poster presentation at the PhD students' day, "Expression of recombinant EDEMs ERAD checkpoint for structural analysis";
- 8th-10th May 2019, Buenos Aires (Argentina): Poster presentation at 3rd Argentinian Symposium of Glycobiology, "Modulation of ERAD for rescue of secretion of glycoprotein responsive mutants in rare disease";
- 18th July 2019, Lecce (Italy): Co-supervisor of a Master's Degree thesis "The SEIPIN gene family of *Arabidopsis thaliana* (L.): physiological role and its commitment in the endoplasmic reticulum stress response";
- 1st-7th September 2019, York (UK): Elevator Pitch at the CCP4/BCA Protein Crystallography Summer School;
- 18th-20th September 2019, Lecce (Italy): 61st National Congress Società Italiana Biochimica.

List of publications

1. July 2018: Scientific peer-reviewed publication: "*In Planta* Preliminary Screening of ER Glycoprotein Folding Quality Control (ERQC) Modulators" Lucia Marti, **Andrea Lia**, Ida-Barbara Reca, Pietro Roversi, Angelo Santino and Nicole Zitzmann. International journal of molecular sciences 19 (7), 2135. `https://doi.org/10.3390/ijms19072135`;

2. December 2018: Scientific peer-reviewed publication: "EFR-Mediated Innate Immune Response in *Arabidopsis thaliana* is a Useful Tool for Identification of Novel ERQC Modulators" **Andrea Lia**, Antonia Gallo, Lucia Marti, Pietro Roversi, Angelo Santino. Genes, 2018. `https://doi.org/10.3390/genes10010015`;

3. October 2019: Review article "Modulation of ERQC and ERAD: A Broad-Spectrum Spanner in the Works of Cancer Cells?" Gábor, **Andrea Lia**, Angelo Santino, Pietro Roversi. Journal of Oncology, 2019. `https://doi.org/10.1155/2019/8384913`;

4. August 2020: Scientific peer-reviewed publication: "Partial catalytic Cys oxidation of human GAPDH to Cys-sulfonic acid" **Andrea Lia**, Adam Dowle, Chris Taylor, Angelo Santino, Pietro Roversi. Wellcome Open Research 5 (114), 114. `https://doi.org/10.12688/wellcomeopenres.15893.2`;

5. April 2021: Scientific peer-reviewed publication: "Clamping, Bending, and Twisting Inter-Domain Motions in the Misfold-Recognising Portion of UDP-Glucose: Glycoprotein Glucosyl-Transferase" Carlos P Modenutti, Juan I Blanco Capurro, Roberta Ibba, Dominic S Alonzi, Mauro N Song, Snežana Vasiljević, Abhinav Kumar, Anu V Chandran, Gabor Tax, Lucia Marti, Johan C Hill, **Andrea Lia**, Mario Hensen, Thomas Waksman, Jonathan Rushton, Simone Rubichi, Angelo Santino, Marcelo A Martí, Nicole Zitzmann, Pietro Roversi. Structure, Volume 29, Issue 4, 1 April 2021, Pages 357-370.e9 `https://dx.doi.org/10.2139/ssrn.3565029`.

Bibliography

[1] Benjamin S. Glick. The secretory pathway. In *Protein Targeting, Transport, and Translocation*, pages 358–376. Elsevier, 2002.

[2] James E. Rothman and Lelio Orci. Molecular dissection of the secretory pathway. *Nature*, 355(6359):409–415, January 1992.

[3] R. Schekman and L. Orci. Coat proteins and vesicle budding. *Science*, 271(5255):1526–1533, March 1996.

[4] J. E. Rothman and F. T. Wieland. Protein sorting by transport vesicles. *Science*, 272(5259):227–234, April 1996.

[5] Sally L. Hanton, Loren A. Matheson, and Federica Brandizzi. Seeking a way out: export of proteins from the plant endoplasmic reticulum. *Trends in Plant Science*, 11(7):335–343, July 2006.

[6] Belinda A. Phillipson, Peter Pimpl, Luis Lamberti Pinto daSilva, Andrew J. Crofts, J. Philip Taylor, Ali Movafeghi, David G. Robinson, and Jürgen Denecke. Secretory bulk flow of soluble proteins is efficient and COPII dependent. *The Plant Cell*, 13(9):2005–2020, September 2001.

[7] C Hammond, I Braakman, and A Helenius. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proceedings of the National Academy of Sciences of the United States of America*, 91(3):913–917, feb 1994.

[8] Stella M. Hurtley and Ari Helenius. Protein oligomerization in the endoplasmic reticulum. *Annual Review of Cell Biology*, 5(1):277–307, November 1989.

[9] R Kornfeld and S Kornfeld. Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry*, 54(1):631–664, June 1985.

[10] Yu Fan, Yu Hu, Cheng Yan, Radoslav Goldman, Yang Pan, Raja Mazumder, and Hayley M. Dingerdissen. Loss and gain of n-linked glycosylation sequons due to single-nucleotide variation in cancer. *Scientific Reports*, 8(1), March 2018.

[11] R Apweiler. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1473(1):4–8, December 1999.

[12] Mukram M Mackeen, Andrew Almond, Michael Deschamps, Ian Cumpstey, Antony J Fairbanks, Clarence Tsang, Pauline M Rudd, Terry D Butters, Raymond A Dwek, and Mark R Wormald. The conformational properties of the Glc3Man unit suggest conformational biasing within the chaperone-assisted glycoprotein folding pathway. *Journal of molecular biology*, 387(2):335–347, mar 2009.

[13] Mark R Wormald, Andrei J Petrescu, Ya-Lan Pao, Ann Glithero, Tim Elliott, and Raymond A Dwek. Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling. *Chemical reviews*, 102(2):371–386, 2002.

[14] Stuart McNicholas and Jon Agirre. Glycoblocks: a schematic three-dimensional representation for glycans and their interactions. *Acta crystallographica. Section D, Structural biology*, 73(Pt 2):187–194, feb 2017.

[15] L. F. Leloir. *The Role of Dolichol in Protein Glycosylation*, pages 9–19. Springer US, Boston, MA, 1977.

[16] Sean P Ferris, Nikita S Jaber, Maurizio Molinari, Peter Arvan, and Randal J Kaufman. UDP-glucose:glycoprotein glucosyltransferase (UGGT1) promotes substrate solubility in the endoplasmic reticulum. *Molecular biology of the cell*, 24(17):2597–2608, sep 2013.

[17] Jianjun Zhang, Jiarui Wu, Linchuan Liu, and Jianming Li. The crucial role of demannosylating asparagine-linked glycans in eradicating misfolded glycoproteins in the endoplasmic reticulum. *Frontiers in Plant Science*, 11:2145, 2021.

[18] Roland Knauer and Ludwig Lehle. The oligosaccharyltransferase complex from yeast. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1426(2):259–273, January 1999.

[19] Jürgen Schweden, Corinna Borgmann, Günter Legler, and Ernst Bause. Characterization of calf liver glucosidase i and its inhibition by basic sugar analogs. *Archives of Biochemistry and Biophysics*, 248(1):335–340, July 1986.

[20] K. Shailubhai, B.S. Pukazhenthi, E.S. Saxena, G.M. Varma, and I.K. Vijay. Glucosidase i, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *Journal of Biological Chemistry*, 266(25):16587–16593, September 1991.

[21] B. Hardt. (arg)3 within the n-terminal domain of glucosidase i contains ER targeting information but is not required absolutely for ER localization. *Glycobiology*, 13(3):159–168, November 2002.

[22] Paola Deprez, Matthias Gautschi, and Ari Helenius. More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. *Molecular Cell*, 19(2):183–195, jul 2005.

[23] Micha A Haeuptle and Thierry Hennet. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Human mutation*, 30(12):1628–1641, dec 2009.

[24] John Samuelson and Phillips W Robbins. Effects of N-glycan precursor length diversity on quality control of protein folding and on protein glycosylation. *Seminars in Cell & Developmental Biology*, 41:121–128, 2015.

[25] Akhlaq Farid, Martin Pabst, Jennifer Schoberer, Friedrich Altmann, Josef Glössl, and Richard Strasser. Arabidopsis thaliana alpha1,2-glucosyltransferase (ALG10) is required for efficient N-glycosylation and leaf growth. *The Plant journal : for cell and molecular biology*, 68(2):314–325, oct 2011.

[26] Giovanna L Gallo, Ayelén Valko, Sofía I Aramburu, Emiliana Etchegaray, Christof Völker, Armando J Parodi, and Cecilia D'Alessio. Abrogation of glucosidase I-mediated glycoprotein deglucosylation results in a sick phenotype in fission yeasts: Model for the human MOGS-CDG disorder. *The Journal of biological chemistry*, 293(52):19957–19973, dec 2018.

[27] De Praeter CM, G J Gerwig, E Bause, L K Nuytinck, J F Vliegenthart, W Breuer, J P Kamerling, M F Espeel, J J Martin, De Paepe AM, N W Chan, G A Dacremont, and Van Coster RN. A novel disorder caused by defective biosynthesis of N-linked oligosaccharides due to glucosidase I deficiency. *American journal of human genetics*, 66(6):1744–1756, jun 2000.

[28] Christof Völker, Claudine M De Praeter, Birgit Hardt, Willi Breuer, Burga Kalz-Füller, Rudy N Van Coster, and Ernst Bause. Processing of N-linked carbohydrate chains in a patient with glucosidase I deficiency (CDG type IIb). *Glycobiology*, 12(8):473–483, 2002.

[29] C. Stewart Gillmor, Patricia Poindexter, Justin Lorieau, Monica M. Palcic, and Chris Somerville. α-glucosidase i is required for cellulose biosynthesis and morphogenesis in arabidopsis. *Journal of Cell Biology*, 156(6):1003–1013, March 2002.

[30] Cecilia D'Alessio, Julio J. Caramelo, and Armando J. Parodi. Udp-glc:glycoprotein glucosyltransferase-glucosidase ii, the ying-yang of the er quality control. *Seminars in Cell Developmental Biology*, 21(5):491 – 499, 2010. Quality control in the endoplasmic reticulum.

[31] Armando J Parodi, Julio J Caramelo, and Cecilia D'Alessio. UDP-Glucose: Glycoprotein Glucosyltransferase 1,2 (UGGT1,2). In *Handbook of Glycosyltransferases and Related Genes*, pages 15–30. Springer Japan, Tokyo, 2014.

[32] Calnexin, calreticulin and the folding of glycoproteins. *Trends in Cell Biology*, 7(5):193 – 200, 1997.

[33] Ursula G. Danilczyk, Myrna F. Cohen-Doyle, and David B. Williams. Functional relationship between calreticulin, calnexin, and the endoplasmic reticulum luminal domain of calnexin. *Journal of Biological Chemistry*, 275(17):13089–13097, April 2000.

[34] Jason D. Oliver, Fimme J. van der Wal, Neil J. Bulleid, and Stephen High. Interaction of the thiol-dependent reductase erp57 with nascent glycoproteins. *Science*, 275(5296):86–88, 1997.

[35] Guennadi Kozlov, Sara Bastos-Aristizabal, Pekka MAAttAnen, Angelika Rosenauer, Fenglin Zheng, April Killikelly, Jean-François Trempe, David Y Thomas, and Kalle Gehring. Structural basis of cyclophilin B binding by the calnexin/calreticulin P-domain. *Journal of Biological Chemistry*, 285(46):35551–35557, nov 2010.

[36] Guennadi Kozlov, Juliana Muñoz-Escobar, Karla Castro, and Kalle Gehring. Mapping the er interactome: The p domains of calnexin and calreticulin as plurivalent adapters for foldases and chaperones. *Structure*, 25(9):1415 – 1422.e3, 2017.

[37] Benjamin M Adams, Michela E Oster, and Daniel N Hebert. Protein Quality Control in the Endoplasmic Reticulum. *The protein journal*, 38(3):317–329, jun 2019.

[38] Chaitanya Patel, Haddas Saad, Marina Shenkman, and Gerardo Z. Lederkremer. Oxidoreductases in glycoprotein glycosylation, folding, and ERAD. *Cells*, 9(9):2138, September 2020.

[39] John G. Elliott, Jason D. Oliver, and Stephen High. The thiol-dependent reductase ERp57 interacts specifically withN-glycosylated integral membrane proteins. *Journal of Biological Chemistry*, 272(21):13849–13855, May 1997.

[40] André Zapun, Nigel J. Darby, Daniel C. Tessier, Marek Michalak, John J. M. Bergeron, and David Y. Thomas. Enhanced catalysis of ribonuclease b folding by the interaction of calnexin or calreticulin with ERp57. *Journal of Biological Chemistry*, 273(11):6009–6012, March 1998.

[41] Y. Saito. Calreticulin functions in vitro as a molecular chaperone for both glycosylated and non-glycosylated proteins. *The EMBO Journal*, 18(23):6718–6729, December 1999.

[42] Peng Wang, Hong Wang, Jiangtao Gai, Xiaoli Tian, Xiaoxiao Zhang, Yongzhi Lv, and Yi Jian. Evolution of protein n-glycosylation process in golgi apparatus which shapes diversity of protein n-glycan structures in plants, animals and fungi. *Scientific Reports*, 7(1), January 2017.

[43] Guennadi Kozlov and Kalle Gehring. Calnexin cycle - structural features of the ER chaperone system. *The FEBS journal*, apr 2020.

[44] Zhihao Sun and Jeffrey L Brodsky. Protein quality control in the secretory pathway. *The Journal of cell biology*, 218(10):3171–3187, 2019.

[45] Pietro Roversi, Lucia Marti, Alessandro T. Caputo, Dominic S. Alonzi, Johan C. Hill, Kyle C. Dent, Abhinav Kumar, Mikail D. Levasseur, Andrea Lia, Thomas Waksman, Souradeep Basu, Yentli Soto Albrecht, Kristin Qian, James Patrick McIvor, Colette B. Lipp, Dritan Siliqi, Snežana Vasiljević, Shabaz Mohammed, Petra Lukacik, Martin A. Walsh, Angelo Santino, and Nicole Zitzmann. Interdomain conformational flexibility underpins the activity of uggt, the eukaryotic glycoprotein secretion checkpoint. *Proceedings of the National Academy of Sciences*, 114(32):8544–8549, 2017.

[46] Carlos P Modenutti, Juan I Blanco Capurro, Roberta Ibba, Dominic S Alonzi, Mauro N Song, Snežana Vasiljević, Abhinav Kumar, Anu V Chandran, Gábor Tax, Lucia Marti, Johan C Hill, Andrea Lia, Mario Hensen, Thomas Waksman, Jonathan Rushton, Simone Rubichi, Angelo Santino, Marcelo A Marti, Nicole Zitzmann, and Pietro Roversi. Clamping, bending, and twisting inter-domain motions in the misfold-recognizing portion of UDP-glucose:glycoprotein glucosyltransferase. *Structure*, December 2020.

[47] Gábor Tax, Andrea Lia, Angelo Santino, and Pietro Roversi. Modulation of ERQC and ERAD: A Broad-Spectrum Spanner in the Works of Cancer Cells? *Journal of oncology*, 2019:8384913, 2019.

[48] Benjamin M Adams, Nathan P Canniff, Kevin P Guay, Ida Signe Bohse Larsen, and Daniel N Hebert. Quantitative glycoproteomics reveals substrate selectivity of the ER protein quality control sensors UGGT1 and UGGT2. *eLife*, 9, December 2020.

[49] Shusaku Daikoku, Akira Seko, Yukishige Ito, and Osamu Kanie. Glycan structure and site of glycosylation in the ER-resident glycoprotein, uridine 5'-diphosphate-glucose: glycoprotein glucosyltransferases 1 from rat, porcine, bovine, and human. *Biochemical and biophysical research communications*, 451(3):356–360, aug 2014.

[50] Shusaku Daikoku, Akira Seko, Sang-Hyun Son, Katsuhiko Suzuki, Yukishige Ito, and Osamu Kanie. The relationship between glycan structures and expression levels of an endoplasmic reticulum-resident glycoprotein, UDP-glucose: Glycoprotein glucosyltransferase 1. *Biochemical and biophysical research communications*, 462(1):58–63, jun 2015.

[51] Diego A Caraballo, Lucila I Buzzi, Carlos P Modenutti, Ana Acosta-Montalvo, Olga A Castro, and María S Rossi. Origin and Evolution of Two Independently Duplicated Genes Encoding UDP- Glucose: Glycoprotein Glucosyltransferases in Caenorhabditis and Vertebrates. *G3 (Bethesda, Md.)*, 10(2):755–768, feb 2020.

[52] FS Fernández, SE Trombetta, U Hellman, and AJ Parodi. Purification to homogeneity of udp-glucose:glycoprotein glucosyltransferase from schizosaccharomyces pombe and apparent absence of the enzyme fro saccharomyces cerevisiae. *The Journal of biological chemistry*, 269(48):30701—30706, December 1994.

[53] S Fanchiotti, F Fernández, C D'Alessio, and A J Parodi. The UDP-Glc:Glycoprotein glucosyltransferase is essential for Schizosaccharomyces pombe viability under conditions of extreme endoplasmic reticulum stress. *The Journal of cell biology*, 143(3):625–635, nov 1998.

[54] Ana B. Herrero, Paula Magnelli, Michael K. Mansour, Stuart M. Levitz, Howard Bussey, and Claudia Abeijon. KRE5 gene null mutant strains of candida albicans are avirulent and have altered cell wall composition and hypha formation properties. *Eukaryotic Cell*, 3(6):1423–1432, December 2004.

[55] Maurizio Molinari, Carmela Galli, Omar Vanoni, Stacey M. Arnold, and Randal J. Kaufman. Persistent glycoprotein misfolding activates the glucosidase ii/ugt1-driven calnexin cycle to delay aggregation and loss of folding competence. *Molecular Cell*, 20(4):503 – 512, 2005.

[56] Peng-Nien Huang, Jia-Rong Jheng, Jamie J Arnold, Jen-Ren Wang, Craig E Cameron, and Shin-Ru Shih. UGGT1 enhances enterovirus 71 pathogenicity by promoting viral RNA synthesis and viral replication. *PLoS pathogens*, 13(5):e1006375, May 2017.

[57] Hua Jin, Zhenyan Yan, Kyoung Hee Nam, and Jianming Li. Allele-specific suppression of a defective brassinosteroid receptor reveals a physiological role of uggt in er quality control. *Molecular cell*, 26(6):821–830, Jun 2007.

[58] Jing Li, Chu Zhao-Hui, Martine Batoux, Vladimir Nekrasov, Milena Roux, Delphine Chinchilla, Cyril Zipfel, and Jonathan D. G. Jones. Specific er quality control components required for biogenesis of the plant innate immune receptor efr. *Proceedings of the National Academy of Sciences of the United States of America*, 106(37):15973–15978, Sep 2009.

[59] Yusuke Saijo, Nico Tintor, Xunli Lu, Philipp Rauf, Karolina Pajerowska-Mukhtar, Heidrun Häweker, Xinnian Dong, Silke Robatzek, and Paul Schulze-Lefert. Receptor quality control in the endoplasmic reticulum for plant innate immunity. *The EMBO journal*, 28(21):3439–3449, Nov 2009.

[60] Andrea Lia, Antonia Gallo, Lucia Marti, Pietro Roversi, and Angelo Santino. Efr-mediated innate immune response in arabidopsis thaliana is a useful tool for identification of novel erqc modulators. *Genes*, 10(1):15, Dec 2018.

[61] F Buttgereit and M D Brand. A hierarchy of ATP-consuming processes in mammalian cells. *Biochemical Journal*, 312(1):163–167, 11 1995.

[62] R D Klausner and R Sitia. Protein degradation in the endoplasmic reticulum. *Cell*, 62(4):611–614, aug 1990.

[63] Annemieke T. van der Goot, Margaret M.P. Pearce, Dara E. Leto, Thomas A. Shaler, and Ron R. Kopito. Redundant and antagonistic roles of XTP3b and OS9 in decoding glycan and non-glycan degrons in ER-associated degradation. *Molecular Cell*, 70(3):516–530.e6, May 2018.

[64] Silvia Olivari, Tito Cali, Kirsi E.H. Salo, Paolo Paganetti, Lloyd W. Ruddock, and Maurizio Molinari. EDEM1 regulates ER-associated degradation by accelerating de-mannosylation of folding-defective polypeptides and by inhibiting their covalent aggregation. *Biochemical and Biophysical Research Communications*, 349(4):1278–1284, November 2006.

[65] Erin M. Quan, Yukiko Kamiya, Daiki Kamiya, Vladimir Denic, Jimena Weibezahn, Koichi Kato, and Jonathan S. Weissman. Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. *Molecular cell*, 32(6):870–877, Dec 2008.

[66] Tadashi Satoh, Yang Chen, Dan Hu, Shinya Hanashima, Kazuo Yamamoto, and Yoshiki Yamaguchi. Structural basis for oligosaccharide recognition of misfolded glycoproteins by OS-9 in ER-associated degradation. *Molecular Cell*, 40(6):905–916, 2010.

[67] Bella Groisman, Marina Shenkman, Efrat Ron, and Gerardo Z Lederkremer. Mannose trimming is required for delivery of a glycoprotein from EDEM1 to XTP3-B and to late endoplasmic reticulum-associated degradation steps. *The Journal of biological chemistry*, 286(2):1292–1300, jan 2011.

[68] Sabine C Horn, Jennifer Hanna, Christian Hirsch, Corinna Volkwein, Anja Schütz, Udo Heinemann, Thomas Sommer, and Ernst Jarosch. Usa1 functions as a scaffold of the HRD-ubiquitin ligase. *Molecular cell*, 36(5):782–793, dec 2009.

[69] Xudong Wu, Marc Siggel, Sergey Ovchinnikov, Wei Mi, Vladimir Svetlov, Evgeny Nudler, Maofu Liao, Gerhard Hummer, and Tom A. Rapoport. Structural basis of er-associated protein degradation mediated by the hrd1 ubiquitin ligase complex. *Science*, 368(6489), 2020.

[70] Marioara B Marin, Simona Ghenea, Laurentiu N Spiridon, Gabriela N Chiritoiu, Andrei-Jose Petrescu, and Stefana-Maria Petrescu. Tyrosinase degradation is prevented when EDEM1 lacks the intrinsically disordered region. *PloS one*, 7(8):e42998, 2012.

[71] Marioara Chiritoiu, Gabriela N Chiritoiu, Cristian V A Munteanu, Florin Pastrama, N Erwin Ivessa, and Stefana M Petrescu. EDEM1 Drives Misfolded Protein Degradation via ERAD and Exploits ER-Phagy as Back-Up Mechanism When ERAD Is Impaired. *International journal of molecular sciences*, 21(10), may 2020.

[72] James H. Cormier, Taku Tamura, Johan C. Sunryd, and Daniel N. Hebert. EDEM1 recognition and delivery of misfolded proteins to the SEL1l-containing ERAD complex. *Molecular Cell*, 34(5):627–633, June 2009.

[73] Nobuko Hosokawa, Linda O Tremblay, Barry Sleno, Yukiko Kamiya, Ikuo Wada, Kazuhiro Nagata, Koichi Kato, and Annette Herscovics. EDEM1 accelerates the trimming of α1,2-linked mannose on the C branch of N-glycans. *Glycobiology*, 20(5):567–575, 01 2010.

[74] Marina Shenkman, Efrat Ron, Rivka Yehuda, Ron Benyair, Isam Khalaila, and Gerardo Z Lederkremer. Mannosidase activity of EDEM1 and EDEM2 depends on an unfolded state of their glycoprotein substrates. *Communications biology*, 1(1):172, 2018.

[75] Satoshi Ninagawa, Tetsuya Okada, Yoshiki Sumitomo, Yukiko Kamiya, Koichi Kato, Satoshi Horimoto, Tokiro Ishikawa, Shunichi Takeda, Tetsushi Sakuma, Takashi Yamamoto, and Kazutoshi Mori. Edem2 initiates mammalian glycoprotein erad by catalyzing the first mannose trimming step. *The Journal of cell biology*, 206(3):347–356, Aug 2014.

[76] Ginto George, Satoshi Ninagawa, Hirokazu Yagi, Taiki Saito, Tokiro Ishikawa, Tetsushi Sakuma, Takashi Yamamoto, Koshi Imami, Yasushi Ishihama, Koichi Kato, Tetsuya Okada, and Kazutoshi Mori. EDEM2 stably disulfide-bonded to TXNDC11 catalyzes the first mannose trimming step in mammalian glycoprotein ERAD. *eLife*, 9, feb 2020.

[77] Simone Clerc, Christian Hirsch, Daniela Maria Oggier, Paola Deprez, Claude Jakob, Thomas Sommer, and Markus Aebi. Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. *The Journal of cell biology*, 184(1):159–172, jan 2009.

[78] Yi-Chang Liu, Danica Galonić Fujimori, and Jonathan S. Weissman. Htm1p–pdi1p is a folding-sensitive mannosidase that marks n-glycoproteins for er-associated protein degradation. *Proceedings of the National Academy of Sciences*, 113(28):E4015–E4024, 2016.

[79] Robert Gauss, Kazue Kanehara, Pedro Carvalho, Davis T W Ng, and Markus Aebi. A complex of Pdi1p and the mannosidase Htm1p initiates clearance of unfolded glycoproteins from the endoplasmic reticulum. *Molecular Cell*, 42(6):782–793, jun 2011.

[80] Ryo Ushioda, Jun Hoseki, Kazutaka Araki, Gregor Jansen, David Y. Thomas, and Kazuhiro Nagata. Erdj5 is required as a disulfide reductase for degradation of misfolded proteins in the er. *Science*, 321(5888):569–572, 2008.

[81] Masatoshi Hagiwara, Ken-Ichi Maegawa, Mamoru Suzuki, Ryo Ushioda, Kazutaka Araki, Yushi Matsumoto, Jun Hoseki, Kazuhiro Nagata, and Kenji Inaba. Structural basis of an ERAD pathway mediated by the ER-resident protein disulfide reductase ERdj5. *Molecular Cell*, 41(4):432–444, feb 2011.

[82] Richard T. Timms, Sam A. Menzies, Iva A. Tchasovnikarova, Lea C. Christensen, James C. Williamson, Robin Antrobus, Gordon Dougan, Lars Ellgaard, and Paul J. Lehner. Genetic dissection of mammalian erad through comparative haploid and crispr forward genetic screens. *Nature Communications*, 7(1):11786, Jun 2016.

[83] Machiko Sakoh-Nakatogawa, Shuh-Ichi Nishikawa, and Toshiya Endo. Roles of protein-disulfide isomerase-mediated disulfide bond formation of yeast Mnl1p in endoplasmic

reticulum-associated degradation. *The Journal of biological chemistry*, 284(18):11815–11825, May 2009.

[84] Geng Tian, Song Xiang, Robert Noiva, William J. Lennarz, and Hermann Schindelin. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell*, 124(1):61–73, 2006.

[85] Ruoyu Xiao, Bonney Wilkinson, Anton Solovyov, Jakob R. Winther, Arne Holmgren, Johanna Lundström-Ljung, and Hiram F. Gilbert. The contributions of protein disulfide isomerase and its homologues to oxidative protein folding in the yeast endoplasmic reticulum. *Journal of Biological Chemistry*, 279(48):49780–49786, November 2004.

[86] Lisa Vincenz-Donnelly and Mark S. Hipp. The endoplasmic reticulum: A hub of protein quality control in health and disease. *Free Radical Biology and Medicine*, 108:383–393, July 2017.

[87] Jane F. Amara, Seng H. Cheng, and Alan E. Smith. Intracellular protein trafficking defects in human disease. *Trends in Cell Biology*, 2(5):145–149, May 1992.

[88] Zhi Hong, Hua Jin, Tzvi Tzfira, and Jianming Li. Multiple Mechanism–Mediated Retention of a Defective Brassinosteroid Receptor in the Endoplasmic Reticulum of Arabidopsis. *The Plant Cell*, 20(12):3418–3429, 12 2008.

[89] Z. Hong, H. Kajiura, W. Su, H. Jin, A. Kimura, K. Fujiyama, and J. Li. Evolutionarily conserved glycan signal to degrade aberrant brassinosteroid receptors in arabidopsis. *Proceedings of the National Academy of Sciences*, 109(28):11437–11442, June 2012.

[90] Vladimir Nekrasov, Jing Li, Martine Batoux, Milena Roux, Zhao-Hui Chu, Severine Lacombe, Alejandra Rougon, Pascal Bittel, Marta Kiss-Papp, Delphine Chinchilla, H Peter van Esse, Lucia Jorda, Benjamin Schwessinger, Valerie Nicaise, Bart P H J Thomma, Antonio Molina, Jonathan D G Jones, and Cyril Zipfel. Control of the pattern-recognition receptor EFR by an ER protein complex in plant immunity. *The EMBO Journal*, 28(21):3428–3438, September 2009.

[91] Jiazhao Huang, Han Yin, Peiqi Yin, Xia Jian, Siqi Song, Junwen Luan, and Leiliang Zhang. SR-BI Interactome Analysis Reveals a Proviral Role for UGGT1 in Hepatitis C Virus Entry. *Frontiers in microbiology*, 10:2043, 2019.

[92] Karin Wisskirchen, Julie Lucifora, Thomas Michler, and Ulrike Protzer. New pharmacological strategies to fight enveloped viruses. *Trends in Pharmacological Sciences*, 35(9):470–478, September 2014.

[93] Andrew C. Sayce, Dominic S. Alonzi, Sarah S. Killingbeck, Beatrice E. Tyrrell, Michelle L. Hill, Alessandro T. Caputo, Ren Iwaki, Kyoko Kinami, Daisuke Ide, J. L. Kiappes, P. Robert

Beatty, Atsushi Kato, Eva Harris, Raymond A. Dwek, Joanna L. Miller, and Nicole Zitzmann. Iminosugars inhibit dengue virus production via inhibition of ER alpha-glucosidases—not glycolipid processing enzymes. *PLOS Neglected Tropical Diseases*, 10(3):e0004524, March 2016.

[94] J. P. Martinez, F. Sasse, M. Brönstrup, J. Diez, and A. Meyerhans. Antiviral drug discovery: broad-spectrum drugs from nature. *Natural Product Reports*, 32(1):29–48, 2015.

[95] Alessandro T. Caputo, Dominic S. Alonzi, Lucia Marti, Ida-Barbara Reca, J. L. Kiappes, Weston B. Struwe, Alice Cross, Souradeep Basu, Edward D. Lowe, Benoit Darlot, Angelo Santino, Pietro Roversi, and Nicole Zitzmann. Structures of mammalian er α-glucosidase ii capture the binding modes of broad-spectrum iminosugar antivirals. *Proceedings of the National Academy of Sciences*, 113(32):E4630–E4638, 2016.

[96] Michel Pieren, Carmela Galli, Angela Denzel, and Maurizio Molinari. The use of calnexin and calreticulin by cellular and viral glycoproteins. *Journal of Biological Chemistry*, 280(31):28265–28271, August 2005.

[97] Veronica T. Chang, Max Crispin, A. Radu Aricescu, David J. Harvey, Joanne E. Nettleship, Janet A. Fennelly, Chao Yu, Kent S. Boles, Edward J. Evans, David I. Stuart, Raymond A. Dwek, E. Yvonne Jones, Raymond J. Owens, and Simon J. Davis. Glycoprotein structural genomics: Solving the glycosylation problem. *Structure*, 15(3):267–273, March 2007.

[98] Stuart T. Perry, Michael D. Buck, Emily M. Plummer, Raju A. Penmasta, Hitesh Batra, Eric J. Stavale, Kelly L. Warfield, Raymond A. Dwek, Terry D. Butters, Dominic S. Alonzi, Steven M. Lada, Kevin King, Brennan Klose, Urban Ramstedt, and Sujan Shresta. An iminosugar with potent inhibition of dengue virus infection in vivo. *Antiviral Research*, 98(1):35–43, April 2013.

[99] Eric J. Stavale, Hong Vu, Aruna Sampath, Urban Ramstedt, and Kelly L. Warfield. In vivo therapeutic protection against influenza a (h1n1) oseltamivir-sensitive and resistant viruses by the iminosugar UV-4. *PLOS ONE*, 10(3):e0121662, March 2015.

[100] Kelly Warfield, Emily Plummer, Dominic Alonzi, Gary Wolfe, Aruna Sampath, Tam Nguyen, Terry Butters, Sven Enterlein, Eric Stavale, Sujan Shresta, and Urban Ramstedt. A novel iminosugar UV-12 with activity against the diverse viruses influenza and dengue (novel iminosugar antiviral for influenza and dengue). *Viruses*, 7(5):2404–2427, May 2015.

[101] Jingjing Ruan, Hussin A. Rothan, Yongwang Zhong, Wenjing Yan, Mark J. Henderson, Feihu Chen, and Shengyun Fang. A small molecule inhibitor of ER-to-cytosol protein dislocation exhibits anti-dengue and anti-zika virus activity. *Scientific Reports*, 9(1), July 2019.

[102] Mohsan Saeed, Ryosuke Suzuki, Noriyuki Watanabe, Takahiro Masaki, Mitsunori Tomonaga, Amir Muhammad, Takanobu Kato, Yoshiharu Matsuura, Haruo Watanabe, Takaji

Wakita, and Tetsuro Suzuki. Role of the endoplasmic reticulum-associated degradation (ERAD) pathway in degradation of hepatitis c virus envelope proteins and production of virus particles∗. $Journal of Biological Chemistry$, 286(43) : 37264 − −37273, $October$2011.

[103] S Nisim, M Lachyankar, and title = Stem cells: Cross–talk and developmental programs journal = royalsocietypublishing.org year=2004' Royal, J Ourednik.

[104] M W Long. Stem cell plasticity: molding the future of tissue development and repair. In *Blood cells, molecules & diseases*, pages 586–589. Department of Pediatrics, Comprehensive Cancer Center, University of Michigan, Ann Arbor, MI 48109, USA. mwlong@umich.edu, May 2001.

[105] Oliver J Harrison, Jonathan L Linehan, Han-Yu Shih, Nicolas Bouladoux, Seong-Ji Han, Margery Smelkinson, Shurjo K Sen, Allyson L Byrd, Michel Enamorado, Chen Yao, Samira Tamoutounour, Francois Van Laethem, Charlotte Hurabielle, Nicholas Collins, Andrea Paun, Rosalba Salcedo, John J O'Shea, and Yasmine Belkaid. Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury. *Science*, 363(6422):eaat6280, January 2019.

[106] Sophie Dekoninck and Cedric Blanpain. Stem cell dynamics, migration and plasticity during wound healing. *Nature cell biology*, 21(1):18–24, January 2019.

[107] Piyush B Gupta, Ievgenia Pastushenko, Adam Skibinski, Cedric Blanpain, and Charlotte Kuperwasser. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell stem cell*, 24(1):65–78, January 2019.

[108] D Hanahan and R A Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, January 2000.

[109] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, March 2011.

[110] Paul Eggleton, Marek Michalak, and Edwin Bremer. Editorial: Endoplasmic Reticulum and Its Role in Tumor Immunity. *Frontiers in oncology*, 5(5):252, 2015.

[111] Norbert Volkmar, Emma Fenech, and John C Christianson. New MAPS for misfolded proteins. *Nature cell biology*, 18(7):724–726, June 2016.

[112] Yanfen Liu and Yihong Ye. Proteostasis regulation at the endoplasmic reticulum: a new perturbation site for targeted cancer therapy. *Cell research*, 21(6):867–883, June 2011.

[113] Multiple myeloma. *Nature Reviews Disease Primers*, 3(1), July 2017.

[114] J. Simoes-Correia, J. Figueiredo, C. Oliveira, J. van Hengel, R. Seruca, F. van Roy, and G. Suriano. Endoplasmic reticulum quality control: a new mechanism of e-cadherin regulation and its implication in cancer. *Human Molecular Genetics*, 17(22):3566–3576, August 2008.

[115] Michiko Sekiya, Akiko Maruko-Otake, Stephen Hearn, Yasufumi Sakakibara, Naoki Fujisaki, Emiko Suzuki, Kanae Ando, and Koichi M. Iijima. EDEM function in ERAD protects against chronic ER proteinopathy and age-related physiological decline in drosophila. *Developmental Cell*, 41(6):652–664.e5, June 2017.

[116] Jian-Xiang Liu, Renu Srivastava, Ping Che, and Stephen H. Howell. Salt stress responses in arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *The Plant Journal*, 51(5):897–909, July 2007.

[117] Maria Anete S. Valente, Jerusa A. Q. A. Faria, Juliana R. L. Soares-Ramos, Pedro A. B. Reis, Guilherme L. Pinheiro, Newton D. Piovesan, Angélica T. Morais, Carlos C. Menezes, Marco A. O. Cano, Luciano G. Fietto, Marcelo E. Loureiro, Francisco J. L. Aragão, and Elizabeth P. B. Fontes. The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *Journal of Experimental Botany*, 60(2):533–546, December 2008.

[118] X.-Y. Jia, C.-Y. Xu, R.-L. Jing, R.-Z. Li, X.-G. Mao, J.-P. Wang, and X.-P. Chang. Molecular cloning and characterization of wheat calreticulin (CRT) gene involved in drought-stressed responses. *Journal of Experimental Botany*, 59(4):739–751, March 2008.

[119] Anthony L. Fink. Chaperone-mediated protein folding. *Physiological Reviews*, 79(2):425–449, April 1999.

[120] Yukako Oda, Nobuko Hosokawa, Ikuo Wada, and Kazuhiro Nagata. Edem as an acceptor of terminally misfolded glycoproteins released from calnexin. *Science*, 299(5611):1394–1397, 2003.

[121] M. Molinari. Role of EDEM in the release of misfolded glycoproteins from the calnexin cycle. *Science*, 299(5611):1397–1400, February 2003.

[122] Anett Pfeiffer, Heike Stephanowitz, Eberhard Krause, Corinna Volkwein, Christian Hirsch, Ernst Jarosch, and Thomas Sommer. A Complex of Htm1 and the Oxidoreductase Pdi1 Accelerates Degradation of Misfolded Glycoproteins. *The Journal of biological chemistry*, 291(23):12195–12207, June 2016.

[123] Stefan Amlacher, Phillip Sarges, Dirk Flemming, Vera van Noort, Ruth Kunze, Damien P Devos, Manimozhiyan Arumugam, Peer Bork, and Ed Hurt. Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*, 146(2):277–289, July 2011.

[124] Thomas Bock, Wei-Hua Chen, Alessandro Ori, Nayab Malik, Noella Silva-Martin, Jaime Huerta-Cepas, Sean T Powell, Panagiotis L Kastritis, Georgy Smyshlyaev, Ivana Vonkova, Joanna Kirkpatrick, Tobias Doerks, Leo Nesme, Jochen Baßler, Martin Kos, Ed Hurt, Teresa Carlomagno, Anne-Claude Gavin, Orsolya Barabas, Christoph W Müller, Vera van Noort,

Martin Beck, and Peer Bork. An integrated approach for genome annotation of the eukaryotic thermophile Chaetomium thermophilum. *Nucleic acids research*, 42(22):13525–13533, November 2014.

[125] Tadashi Satoh, Takayasu Toshimori, Gengwei Yan, Takumi Yamaguchi, and Koichi Kato. Structural basis for two-step glucose trimming by glucosidase II involved in ER glycoprotein quality control. *Scientific Reports*, 6(1), February 2016.

[126] Silvia Hüttner, Christiane Veit, Ulrike Vavra, Jennifer Schoberer, Eva Liebminger, Daniel Maresch, Josephine Grass, Friedrich Altmann, Lukas Mach, and Richard Strasser. Arabidopsis class i -mannosidases MNS4 and MNS5 are involved in endoplasmic reticulum–associated degradation of misfolded glycoproteins. *The Plant Cell*, 26(4):1712–1728, April 2014.

[127] M C Sousa, M A Ferrero-Garcia, and A J Parodi. Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry*, 31(1):97–105, 1992.

[128] Ari Helenius and Christiane Ritter. *Nature Structural Biology*, 7(4):278–280, April 2000.

[129] Grégory Vert, Jennifer L. Nemhauser, Niko Geldner, Fangxin Hong, and Joanne Chory. MOLECULAR MECHANISMS OF STEROID HORMONE SIGNALING IN PLANTS. *Annual Review of Cell and Developmental Biology*, 21(1):177–201, November 2005.

[130] Mari-Anne Newman, Thomas Sundelin, Jon T. Nielsen, and Gitte Erbs. MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Frontiers in Plant Science*, 4, 2013.

[131] Gernot Kunze, Cyril Zipfel, Silke Robatzek, Karsten Niehaus, Thomas Boller, and Georg Felix. The n terminus of bacterial elongation factor tu elicits innate immunity in arabidopsis plants. *The Plant Cell*, 16(12):3496–3507, November 2004.

[132] Toshio Murashige and Folke Skoog. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, 15(3):473–497, July 1962.

[133] M. W. Pfaffl. Relative expression software tool (REST(c)) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research*, 30(9):36e–36, May 2002.

[134] J.S. Farrar and C.T. Wittwer. High-resolution melting curve analysis for molecular diagnostics. In *Molecular Diagnostics*, pages 79–102. Elsevier, 2017.

[135] Simone Ferrari, Roberta Galletti, Donatella Vairo, Felice Cervone, and Giulia De Lorenzo. Antisense expression of the arabidopsis thaliana AtPGIP1 gene reduces polygalacturonase-inhibiting protein accumulation and enhances susceptibility to botrytis cinerea. *Molecular Plant-Microbe Interactions®*, 19(8):931–936, August 2006.

[136] Damir Kalpić, Nikica Hlupić, and Miodrag Lovrić. Student's t-tests. In *International Encyclopedia of Statistical Science*, pages 1559–1563. Springer Berlin Heidelberg, 2011.

[137] Tae Kyun Kim. Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology*, 70(1):22, 2017.

[138] Steven J. Clough and Andrew F. Bent. Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. *The Plant Journal*, 16(6):735–743, December 1998.

[139] PD Evans, SN Cook, PD Riggs, and CJ Noren. Litmus: multipurpose cloning vectors with a novel system for bidirectional in vitro transcription. *BioTechniques*, 19(1):130—135, July 1995.

[140] A Radu Aricescu, Weixian Lu, and E Yvonne Jones. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 10):1243–1250, oct 2006.

[141] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, April 2009.

[142] Nicola Portolano, Peter J. Watson, Louise Fairall, Christopher J. Millard, Charles P. Milano, Yun Song, Shaun M. Cowley, and John W.R. Schwabe. Recombinant protein expression for structural biology in HEK 293f suspension cells: A novel and accessible approach. *Journal of Visualized Experiments*, (92), October 2014.

[143] Warren Strober. Trypan blue exclusion test of cell viability. *Current Protocols in Immunology*, 111(1), November 2015.

[144] Gerald R. Grimsley and C. Nick Pace. Spectrophotometric determination of protein concentration. *Current Protocols in Protein Science*, 33(1), August 2003.

[145] Fabrice Gorrec. The MORPHEUS protein crystallization screen. *Journal of applied crystallography*, 42(Pt 6):1035–1042, December 2009.

[146] Fabrice Gorrec. The MORPHEUS II protein crystallization screen. *Acta crystallographica. Section F, Structural biology communications*, 71(Pt 7):831–837, July 2015.

[147] D N Perkins, D J Pappin, D M Creasy, and J S Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, December 1999.

[148] Tristan Bepler, Kotaro Kelley, Alex J Noble, and Bonnie Berger. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nature communications*, 11(1):5208, October 2020.

[149] Sjors H W Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, 180(3):519–530, December 2012.

[150] Lukas Zimmermann, Andrew Stephens, Seung-Zin Nam, David Rau, Jonas Kübler, Marko Lozajic, Felix Gabler, Johannes Söding, Andrei N Lupas, and Vikram Alva. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, 430(15):2237–2243, July 2018.

[151] A Sali, L Potterton, F Yuan, H van Vlijmen, and M Karplus. Evaluation of comparative protein modeling by MODELLER. *Proteins*, 23(3):318–326, November 1995.

[152] Marie-Christine Klein, Katharina Zimmermann, Stefan Schorr, Martina Landini, Patrick A W Klemens, Jacqueline Altensell, Martin Jung, Elmar Krause, Duy Nguyen, Volkhard Helms, Jens Rettig, Claudia Fecher-Trost, Adolfo Cavalié, Markus Hoth, Ivan Bogeski, H Ekkehard Neuhaus, Richard Zimmermann, Sven Lang, and Ilka Haferkamp. AXER is an ATP/ADP exchanger in the membrane of the endoplasmic reticulum. *Nature communications*, 9(1):3489, August 2018.

[153] Stefan Schoebel, Wei Mi, Alexander Stein, Sergey Ovchinnikov, Ryan Pavlovicz, Frank DiMaio, David Baker, Melissa G Chambers, Huayou Su, Dongsheng Li, Tom A Rapoport, and Maofu Liao. Cryo-EM structure of the protein-conducting ERAD channel Hrd1 in complex with Hrd3. *Nature*, 548(7667):352–355, August 2017.

[154] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF chimera?a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[155] Tali Gidalevitz, Fred Stevens, and Yair Argon. Orchestration of secretory protein folding by er chaperones. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(11):2410–2424, 2013. Functional and structural diversity of the endoplasmic reticulum.

[156] Dominic S. Alonzi, Kathryn A. Scott, Raymond A. Dwek, and Nicole Zitzmann. Iminosugar antivirals: the therapeutic sweet spot. *Biochemical Society Transactions*, 45(2):571–582, April 2017.

[157] Terry D. Butters, Howard R. Mellor, Keishi Narita, Raymond A. Dwek, and Frances M. Platt. Small–molecule therapeutics for the treatment of glycolipid lysosomal storage disorders. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1433):927–945, May 2003.

[158] Daniele Vergara, Stefania de Domenico, Michele Maffia, Gabriella Piro, and Gian-Pietro Sansebastiano. Transgenic plants as low-cost platform for chemotherapeutic drugs screening. *International Journal of Molecular Sciences*, 16(1):2174–2186, January 2015.

[159] Xunli Lu, Nico Tintor, Tobias Mentzel, Erich Kombrink, Thomas Boller, Silke Robatzek, Paul Schulze-Lefert, and Yusuke Saijo. Uncoupling of sustained MAMP receptor signaling from early outputs in an arabidopsis endoplasmic reticulum glucosidase II allele. *Proceedings of the National Academy of Sciences*, 106(52):22522–22527, December 2009.

[160] Catalin Lazar, David Durantel, Alina Macovei, Nicole Zitzmann, Fabien Zoulim, Raymond A. Dwek, and Norica Branza-Nichita. Treatment of hepatitis b virus-infected cells with α-glucosidase inhibitors results in production of virions with altered molecular composition and infectivity. *Antiviral Research*, 76(1):30–37, 2007.

[161] Eike Steinmann, Thomas Whitfield, Stephanie Kallis, Raymond A. Dwek, Nicole Zitzmann, Thomas Pietschmann, and Ralf Bartenschlager. Antiviral effects of amantadine and iminosugar derivatives against hepatitis c virus. *Hepatology*, 46(2):330–338, 2007.

[162] Stephanie Pollock, Raymond A Dwek, Dennis R Burton, and Nicole Zitzmann. N-butyldeoxynojirimycin is a broadly effective anti-HIV therapy significantly enhanced by targeted liposome delivery. *AIDS*, 22(15):1961–1969, October 2008.

[163] Kelly L. Warfield, Emily M. Plummer, Andrew C. Sayce, Dominic S. Alonzi, William Tang, Beatrice E. Tyrrell, Michelle L. Hill, Alessandro T. Caputo, Sarah S. Killingbeck, P. Robert Beatty, Eva Harris, Ren Iwaki, Kyoko Kinami, Daisuke Ide, J.L. Kiappes, Atsushi Kato, Michael D. Buck, Kevin King, William Eddy, Mansoora Khaliq, Aruna Sampath, Anthony M. Treston, Raymond A. Dwek, Sven G. Enterlein, Joanna L. Miller, Nicole Zitzmann, Urban Ramstedt, and Sujan Shresta. Inhibition of endoplasmic reticulum glucosidases is required for in vitro and in vivo dengue antiviral activity by the iminosugar UV-4. *Antiviral Research*, 129:93–98, May 2016.

[164] Beatrice Ellen Tyrrell, Andrew Cameron Sayce, Kelly Lyn Warfield, Joanna Louise Miller, and Nicole Zitzmann. Iminosugars: Promising therapeutics for influenza infection. *Critical Reviews in Microbiology*, 43(5):521–545, December 2016.

[165] Graeme F. Wilkinson and Kevin Pritchard. In vitro screening for drug repositioning. *Journal of Biomolecular Screening*, 20(2):167–179, December 2014.

[166] Lionel Navarro, Cyril Zipfel, Owen Rowland, Ingo Keller, Silke Robatzek, Thomas Boller, and Jonathan D.G. Jones. The transcriptional innate immune response to flg22. interplay and overlap with avr gene-dependent defense responses and bacterial pathogenesis. *Plant Physiology*, 135(2):1113–1128, June 2004.

[167] Cyril Zipfel, Gernot Kunze, Delphine Chinchilla, Anne Caniard, Jonathan D.G. Jones, Thomas Boller, and Georg Felix. Perception of the bacterial pamp ef-tu by the receptor efr restricts agrobacterium-mediated transformation. *Cell*, 125(4):749–760, 2006.

[168] Lourdes Gómez-Gómez and Thomas Boller. FLS2. *Molecular Cell*, 5(6):1003–1011, June 2000.

[169] G.B. Karlsson, T.D. Butters, R.A. Dwek, and F.M. Platt. Effects of the imino sugar n-butyldeoxynojirimycin on the n-glycosylation of recombinant gp120. *Journal of Biological Chemistry*, 268(1):570–576, January 1993.

[170] Frances M. PLATT, Gunilla B. KARLSSON, and Gary S. JACOB. Modulation of cell-surface transferrin receptor by the imino sugar n-butyldeoxynojirimycin. *European Journal of Biochemistry*, 208(1):187–193, August 1992.

[171] Francisca Blanco-Herrera, Adrián A Moreno, Rodrigo Tapia, Francisca Reyes, Macarena Araya, Cecilia D'Alessio, Armando Parodi, and Ariel Orellana. The UDP-glucose: glycoprotein glucosyltransferase (UGGT), a key enzyme in ER quality control, plays a significant role in plant growth as well as biotic and abiotic stress in Arabidopsis thaliana. *BMC plant biology*, 15(1):127, may 2015.

[172] Pravina Soussillane, Cecilia D'Alessio, Thomas Paccalet, Anne-Catherine Fitchette, Armando J. Parodi, Richard Williamson, Carole Plasson, Loïc Faye, and Véronique Gomord. N-glycan trimming by glucosidase II is essential for arabidopsis development. *Glycoconjugate Journal*, 26(5):597–607, October 2008.

[173] Matteo Gravino, Daniel Valentin Savatin, Alberto Macone, and Giulia De Lorenzo. Ethylene production in botrytis cinerea-and oligogalacturonide-induced immunity requires calcium-dependent protein kinases. *The Plant Journal*, 84(6):1073–1086, November 2015.

[174] Yueh Cho and Kazue Kanehara. Endoplasmic reticulum stress response in arabidopsis roots. *Frontiers in Plant Science*, 8, March 2017.

[175] Ulrich Schubert, Luis C. Antón, James Gibbs, Christopher C. Norbury, Jonathan W. Yewdell, and Jack R. Bennink. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature*, 404(6779):770–774, April 2000.

[176] Inki Kim, Wenjie Xu, and John C. Reed. Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities. *Nature Reviews Drug Discovery*, 7(12):1013–1030, December 2008.

[177] Lijing Liu, Feng Cui, Qingliang Li, Bojiao Yin, Huawei Zhang, Baoying Lin, Yaorong Wu, Ran Xia, Sanyuan Tang, and Qi Xie. The endoplasmic reticulum-associated degradation is necessary for plant salt tolerance. *Cell Research*, 21(6):957–969, December 2010.

[178] W. Su, Y. Liu, Y. Xia, Z. Hong, and J. Li. Conserved endoplasmic reticulum-associated degradation system to eliminate mutated receptor-like kinases in arabidopsis. *Proceedings of the National Academy of Sciences*, 108(2):870–875, December 2010.

[179] Silvia Hüttner, Christiane Veit, Jennifer Schoberer, Josephine Grass, and Richard Strasser. Unraveling the function of arabidopsis thaliana OS9 in the endoplasmic reticulum-associated degradation of glycoproteins. *Plant Molecular Biology*, 79(1-2):21–33, February 2012.

[180] Eva Liebminger, Silvia Hüttner, Ulrike Vavra, Richard Fischl, Jennifer Schoberer, Josephine Grass, Claudia Blaukopf, Georg J. Seifert, Friedrich Altmann, Lukas Mach, and Richard Strasser. Class i -mannosidases are required for n-glycan processing and root development in arabidopsis thaliana. *The Plant Cell*, 21(12):3850–3867, December 2009.

[181] J.L. Risler, M.O. Delorme, H. Delacroix, and A. Henaut. Amino acid substitutions in structurally related proteins a pattern recognition approach: Determination of a new and efficient scoring matrix. *Journal of Molecular Biology*, 204(4):1019–1029, 1988.

[182] Xavier Robert and Patrice Gouet. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1):W320–W324, 04 2014.

[183] Schrödinger, LLC. The `PyMOL` molecular graphics system.

[184] D. J. DE ROSIER and A. KLUG. Reconstruction of three dimensional structures from electron micrographs. *Nature*, 217(5124):130–134, January 1968.

[185] Marta Artal-Sanz, Liesbeth de Jong, and Nektarios Tavernarakis. Caenorhabditis elegans: A versatile platform for drug discovery. *Biotechnology Journal*, 1(12):1405–1418, December 2006.

[186] Martin Dalziel, Max Crispin, Christopher N. Scanlan, Nicole Zitzmann, and Raymond A. Dwek. Emerging principles for the therapeutic exploitation of glycosylation. *Science*, 343(6166), 2014.

[187] Joanne E. Burn, Ursula A. Hurley, Rosemary J. Birch, Tony Arioli, Ann Cork, and Richard E. Williamson. The cellulose-deficient arabidopsis mutant rsw3 is defective in a gene encoding a putative glucosidase II, an enzyme processing n-glycans during ER quality control. *The Plant Journal*, 32(6):949–960, December 2002.

[188] Mayuriben Parmar, Shaun Rawson, Charlotte A Scarff, Adrian Goldman, Timothy R Dafforn, Stephen P Muench, and Vincent L G Postis. Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochimica et biophysica acta. Biomembranes*, 1860(2):378–383, February 2018.

[189] Ganesh P Subedi, Roy W Johnson, Heather A Moniz, Kelley W Moremen, and Adam Barb. High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. *Journal of visualized experiments : JoVE*, (106):e53568, December 2015.

[190] Svetlana V Antonyuk, Robert R Eady, Richard W Strange, and S Samar Hasnain. The structure of glyceraldehyde 3-phosphate dehydrogenase from Alcaligenes xylosoxidans at 1.7 A resolution. *Acta crystallographica. Section D, Biological crystallography*, 59(Pt 5):835–842, May 2003.

[191] Luigi Di Costanzo, German A Gomez, and David W Christianson. Crystal structure of lactaldehyde dehydrogenase from Escherichia coli and inferences regarding substrate and cofactor specificity. *Journal of molecular biology*, 366(2):481–493, February 2007.

[192] D M Needham and R K Pillai. The coupling of oxido-reductions and dismutations with esterification of phosphate in muscle. *The Biochemical journal*, 31(10):1837–1851, October 1937.

[193] William F Martin and Rüdiger Cerff. Physiology, phylogeny, early evolution, and GAPDH. *Protoplasma*, 254(5):1823–1834, September 2017.

[194] A Tarze, A Deniaud, M Le Bras, E Maillier, D Molle, N Larochette, N Zamzami, G Jan, G Kroemer, and C Brenner. GAPDH, a novel regulator of the pro-apoptotic mitochondrial membrane permeabilization. *Oncogene*, 26(18):2606–2620, April 2007.

[195] Diana Zala, Maria-Victoria Hinckelmann, Hua Yu, Marcel Menezes Lyra da Cunha, Géraldine Liot, Fabrice P Cordelières, Sergio Marco, and Frédéric Saudou. Vesicular glycolysis provides on-board energy for fast axonal transport. *Cell*, 152(3):479–491, January 2013.

[196] Giovanna Butera, Nidula Mullappilly, Francesca Masetto, Marta Palmieri, Maria Teresa Scupoli, Raffaella Pacchiana, and Massimo Donadelli. Regulation of Autophagy by Nuclear GAPDH and Its Aggregates in Cancer and Neurodegenerative Disorders. *International journal of molecular sciences*, 20(9), April 2019.

[197] Polina V Danshina, Weidong Qu, Brenda R Temple, Rafael J Rojas, Michael J Miley, Mischa Machius, Laurie Betts, and Deborah A O'Brien. Structural analyses to identify selective inhibitors of glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme. *Molecular human reproduction*, 22(6):410–426, June 2016.

[198] Otto Warburg and Walter Christian. *Isolierung und Kristallisation des Proteins des oxydierenden Gärungsferments*. Springer, 1939.

[199] Gerty T Cori, Milton W Slein, and Carl F Cori. Crystalline d-glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle. *J. biol. Chem*, 173:605–618, 1948.

[200] H C Watson, E Duée, and W D Mercer. Low resolution structure of glyceraldehyde 3-phosphate dehydrogenase. *Nature: New biology*, 240(100):130–133, November 1972.

[201] M G Rossmann, G C Ford, H C Watson, and L J Banaszak. Molecular symmetry of glyceraldehyde-3-phosphate dehydrogenase. *Journal of Molecular Biology*, 64(1):237–245, February 1972.

[202] Sandra W Cowan-Jacob, Markus Kaufmann, Anthony N Anselmo, Wilhelm Stark, and Markus G Grütter. Structure of rabbit-muscle glyceraldehyde-3-phosphate dehydrogenase. *Acta crystallographica. Section D, Biological crystallography*, 59(Pt 12):2218–2227, December 2003.

[203] Jan Frayne, Abby Taylor, Gus Cameron, and Andrea T Hadfield. Structure of insoluble rat sperm glyceraldehyde-3-phosphate dehydrogenase (GAPDH) via heterotetramer formation with Escherichia coli GAPDH reveals target for contraceptive design. *The Journal of biological chemistry*, 284(34):22703–22712, August 2009.

[204] Christophe Moreau, Rémi Terrasse, Nicole M Thielens, Thierry Vernet, Christine Gaboriaud, and Anne Marie Di Guilmi. Deciphering Key Residues Involved in the Virulence-promoting Interactions between Streptococcus pneumoniae and Human Plasminogen. *The Journal of biological chemistry*, 292(6):2217–2225, February 2017.

[205] Luis M de Pablos, Tiago R Ferreira, Adam A Dowle, Sarah Forrester, Ewan Parry, Katherine Newling, and Pegine B Walrad. The mRNA-bound Proteome of Leishmania mexicana: Novel Genetic Insight into an Ancient Parasite. *Molecular & cellular proteomics : MCP*, 18(7):1271–1284, July 2019.

[206] Ngoc Hieu Tran, Rui Qiao, Lei Xin, Xin Chen, Chuyi Liu, Xianglilan Zhang, Baozhen Shan, Ali Ghodsi, and Ming Li. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nature methods*, 16(1):63–66, January 2019.

[207] Marshall Bern, Yong J Kil, and Christopher Becker. Byonic: advanced peptide and protein identification software. *Current protocols in bioinformatics*, Chapter 13:Unit13.20, December 2012.

[208] Ravali Adusumilli and Parag Mallick. Data conversion with ProteoWizard msConvert. In *Methods in Molecular Biology*, pages 339–368. Springer New York, 2017.

[209] Harald Barsnes and Marc Vaudel. SearchGUI: A highly adaptable common interface for proteomics search and de novo engines. *Journal of Proteome Research*, 17(7):2552–2555, May 2018.

[210] Marc Vaudel, Julia M Burkhart, René P Zahedi, Eystein Oveland, Frode S Berven, Albert Sickmann, Lennart Martens, and Harald Barsnes. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature Biotechnology*, 33(1):22–24, January 2015.

[211] Clemens Vonrhein, Claus Flensburg, Peter Keller, Andrew Sharff, Oliver Smart, Wlodek Paciorek, Thomas Womack, and Gérard Bricogne. Data processing and analysis with the autoPROC toolbox. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4):293–302, April 2011.

[212] Alexei Vagin and Alexei Teplyakov. Molecular replacement with MOLREP. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 1):22–25, January 2010.

[213] E Blanc, P Roversi, C Vonrhein, C Flensburg, S M Lea, and G Bricogne. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1):2210–2221, December 2004.

[214] Gérard Bricogne, E Blanc, Brandl M, C Flensburg, Peter Keller, Wlodek Paciorek, Pietro Roversi, Andrew Sharff, Oliver S Smart, C Vonrhein, and Womack TO. BUSTER 2.10.3 . In *BUSTER 2.10.3*. January 2017.

[215] Oliver S Smart, Thomas O Womack, Claus Flensburg, Peter Keller, Włodek Paciorek, Andrew Sharff, Clemens Vonrhein, and Gérard Bricogne. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta crystallographica. Section D, Biological crystallography*, 68(Pt 4):368–380, April 2012.

[216] M Buehner, G C Ford, D Moras, K W Olsen, and M G Rossman. D-glyceraldehyde-3-phosphate dehydrogenase: three-dimensional structure and evolutionary significance. *Proceedings of the National Academy of Sciences*, 70(11):3052–3054, November 1973.

[217] M Buehner, G C Ford, K W Olsen, D Moras, and M G Rossman. Three-dimensional structure of D-glyceraldehyde-3-phosphate dehydrogenase. *Journal of Molecular Biology*, 90(1):25–49, November 1974.

[218] M Buehner, G C Ford, D Moras, K W Olsen, and M G Rossmann. Structure determination of crystalline lobster D-glyceraldehyde-3-phosphate dehydrogenase. *Journal of Molecular Biology*, 82(4):563–585, February 1974.

[219] Jermaine L Jenkins and John J Tanner. High-resolution structure of human D-glyceraldehyde-3-phosphate dehydrogenase. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 3):290–301, March 2006.

[220] S A Ismail and H W Park. Structural analysis of human liver glyceraldehyde-3-phosphate dehydrogenase. *Acta crystallographica. Section D, Biological crystallography*, 61(Pt 11):1508–1513, November 2005.

[221] W D Mercer, S I Winn, and H C Watson. Twinning in crystals of human skeletal muscle D-glyceraldehyde-3-phosphate dehydrogenase. *Journal of Molecular Biology*, 104(1):277–283, June 1976.

[222] Michael R White, Mohd M Khan, Daniel Deredge, Christina R Ross, Royston Quintyn, Beth E Zucconi, Vicki H Wysocki, Patrick L Wintrode, Gerald M Wilson, and Elsa D Garcin. A dimer interface mutation in glyceraldehyde-3-phosphate dehydrogenase regulates its binding to AU-rich RNA. *The Journal of biological chemistry*, 290(3):1770–1785, January 2015.

[223] Michael R White, Mohd M Khan, Daniel Deredge, Christina R Ross, Royston Quintyn, Beth E Zucconi, Vicki H Wysocki, Patrick L Wintrode, Gerald M Wilson, and Elsa D Garcin. A dimer interface mutation in glyceraldehyde 3-phosphate dehydrogenase regulates its binding to AU-rich RNA. *The Journal of biological chemistry*, 290(7):4129, February 2015.

[224] Jun Bae Park, Hayeong Park, Jimin Son, Sang-Jun Ha, and Hyun-Soo Cho. Structural Study of Monomethyl Fumarate-Bound Human GAPDH. *Molecules and cells*, 42(8):597–603, August 2019.

[225] Apirat Chaikuad, Naeem Shafqat, Ruby Al-Mokhtar, Gus Cameron, Anthony R Clarke, R Leo Brady, Udo Oppermann, Jan Frayne, and Wyatt W Yue. Structure and kinetic characterization of human sperm-specific glyceraldehyde-3-phosphate dehydrogenase, GAPDS. *Biochemical Journal*, 435(2):401–409, April 2011.

[226] Chia-Hsien Feng and Chi-Yu Lu. Modification of major plasma proteins by acrylamide and glycidamide: Preliminary screening by nano liquid chromatography with tandem mass spectrometry. *Analytica chimica acta*, 684(1-2):80–86, January 2011.

[227] Carlos Tristan, Neelam Shahani, Thomas W Sedlak, and Akira Sawa. The diverse functions of GAPDH: views from different subcellular compartments. *Cellular signalling*, 23(2):317–323, February 2011.

[228] Peter H Frederikse, Anoop Nandanoor, and Chinnaswamy Kasinathan. "Moonlighting" GAPDH Protein Localizes with AMPA Receptor GluA2 and L1 Axonal Cell Adhesion Molecule at Fiber Cell Borders in the Lens. *Current eye research*, 41(1):41–49, 2016.

[229] Toshiaki Nakano, Shigeru Goto, Yuki Takaoka, Hui-Peng Tseng, Takashi Fujimura, Seiji Kawamoto, Kazuhisa Ono, and Chao-Long Chen. A novel moonlight function of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for immunomodulation. *BioFactors (Oxford, England)*, 44(6):597–608, November 2018.

[230] Himanshu Malhotra, Manoj Kumar, Anoop Singh Chauhan, Asmita Dhiman, Surbhi Chaudhary, Anil Patidar, Pallavi Jaiswal, Kapil Sharma, Navdeep Sheokand, Chaaya Iyengar Raje, and Manoj Raje. Moonlighting Protein Glyceraldehyde-3-Phosphate Dehydrogenase: A Cellular Rapid-Response Molecule for Maintenance of Iron Homeostasis in Hypoxia. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 52(3):517–531, 2019.

[231] Diego Muñoz-Provencio, Gaspar Pérez-Martínez, and Vicente Monedero. Identification of Surface Proteins from Lactobacillus casei BL23 Able to Bind Fibronectin and Collagen. *Probiotics and antimicrobial proteins*, 3(1):15–20, March 2011.

[232] Philippe Vanden Bergh, Manfred Heller, Sophie Braga-Lagache, and Joachim Frey. The Aeromonas salmonicida subsp. salmonicida exoproteome: global analysis, moonlighting pro-

teins and putative antigens for vaccination against furunculosis. *Proteome science*, 11(1):44, October 2013.

[233] Revathi Nagarajan and Karthe Ponnuraj. Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of glyceraldehyde-3-phosphate dehydrogenase from Streptococcus agalactiae NEM316. *Acta crystallographica. Section F, Structural biology communications*, 70(Pt 7):938–941, July 2014.

[234] Rebeca Martín, Borja Sánchez, Maria C Urdaci, Philippe Langella, Juan E Suárez, and Luis G Bermúdez-Humarán. Effect of iron on the probiotic properties of the vaginal isolate Lactobacillus jensenii CECT 4306. *Microbiology (Reading, England)*, 161(Pt 4):708–718, April 2015.

[235] H Borgdorff, R Gautam, S D Armstrong, D Xia, G F Ndayisaba, N H van Teijlingen, T B H Geijtenbeek, J M Wastling, and J H H M van de Wijgert. Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. *Mucosal immunology*, 9(3):621–633, May 2016.

[236] Esther Frohnmeyer, Paulina Deptula, Tuula A Nyman, Pia K S Laine, Helena Vihinen, Lars Paulin, Petri Auvinen, Eija Jokitalo, Vieno Piironen, Pekka Varmanen, and Kirsi Savijoki. Secretome profiling of Propionibacterium freudenreichii reveals highly variable responses even among the closely related strains. *Microbial biotechnology*, 11(3):510–526, May 2018.

[237] Juliane Grimmer and Roger Dumke. Organization of multi-binding to host proteins: The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of Mycoplasma pneumoniae. *Microbiological research*, 218:22–31, January 2019.

[238] Nicolas Ugarte, Romain Ladouce, Sabrina Radjei, Monique Gareil, Bertrand Friguet, and Isabelle Petropoulos. Proteome alteration in oxidative stress-sensitive methionine sulfoxide reductase-silenced HEK293 cells. *Free radical biology & medicine*, 65:1023–1036, December 2013.

[239] Phillip A Wages, Katelyn S Lavrich, Zhenfa Zhang, Wan-Yun Cheng, Elizabeth Corteselli, Avram Gold, Philip Bromberg, Steven O Simmons, and James M Samet. Protein Sulfenylation: A Novel Readout of Environmental Oxidant Stress. *Chemical research in toxicology*, 28(12):2411–2418, December 2015.

[240] Vinayak Gupta and Kate S Carroll. Sulfenic acid chemistry, detection and cellular lifetime. *Biochimica et biophysica acta*, 1840(2):847–875, February 2014.

[241] R Ehring and S P Colowick. The two-step formation and inactivation of acylphosphatase by agents acting on glyceraldehyde phosphate dehydrogenase. *Journal of Biological Chemistry*, 244(17):4589–4599, September 1969.

[242] W S Allison and M J Connors. The activation and inactivation of the acyl phosphatase activity of glyceraldehyde-3-phosphate dehydrogenase. *Archives of biochemistry and biophysics*, 136(2):383–391, February 1970.

[243] W S Allison and L V Benitez. An adenosine triphosphate-phosphate exchange catalyzed by a soluble enzyme couple inhibited by uncouplers of oxidative phosphorylation. *Proceedings of the National Academy of Sciences*, 69(10):3004–3008, October 1972.

[244] Francesca Zappa, Mario Failli, and Maria Antonietta De Matteis. The Golgi complex in disease and therapy. *Current Opinion in Cell Biology*, 50:102–116, February 2018.

[245] Lydia Lamriben, Jill B Graham, Benjamin M Adams, and Daniel N Hebert. N-Glycan-based ER Molecular Chaperone and Protein Quality Control System: The Calnexin Binding Cycle. *Traffic (Copenhagen, Denmark)*, 17(4):308–326, April 2016.

[246] Daniel N Hebert, Lydia Lamriben, Evan T Powers, and Jeffery W Kelly. The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nature Chemical Biology*, 10(11):902–910, October 2014.

[247] Hanna J Clarke, Joseph E Chambers, Elizabeth Liniker, and Stefan J Marciniak. Endoplasmic reticulum stress in malignancy. *Cancer cell*, 25(5):563–573, May 2014.

[248] Hana Kim, Asmita Bhattacharya, and Ling Qi. Endoplasmic reticulum quality control in cancer: Friend or foe. *Seminars in cancer biology*, 33:25–33, August 2015.

[249] Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, and Fredrik Pontén. Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220):1260419–1260419, January 2015.

[250] Khalid El Bairi, Mariam Amrani, and Said Afqir. The Human Cancer Pathology Atlas: An open-access source for more than 900,000 Kaplan-Meier plots and 5 million cancer tissue images. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 96:1438–1439, December 2017.

[251] Philip Sedgwick. How to read a Kaplan-Meier survival plot. *BMJ (Clinical research ed.)*, 349(sep12 4):g5608–g5608, September 2014.

[252] Simon A Forbes, David Beare, Harry Boutselakis, Sally Bamford, Nidhi Bindal, John Tate, Charlotte G Cole, Sari Ward, Elisabeth Dawson, Laura Ponting, Raymund Stefancsik, Bhavana Harsha, Chai Yin Kok, Mingming Jia, Harry Jubb, Zbyslaw Sondka, Sam Thompson, Tisham De, and Peter J Campbell. COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research*, 45(D1):D777–D783, January 2017.

[253] Bülent Arman Aksoy, Emek Demir, Özgün Babur, Weiqing Wang, Xiaohong Jing, Nikolaus Schultz, and Chris Sander. Prediction of individualized therapeutic vulnerabilities in cancer from genomic profiles. *Bioinformatics (Oxford, England)*, 30(14):2051–2059, July 2014.

[254] Maureen E Higgins, Martine Claremont, John E Major, Chris Sander, and Alex E Lash. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic acids research*, 35(Database issue):D721–6, January 2007.

[255] Teresa Davoli, Andrew Wei Xu, Kristen E Mengwasser, Laura M Sack, John C Yoon, Peter J Park, and Stephen J Elledge. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*, 155(4):948–962, November 2013.

[256] Pora Kim, Feixiong Cheng, Junfei Zhao, and Zhongming Zhao. ccmGDB: a database for cancer cell metabolism genes. *Nucleic acids research*, 44(D1):D959–68, January 2016.

[257] Mohammed A Sadat, Susan Moir, Tae-Wook Chun, Paolo Lusso, Gerardo Kaplan, Lynne Wolfe, Matthew J Memoli, Miao He, Hugo Vega, Leo J Y Kim, Yan Huang, Nadia Hussein, Elma Nievas, Raquel Mitchell, Mary Garofalo, Aaron Louie, Derek C Ireland, Claire Grunes, Raffaello Cimbro, Vyomesh Patel, Genevieve Holzapfel, Daniel Salahuddin, Tyler Bristol, David Adams, Beatriz E Marciano, Madhuri Hegde, Yuxing Li, Katherine R Calvo, Jennifer Stoddard, J Shawn Justement, Jerome Jacques, Debra A Long Priel, Danielle Murray, Peter Sun, Douglas B Kuhns, Cornelius F Boerkoel, John A Chiorini, Giovanni Di Pasquale, Daniela Verthelyi, and Sergio D Rosenzweig. Glycosylation, hypogammaglobulinemia, and resistance to viral infections. *The New England journal of medicine*, 370(17):1615–1625, April 2014.

[258] Stacey M Arnold and Randal J Kaufman. The noncatalytic portion of human UDP-glucose: glycoprotein glucosyltransferase I confers UDP-glucose binding and transferase function to the catalytic domain. *Journal of Biological Chemistry*, 278(44):43320–43328, oct 2003.

[259] Yoichi Takeda, Akira Seko, Masakazu Hachisu, Shusaku Daikoku, Masayuki Izumi, Akihiko Koizumi, Kohki Fujikawa, Yasuhiro Kajihara, and Yukishige Ito. Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology*, 24(4):344–350, 2014.

[260] Ching-Chi Chiu, Chien-Yu Lin, Li-Yu Lee, Yin-Ju Chen, Ya-Ching Lu, Hung-Ming Wang, Chun-Ta Liao, Joseph Tung-Chieh Chang, and Ann-Joy Cheng. Molecular chaperones as a common set of proteins that regulate the invasion phenotype of head and neck cancer. *Clinical Cancer Research*, 17(14):4629–4641, June 2011.

[261] Hayley M Dingerdissen, John Torcivia-Rodriguez, Yu Hu, Ting-Chia Chang, Raja Mazumder, and Robel Kahsay. BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. *Nucleic Acids Research*, 46(D1):D1128–D1136, October 2017.

[262] Easwari Kumaraswamy, Andrey Malykh, Konstantin V. Korotkov, Sergei Kozyavkin, Yajun Hu, So Yeon Kwon, Mohamed E. Moustafa, Bradley A. Carlson, Marla J. Berry, Byeong J. Lee, Dolph L. Hatfield, Alan M. Diamond, and Vadim N. Gladyshev. Structure-expression relationships of the 15-kDa selenoprotein gene. *Journal of Biological Chemistry*, 275(45):35540–35547, November 2000.

[263] Sinoula Apostolou, Julian O Klein, Yasuhiro Mitsuuchi, Justin N Shetler, Poulikos I Poulikakos, Suresh C Jhanwar, Warren D Kruger, and Joseph R Testa. Growth inhibition and induction of apoptosis in mesothelioma cells by selenium and dependence on selenoprotein SEP15 genotype. *Oncogene*, 23(29):5032–5040, April 2004.

[264] Bingyu Ren, Min Liu, Jiazuan Ni, and Jing Tian. Role of selenoprotein f in protein folding and secretion: Potential involvement in human disease. *Nutrients*, 10(11):1619, November 2018.

[265] Sun Hee Yim, Robert A. Everley, Frank A. Schildberg, Sang-Goo Lee, Andrea Orsi, Zachary R. Barbati, Kutay Karatepe, Dmitry E. Fomenko, Petra A. Tsuji, Hongbo R. Luo, Steven P. Gygi, Roberto Sitia, Arlene H. Sharpe, Dolph L. Hatfield, and Vadim N. Gladyshev. Role of selenof as a gatekeeper of secreted disulfide-rich glycoproteins. *Cell Reports*, 23(5):1387–1398, May 2018.

[266] Vyacheslav M. Labunskyy, Andrew D. Ferguson, Dmitri E. Fomenko, Yogarany Chelliah, Dolph L. Hatfield, and Vadim N. Gladyshev. A novel cysteine-rich domain of sep15 mediates the interaction with UDP-glucose:glycoprotein glucosyltransferase. *Journal of Biological Chemistry*, 280(45):37839–37845, August 2005.

[267] Andrew D. Ferguson, Vyacheslav M. Labunskyy, Dmitri E. Fomenko, Demet Araç, Yogarany Chelliah, Carlos A. Amezcua, Josep Rizo, Vadim N. Gladyshev, and Johann Deisenhofer. NMR structures of the selenoproteins sep15 and SelM reveal redox activity of a new thioredoxin-like family∗. $Journal of Biological Chemistry, 281(6) : 3536 - -3543, February 2006$.

[268] Jiantang Zhu, Guangling Wang, Cuiling Li, Qingqing Li, Yankun Gao, Fanguo Chen, and Guangmin Xia. Maize sep15 -like functions in endoplasmic reticulum and reactive oxygen species homeostasis to promote salt and osmotic stress resistance. *Plant, Cell & Environment*, 42(5):1486–1502, February 2019.

[269] Yukishige Ito, Yoichi Takeda, Akira Seko, Masayuki Izumi, and Yasuhiro Kajihara. Functional analysis of endoplasmic reticulum glucosyltransferase (UGGT): Synthetic chemistry's initiative in glycobiology. *Seminars in Cell & Developmental Biology*, 41:90–98, May 2015.

[270] Robert Irons, Petra A. Tsuji, Bradley A. Carlson, Ping Ouyang, Min-Hyuk Yoo, Xue-Ming Xu, Dolph L. Hatfield, Vadim N. Gladyshev, and Cindy D. Davis. Deficiency in the 15-kDa selenoprotein inhibits tumorigenicity and metastasis of colon cancer cells. *Cancer Prevention Research*, 3(5):630–639, April 2010.

[271] Xiuwen Lan, Jun Xing, Hongyu Gao, Sen Li, Lina Quan, Yang Jiang, Shaohua Ding, and Yingwei Xue. Decreased expression of selenoproteins as a poor prognosticator of gastric cancer in humans. *Biological Trace Element Research*, 178(1):22–28, December 2016.

[272] Sarah P. Short and Christopher S. Williams. Selenoproteins in tumorigenesis and cancer progression. In *Advances in Cancer Research*, pages 49–83. Elsevier, 2017.

[273] Jeyoung Bang, Jang Hoe Huh, Ji-Woon Na, Qiao Lu, Bradley A. Carlson, Ryuta Tobe, Petra A. Tsuji, Vadim N. Gladyshev, Dolph L. Hatfield, and Byeong Jae Lee. Cell proliferation and motility are inhibited by g1 phase arrest in 15-kDa selenoprotein-deficient chang liver cells. *Molecules and Cells*, 38(5):457–465, February 2015.

[274] Ewa Jablonska, Jolanta Gromadzinska, Wojciech Sobala, Edyta Reszka, and Wojciech Wasowicz. Lung cancer risk associated with selenium status is modified in smoking individuals by sep15 polymorphism. *European Journal of Nutrition*, 47(1):47–54, January 2008.

[275] Rafał Watrowski, Dan Cacsire Castillo-Tong, Gerhild Fabjani, Eva Schuster, Michael Fischer, and Robert Zeillinger. The 811 c/t polymorphism in the 3′ untranslated region of the selenoprotein 15-kDa (sep15) gene and breast cancer in caucasian women. *Tumor Biology*, 37(1):1009–1015, August 2015.

[276] E.S. Trombetta. Glycoprotein reglucosylation and nucleotide sugar utilization in the secretory pathway: identification of a nucleoside diphosphatase in the endoplasmic reticulum. *The EMBO Journal*, 18(12):3282–3292, June 1999.

[277] Barbara Hadley, Andrea Maggioni, Angel Ashikov, Christopher J. Day, Thomas Haselhorst, and Joe Tiralongo. Structure and function of nucleotide sugar transporters: Current progress. *Computational and Structural Biotechnology Journal*, 10(16):23–32, June 2014.

[278] Julio J. Mulero, George Yeung, Sarah T. Nelken, and John E. Ford. CD39-l4 is a secreted human apyrase, specific for the hydrolysis of nucleoside diphosphates. *Journal of Biological Chemistry*, 274(29):20064–20067, July 1999.

[279] Min Fang, Zhirong Shen, Song Huang, Liping Zhao, She Chen, Tak W. Mak, and Xiaodong Wang. The ER UDPase ENTPD5 promotes protein n-glycosylation, the warburg effect, and proliferation in the PTEN pathway. *Cell*, 143(5):711–724, November 2010.

[280] J. Paez, Juan Recio, Ana Rouzaut, and Vicente Notario. Identity between the PCPH proto-oncogene and the CD39l4 (ENTPD5) ectonucleoside triphosphate diphosphohydrolase gene. *International Journal of Oncology*, December 2001.

[281] Joaquín Villar, María Isabel Arenas, Caitlin M. MacCarthy, María José Blánquez, Oscar M. Tirado, and Vicente Notario. PCPH/ENTPD5 expression enhances the invasiveness of human prostate cancer cells by a protein kinase c–dependent mechanism. *Cancer Research*, 67(22):10859–10868, November 2007.

[282] Joaquín Villar, Humair S. Quadri, Insun Song, York Tomita, Oscar M. Tirado, and Vicente Notario. PCPH/ENTPD5 expression confers to prostate cancer cells resistance against cisplatin-induced apoptosis through protein kinase c–mediated bcl-2 stabilization. *Cancer Research*, 69(1):102–110, December 2008.

[283] Masahiro Shuda, Nobuo Kondoh, Nobuo Imazeki, Kenji Tanaka, Tetsuya Okada, Kazutoshi Mori, Akiyuki Hada, Masaaki Arai, Toru Wakatsuki, Osamu Matsubara, Naoki Yamamoto, and Mikio Yamamoto. Activation of the ATF6, XBP1 and grp78 genes in human hepatocellular carcinoma: a possible involvement of the ER stress pathway in hepatocarcinogenesis. *Journal of Hepatology*, 38(5):605–614, May 2003.

[284] Yoon Seon Yoo, Hye Gyeong Han, and Young Joo Jeon. Unfolded protein response of the endoplasmic reticulum in tumor progression and immunogenicity. *Oxidative Medicine and Cellular Longevity*, 2017:1–18, December 2017.

[285] Y. C. Tsai and A. M. Weissman. The unfolded protein response, degradation from the endoplasmic reticulum, and cancer. *Genes & Cancer*, 1(7):764–778, July 2010.

[286] Xi Chen, Dimitrios Iliopoulos, Qing Zhang, Qianzi Tang, Matthew B. Greenblatt, Maria Hatziapostolou, Elgene Lim, Wai Leong Tam, Min Ni, Yiwen Chen, Junhua Mai, Haifa Shen, Dorothy Z. Hu, Stanley Adoro, Bella Hu, Minkyung Song, Chen Tan, Melissa D. Landis, Mauro Ferrari, Sandra J. Shin, Myles Brown, Jenny C. Chang, X. Shirley Liu, and Laurie H. Glimcher. XBP1 promotes triple-negative breast cancer by controlling the HIF1 pathway. *Nature*, 508(7494):103–107, March 2014.

[287] Amanda Croft, Kwang H. Tay, Suzanah C. Boyd, Su T. Guo, Chen C. Jiang, Fritz Lai, Hsin-Yi Tseng, Lei Jin, Helen Rizos, Peter Hersey, and Xu D. Zhang. Oncogenic activation of MEK/ERK primes melanoma cells for adaptation to endoplasmic reticulum stress. *Journal of Investigative Dermatology*, 134(2):488–497, February 2014.

[288] A.-H. Lee, N. N. Iwakoshi, K. C. Anderson, and L. H. Glimcher. Proteasome inhibitors disrupt the unfolded protein response in myeloma cells. *Proceedings of the National Academy of Sciences*, 100(17):9946–9951, August 2003.

[289] Rozita Bagheri-Yarmand, Krishna M. Sinha, Ling Li, Yue Lu, Gilbert J. Cote, Steven I. Sherman, and Robert F. Gagel. Combinations of tyrosine kinase inhibitor and ERAD inhibitor promote oxidative stress–induced apoptosis through ATF4 and KLF9 in medullary thyroid cancer. *Molecular Cancer Research*, 17(3):751–760, December 2018.

[290] A.D. Elbein, J.E. Tropea, M. Mitchell, and G.P. Kaushal. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase i. *Journal of Biological Chemistry*, 265(26):15599–15605, September 1990.

[291] Karen Legler, Ricarda Rosprim, Tosca Karius, Kathrin Eylmann, Maila Rossberg, Ralph M Wirtz, Volkmar Müller, Isabell Witzel, Barbara Schmalfeldt, Karin Milde-Langosch, and Leticia Oliveira-Ferrer. Reduced mannosidase MAN1a1 expression leads to aberrant n-glycosylation and impaired survival in breast cancer. *British Journal of Cancer*, 118(6):847–856, January 2018.

[292] Yi Lu, Ying-Ying Xu, Kai-Yi Fan, and Zong-Hou Shen. 1-deoxymannojirimycin, the 1, 2-mannosidase inhibitor, induced cellular endoplasmic reticulum stress in human hepatocarcinoma cell 7721. *Biochemical and Biophysical Research Communications*, 344(1):221–225, May 2006.

[293] Shujuan Pan, Xiaoyun Cheng, Hongan Chen, Patricia D. Castro, Michael M. Ittmann, Anne W. Hutson, Susan K. Zapata, and Richard N. Sifers. ERManI is a target of miR-125b and promotes transformation phenotypes in hepatocellular carcinoma (HCC). *PLoS ONE*, 8(8):e72829, August 2013.

[294] Mannosidase ia is in quality control vesicles and participates in glycoprotein targeting to erad. *Journal of Molecular Biology*, 428(16):3194–3205, 2016. Glycosylation in Cellular Processes and Disease.

[295] Tianhua Liu, Shu Zhang, Jie Chen, Kai Jiang, Qinle Zhang, Kun Guo, and Yinkun Liu. The transcriptional profiling of glycogenes associated with hepatocellular carcinoma metastasis. *PLoS ONE*, 9(9):e107941, September 2014.

[296] Alexandra Papaioannou, Arisa Higa, Gwénaële Jégou, Florence Jouan, Raphael Pineau, Laure Saas, Tony Avril, Olivier Pluquet, and Eric Chevet. Alterations of EDEM 1 functions enhance ATF 6 pro-survival signaling. *The FEBS Journal*, 285(22):4146–4164, October 2018.

[297] Pekka Määttänen, Kalle Gehring, John J.M. Bergeron, and David Y. Thomas. Protein quality control in the ER: The recognition of misfolded proteins. *Seminars in Cell & Developmental Biology*, 21(5):500–511, July 2010.

[298] Lydia Lamriben, Michela E. Oster, Taku Tamura, Weihua Tian, Zhang Yang, Henrik Clausen, and Daniel N. Hebert. EDEM1's mannosidase-like domain binds ERAD client proteins in a redox-sensitive manner and possesses catalytic activity. *Journal of Biological Chemistry*, 293(36):13932–13945, September 2018.

[299] M Corazzari, P E Lovat, J L Armstrong, G M Fimia, D S Hill, M Birch-Machin, C P F Redfern, and M Piacentini. Targeting homeostatic mechanisms of endoplasmic reticulum

stress to increase susceptibility of cancer cells to fenretinide-induced apoptosis: the role of stress proteins ERdj5 and ERp57. *British Journal of Cancer*, 96(7):1062–1071, March 2007.

[300] Christophoros G. Thomas and Giannis Spyrou. ERdj5 sensitizes neuroblastoma cells to endoplasmic reticulum stress-induced apoptosis. *Journal of Biological Chemistry*, 284(10):6282–6290, March 2009.

[301] Sang Yeon Cho, Sungha Kim, Mi-Ju Son, Woo Sun Rou, Seok Hyun Kim, Hyuk Soo Eun, and Byung Seok Lee. Clinical significance of the thioredoxin system and thioredoxin-domain-containing protein family in hepatocellular carcinoma. *Digestive Diseases and Sciences*, 64(1):123–136, October 2018.

[302] John C. Christianson, Thomas A. Shaler, Ryan E. Tyler, and Ron R. Kopito. OS-9 and GRP94 deliver mutant 1-antitrypsin to the hrd1–SEL1l ubiquitin ligase complex for ERAD. *Nature Cell Biology*, 10(3):272–282, February 2008.

[303] Marjolein Kikkert, Ram Doolman, Min Dai, Rachel Avner, Gerco Hassink, Sjaak van Voorden, Swapna Thanedar, Joseph Roitelman, Vincent Chau, and Emmanuel Wiertz. Human HRD1 is an e3 ubiquitin ligase involved in degradation of proteins from the endoplasmic reticulum. *Journal of Biological Chemistry*, 279(5):3525–3534, January 2004.

[304] Pedro Carvalho, Veit Goder, and Tom A. Rapoport. Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. *Cell*, 126(2):361–373, July 2006.

[305] Y. Kimura, M. Nakazawa, and M. Yamada. Cloning and characterization of three isoforms of OS-9 cDNA and expression of the OS-9 gene in various human tumor cell lines. *Journal of Biochemistry*, 123(5):876–882, May 1998.

[306] Kiyoshi Yanagisawa, Hiroyuki Konishi, Chinatsu Arima, Shuta Tomida, Toshiyuki Takeuchi, Yukako Shimada, Yasushi Yatabe, Tetsuya Mitsudomi, Hirotaka Osada, and Takashi Takahashi. Novel metastasis-related gene CIM functions in the regulation of multiple cellular stress–response pathways. *Cancer Research*, 70(23):9949–9958, November 2010.

[307] Y-W Sun, Y-F Chen, J Li, Y-M Huo, D-J Liu, R Hua, J-F Zhang, W Liu, J-Y Yang, X-L Fu, T Yan, J Hong, and H Cao. A novel long non-coding RNA ENST00000480739 suppresses tumour cell invasion by regulating OS-9 and HIF-1 in pancreatic ductal adenocarcinoma. *British Journal of Cancer*, 111(11):2131–2141, October 2014.

[308] Martin Mehnert, Thomas Sommer, and Ernst Jarosch. ERAD ubiquitin ligases. *BioEssays*, 32(10):905–913, August 2010.

[309] John C Christianson and Yihong Ye. Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nature Structural & Molecular Biology*, 21(4):325–335, April 2014.

[310] Dimitrios Zattas, Jason M. Berk, Stefan G. Kreft, and Mark Hochstrasser. A conserved c-terminal element in the yeast doa10 and human MARCH6 ubiquitin ligases required for selective substrate degradation. *Journal of Biological Chemistry*, 291(23):12105–12118, June 2016.

[311] Yilin Ye, Suk-Hwan Baek, Yihong Ye, and Ting Zhang. Proteomic characterization of endogenous substrates of mammalian ubiquitin ligase hrd1. *Cell & Bioscience*, 8(1), August 2018.

[312] Masayuki Kaneko, Masataro Ishiguro, Yoshifumi Niinuma, Mai Uesugi, and Yasuyuki Nomura. Human HRD1 protects against ER stress-induced apoptosis through ER-associated degradation1. *FEBS Letters*, 532(1-2):147–152, November 2002.

[313] Xueming Tan, Xiaolu He, and Zhining Fan. Upregulation of HRD1 promotes cell migration and invasion in colon cancer. *Molecular and Cellular Biochemistry*, 454(1-2):1–9, October 2018.

[314] S. Fang, M. Ferrone, C. Yang, J. P. Jensen, S. Tiwari, and A. M. Weissman. The tumor autocrine motility factor receptor, gp78, is a ubiquitin protein ligase implicated in degradation from the endoplasmic reticulum. *Proceedings of the National Academy of Sciences*, 98(25):14422–14427, November 2001.

[315] Jasmin Schulz, Dönem Avci, Markus A. Queisser, Aljona Gutschmidt, Lena-Sophie Dreher, Emma J. Fenech, Norbert Volkmar, Yuki Hayashi, Thorsten Hoppe, and John C. Christianson. Conserved cytoplasmic domains promote hrd1 ubiquitin ligase complex formation for ER-associated degradation (ERAD). *Journal of Cell Science*, 130(19):3322–3335, October 2017.

[316] Yue-Mei Xu, Hong-Jiang Wang, Fang Chen, Wan-Hua Guo, Yan-Yang Wang, Hang-Yu Li, Jin-Hai Tang, Ying Ding, Ya-Chen Shen, Min Li, Wen-Ying Xuan, Lin-Hui Liu, Jia Wang, Xue-Rong Wang, Ze-Jun Gao, Xiu-Bin Liang, and Dong-Ming Su. HRD1 suppresses the growth and metastasis of breast cancer cells by promoting IGF-1r degradation. *Oncotarget*, 6(40):42854–42867, October 2015.

[317] YanYang Wang, AiBin Guo, XiuBin Liang, Min Li, Ming Shi, Yan Li, Gareth Jenkins, XiaWen Lin, XueFei Wei, ZhiJun Jia, XueFeng Feng, DongMing Su, and WanHua Guo. Hrd1 sensitizes breast cancer cells to tamoxifen by promoting s100a8 degradation. *Oncotarget*, 8(14):23564–23574, 2017.

[318] Jill E. Slansky and Paul T. Spellman. Alternative splicing in tumors — a path to immunogenicity? *New England Journal of Medicine*, 380(9):877–880, 2019. PMID: 30811916.

[319] T. P. J. Garrett, A. W. Burgess, H. K. Gan, R. B. Luwor, G. Cartwright, F. Walker, S. G. Orchard, A. H. A. Clayton, E. C. Nice, J. Rothacker, B. Catimel, W. K. Cavenee, L. J. Old,

E. Stockert, G. Ritter, T. E. Adams, P. A. Hoyne, D. Wittrup, G. Chao, J. R. Cochran, C. Luo, M. Lou, T. Huyton, Y. Xu, W. D. Fairlie, S. Yao, A. M. Scott, and T. G. Johns. Antibodies specifically targeting a locally misfolded region of tumor associated EGFR. *Proceedings of the National Academy of Sciences*, 106(13):5082–5087, March 2009.

[320] Amit Kumar and Vikas Jaitak. Natural products as multidrug resistance modulators in cancer. *European Journal of Medicinal Chemistry*, 176:268–291, August 2019.

[321] Silvia Santamaria, Marisa Delgado, Leonor Kremer, and Jose A. Garcia-Sanz. Will a mAb-based immunotherapy directed against cancer stem cells be feasible? *Frontiers in Immunology*, 8, November 2017.

[322] C. Willemien Menke van der Houven van Oordt, Carlos Gomez-Roca, Carla van Herpen, Andrew L. Coveler, Devalingam Mahalingam, Henk M. W. Verheul, Winette T. A. van der Graaf, Randolph Christen, Dominik Rüttinger, Stefan Weigand, Michael A. Cannarile, Florian Heil, Michael Brewster, Antje-Christine Walz, Tapan K. Nayak, Ernesto Guarin, Valerie Meresse, and Christophe Le Tourneau. First-in-human phase i clinical trial of RG7356, an anti-CD44 humanized antibody, in patients with advanced, CD44-expressing solid tumors. *Oncotarget*, 7(48):80046–80058, August 2016.

[323] Andrew M. Scott, Jedd D. Wolchok, and Lloyd J. Old. Antibody therapy of cancer. *Nature Reviews Cancer*, 12(4):278–287, March 2012.

[324] Nick R P Dekker, Jos A Stigt, Otto Visser, F J Sabine Netters, Rutger H T Koornstra, and Jan Willem B de Groot. [immunotherapy for cancer: from medical breakthrough to application in daily practice]. *Nederlands tijdschrift voor geneeskunde*, 163, May 2019.

[325] Anne Bordron, Cristina Bagacean, Adrian Tempescul, Christian Berthou, Eléonore Bettacchioli, Sophie Hillion, and Yves Renaudineau. Complement system: a neglected pathway in immunotherapy. *Clinical Reviews in Allergy & Immunology*, 58(2):155–171, May 2019.

[326] Tomonori Kaifu and Akira Nakamura. Polymorphisms of immunoglobulin receptors and the effects on clinical outcome in cancer immunotherapy and other immune diseases: a general review. *International Immunology*, 29(7):319–325, July 2017.

[327] Sophie Sibéril, Charles-Antoine Dutertre, Wolf-Herman Fridman, and Jean-Luc Teillaud. Fcr: The key to optimize therapeutic antibodies? *Critical Reviews in Oncology/Hematology*, 62(1):26–33, April 2007.