The AGR2 interactome

The AGR2 interactome

Frederic Delom 1,2,3*, M. Aiman Mohtar 4*, Ted Hupp 5,6 and Delphine Fessart.1,2,3*

[1]Univ. Bordeaux, ACTION, F-33000 Bordeaux, France; [2]INSERM U1218, F-33000 Bordeaux, France; [3]Institut Bergonié, F-33000 Bordeaux, France; [4]UKM Medical Molecular Biology Institute (UMBI), The National University of Malaysia, Cheras 56000 Kuala Lumpur, Malaysia. [5]University of Edinburgh, Institute of Genetics and Molecular Medicine, Edinburgh, Scotland, United Kingdom, EH4 2XR; [6]University of Gdansk, International Centre for Cancer Vaccine Science, ul. Wita Stwosza 63, 80-308 Gdansk, Poland.

* Corresponding authors:

Delphine Fessart - Bordeaux University - 146, Léo-Saignat – 33076 Bordeaux – France. Phone: (00 33) +5 56 33 04 31. Fax: (00 33) +5 56 33 33 30. E-mail: delphine.fessart@yahoo.fr

Frederic Delom - Bordeaux University - 146, Léo-Saignat – 33076 Bordeaux – France. Phone: (00 33) +5 56 33 04 31. Fax: (00 33) +5 56 33 33 30. E-mail: frederic.delom@yahoo.fr

M. Aiman Mohtar - UKM Medical Molecular Biology Institute (UMBI), The National University of Malaysia, Cheras 56000 Kuala Lumpur, Malaysia. Phone: +60391459062 Fax: +603 9171 7185. Email: m.aimanmohtar@ppukm.ukm.edu.my

Running Head: AGR2-proteins interaction network

Keywords: AGR2, monomer, homodimer, binding partners

Abstract (250 words)

The Anterior Gradient-2 (AGR2) is an Endoplasmic Reticulum (ER)-resident protein belonging to the Protein Disulfide Isomerase (PDI) family which mediates the formation of disulfide bonds, and assists the protein quality control in the ER. In addition to its role in proteostasis, extracellular AGR2 is responsible for various cellular effects in many types of cancer, including cell proliferation, survival, and metastasis. Various OMICs approaches have been used to identify AGR2 binding partners and to investigate the functions of AGR2 in the ER and outside the cell. Emerging data showed that AGR2 exists not only as monomer, but it can also form homodimeric structure and thus interact with different partners, yielding different biological outcomes. In this review, we summarize the AGR2 '*interactome*' and discuss the pathological and physiological role of such AGR2 interactions.

Introduction

Anterior gradient protein 2 (AGR2), also known as secreted cement gland protein XAG-2 homolog, is a protein encoded in humans by the AGR2 gene, which belongs to the AGR subfamily (AGR2, AGR3 and TXNDC12 (AGR1)). AGR2 is located on chromosome 7p21, a region known to have frequent genetic alterations (42). AGR2 was originally discovered in Xenopus laevis (XAG-2 gene) through dissection of different aged embryos (51). XAG-2 expression gathers at the anterior region of the dorsal ectoderm, which corresponds to the cement gland anlage (1). In another amphibian model, Salamander, AGR2 allows the regeneration of limbs from dedifferentiated cells and stem cells (28). AGR2 was first identified in humans in estrogen-receptor-positive breast cancer cells(52). Later studies demonstrated high expression of AGR2 in adenocarcinomas of the pancreas (11), esophagus (21, 44), lung (14, 43) and prostate (59). It is expressed strongly in tissues that secrete mucus or function as endocrine organs, including the lung, stomach, colon, prostate and small intestine (7). Furthermore, high expression of AGR2 protein in tissues that contain mucus-secreting cells and/or cells operative in endocrine organs supports a role for this protein in the epithelial barrier function. The positive regulation of the AGR2 promoter activity by Foxa1 and Foxa2, two transcription factors typical of epithelial goblet cells (7), also supports this hypothesis.

Studies have shown that the AGR2 protein can be localized to subcellular locations other than its typical ER residency. This AGR2 mislocalization results in interaction with a variety of cellular proteins involved in proteostasis control and signal transduction, and thus in cellular response. In Xenopus, AGR2 plays a role in cement gland differentiation (1), whereas in human cancer cell lines, high levels of AGR2 correlate with downregulation of the p53 response (44), cell migration, and cell transformation (57). AGR2's physiological function might be that of a redox chaperone as it affects the folding of secreted proteins such as Alpha-1 antitrypsin and mucins (22), and the forming of mixed disulfides with protein substrates (39). Moreover, using an AGR2 knockout mouse lines, it has been demonstrated that AGR2 is essential for the secretion of the mucin MUC2 in intestinal Paneth and goblet cells (39, 60). Several laboratories have shown that this protein can act as either a monomer or a homodimer (34), depending on cellular demands or disease states. AGR2 displays a complex

biological network, as we discuss here. Furthermore, AGR2 expression levels and functionality have been altered in some pathologies, which leads us to argue that these changes may contribute to the onset or development of diseases such as cancer and inflammatory bowel disease, among others.

A - AGR2 is a multidomain protein

AGR2 has different roles that can be attributed to its multidomain structure. AGR2 is a protein composed of 175 amino acids with a molecular weight of 17 kDa (Figure 1). The encoded protein has an N-terminal ER-signal sequence, a catalytically active thioredoxin domain, and a C-terminal ER-retention sequence (Figure 1). As an ER-localized molecular chaperone, it plays a role in the folding, trafficking, and assembly of cysteine-rich transmembrane receptors and the cysteine-rich intestinal glycoprotein mucin. From a structural point of view, it includes a particular signal sequence cleavable N-terminal to translocate into the ER (AA from 1 to 21) (Figure 1). AGR2 is believed to participate in protein folding, and it has a KTEL C-terminal motif (AA from 172 to 175 AA) (Figure 1) similar to KDEL and KVEL ER retention sequences (20). AGR2 protein belongs to the protein disulfide isomerase (PDI) family of ER proteins that catalyze protein folding and thiol-disulfide interchange reactions. Examination of AGR2 amino acid sequence reveals a single CXXS active domain motif (AA 81 to 84) (Figure 1) for oxidation and reduction reactions (17, 41), and the N-terminal region (AA 21 to 44) (Figure 1), which is responsible for the cell adhesion properties of AGR2 (18). In addition to being an ER-resident protein, AGR2 is also localized in cytoplasm (58), nucleus (30, 33), mitochondria (32), cell surface (12), extracellular matrix (14), and blood and urine (6, 13, 14), although the mechanism by which AGR2 accesses the cytoplasm, nucleus and extracellularly is unclear.

B - Oligomeric Organizations and Interactome of AGR2

B1. AGR-AGR2 interaction (monomer *vs* homodimer)

The nature of AGR2 protein structure was not fully understood until the development of AGR2 crystal structure, which gives a clearer picture of AGR2 protein architecture. The Nuclear Magnetic Resonance (NMR) crystallography showed that $AGR2_{41\text{-}175}$ can exist in homodimer structure through a

motif EALYK at position E60-K64 (40). The salt bridge between the carboxylate group of E60 and the ammonium group of K64 forms the major interaction between the two AGR2 subunits in a reverse parallel fashion. Mutagenesis at this dimerization motif (E60A) abolishes the homodimeric structure. It is important to note that the dimerization interface of AGR2 is held far from the catalytic CXXS motif (opposing faces), where enhanced disulfide exchange in client proteins might occur. Although AGR2 has only one active cysteine residue (pseudo-thioredoxin domain) (Figure 1), the homodimeric structure of AGR2 might provide equivalent redox capacity.

Alternatively, AGR2 can form a shaped homodimeric structure in which oxidation-dependent homodimerization occurs through C81-mediated disulfide bond formation (Figure 1), which would re-orientate the homodimer into a different formation (49). Indeed, Clarke *et al.* showed that AGR2 can homodimerize in a similar fashion via the covalent bond (C81-C81) oxidation upon chemical peroxide exposure (8). The mutation of the single cysteine (C81) to alanine prevents hydrogen peroxide-catalyzed homodimerization of AGR2. Additionally, mass spectrometry analysis reveals that low levels of a chemical oxidant promote an intermolecular disulfide bond through the formation of a labile sulfenic acid intermediate. However, higher levels of oxidant promote sulfinic or sulfonic acid formation, thus preventing covalent homodimerization of AGR2. The single cysteine (C81) of the AGR2 therefore acts as an oxidant-responsive moiety that regulates its tendency for oxidation and its monomeric-homodimeric state.

AGR2 homodimerization is also associated with its N-terminal. The NMR study revealed that N-terminal amino acid region 21-40 (Figure 1) is unfolded (40). Deletion of the N-terminal disordered region (AA 21-45) (Figure 1) creates more stable AGR2 homodimer (18). Indeed, recombinant AGR2, having this deletion, demonstrated a spontaneous homodimer in solution, and this mutant showed enhanced binding to its client protein Reptin compared to wild type AGR2 (18). AGR2 was also shown to be capable of promoting cell adhesion by its N terminal domain (AA 21 to 44) (Figure 1). The monomeric AGR2 (E60A) or the native homodimeric AGR2 showed relatively similar cell adhesion properties, suggesting that homodimerization is not important for this property. As such, the

N-terminal could be an anchor, for example to the plastic surface (40), while the other part of AGR2 interacts with cell receptors through thioredoxin motif or the facilitation of cell adhesion.

Physiologically, this suggests that the monomer to homodimer equilibrium is important in AGR2 signaling. It also dictates the protein geometry and synergy between the homodimer interface and the catalytic thioredoxin motif, given that AGR2 has the sole C81 for thiol-disulfide interchange to assemble correctly folded disulfide bridges in client cargo proteins. High levels of homodimeric AGR2 might exist in the ER, which is important for cysteine-rich client protein folding. Recombinant AGR2 readily exists in homodimeric state in solution (8) consistent with the calculated monomer-homodimer equilibrium dissociation constant, Kd of 8.83 μM (40). This provides an indication that homodimeric AGR2 might be a pre-dominant form in physiological condition. In addition, AGR2 can respond to oxidizing conditions which could affect the homodimer and monomer equilibrium. The ER environment, especially in the ER of cancer, is enriched in redox signaling mediators that in turn modulate the reactive oxygen species (ROS) (31). This creates a redox imbalance environment that can affect the monomeric and homodimeric equilibrium of AGR2 protein and this could give rise to AGR2 oncogenic functions.

We have recently shown that the control of AGR2 homodimerization participates in the establishment of pro-inflammatory responses, which suggests that the AGR2 homodimer would behave as a sensor that contributes to the regulation of ER proteostasis (34). Using a protein–protein interaction screen to identify cellular regulators of AGR2 homodimerization, we discovered TMED2 as a specific enhancer of AGR2 homodimerization. Further, we have characterized the interaction *in vitro* and *in vivo* and shown that upon ER stress, AGR2 and TMED2 dissociate. Finally, by a molecular modeling approach, we identified K66 and Y111 amino acid residues are involved in the TMED2-AGR2 interaction (Table 1, (34))

This part of AGR2 functions are yet to be fully understood: (**1**) *What are the roles and ratios of monomeric and homodimeric AGR2 in normal and disease conditions (e.g. cancer)?*, (**2**) *How does the oxidation and monomeric/homodimeric state of AGR2 impact on client protein binding and client protein maturation, especially cysteine-rich proteins like MUC2?*, (**3**) *Could this affect the protein*

folding capacity of AGR2 and shift its function to chaperone activity?, and (**4**) *How does homodimeric AGR2 contribute to the progression of diseases such as cancer?*

B2. Insight into further AGR2 interactions

Existing data shows that AGR2 can bind to a wide range of proteins such as nuclear, cytosolic and plasma membrane proteins. AGR2 interacting proteins validated using protein-protein interaction assays published in the literature largely belong to plasma membrane proteins (Table 1). This is not surprising since AGR2 is a PDI involved in the secretory pathway and therefore in the regulation of proteostasis, as we have reviewed recently (9). Such a function is supported by the observation that AGR2 can act as a PDI molecule critical for mucin biosynthesis (39). Another example is the AGR2 binding partner, EGFR (Table 1), which can form mixed-disulfide bonds with AGR2. In this way, AGR2 controls EGFR cell surface delivery (10). This raises the question of how AGR2 can catalyze the folding, assembly, and trafficking of receptors such as EGFR to the cell surface. Moreover, one recent study showed that gene expression of EGFR ligand TGFA promotes the growth of mutant $EGFR^{L858R}$ lung tumor cells. AGR2 was found to be the key regulator as it was highly expressed in the TGFA-induced *EGFR*-mutant lung adenocarcinoma (55). AGR2 has also been shown to interact with cytoplasmic protein TAB2 (TGF-beta activated kinase 1) (58), ER and Golgi protein TMED2 (34), mitochondrial protein UNG1 (32), and nuclear transcription factor HIF-1α and Reptin (30, 33) (Table 1). Some of AGR2 protein-protein interactions in the table 1 have been reviewed elsewhere (5, 7, 9).

Murray *et al*. used an *in silico* study to expand AGR2 interactome by exploiting the sequence-specific peptide motif that binds to AGR2 which was previously identified using peptide-phage display (36). To our knowledge, AGR2 is the only PDI with such a specific peptide-binding activity. Attempts to define a consensus peptide binding site for AGR3 have been unsuccessful (data not shown). We hypothesize that if AGR2 can bind to this sequence-specific penta-peptide motif, it can also bind to proteins that have the same motif. To this end, a consensus AGR2 peptide motif was developed using synthetic peptide libraries. The consensus motif Tx[IL][YF[YF] was formed and this motif has been used to mine the human protein database to search for proteins harboring a similar motif. The data revealed that this motif is enriched in membrane-associated proteins, which suggest a role for AGR2 in this type of protein. Interestingly, none of the AGR2 client proteins identified in

previous studies overlap with our database results, suggesting that this method is a first in terms of determining novel AGR2 client proteins and complements the classical PPI methods. One key receptor protein, EpCAM (Table 1), was validated as an AGR2 client protein using both *in vitro* and cell-based assays and maps of the binding interface using hydrogen deuterium mass spectrometry. The data suggests one plausible role of AGR2 in membrane trafficking – and thus in proteostasis – by ensuring the proper maturation of this class of protein type to their final destination through secretory pathway. However, it remains to be studied whether extracellular AGR2 (eAGR2) has extracellular activity for EpCAM since the 'docking motif' is located at the extracellular stalk of EpCAM. Other potential AGR2 proteins from this transmembrane database mining, including MKS3 (35) (Figure 3), are yet to be fully validated. Consistent with this EpCAM link to AGR2, quantitative shotgun proteomics was used to demonstrate that AGR2 and EpCAM were among the top upregulated cancer-associated proteins, and immunohistochemistry in esophageal adenocarcinoma tissues showed high expression of both proteins, thus presenting a novel biomarker pathway for this type of cancer (38).

The ability of AGR2 to interact with receptor proteins is not surprising as AGR2 was shown to be present in both ER (intracellular AGR2 (iAGR2)) and on the surface (extracellular AGR2 (eAGR2)) of pancreatic cancer cells using immunofluorescence and flow cytometry (12). Localization of eAGR2 at the surface of the cell could suggest that AGR2 is attached to a client receptor protein. Moreover, eAGR2 can be secreted in the microenvironment and interact directly with the extracellular matrix (14). eAGR2 is also present in the extracellular mediums of pancreatic (46), colon (53) and lung (14) cancer cells, in the urine of prostate cancer patients (6), in the plasma of ovarian cancer patients (13), and in the serum of pituitary adenomas (54). In addition, atypical roles for eAGR2 have been reported in the tumor microenvironment (9). For example, the addition of exogenous AGR2 has been shown to promote angiogenesis through the regulation of hypoxia induced factor-1 (HIF-1) (23). We have previously demonstrated that eAGR2 performs an extracellular role independent of its ER function as an important microenvironmental pro-oncogenic regulator of epithelial morphogenesis and tumorigenesis (14). Furthermore, eAGR2 can stimulate the insulin-like growth factor-1 receptor (IGF-1)-induced cell proliferation, migration and EMT in breast cancer cells through interaction with membrane ER-α (29). In colorectal cancer, eAGR2 can promote migration and metastasis through

196 Wnt11 signaling (53). This extracellular role of AGR2 has shifted the paradigm and current focus of
197 how an ER-resident protein can exit the ER, be secreted to extracellular space and still retain its pro-
198 oncogenic effects.
199
200 **C- Physiological and pathological role of AGR interactome**
201 The biological network of AGR2 highlights the fact that AGR2 lies at the crossroads of
202 multiple signaling pathways (Figure 2): Cell signaling, Proteostasis, Chemoresistance, Tumor cell
203 dissemination and Protein trafficking. However, a number of important issues remain unsolved: (**1**)
204 *How are these interactions modulated?*, (**2**) *Which of them are relevant in physiological and*
205 *pathological situations characterized by alterations in AGR2 expression levels, localization and/or*
206 *function?*, (**3**) *Are these new interactions of AGR2 important to its role in key biological functions?* In
207 the following part, we will focus on these AGR2 reported physiological functions (Figure 2).
208
209 **C.1 AGR2 and proteostasis**
210 AGR2, as a PDI, is involved in the maturation of various proteins through forming disulfide
211 bonds with mucins (Table 1), and directly regulating their processing and secretion. PDIs function to
212 attain the highest quality and productivity of the ER and to therefore maintain ER protein homeostasis
213 (also called proteostasis). In a healthy intestine, AGR2 acts as a critical modulator of intestinal
214 homeostasis presenting within the ER of intestinal secretory epithelial cells (34) and regulating
215 production of intestinal mucins via the formation of mixed disulfide bonds (60). Mice lacking AGR2
216 expression are unable to produce intestinal mucins and are highly susceptible to colitis.
217 Recently, AGR2 has also been shown to interact with UNG1 (Table 1) (32). UNG1 is
218 important for initiating base-excision repair in mitochondria. Due to UNG1 being located in close
219 proximity to the respiration chain, UNG1 is subject to oxidative damage. Liu *et al.* have demonstrated
220 that overexpression of UNG1 enhances cell resistance to oxidative stress and protects mtDNA from
221 oxidation. Thus, it has been suggested that the interaction of UNG1 and AGR2 stabilizes UNG1,
222 through its PDI function, and enhances its enzymatic activity in DNA repair. However, more studies
223 are needed to examine the nature and the physiological role of UNG1-AGR2 interaction.

We have previously demonstrated that AGR2 represents a mechanistic intermediate between endoplasmic reticulum quality control (ERQC) and tumor development (7, 22). Thus, "*AGR2 proteotasis*" interactome in cancer could enhance proteostasis (Figure 2) thereby allowing tumor cells to cope with abnormal protein production and secretion and contributing to the tumor development.

C.2 AGR2, a HIF-1α-binding stabilizer for chemoresistance properties

In breast cancer, AGR2 has been shown to bind and stabilize HIF-1α (Table 1) (30), stabilize HIF-1α and delay its proteasomal degradation. It appears that AGR2 regulates the chemical hypoxia-induced doxorubicin resistance in breast cancer cells, which explains the variable levels of chemoresistance in breast cancers. This further validates its role as a potential breast cancer therapeutic target. Hence, given the important role of AGR2/HIF-1α interactome in increasing chemoresistance in breast cancer, it would therefore be interesting to investigate the AGR2/HIF-1α pathway to develop strategies to block chemoresistance AGR2 properties (Figure 2) in breast cancer.

C.3 AGR2 and tumor cell dissemination

Reptin was identified by the yeast two-hybrid screen (Table 1) and validated as an interacting protein of AGR2 in human cells. Reptin is a highly conserved member of the AAA+ family that can be found in numerous multiprotein complexes linked to transcription, DNA damage response, and nonsense-mediated RNA decay (16, 24-27, 47). This protein is a member of the RuvBl1/2 superfamily containing ATP binding motifs and has the ability to form biologically relevant protein–protein interactions with proteins involved in cancer, including Myc, Tip60, APPL1, Pontin, and telomerase holoenzyme complexes (3, 4, 45, 47, 56). The ATP binding motifs of Reptin and the proposed substrate-binding loop of AGR2 are determinants that drive a specific complex between AGR2 and Reptin proteins (33). Reptin was also shown to be overproduced in primary breast cancer biopsy specimens, suggesting a role in cancer growth *in vivo*. Because Reptin can also function as a prometastatic transcription protein, future research will inform whether AGR2 uses its substrate-binding loop to chaperone Reptin into inactive and activated transcriptional states and/or whether the allosteric ATP-binding motifs of Reptin regulate AGR2 function as a prometastatic factor in cancer.

Orphan glycosyl-phosphatidyl-inositol-linked receptor, C4.4A, has been reported to interact with eAGR2 to induce cell proliferation, migration and chemoresistance in pancreatic ductal adenocarcinoma (2). The secreted ~~proteins~~ metastasis-associated GPI-anchored C4.4A protein and the extracellular domain of α-dystroglycan have been shown to directly interact with eAGR2, indicating potential mechanisms for eAGR2 in the promotion of tumor metastasis via the regulation of receptor adhesion and the interaction with the extracellular matrix (59, 60).

Interestingly, C4.4A is a structural homologue of the urokinase-type plasminogen activator receptor (uPAR), which, along with CD59, is the most closely related human homologue of Prod1. In salamander limb regeneration, eAGR2 interacts with a cell surface receptor on adjacent cells identified as Prod1. Prod1 is a member of the three-finger protein (TFP) superfamily that is GPI-anchored at the cell surface. Whether CD59 is responsible for transducing the effect of eAGR2 is at present unknown. Prostate basal cells are immuno-stained by anti-CD59 and AGR2 (31). Therefore, it is possible to speculate that cells without CD59 might be resistant to the effect of eAGR2. Thus, eAGR2 interacts with the cell surface in order to modulate adhesion and promote tumor cell dissemination.

A better understanding of the molecular mechanisms that promote the dissemination of tumor cells is essential for the development of novel detection and therapeutic strategies. For lung cancer, we have demonstrated that AGR2 induced the EMT and contributed to cancer dissemination (14). As AGR2 promotes the dissemination of tumor cells, we therefore suggest that "*AGR2 tumor cell dissemination*" interactome targeting may permit new strategies for therapeutic management against tumor cell dissemination capacity.

C.4 AGR2 and regulation – the trafficking of proteins destined for the secretory pathway

It has been established that AGR2 interacts physically with the Epithelial Growth Factor Receptor (EGFR) within the ER, and that this interaction is necessary before the receptor can progress to the Golgi apparatus and the plasma membrane (10). In pancreatic disease, it has been hypothesized that this AGR2-induced EGFR signaling forms a common and essential link between injury-induced tissue regeneration and the development of pancreatic cancer. AGR2 expression represents an early initiating event necessary for EGFR signaling, and thus serves as a potential target for the treatment of chronic and neoplastic pancreatic disease. In Pancreatic Ductal Adenocarcinoma (PDAC), AGR2 is

found activated downstream of mutant $Kras^{G12D}$ (11) and its important role in PDAC development was shown using both orthotopic and Kras-driven pancreatic cancer mouse model with the deletion of Smad4, a tumor suppressor gene frequently lost in the late stage (PanIN-3) of pancreatic carcinogenesis (11).

Screening the human proteome for proteins that harbor a specific peptide motif (Tx[IL][YF][YF]) revealed an EpCAM-interaction (35). AGR2 binding to EpCAM seems to be in the ER, since the recombinant AGR2 and EpCAM binds *in vitro*. This recombinant protein is made in bacterial system, which lacks glycosylation – an event before post-translational modification that supports the hypothesis that AGR2-EpCAM binding is in the ER. EpCAM forms a model client protein for AGR2, which suggests that AGR2 is involved in regulating the trafficking of specific proteins destined for the secretory pathway, as has been demonstrated for EGFR.

Thus, AGR2 not only plays an essential role in the secretory pathway via the "*AGR2 proteostasis*" interactome (see C.1), but also via its protein-trafficking function. Thus, a protein trafficking network controlled by AGR2 could be a future target for anti-cancer therapy.

C.5 AGR2-enhanced VEGF and FGF signaling

We have demonstrated that AGR2 is a signaling molecule that is found outside cancer cells (eAGR2) and one that makes these cells more aggressive. Indeed, eAGR2 has several extracellular functions (9), including the promotion of angiogenesis- and fibroblast- coordinated tumor cell invasion. Indeed, eAGR2 has been reported to directly bind to VEGF and FGF2, are the major players in tumor angiogenesis, and enhance their activities (19). This enhancement is dependent on both the AGR2 self-homodimerization region and the signaling pathways of VEGF and FGF2. It has been shown, both *in vivo* and *in vitro*, that a monoclonal antibody targeting the AGR2 self-homodimerization domain can partially abrogate eAGR2 activity. Thus, eAGR2 acts at multiple signaling crossroads, suggesting that eAGR2 is potentially a therapeutic target.

Concluding remarks

The increasing amount of data on AGR2 'interactome' identifies this PDI as a signaling node involved in the regulation of cellular homeostasis. The intricacy of this network and the involvement of this protein in cancer-related processes has led us to argue that changes in AGR2 levels, localization and/or function have important roles to play in oncogenesis. This makes the AGR2 protein and its pathway a potential diagnostic marker and therapeutic target in cancer. Nevertheless, the biochemical mechanisms that explain how the AGR2 interactome is orchestrated based on the physiological context remains to be defined. The modulation of these potential AGR2 interactions, in addition to the overall physiological and pathological implications emerging from this AGR2 interaction map in disease context, are remain uncovered. We believe the AGR2 protein has a dual role, intracellular (iAGR2) and extracellular (eAGR2) (Figure 3). How AGR2 localization is regulated in normal and disease conditions, and how this impacts on client protein trafficking, are beginning to be defined. In summary, there is an intracellular chaperone role for AGR2 (iAGR2) in stimulating receptor maturation of pro-oncogenic proteins (EpCAM, EGFR) that can contribute to cancer growth. It appears that iAGR2 can stimulate receptor expression via transient binding to its sequence specific Tx[IL][YF][YF] motif in client proteins (docking-dependent) and through more stable interaction such as mixed disulfide bonding (docking independent) (Figure 3). Additional receptors identified with the peptide-binding motif include MUC16, FITM1, KCNC3, CXCR5 and OR2M4 (34). In addition, an extracellular role for AGR2 (eAGR2) may accelerate unidentified protein-protein interactions, which may be these same receptors as they function outside the cells that regulate cell-autonomy, as well as cell-cell interactions conducive to cancer cell proliferation.

In this context, increased understanding of the mechanisms and signals governing the expression, localization and function of AGR2 and other AGRs (AGR3 and TXNDC12) is crucial. Likewise, future investigations are needed to evaluate the effect of such changes on the complex coordinated network of AGR cellular functions as it is likely that altered AGR2 expression would differentially affect the function of its interacting partners and impair homeostasis/proteostasis in distinct ways.

In this context, the availability of specific AGR2 peptides will be advantageous in helping to clearly establish the physiological relevance of AGR2 'interactome'. New understanding of the AGR2

335 structure and the emerging identification of interaction domains will help to fine-tune and design
336 peptides or molecules that are able to impair or disrupt such interactions selectively in order to target
337 specific intracellular and extracellular functions of this PDI.
338

Bibliography

1. **Aberger F, Weidinger G, Grunz H, and Richter K**. Anterior specification of embryonic ectoderm: the role of the Xenopus cement gland-specific gene XAG-2. *Mechanisms of development* 72: 115-130, 1998.
2. **Arumugam T, Deng D, Bover L, Wang H, Logsdon CD, and Ramachandran V**. New Blocking Antibodies against Novel AGR2-C4.4A Pathway Reduce Growth and Metastasis of Pancreatic Tumors and Increase Survival in Mice. *Molecular cancer therapeutics* 14: 941-951, 2015.
3. **Bauer A, Chauvet S, Huber O, Usseglio F, Rothbacher U, Aragnol D, Kemler R, and Pradel J**. Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *The EMBO journal* 19: 6121-6130, 2000.
4. **Bellosta P, Hulf T, Balla Diop S, Usseglio F, Pradel J, Aragnol D, and Gallant P**. Myc interacts genetically with Tip48/Reptin and Tip49/Pontin to control growth and proliferation during Drosophila development. *Proceedings of the National Academy of Sciences of the United States of America* 102: 11799-11804, 2005.
5. **Brychtova V, Mohtar A, Vojtesek B, and Hupp TR**. Mechanisms of anterior gradient-2 regulation and function in cancer. *Seminars in cancer biology* 33: 16-24, 2015.
6. **Bu H, Bormann S, Schafer G, Horninger W, Massoner P, Neeb A, Lakshmanan VK, Maddalo D, Nestl A, Sultmann H, Cato AC, and Klocker H**. The anterior gradient 2 (AGR2) gene is overexpressed in prostate cancer and may be useful as a urine sediment marker for prostate cancer detection. *The Prostate* 71: 575-587, 2011.
7. **Chevet E, Fessart D, Delom F, Mulot A, Vojtesek B, Hrstka R, Murray E, Gray T, and Hupp T**. Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. *Oncogene* 32: 2499-2509, 2013.
8. **Clarke DJ, Murray E, Faktor J, Mohtar A, Vojtesek B, MacKay CL, Smith PL, and Hupp TR**. Mass spectrometry analysis of the oxidation states of the pro-oncogenic protein anterior gradient-2 reveals covalent dimerization via an intermolecular disulphide bond. *Biochimica et biophysica acta* 1864: 551-561, 2016.
9. **Delom F, Nazaraliyev A, and Fessart D**. The role of protein disulphide isomerase AGR2 in the tumour niche. *Biology of the cell* 110: 271-282, 2018.
10. **Dong A, Wodziak D, and Lowe AW**. Epidermal growth factor receptor (EGFR) signaling requires a specific endoplasmic reticulum thioredoxin for the post-translational control of receptor presentation to the cell surface. *The Journal of biological chemistry* 290: 8016-8027, 2015.
11. **Dumartin L, Alrawashdeh W, Trabulo SM, Radon TP, Steiger K, Feakins RM, di Magliano MP, Heeschen C, Esposito I, Lemoine NR, and Crnogorac-Jurcevic T**. ER stress protein AGR2 precedes and is involved in the regulation of pancreatic cancer initiation. *Oncogene* 36: 3094-3103, 2017.
12. **Dumartin L, Whiteman HJ, Weeks ME, Hariharan D, Dmitrovic B, Iacobuzio-Donahue CA, Brentnall TA, Bronner MP, Feakins RM, Timms JF, Brennan C, Lemoine NR, and Crnogorac-Jurcevic T**. AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. *Cancer research* 71: 7091-7102, 2011.
13. **Edgell TA, Barraclough DL, Rajic A, Dhulia J, Lewis KJ, Armes JE, Barraclough R, Rudland PS, Rice GE, and Autelitano DJ**. Increased plasma concentrations of anterior gradient 2 protein are positively associated with ovarian cancer. *Clinical science* 118: 717-725, 2010.
14. **Fessart D, Domblides C, Avril T, Eriksson LA, Begueret H, Pineau R, Malrieux C, Dugot-Senant N, Lucchesi C, Chevet E, and Delom F**. Secretion of protein disulphide isomerase AGR2 confers tumorigenic properties. *eLife* 5: 2016.
15. **Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA, Daviet L, Legrain P, Parekh R, Harris AL, and Terrett JA**. hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. *British journal of cancer* 88: 579-585, 2003.
16. **Gallant P**. Control of transcription by Pontin and Reptin. *Trends in cell biology* 17: 187-192, 2007.

17. **Galligan JJ, and Petersen DR**. The human protein disulfide isomerase gene family. *Human genomics* 6: 6, 2012.
18. **Gray TA, Murray E, Nowicki MW, Remnant L, Scherl A, Muller P, Vojtesek B, and Hupp TR**. Development of a fluorescent monoclonal antibody-based assay to measure the allosteric effects of synthetic peptides on self-oligomerization of AGR2 protein. *Protein science : a publication of the Protein Society* 22: 1266-1278, 2013.
19. **Guo H, Zhu Q, Yu X, Merugu SB, Mangukiya HB, Smith N, Li Z, Zhang B, Negi H, Rong R, Cheng K, Wu Z, and Li D**. Tumor-secreted anterior gradient-2 binds to VEGF and FGF2 and enhances their activities by promoting their homodimerization. *Oncogene* 36: 5098-5109, 2017.
20. **Gupta A, Dong A, and Lowe AW**. AGR2 gene function requires a unique endoplasmic reticulum localization motif. *The Journal of biological chemistry* 287: 4773-4782, 2012.
21. **Hao Y, Triadafilopoulos G, Sahbaie P, Young HS, Omary MB, and Lowe AW**. Gene expression profiling reveals stromal genes expressed in common between Barrett's esophagus and adenocarcinoma. *Gastroenterology* 131: 925-933, 2006.
22. **Higa A, Mulot A, Delom F, Bouchecareilh M, Nguyen DT, Boismenu D, Wise MJ, and Chevet E**. Role of pro-oncogenic protein disulfide isomerase (PDI) family member anterior gradient 2 (AGR2) in the control of endoplasmic reticulum homeostasis. *The Journal of biological chemistry* 286: 44855-44868, 2011.
23. **Hong XY, Wang J, and Li Z**. AGR2 expression is regulated by HIF-1 and contributes to growth and angiogenesis of glioblastoma. *Cell Biochem Biophys* 67: 1487-1495, 2013.
24. **Izumi N, Yamashita A, Iwamatsu A, Kurata R, Nakamura H, Saari B, Hirano H, Anderson P, and Ohno S**. AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. *Science signaling* 3: ra27, 2010.
25. **Jha S, and Dutta A**. RVB1/RVB2: running rings around molecular biology. *Molecular cell* 34: 521-533, 2009.
26. **Kim JH, Choi HJ, Kim B, Kim MH, Lee JM, Kim IS, Lee MH, Choi SJ, Kim KI, Kim SI, Chung CH, and Baek SH**. Roles of sumoylation of a reptin chromatin-remodelling complex in cancer metastasis. *Nature cell biology* 8: 631-639, 2006.
27. **Kim JH, Kim B, Cai L, Choi HJ, Ohgi KA, Tran C, Chen C, Chung CH, Huber O, Rose DW, Sawyers CL, Rosenfeld MG, and Baek SH**. Transcriptional regulation of a metastasis suppressor gene by Tip60 and beta-catenin complexes. *Nature* 434: 921-926, 2005.
28. **Kumar A, Godwin JW, Gates PB, Garza-Garcia AA, and Brockes JP**. Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. *Science* 318: 772-777, 2007.
29. **Li Z, Zhu Q, Chen H, Hu L, Negi H, Zheng Y, Ahmed Y, Wu Z, and Li D**. Binding of anterior gradient 2 and estrogen receptor-alpha: Dual critical roles in enhancing fulvestrant resistance and IGF-1-induced tumorigenesis of breast cancer. *Cancer letters* 377: 32-43, 2016.
30. **Li Z, Zhu Q, Hu L, Chen H, Wu Z, and Li D**. Anterior gradient 2 is a binding stabilizer of hypoxia inducible factor-1alpha that enhances CoCl2 -induced doxorubicin resistance in breast cancer cells. *Cancer science* 106: 1041-1049, 2015.
31. **Liou GY, and Storz P**. Reactive oxygen species in cancer. *Free radical research* 44: 479-496, 2010.
32. **Liu Z, Hu Y, Gong Y, Zhang W, Liu C, Wang Q, and Deng H**. Hydrogen peroxide mediated mitochondrial UNG1-PRDX3 interaction and UNG1 degradation. *Free radical biology & medicine* 99: 54-62, 2016.
33. **Maslon MM, Hrstka R, Vojtesek B, and Hupp TR**. A divergent substrate-binding loop within the pro-oncogenic protein anterior gradient-2 forms a docking site for Reptin. *Journal of molecular biology* 404: 418-438, 2010.
34. **Maurel M, Obacz J, Avril T, Ding YP, Papadodima O, Treton X, Daniel F, Pilalis E, Horberg J, Hou W, Beauchamp MC, Tourneur-Marsille J, Cazals-Hatem D, Sommerova L, Samali A, Tavernier J, Hrstka R, Dupont A, Fessart D, Delom F, Fernandez-Zapico ME, Jansen G, Eriksson LA, Thomas DY, Jerome-Majewska L, Hupp T, Chatziioannou A, Chevet E, and Ogier-Denis E**. Control of anterior

441 GRadient 2 (AGR2) dimerization links endoplasmic reticulum proteostasis to inflammation. *EMBO*
442 *molecular medicine* 2019.
443 35. **Mohtar MA, Hernychova L, O'Neill JR, Lawrence ML, Murray E, Vojtesek B, and Hupp TR**.
444 The Sequence-specific Peptide-binding Activity of the Protein Sulfide Isomerase AGR2 Directs Its
445 Stable Binding to the Oncogenic Receptor EpCAM. *Molecular & cellular proteomics : MCP* 17: 737-
446 763, 2018.
447 36. **Murray E, McKenna EO, Burch LR, Dillon J, Langridge-Smith P, Kolch W, Pitt A, and Hupp TR**.
448 Microarray-formatted clinical biomarker assay development using peptide aptamers to anterior
449 gradient-2. *Biochemistry* 46: 13742-13751, 2007.
450 37. **Norris AM, Gore A, Balboni A, Young A, Longnecker DS, and Korc M**. AGR2 is a SMAD4-
451 suppressible gene that modulates MUC1 levels and promotes the initiation and progression of
452 pancreatic intraepithelial neoplasia. *Oncogene* 32: 3867-3876, 2013.
453 38. **O'Neill JR, Pak HS, Pairo-Castineira E, Save V, Paterson-Brown S, Nenutil R, Vojtesek B,**
454 **Overton I, Scherl A, and Hupp TR**. Quantitative Shotgun Proteomics Unveils Candidate Novel
455 Esophageal Adenocarcinoma (EAC)-specific Proteins. *Molecular & cellular proteomics : MCP* 16: 1138-
456 1150, 2017.
457 39. **Park SW, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ, Killeen N, and Erle**
458 **DJ**. The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. *Proceedings*
459 *of the National Academy of Sciences of the United States of America* 106: 6950-6955, 2009.
460 40. **Patel P, Clarke C, Barraclough DL, Jowitt TA, Rudland PS, Barraclough R, and Lian LY**.
461 Metastasis-promoting anterior gradient 2 protein has a dimeric thioredoxin fold structure and a role
462 in cell adhesion. *Journal of molecular biology* 425: 929-943, 2013.
463 41. **Persson S, Rosenquist M, Knoblach B, Khosravi-Far R, Sommarin M, and Michalak M**.
464 Diversity of the protein disulfide isomerase family: identification of breast tumor induced Hag2 and
465 Hag3 as novel members of the protein family. *Molecular phylogenetics and evolution* 36: 734-740,
466 2005.
467 42. **Petek E, Windpassinger C, Egger H, Kroisel PM, and Wagner K**. Localization of the human
468 anterior gradient-2 gene (AGR2) to chromosome band 7p21.3 by radiation hybrid mapping and
469 fluorescencein situ hybridisation. *Cytogenetics and cell genetics* 89: 141-142, 2000.
470 43. **Pizzi M, Fassan M, Balistreri M, Galligioni A, Rea F, and Rugge M**. Anterior gradient 2
471 overexpression in lung adenocarcinoma. *Applied immunohistochemistry & molecular morphology :*
472 *AIMM* 20: 31-36, 2012.
473 44. **Pohler E, Craig AL, Cotton J, Lawrie L, Dillon JF, Ross P, Kernohan N, and Hupp TR**. The
474 Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage.
475 *Molecular & cellular proteomics : MCP* 3: 534-547, 2004.
476 45. **Qi D, Jin H, Lilja T, and Mannervik M**. Drosophila Reptin and other TIP60 complex
477 components promote generation of silent chromatin. *Genetics* 174: 241-251, 2006.
478 46. **Ramachandran V, Arumugam T, Wang H, and Logsdon CD**. Anterior gradient 2 is expressed
479 and secreted during the development of pancreatic cancer and promotes cancer cell survival. *Cancer*
480 *Res* 68: 7811-7818, 2008.
481 47. **Rashid S, Pilecka I, Torun A, Olchowik M, Bielinska B, and Miaczynska M**. Endosomal
482 adaptor proteins APPL1 and APPL2 are novel activators of beta-catenin/TCF-mediated transcription.
483 *The Journal of biological chemistry* 284: 18115-18128, 2009.
484 48. **Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, and Ruddock L**. A
485 molecular specificity code for the three mammalian KDEL receptors. *The Journal of cell biology* 179:
486 1193-1204, 2007.
487 49. **Ryu J, Park SG, Lee PY, Cho S, Lee DH, Kim GH, Kim JH, and Park BC**. Dimerization of pro-
488 oncogenic protein Anterior Gradient 2 is required for the interaction with BiP/GRP78. *Biochemical*
489 *and biophysical research communications* 430: 610-615, 2013.
490 50. **Schroeder BW, Verhaeghe C, Park SW, Nguyenvu LT, Huang X, Zhen G, and Erle DJ**. AGR2 is
491 induced in asthma and promotes allergen-induced mucin overproduction. *Am J Respir Cell Mol Biol*
492 47: 178-185, 2012.

51. **Sive HL, Hattori K, and Weintraub H**. Progressive determination during formation of the anteroposterior axis in Xenopus laevis. *Cell* 58: 171-180, 1989.
52. **Thompson DA, and Weigel RJ**. hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with EGFR the cell s receptor in breast cancer cell lines. *Biochemical and biophysical research communications* 251: 111-116, 1998.
53. **Tian S, Hu J, Tao K, Wang J, Chu Y, Li J, Liu Z, Ding X, Xu L, Li Q, Cai M, Gao J, Shuai X, Wang G, Wang L, and Wang Z**. Secreted AGR2 promotes invasion of colorectal cancer cells via Wnt11-mediated non-canonical Wnt signaling. *Experimental cell research* 364: 198-207, 2018.
54. **Tohti M, Li J, Tang C, Wen G, Abdujilil A, Yizim P, and Ma C**. Serum AGR2 as a useful biomarker for pituitary adenomas. *Clin Neurol Neurosurg* 154: 19-22, 2017.
55. **Tomoshige K, Guo M, Tsuchiya T, Fukazawa T, Fink-Baldauf IM, Stuart WD, Naomoto Y, Nagayasu T, and Maeda Y**. An EGFR ligand promotes EGFR-mutant but not KRAS-mutant lung cancer in vivo. *Oncogene* 2018.
56. **Venteicher AS, Meng Z, Mason PJ, Veenstra TD, and Artandi SE**. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* 132: 945-957, 2008.
57. **Wang Z, Hao Y, and Lowe AW**. The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation. *Cancer research* 68: 492-497, 2008.
58. **Yu H, Zhao J, Lin L, Zhang Y, Zhong F, Liu Y, Yu Y, Shen H, Han M, He F, and Yang P**. Proteomic study explores AGR2 as pro-metastatic protein in HCC. *Molecular bioSystems* 8: 2710-2718, 2012.
59. **Zhang JS, Gong A, Cheville JC, Smith DI, and Young CY**. AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. *Genes, chromosomes & cancer* 43: 249-259, 2005.
60. **Zhao F, Edwards R, Dizon D, Afrasiabi K, Mastroianni JR, Geyfman M, Ouellette AJ, Andersen B, and Lipkin SM**. Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in Agr2-/- mice. *Developmental biology* 338: 270-279, 2010.

Figures legends

Figure 1. Primary structure of the AGR2 protein.

The colored boxes indicate the identified functional domains and amino acids involved in the regulation of AGR2 function.

Figure 2. Physiological functions of AGR2.

Schematic representation of AGR2 interactome that underlies its participation in key biological functions.

Figure 3. Model of iAGR2 and eAGR2 functions in cancer.

Schematic representation of iAGR2 and eAGR2 functions in cancer. AGR2 protein primarily resides in the ER, which helps to maintain proteostasis. Biochemical dissections have not yet been performed to explain how iAGR2 engages receptors as they enter the ER, the di-sulfide shuffling that is required to produce a fully folded receptor, and how vesicles can travel to the plasma membrane to deposit membrane receptors on the cell surface. Nevertheless, this model presents a framework for designing experimental protocols to answer such questions. There are a few AGR2 interactome assays that have been used to discover client receptor proteins and this was mainly discovered by protein-protein interaction assays. For example, classical co-immunoprecipitation assays have discovered EGFR (red) and Mucin (red) as an AGR2 interaction partner in which AGR2 can form mixed disulfide bond through its CXXS pseudo-thioredoxin motif. Additionally, a recent study demonstrated that AGR2 can bind to proteins containing an AGR2 sequence-specific peptide motif (Tx[IL][YF][YF]) and this was experimentally validated in the case of EpCAM (blue). Other proteins discovered from this database mining that harbor this peptide motif, including MKS3 (TMEM67), are yet to be fully validated. Thus, receptors like EpCAM (blue) and TMEM67 (MKS3; blue) appear to exploit this specific peptide-binding function of AGR2 to accelerate membrane expression. We consider these receptors iAGR2 ‘docking-dependent’ since mutation of the consensus-binding motif precludes proper plasma membrane expression of EpCAM (34) and TMEM67 (35). AGR2 can also be secreted and this eAGR2 was recently shown to play a critical role in regulating cell growth control promoting cancer development.

Acknowledgments

This work was supported in part by grants from *La Ligue contre le Cancer, comité Charente* and by the *Fondation ARC pour la recherche sur le cancer*, who we thank for their financial support. We apologize to all our colleagues whose work could not be cited due to space limitations. MAM is supported by the Young Researcher Grant GGPM-2018-043 from UKM and by the FRGS grant from the Ministry of Education Malaysia (FRGS/1/2018/STG04/UKM/03/1). The International Centre for Cancer Vaccine Science project is carried out within the International Research Agendas program of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

561 **Table 1.** List of AGR2 interacting proteins validated by protein-protein interaction assays.
562 Each color codes for a specific cellular component.
563

564

Cellular Component	Protein name	PPI methods	AGR2 interacting Domain	Reference
Golgi, Endoplasmic reticulum	KDELR	Co-IP	KTEL motif (AA 172 to 175)	(48)
Golgi, Endoplasmic reticulum	TMED2	ERMIT, Co-IP	N-terminal/ dimer interface regions (K66 and Y111)	(34)
Plasma membrane	DAG1	Y2H	-	(15)
Plasma membrane	LYPD3	Y2H	-	(15)
Plasma membrane	MUC1	Co-IP	-	(37)
Plasma membrane	MUC2	Co-IP	CXXS motif (AA 81 to 84)	(39, 60)
Plasma membrane	Immature MUC5AC	Co-IP	-	(50)
Plasma membrane	MUC5B	Co-localization	-	(50)
Plasma membrane	PROD1 (Axolotl homolog of human CD59)	Y2H	-	(28)
Plasma membrane	EGFR	Co-IP	CXXS motif (AA 81 to 84)	(10)
Plasma membrane	EpCAM	ELISA, Co-localization, Proximity ligation assay,	Structural loop VDPSL (AA 131 to 135)	(35)
Mitochondria	UNG1 (uracil-DNA glycosylase)	Ni-NTA Affinity Purification	-	(32)
Cytoplasm	TAB2 (TGF-beta-activated kinase 1)	Tandem Affinity Purification (TAP)	-	(58)
Cytoplasm	REPTIN	Y2H	Domain (AA 104 to 111)	(33)

Figure 1

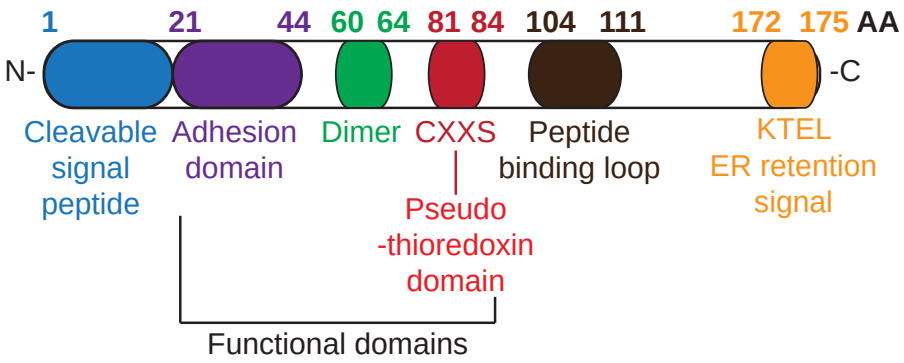

Figure 2