International PhD Program in Biomolecular Sciences
Department of Cellular, Computational
and Integrative Biology – CIBIO
34 Cycle

"Structural characterization of PTG and PTG/PP1 protein complex"

Tutor/Advisor

Graziano Lolli, PhD

Paola Storici, PhD

Ph.D. Thesis of

Marta Stefania Semrau

Academic Year 2018-2021

Table of Contents

1. ABSTRACT

PTG (protein targeting to glycogen) is a scaffolding protein that is involved in the activation of glycogen synthesis by bringing PP1 (type 1 protein phosphatase) to its substrates. It is proposed as a target for the treatment of Lafora disease (LD), a genetic disorder manifested by catastrophic teenage onset of progressive myoclonus epilepsy. In healthy neurons, PTG is downregulated by the laforin-malin complex resulting in very low glycogen production. Mutations in malin or laforin causes accumulation of PTG, which promotes glycogen synthesis by directing PP1 to glycogen synthase and glycogen phosphorylase. This results in the appearance of neurotoxic inclusion bodies formed by insoluble polyglucosans called Lafora bodies (LB), which are ultimately responsible for Lafora disease (LD). In LD mice models knocking out PTG resulted in a nearly complete disappearance of LB and resolution of neurodegeneration and myoclonic epilepsy, indicating that small molecules interfering with the PTG/PP1 interaction emerges as an excellent therapeutic strategy for LD. Up to date, there was no structural data of PTG and PTG/PP1 complex allowing for identification of potential druggable pockets. We present our efforts to obtain structural information of PTG-CBM21 and PTG/PP1 complex. We were able to set up protein production and crystallization protocols, which resulted in obtaining good quality crystals allowing for structural studies. Herein we report the first structures of human PTG: PTG-CBM21 in complex with β-cyclodextrin, PP1 in complex with PTG N-terminal peptide containing the conserved binding motif RVXF and finally PTG/PP1 holoenzyme. Our findings contribute to elucidating the interplay mechanism between PTG and PP1 and provide the basis for further structural analysis in order to identify druggable pocket.

2. INTRODUCTION

2.1. Glycogen structure and function

Glycogen is a branched glucose polymer that acts as a primary storage form of carbohydrates in many organisms, from bacteria to humans. In plants, it corresponds to starch composed of amylopectin, a highly branched polysaccharide similar to glycogen, and amylose, a straight linear polymer of glucose molecules (1, 2). In mammals, a significant amount of glycogen is present mainly in the liver and skeletal muscle (4); however, many other cell types, such as cardiac and smooth muscle, the kidney, the brain, and adipose tissue, are also capable of glycogen synthesis (1). Muscle glycogen serves as a rapid glucose fount only for muscle cells, not supplying any other cells, unlike liver glycogen, which is crucial for blood glucose homeostasis. It acts as a fast energy source for tissues when the glucose level in the organism decreases, releasing glucose into the bloodstream upon glycogen breakdown.

The whole molecule is a hydrophilic sphere, organized in 'tiers' of carbohydrate, allowing mammalian cells to store large amounts of energy in a soluble, rapidly accessible form. It can contain up to 55,000 glucose residues (1, 3, 7). The glucose residues are linked by α-1,4-glycosidic bonds forming a chain of 8-12 units (6), connected by α-1,6-glycosidic linkages forming branches. Each branch finishes in non-reducing ends that are the sites where both synthesis and degradation occur. Classical structural biology methods cannot determine the detailed structure of glycogen as it is present in different sizes and contains the different locations of branch points. A model for glycogen structure has been proposed that divides the chains into two categories: B-chains, present in the inner part of the particles, which would typically contain two branch points, and outer A-chains, which are unbranched (16). Chemical analysis of mammalian glycogen suggests that the average chain length is ~13 residues (6). In this model, glycogen would be organized in a series of tiers. The amount of glucose residues grows exponentially with each tier, and the outermost tier of any molecule would contain 50% of the total glucose of the particle as unbranched A-chains (Fig.1). Glycogen is highly hydrophilic and is found in the cytosol of cells (4).
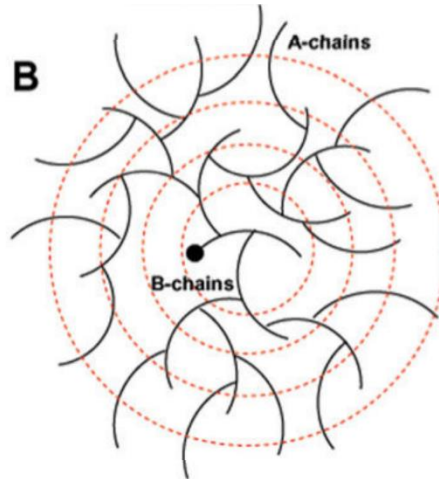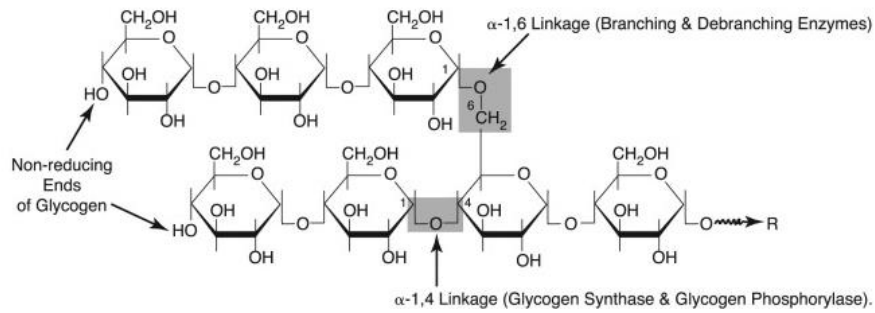
Fig 1. Glycogen structure: (A) Polymerizing α-1,4-glycosidic linkages and a branching α-1,6-glycosidic linkage are shown. (B) The tiered model for glycogen organization in which inner B-chains on average carry two branches and the outer A-chains are unbranched. The black circle denotes glycogenin

There have been three types of glycogen particles identified by EM: α- and β-granules and γ-particles (Fig 2) (11). The β-granules include a central protein-glycogenin (GN). The particles can reach ~10–44 nm in diameter and ~106–107 in molecular weight and can be found mainly in muscles and the brain. In contrast, the liver α-granules are composed of several β-granules forming rossetta shaped aggregations and are considered a slower energy source. They can be as large as 300 nm in diameter and >108 in molecular weight (1, 3, 12, 13). The process of how the α-granules are formed is still unclear and probably involves glycogenin in the linkage mechanism (15). The γ-particles are the subunits of both α- and β-structures (14).
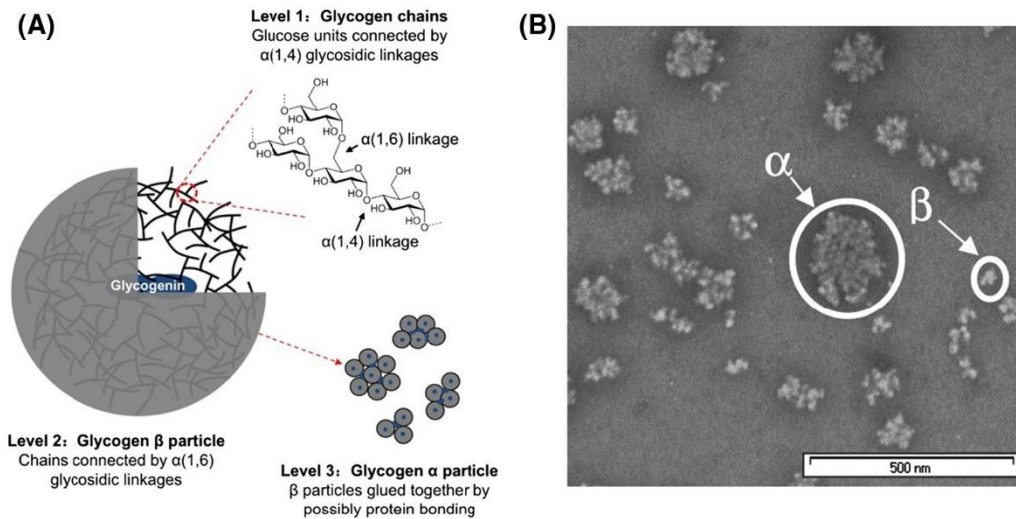
Fig 2. A scheme showing the three levels of glycogen hierarchical structures (A) and a TEM image of mouse liver glycogen showing glycogen α and β particles (B) (10)

Glycogen is known to interact with associated proteins forming assemblies called 'glycogen particles.' Those proteins can bind to glycogen, sometimes to each other and membranes (17, 18, 19). Up to date, the known glycogen-associated proteins are the glycogen synthesis initiator- glycogenin (GN), the metabolic enzymes as glycogen synthase (GS), glycogen phosphorylase (PH) and the debranching enzyme (DBE), and several modulators as phosphorylase kinase (PHK), protein phosphatase 1 (PP1) and its regulatory subunits (PPP1R). Other proteins such as laforin, Stbd1 (starch-binding domain protein 1)/genethonin 1, and AMPK (AMP-activated protein kinase) contain CBM20 (carbohydrate-binding module 20) domain and has been shown to bind glycogen (20, 21, 22, 23, 24) (Fig 3).

Fig 3. Glycogen particle with associated proteins: GN- glycogenin; GS-glycogen synthase; PH-glycogen phosphorylase; DBE-debranching enzyme; PP1-protein phosphatase 1; PP1 regulatory subunits: RGL, GL, PTG; LF-laforin; PH kinase-glycogen phosphorylase kinase; AMPK- AMP-activated protein kinase; Stbd1-starch-binding domain protein 1)/genethonin 1 (1).
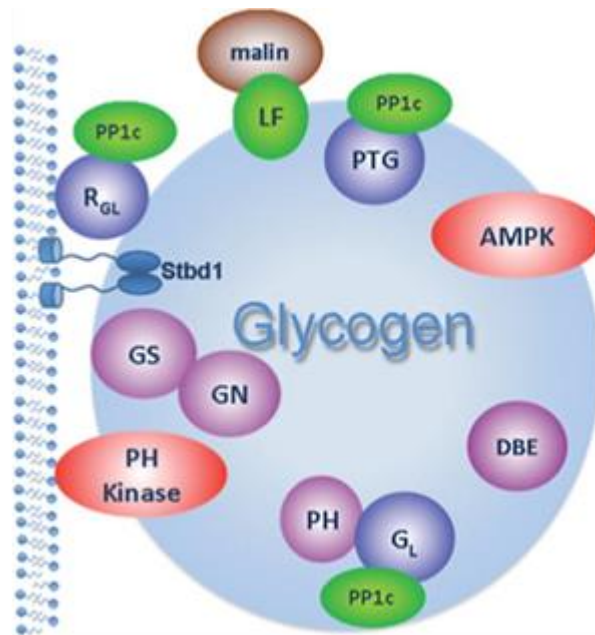
2.2. Glycogen synthesis

Glycogen metabolism is tightly regulated through tissue-specific mechanisms responding to differentiated missions, i.e., blood glucose homeostasis in the liver and mechanical work in muscles. Different organ-restricted isozymes participate in glycogen metabolism, being variably regulated both allosterically and post-translationally.

The glycogen synthesis is triggered by the presence of the glucose molecules derived from consumed carbohydrates or gluconeogenic precursors, such as lactate and alanine, from a process called gluconeogenesis. During the first step of the synthesis, glucose molecules are transformed into glucose-6-phosphate (Glc6P) by glucokinase in the liver and hexokinase in other tissues. Next, a reversible enzyme phosphoglucomutase (PGM) transfers the phosphoryl group from C6 to C1, converting Glc6P into glucose-1-phosphate (Glc1P). Further, Glc1P is transformed into uridine diphosphate-glucose (UDP-glucose) in the reaction catalyzed by UDP-glucose pyrophosphorylase. The next step is the generation of the primer molecule, which is initiated by a glucosyl transferase enzyme called glycogenin (GN). GN dimerizes and autoglycosylates by attaching a glucose molecule from UDP-glucose to its tyrosine residue (Tyr-194) and extends the

9

chain by adding additional 6 or 7 glucose molecules via alpha α-1,4-glycosidic bonds (5, 25). At that moment, glycogen synthase (GYS) associates with glycogen branching enzyme (GBE), elongating the glucan polymer and building the glycogen particle on the protein backbone (Fig. 4).
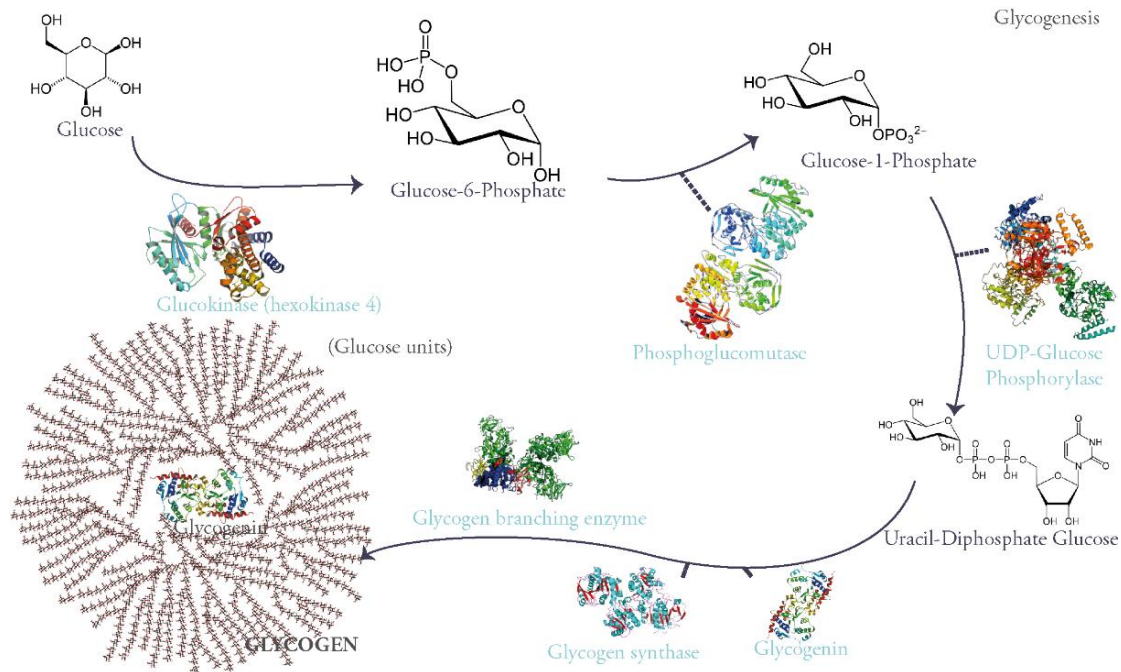
Fig 4. Glycogenesis

2.3. Glycogen degradation

Glycogen degradation, or glycogenolysis, requires the coordinated action of two key enzymes: glycogen phosphorylase (GP), which cleaves the α-1,4 glycosidic bond between the glucosyl residues at the non-reducing ends of the glycogen chain, and debranching enzyme (GDE), which is responsible for the α-1,6 glycosidic bonds breakdown (1-5). GP proceeds until the external chains arrive at a branch point, where GDE intervenes, acting as a transferase and moving the outer three glucosyl units to the non-reducing end of the glycogen molecule, enabling further GP activity. Upon glycogen degradation, Glc1P is released and gets converted to Glc6P by phosphoglucomutase. The metabolic fate of Glc6P depends on the tissue; in most cases, it is used internally as an energy source; however, in liver, it gets dephosphorylated to glucose by glucose-6-phosphatase and secreted by the cell into the bloodstream (5).

2.4. Glycogen regulation

Glycogen metabolism is regulated by a complex interplay involving various allosteric factors and a series of synchronized phosphorylation events. Glycogen synthesis is stimulated by the presence of insulin, whereas the degradation is triggered by epinephrine (muscles) and glucagon (liver). The key players in the regulation are glycogen synthases and phosphorylases. GS is regulated by glucose-6-phosphate and by reversible phosphorylation of at least nine different sites catalyzed by a range of kinases, including glycogen synthase kinase 3 (GSK3), protein kinase A (PKA), or phosphorylase kinase (PhK) (59-61). The phosphorylated form of GS is inactive; on the contrary, GP gets activated by phosphorylation, mediated by phosphorylase kinase. Moreover, GP gets stimulated by AMP and inhibited by glucose-6-phosphate and ATP. Shortly, phosphorylation of both enzymes results in glycogen degradation. On the other hand, to initiate glycogen synthesis, GS needs to get 'switched on' and GP 'switched off', which occurs by dephosphorylation events mediated by protein phosphatase 1 (PP1) (Fig. 5).
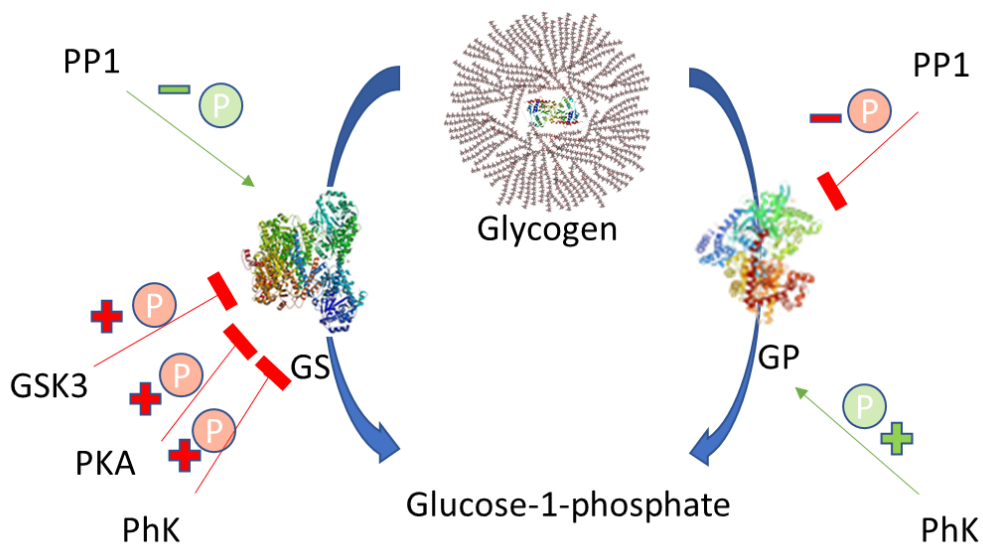
Fig 5. Schematic overview of glycogen metabolism regulation (GS glycogen synthase, GP glycogen phosphorylase, PP1 protein phosphatase 1, GSK3 glycogen synthase kinase 3, PhK phosphorylase kinase)

PP1 is a serine/threonine phosphatase that is involved in many other cellular processes like cell growth and protein synthesis, other than glycogen metabolism. Its specificity is driven by tight interaction with more than 200 known targeting proteins (62, 63). The

regulatory subunits directing PP1 towards GS and GP are called PPP1R3. In humans, there have been seven different glycogen-related subunits identified, and none of them share more than 40% of aminoacidic identity (63-66). They are characterized by different tissue distribution. The most known subunits are GM of 124-126kDa, encoded by PPP1R3A, present in the skeletal muscles and heart, GL of 33kDa, encoded by PPP1R3B, specific for liver and PTG (protein targeting to glycogen) of 36kDa, encoded by PPP1R3C which is ubiquitous among tissues, with prevalence in astrocytes. The role of the remaining subunits is poorly understood. Overexpression of PTG, GM, and GL leads to glycogen accumulation (69).

All of the mentioned proteins contain a canonical PP1 binding motif, RVXF, and carbohydrate binding domain CBM21 (67, 68), which is crucial for the localization of the PP1 phosphatase to glycogenic substrates.

2.5. Glycogen in neurons

In the brain, a relatively small quantity of glycogen has been identified compared to other organs such as the liver or muscles, suggesting that it is unlikely to act as a conventional energy reserve (73). Interestingly, glycogen is almost exclusively stored in astrocytes and only in the marginal amount in a neuron. For that reason, all the glycogen functions in the brain have been associated with astrocytes. Neurons exhibit higher energy consumption than glial cells, suggesting that they take up most glucose (75). Thus, astrocytes would be expected to metabolize glycogen and provide glucose to fulfill energetic neuronal needs.

Nonetheless, glucose-6-phosphatase is not present in astrocytes; hence, they cannot release free glucose for the benefit of neurons. However, during high energy demand, they can use it internally. Another possible pathway has been discussed, which connects glycogen degradation to lactate production, that later can be utilized by neurons as an energy source. The hypothesis is called the astrocyte-neuron lactate shuttle (ANLS) (Fig. 6). In this scenario, astrocytes metabolize glucose to lactate, which is subsequently shuttled to neurons, used for energy production. Shortly, glucose is taken up by astrocytes and can either be stored as glycogen or be metabolized to pyruvate, which further is converted by the enzyme lactate dehydrogenase into lactate and finally transferred to neurons, where is converted by the enzyme lactate dehydrogenase

isoform 1 (LDH-1) back into pyruvate and further metabolized producing energy. Withal, this interpretation is still widely debated.
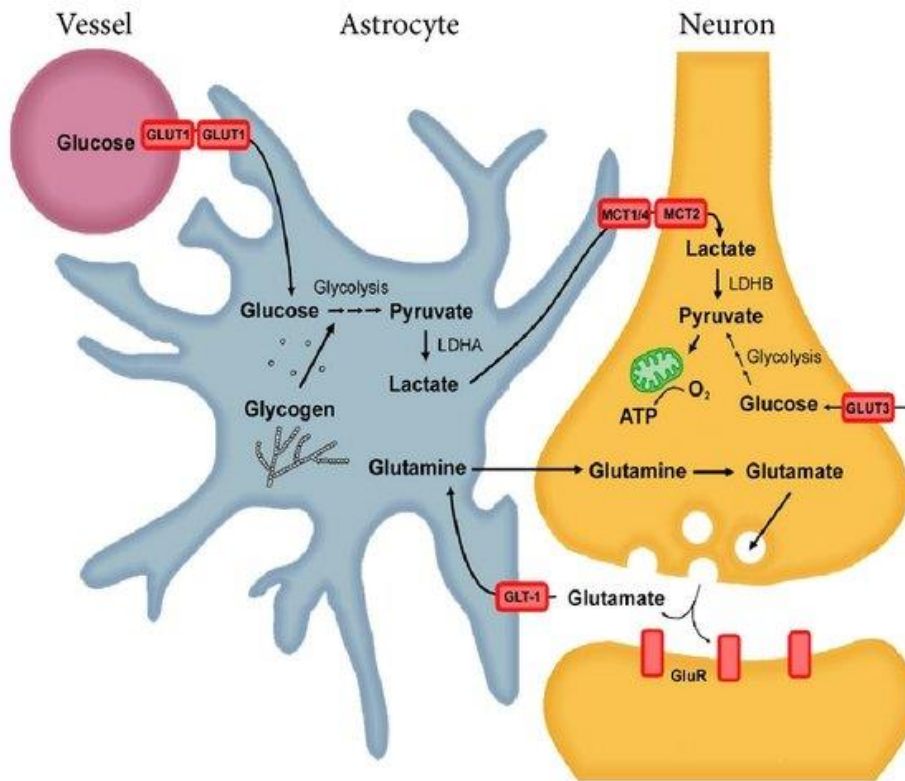
Fig. 6. Astrocyte-neuron lactate shuttle (ANLS) hypothesis (76).

Interestingly, neurons possess their own active machinery for glycogen synthesis (70), which is suppressed by inactivation of GS through extensive phosphorylation and an additional regulatory pathway involving laforin- a dual-specificity phosphatase and malin-E3 ubiquitin ligase (72).

Human laforin acts on threonine/serine and tyrosine residues, and EPM2A encodes it. It contains an N-terminal carbohydrate-binding module (CBM, family 20), which dictates its location in the vicinity of glycogen metabolizing enzymes (84). It is associated with malin-an E3 ubiquitin ligase carrying a RING and six NHL-repeat domains that are encoded by EPM2B. Laforin-malin complex is shown to mainly attach K63-linked polyubiquitin chains, which suggests targeting for autophagic degradation (85, 86). However, there is also evidence of K48-linked polyubiquitination by malin, which targets proteasomal breakdown (88). Malin has been shown to ubiquitinate laforin (90), GS (89), GDE (91),

and also a PP1 targeting subunit PTG (87), which all are involved in glycogen metabolism.

Laforin, on the other hand, has been linked to several substrates that include GSK3, malin, GS, and PTG, as well as glycogen itself. It has been reported that it acts as glucan phosphatase, preventing glycogen hyperphosphorylation (92-94). Studies show that malin requires laforin to ubiquitinate GS and PTG, suggesting a role of laforin as a scaffold protein directing malin to its substrates (89, 91, 103). This observation indicates that the laforin–malin complex has a crucial role in the control of glycogen regulation by targeting glycogen proteins for degradation. The laforin-malin PTG degradation is linked to AMPK signaling, where phosphorylation of PTG by AMPK increases its ubiquitination (104).

In summary, the activity of malin-laforin complex contributes to the downregulation of glycogen synthesis by inactivation of GS and PTG (Fig 7).
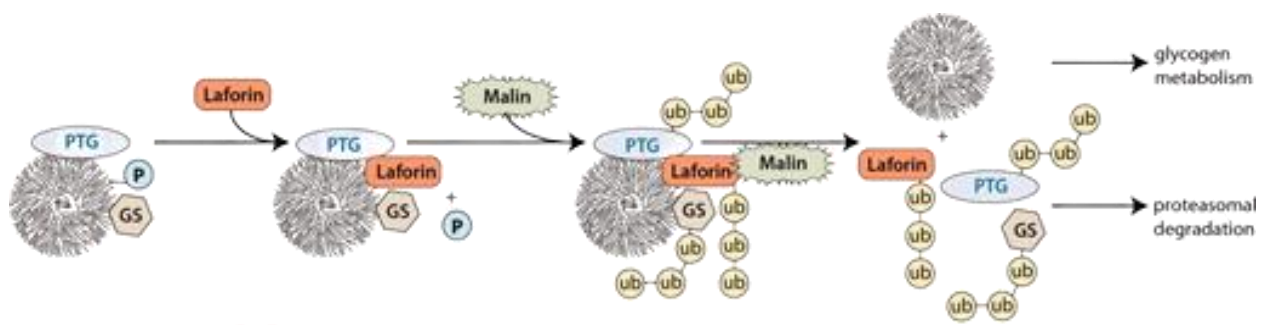
Fig. 7 A scheme of glycogen regulation by laforin-malin complex. In short, PTG and GS bind to glycogen. Laforin interacts with glycogen and cleaves one of its phosphates. Next, malin recognises laforin and ubiquinates it together with PTG and GS, which are finally targeted for proteosomal degradation (95)

Recent studies reported that neurons solely express the brain's GP isomerase, while astrocytes have two forms: brain and muscle (74). They undergo different activation modes. Muscle isoform is regulated mainly through phosphorylation as a reaction to extracellular factors, while the brain demonstrates sensitivity to AMP levels, suggesting a response to energetic internal needs (71).

The reason why neurons can synthesize glycogen is not yet apparent. Recent findings have linked neuronal glycogen to protective response during hypoxia stress (71). The mechanism behind that phenomenon is still unknown. Moreover, glycogen has been shown to play an essential role in memory formation and learning. A mouse model

knocked-out for GS in forebrain neurons demonstrated significant impairment in learning capacity and acquisition of new motor and cognitive properties (77). Still, little is known about glycogen's role in the brain, leaving many aspects of its relevance and regulation obscure. However, its altered metabolism is associated with neural apoptosis and severe disorders, such as Lafora Disease, among others (76-78).

2.6. Lafora disease

Lafora disease (LD) is a fatal autosomal recessive neurodegenerative disorder characterized by the formation of abnormally branched glycogen inclusions, called Lafora bodies. It is characterized by progressive myoclonus epilepsies and has a poor prognosis. It usually manifests in previously healthy adolescents, whose condition rapidly deteriorates. Early epileptic symptoms can be managed with antiepileptic drugs; however, their efficacy drops down with disease progression. LD patients typically die within ten years of the onset of the symptoms due to complications related to nervous system degeneration (87, 90, 95, 96).

Most of the cases have been diagnosed in the Mediterranean countries as Spain, France, and Italy and central Asia, India, Pakistan, northern Africa, and the Middle East. It is a rare orphan disease, and no cure is available. LD has been linked to mutations in the EPM2A (epilepsy, progressive myoclonic type 2A) or EPM2B genes, encoding for laforin and malin, respectively (Fig. 8) (97, 98, 100).
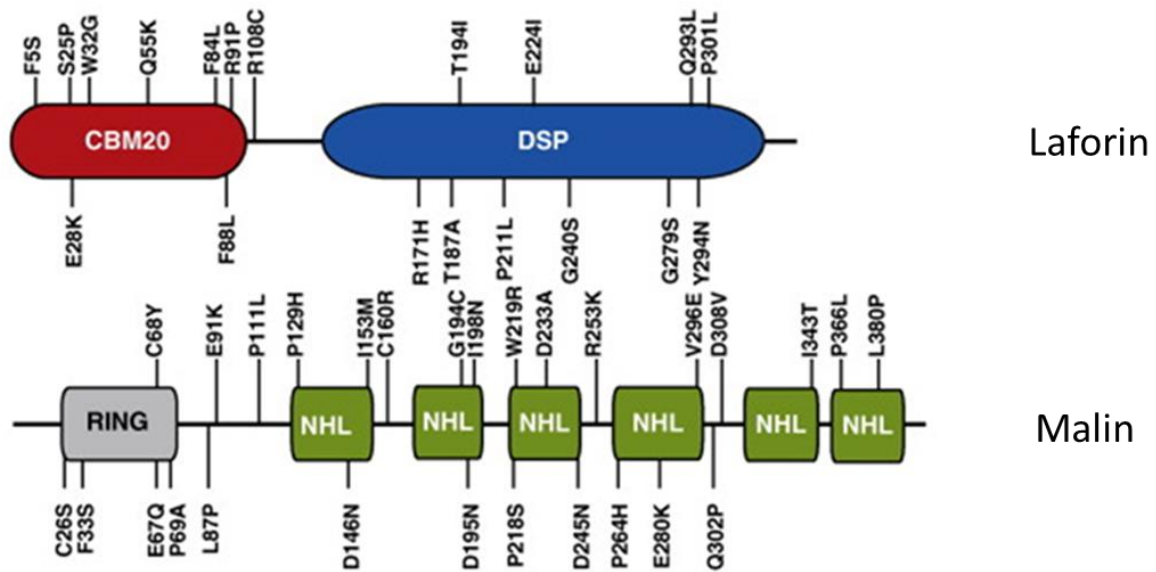
Fig 8. A schematic representation of laforin and malin with their mutations found in LD. Laforin contains N-terminal CBM (carbohydrate binding domain) and C-terminal DSP (dual specificity phosphatase domain. Malin contains N-terminal RING domain that bind ubiquitination enzymes and their substrates, and 6 NHL repeats. NHL domains are usually involved in protein-protein interactions (96).

Malfunction of laforin-malin complex causes synthesis of insoluble glucose polysaccharides that precipitate and aggregate into toxic inclusions called Lafora bodies (LB) present in the cytoplasm of nearly all cell types (Fig. 19) (79, 80). The mechanism of the aggregation is not precise. LB is characterized by fewer α-1,6-glycosidic branches, which promotes insolubility. Moreover, LBs contain much higher phosphate quantities than glycogen (80). A recent report suggested that laforin involvement in phosphate removal from glycogen could be a damage control mechanism (101).

Fig. 9. Schematic representation of Lafora disease regulation by laforin-malin complex in neurons.

Recent studies, however, demonstrate that the phosphatase activity of laforin, does not prevent the disease. The mouse model expressing inactive laforin does not manifest LD phenotype (102). Moreover, it is reported that muscle and brain glycogen from those mic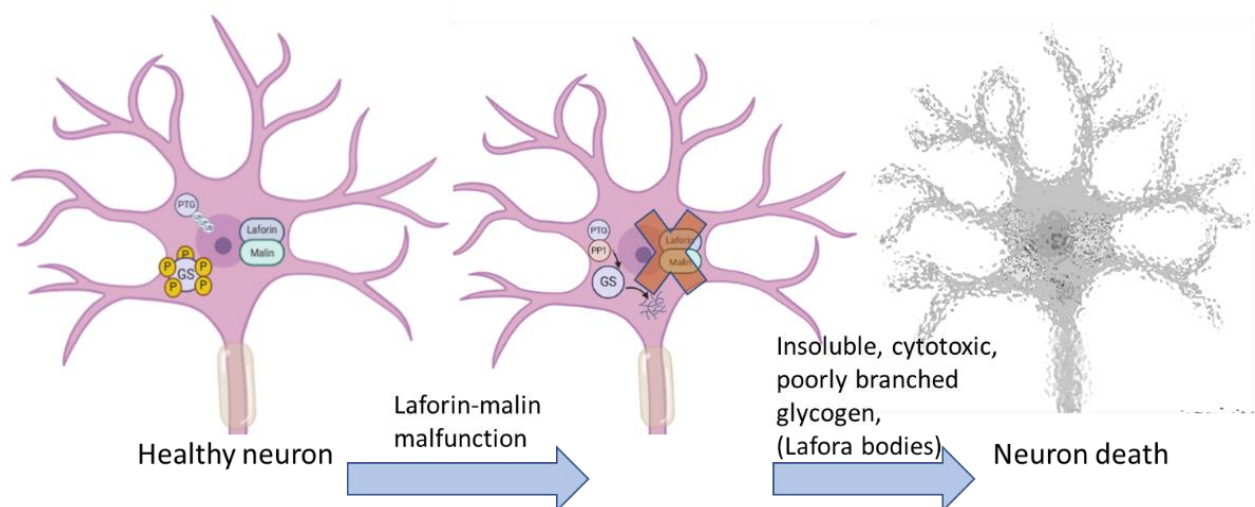e are still hyperphosphorylated, but the chain length of glycogen is completely restored in both tissues. These results open a question of the cause for LB formation (Fig. 10)
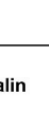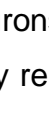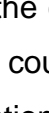

	Glycogen and phenotypic parameters in different mouse lines				
Genotype	Wild-type	*Epm2a*$^{-/-}$	*Epm2b*$^{-/-}$	*Epm2a*$^{-/-}$.C266Slaf	*Epm2b*$^{-/-}$.C266Slaf
Expression of laforin and malin	Laforin Malin	Malin	Laforin	Laforin Malin	Laforin Laforin
C6 phosphate in brain glycogen	normal	elevated	elevated	elevated	elevated
Chain length distribution in brain glycogen	normal	abnormal	abnormal	normal	abnormal
Lafora bodies in the hippocampus	none	plenty	plenty	none	plenty

Fig. 10. LD phenotypes correlation to different laforin-malin expression form. A phosphatase inactive laforin (C266S) together with functional malin cause glycogen hyperphosphorylation, however does not lead to LB accumulation, which suggest that hyperphosphorylation does not cause LB formation. (102)

Many different mutations in EPM2A and EPM2B genes have been identified in patients (Fig. 8) who manifests similar clinical phenotype. Mutations in EPM2A can be divided into two types, one linked to Laforin phosphatase activity impairment and the second one being associated with the inability to interact with other proteins as malin and PTG. On the other hand, all mutations in EPM2B cause loss of function. Mutations associated with EPM2A lead to faster disease progression than in patients with EPM2B alternations (108).

Another puzzling characteristic of LD is the fact that LBs accumulate in many cell types; however, only neurons undergo apoptosis. One possibility could be involved in a disruption of energy release by the interference of trapped energy accumulation in the form of LBs. The second explanation is linked to the trafficking pathway, where LBs cause blockage in the cytoplasm, which triggers death because of their important size. Neuronal apoptosis could also be attributed to their age, as the toxicity of LBs increases with their accumulation over time. Other cell types have a much shorter lifespan than neurons, meaning that they cannot live long enough to experience the toxic LBs effect (105).

Another interesting scenario proposes that LBs are not a cause of LD, but a defensive mechanism of the cell towards glycogen metabolism alternations, suggesting that LD originates from a malfunction in autophagy or proteasomal degradation (106, 107). Autophagy is a degradation process that removes unnecessary cellular content as proteins, protein complexes/oligomers, organelles, and certain intracellular pathogens via autophagosomes and targets them into a lysosomal breakdown. It is crucial in maintaining cellular homeostasis by disposing of disease-associated aggregates (109). It has been reported that laforin is a positive regulator of autophagy, via the mammalian target of rapamycin kinase (mTOR), which gets inhibited during laforin overexpression. In in vitro and in vivo studies, a decreased autophagosome formation was observed when laforin activity was impaired (110).
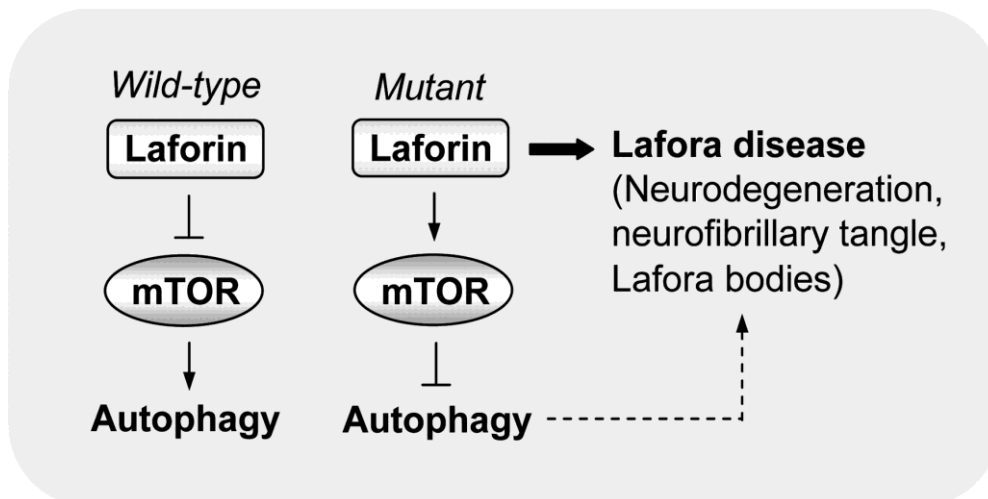

Fig. 11. Interplay between laforin and mTOR in autophagy (110).

2.7. PTG as emerging target in LD therapy development

Protein targeting to glycogen (PTG), described in the previous section, is one of the regulatory subunits of PP1 and is responsible for glycogen synthesis activation by bringing PP1 to its glycogenic substrates GS and GP. PTG is composed of 317 amino

acids protein containing carbohydrate binding domain CBM21 and N-terminal canonical PP1 recognition signature sequence RVxF (Fig. 12).
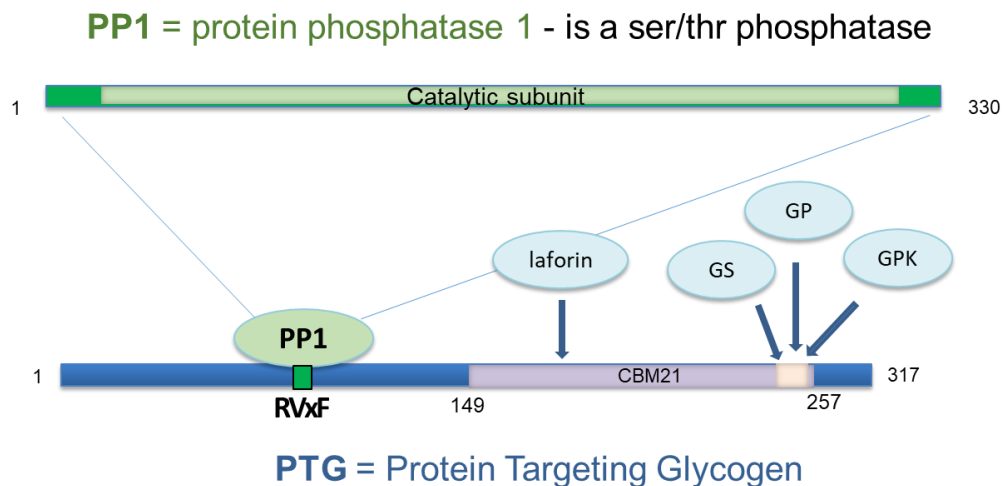
Fig. 12. Schematic representation of PTG protein with proposed interaction domains. PP1-protein phosphatase 1, GS-glycogen synthase, GP-glycogen phosphorylase, GPK-glycogen phosphorylase kinase. RVxF-conserved PP1 interaction motif.

PTG has a relevant role in glycogen metabolism in the brain, where its levels are post-transcriptionally regulated. It binds to glycogen and other glycogen-associated enzymes as GP and GS (111). Overexpression of PTG in neurons induces activation of GS, which is dephosphorylated by PP1 and causes an increase in glycogen accumulation. This result suggests that in neurons, PTG is downregulated in order to prevent glycogen synthesis. Laforin-malin complex plays an essential role in maintaining low levels of PTG in neurons by targeting it to proteasomal degradation via ubiquitination.

In Lafora Disease, that regulation is interrupted by alternation in laforin-malin complex activity, leading to extension of glycogen chains. As there is no active laforin present, the glycogen remains hyperphosphorylated, which causes it to unfold and precipitate. GS stays bound to the precipitated glycogen and continues chain elongation, while glycogen branching enzymes fails to associate tightly, thus no branching occurs and insoluble polyglucosan forms (Fig. 9).

One of strategies for treatment development for LD would be directly decreasing GS activity, which would cause important side effects as hyperglycemia and cardiomyopathy. Another approach could be inhibiting PP1 as it is responsible for GS

activation. However, PP1 is an essential enzyme involved in many cellular processes other than glycogen metabolism (from cell growth to protein synthesis and regulation of membrane receptors and channels). Its inhibitors, used for research purposes, are carcinogenic and PP1 is not a suitable pharmacological target. Nonetheless, PP1 requires PTG in order to activate GS. Knocking out PTG in LD mouse model resulted in a nearly complete disappearance of LB and resolution of neurodegeneration and myoclonic epilepsy, revealing PTG as a promising target for the inhibition (26,27). In addition, aiming specifically for PTG would not inhibit GS activity in muscles or liver at the same level as in the brain, as those organs have their own PP1 regulatory subunits and are able to compensate PTG inhibition and reduce side-effects. This thesis is supported by observation, that in PTG knockout mice the production of glycogen in skeletal muscles was maintained at the level of 70% of the normal condition (26, 27). All the above findings point out PTG as a promising target for small molecule development, that would aim at interference with PTG mediated glycogen synthesis activation cascade (Fig. 13).
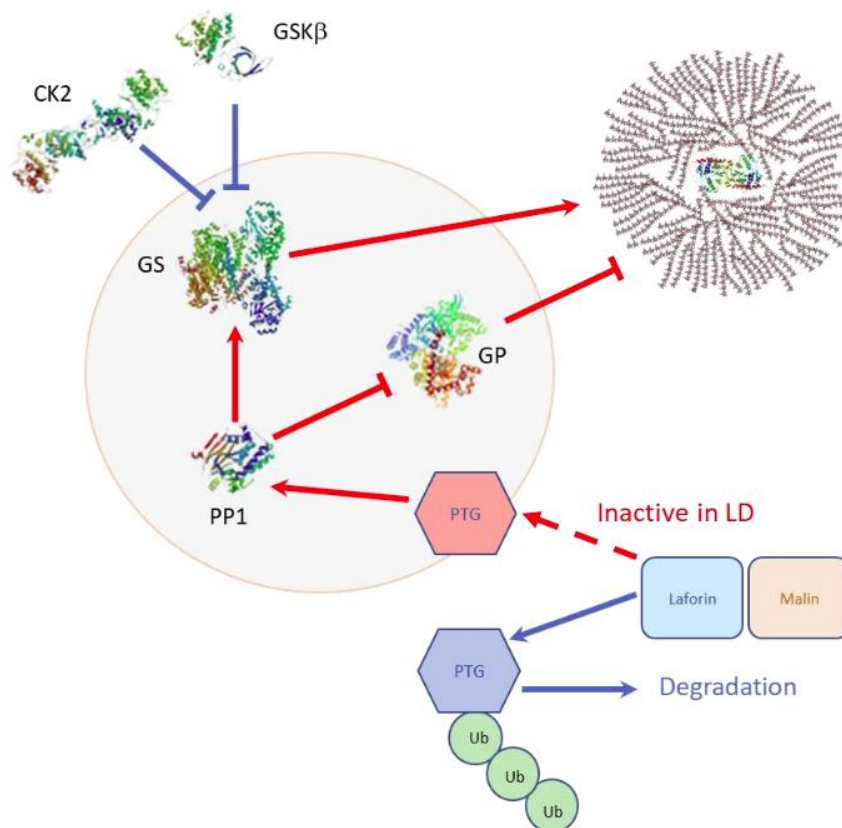

Fig. 13. An overview of possible LD therapy targeted interaction, with main focus on disrupting interactions of PTG/PP1/GS and PTG/PP1/GP assembly.

3. AIM OF THE STUDY

Currently there is no cure available for Lafora disease (LD) and the treatment is based on the signs and symptoms present in each person, mainly helping with seizures managing. Recent studies show that knocking out PPP1R3C, also named protein targeting to glycogen (PTG), in LD mice models resulted in a nearly complete disappearance of Lafora Bodies (LB) and resolution of neurodegeneration and myoclonic epilepsy, revealing PTG as a promising drug target (26,27). PTG belongs to a family of regulatory subunits of protein phosphatase 1 (PP1), a glycogen synthase (GS) activator in the glycogen synthesis pathway. PTG is ubiquitous; however, in case of its inhibition, mechanisms exist in various organs (but not in the brain) that can compensate for the loss of PTG-mediated PP1 recruitment to GS, as they have specific PP1 regulatory subunits: G_M, present in skeletal muscle and heart and G_L, most abundantly expressed in liver. Those observations open the window of opportunity for developing LD therapy that would limit severe side effects.

Little is known about the PTG/PP1/GS interaction mechanism and detailed structural investigation is essential to unlock that missing information. Up to date, there is no structural data reported of PTG protein alone or in the complex with PP1, which would help understand the basis of the interplay between PTG and its partners. The crystallographic analysis is crucial for the identification of potential druggable pockets for small molecule design.

The project aims at obtaining soluble PTG protein and PTG/PP1 complex suitable for structural studies, starting from construct design, cloning, expression and purification with the final focus on their biophysical and structural characterization by X-ray crystallography and SAXS. Understanding the PTG protein architecture and PTG/PP1 interaction mechanism would provide the basis for the rational design of selective compounds targeting PTG interaction with its partners.

4. MATERIALS AND METHODS

4.1. Construct design

The PTG sequence was analyzed using PSIPRED server (32) for protein secondary structure prediction and three expression constructs were designed: one longer (aa 70-264) containing PP1 binding site and carbohydrate binding motif (CBM) and 2 shorter containing CBM (aa 132-264 and 142-264) (Fig. 13).
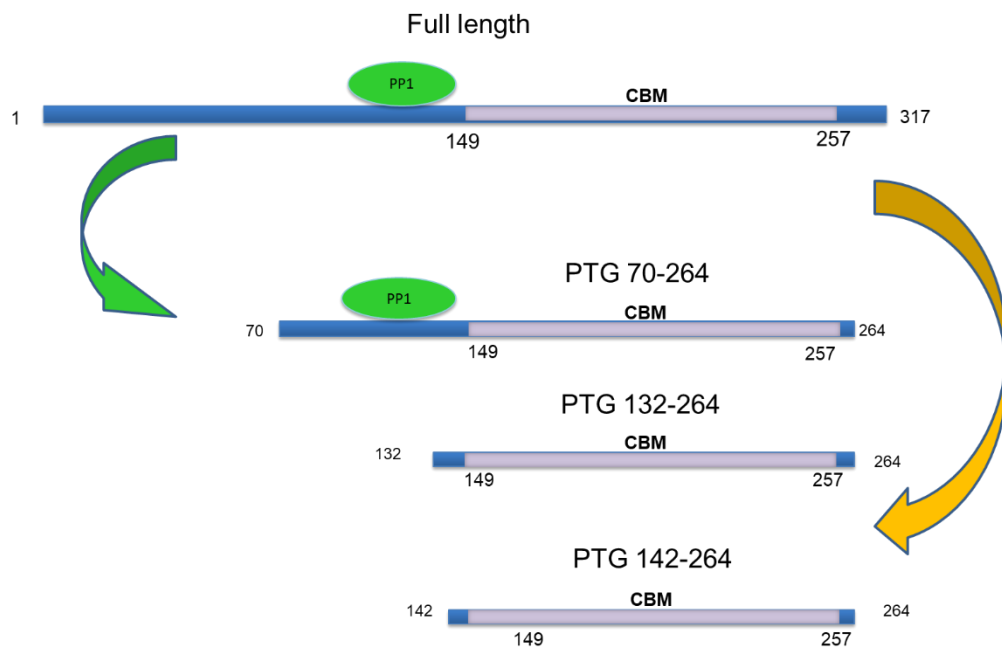
Fig. 14. Schematic description of different designs of PTG constructs

4.2. Protein expression and purification

4.2.1.   PTG 70 264

The *E. coli* codon optimized gene encoding for human PTG (aa 70-264) was cloned into pGEX-6-3 vector using BamHI and NotI and purchased from GenScript. The construct contains N-terminal GST tag cleavable by GST-3C (PreScission®) protease.

GST-3C-PTG 70-264

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYI
DGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETL
KVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPK
LVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQ/GPLGSSQ
NDWKCSHNQAKKRVVFADSKGLSLTAIHVFSDLPEEPAWDLQFDLLDLNDISSALKHH
EEKNLILDFPQPSTDYLSFRSHFQKNFVCLENCSLQERTVTGTVKVKNVSFEKKVQIRI
TFDSWKNYTDVDCVYMKNVYGGTDSDTFSFAIDLPPVIPTEQKIEFCISYHANGQVFW
DNNDGQNYRIVHVQWKPD
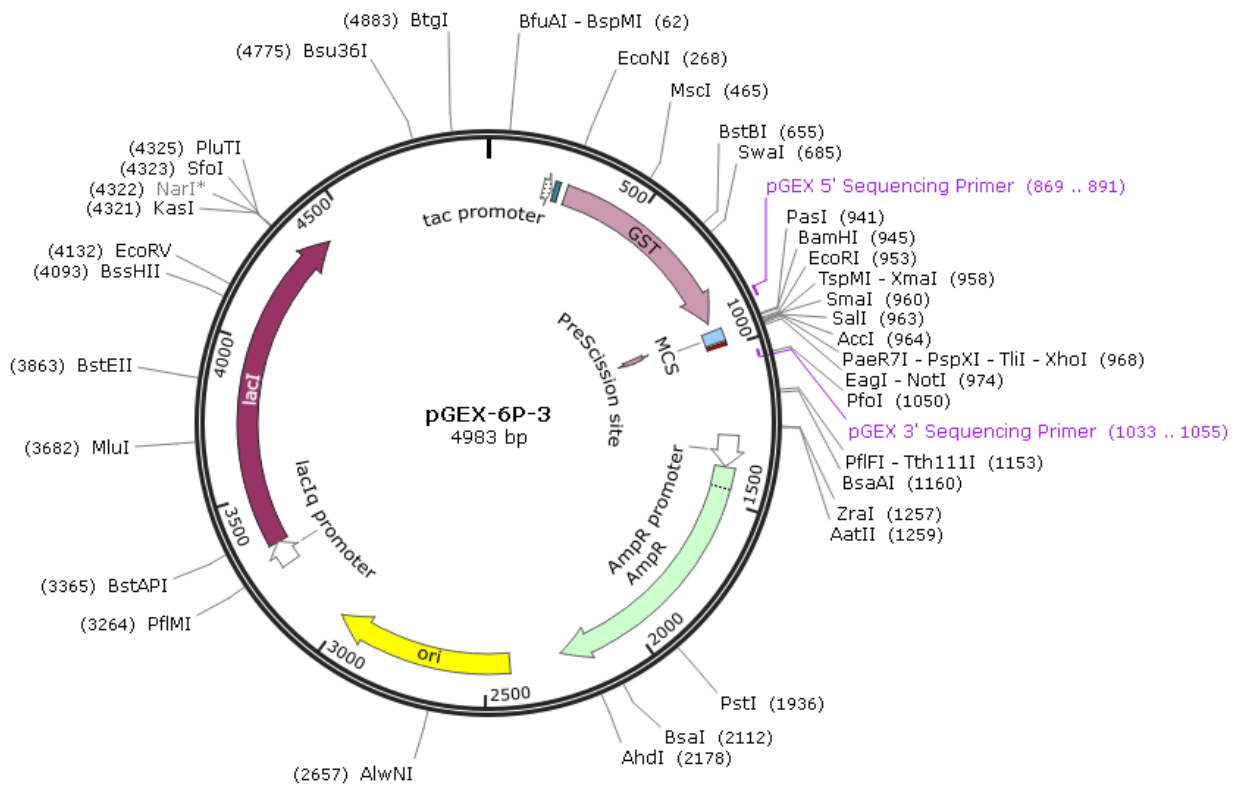
Expression

Extensive small-scale expression trials were performed in 24 deep well plate in 3 ml cell culture testing different conditions:

- Cell strains: DH5alpha, BL21(DE3), BL21(DE3) pLysS, C41(DE3)
- Medium: LB, TB, 2xYT, autoinduction
- IPTG concentrations: 0.1mM and 0.5mM
- Temperature: 37°C and 18°C

After failing in obtaining soluble protein further optimisation was introduced by co-expressing the protein with chaperonin GroE complex that is composed of 2 subunits GroES (10 kDa) and GroEL (57 kDa). GroE supports protein expression and folding ability, helping proteins to form tertiary structures upon translation. Bl21(DE3) competent cells were transform with pGro7 vector (Takara Bio) and new stock of chemically competent cells was prepared resulting in strain BL21 (DE3) pGro7. Expression trials were performed in 24 deep well plate in 3ml of cell culture testing:

- Medium: LB, TB
- IPTG concentrations: 0.1mM, 0.5mM and 1mM
- Temperature: 37°C and 18°C
- Arabinose concentrations: 0.5mg/ml and 2mg/ml
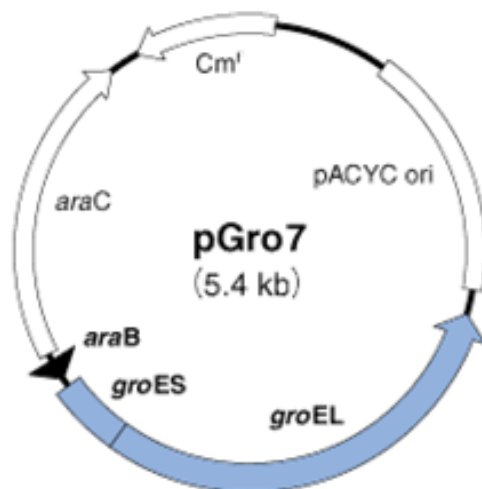
Scale up protein expression was performed in LB medium supplemented with 100ug/ml of ampicillin, 34 ug/ml of chloramphenicol and 0,5mg/ml arabinose for chaperonin complex expression induction. The PTG expression was induced with 0,1mM 1-thio-β-D-galactopyranoside (IPTG) at OD_{600}~ 0,6 and cell culture was grown at 18°C for 18-20 hours. Cell culture was harvested by centrifugation at 3000xg at 4°C for 30 minutes, washed with cold 1xPBS+10% glycerol, centrifuged at 3000xg at 4°C for 30 minutes and stored at -80°C until further use.

Purification

Cell pellet was resuspended in lysis buffer (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 10mM DTT, 10mM $MgCl_2$, 10μg/ml DNase I, 1x complete protease inhibitor cocktail (Roche)) and lysed by 2 cycles of homogenization at ~ 1000bar (PandaPLUS 2000, GEA). Soluble fractions were separated by centrifugation at 30.000xg at 4°C for 1 hour. The supernatant was incubated with Protino Glutathione Agarose 4B resin (MACHEREY-NAGEL) for 1 hour at 4°C in mild agitation. The resin was loaded on polypropylene column (Bio-Rad) and extensively washed with wash buffer 1 (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 2mM DTT), followed by 2-3 column wash with buffer 2 (100mM Tris pH 7.5, 50mM sucrose, 10mM KCl, 25% glycerol, 20mM MgCl2, 5mM ATP, 1mM DTT) in order to wash off the excess of chaperonin. Further the resin was rinsed with additional 10 column volumes (CV) of buffer 1. GST-tag was removed on-column by addition of GST-3C (PreScission®) protease directly to the resin in 1 CV of wash buffer 1 and incubated over-night at 4°C. Flow-through fraction containing cleaved PTG was collected and concentrated using 10k MWCO centrifugal concentrator (Pierce™, ThermoScientific) and further purified on size-exclusion chromatography column Superdex 200 10 300 GL (Cytiva). The fractions containing PTG protein were pooled together, concentrated using 10k MWCO concentrator, flash frozen in liquid nitrogen and stored at -80°C.

4.2.2.   PTG CBM 132-264 and 142-264

Cloning

The DNA sequence containing CBM21 (residues 132-264 and 142-264) were PCR amplified from the synthetic PTG 70-264 template and inserted by LIC cloning into the set of vectors:

- pGTvL1: N-GST-TEV

- pNH-Trxt: N-6xHis-Trxt-TEV

- pNIC_28_Bse4: N-6xHis-TEV

- pNIC_ZB: N-6xHis-ZB-TEV

Primers for LIC cloning were ordered from Eurofins Genomics:

PTG CBM 132-264	sequence
Forward LIC1	TACTTCCAATCCAAGAACCTGATTCTGGATTTCCCG
Reverse LIC1	TATCCACCTTTACTGttaATCCGGTTTCCATTGAACGTG
PTG CBM 142-264	
Forward LIC1	TACTTCCAATCCAGCACCGACTACCTGAGCTTCCG
Reverse LIC1	TATCCACCTTTACTGttaATCCGGTTTCCATTGAACGTG

In red: LIC1 sequence, in green: stop codon

Expression

Extensive small-scale expression trials were performed in 24 deep well plate in 3 ml cell culture testing different conditions:

- Cell strains: BL21(DE3), BL21(DE3) pLysS, C41(DE3)
- Medium: autoinduction ZYM-5052 (Studier, 2005)
- Temperature: 18°C overnight

For the scale-up PTG CBM 132-264 construct cloned in pNIC28-Bsa4 vector with N-terminal 6xHis-tag and TEV cleavage site was chosen. Expression was performed using BL21 (DE3) competent cells grown in autoinduction medium ZYM-5052 at 18°C for 18-20h.

Purification

Cell culture was harvested by centrifugation at 3000xg at 4°C and washed with cold 1xPBS+10% glycerol. The pellet was resuspended in binding buffer (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 10mM imidazole, 1x complete protease inhibitor cocktail EDTA-free (Roche), 10mM MgCl2, 10ug/ml DNaseI) and lysed by 2 cycles of homogenization at ~1000bar (PandaPLUS 2000, GEA). Soluble fraction was separated by centrifugation at 30000xg at 4°C for 1 hour. The supernatant was incubated with Ni-NTA Agarose resin (Qiagen) for 1h at 4°C, the flow-through was collected by gravity, the resin was washed extensively with wash buffer (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 10mM imidazole) and the protein was eluted with the same buffer containing 300mM imidazole. Eluted protein was dialysed O/N at 4°C against 25mM Hepes pH 7.4, 0.2M NaCl, 1mM DTT in presence of TEV protease in order to remove 6x His-tag. The cleaved protein was further purified by negative affinity using Ni-NTA Agarose resin and loaded on anion exchange column 5ml HiTrap Q HP (Cytiva). FT containing the protein was further concentrated on 10k MWCO centrifugal concentrator (PierceTM, ThermoScientific) and loaded on size-exclusion chromatography column Superdex 75 16/60 in 25mM Hepes pH 7.4, 0.2M NaCl, 5% glycerol, 1mM DTT. Fractions containing PTG CBM21 were pooled together, flash frozen and stored at -80°C.

4.2.3.   PP17-300 and PP1 7-330

Expression

PP1 7-330 vector with N-terminal 6xHis-tag and TEV cleavage site was a gift from Wolfgang Peti (Addgene plasmid # 26566; http://n2t.net/addgene:26566; RRID: Addgene_26566). In order to obtain construct 7-300, a stop codon was introduced by site directed mutagenesis using Q5 polymerase (NEB) and following provided protocol.

Stop codon introduction	Oligo sequence
Forward	GCC GGC CGA CTG ATA AAG AAC AAG GG
Reverse	TTA AGG ATC TGG AAG GAG C

in green: stop codon

The protein was expressed in BL21 (DE3) pGro7 cells in LB supplemented with antibiotics, 0,5mg/ml of arabinose and 1mM $MnCl_2$. Induction and cell harvest were performed as described for PTG 70-264.

Purification

Cell pellet was resuspended in lysis buffer (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 1mM MnCl2, 10mM MgCl2, 10µg/ml DNase I, 1x complete protease inhibitor cocktail (Roche)) and lysed by 2 cycles of homogenization at ~ 1000bar (PandaPLUS 2000, GEA). The supernatant was collected by centrifugation at 30000xg at 4°C for 1 hour and incubated with Ni-NTA Agarose resin (Qiagen) for 1h at 4°C. The flow-through was collected by gravity, the resin was washed extensively with wash buffer (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 1mM MnCl2, 10mM imidazole) and the protein was eluted with the same buffer containing 300mM imidazole. His-tag was removed by TEV protease O/N at 4°C in dialysis against the buffer 50mM Tris pH8, 0.5M NaCl, 10% glycerol, 1mM $MnCl_2$. The next day precipitation was observed and removed by centrifugation at 30000xg, 20min at 4°C. The protein was later loaded on SEC column Superdex 200 26/60 (Cytiva) equilibrated with dialysis buffer. Fractions containing $PP1^{7-300}$ were pooled together and flash frozen. Construct $PP1^{7-330}$ was produced following the same protocol.

4.2.4. PTG 70-264/PP1 complex

Expression

For complex production BL21 (DE3) pGro7 cells containing $PP1^{7-300}$ were rendered competent and subsequently transformed with PTG^{70-264}. Expression was performed as described for $PP1^{7-300}$. Shortly, LB medium was supplemented with chloramphenicol (34µg/ml) selecting for pGro7 vector, kanamycin (50µg/ml) selecting for $PP1^{7-300}$ and ampicillin (100µg/ml) selecting for PTG^{70-264}, arabinose (0,5mg/ml) for chaperonin complex expression induction and 1mM $MnCl_2$. The PTG/PP1 expression was induced with 0,1mM IPTG at OD_{600}~ 0,6 and cell culture was grown at 18°C for 18-20 hours.

Purification

Cell pellet was resuspended in lysis buffer (50mM Tris pH8, 0.5M NaCl, 10% glycerol, 1mM MnCl2, 0,5mM TCEP, 10mM MgCl2, 10µg/ml DNase I, 1x complete protease inhibitor cocktail (Roche)) and lysed by 2 cycles of homogenization at ~ 1000bar (PandaPLUS 2000, GEA). Soluble fractions were separated by centrifugation at 30000xg at 4°C for 1 hour and purification was performed as described for PTG 70-264 with buffers supplemented with 1mM MnCl2 and 0,5mM TCEP instead of DTT. The tags were removed on-column by addition of 3C (PreScission®) protease and TEV protease directly to the resin in 1 CV of wash buffer 1 and incubated over-night at 4°C. Flow-through (FT) fraction containing cleaved PTG/PP1 complex was further incubated with Ni-NTA Agarose resin in order to remove His-tag and uncleaved PP1 7-300. The FT was collected and concentrated using 10k MWCO centrifugal concentrator (PierceTM, ThermoScientific) and further purified on size-exclusion chromatography column Superdex 200 10 300 GL (Cytiva). The fractions containing PTG/PP1 protein complex were pooled together and flash frozen in liquid nitrogen and stored at -80°C.

For SAXS analysis PTG$^{70-264}$/PP1$^{7-330}$ complex was produced as described above eliminating His-tag removal and Ni-NTA negative affinity steps.

4.2.5.   PTG N-terminal peptides

PTG[81-107] and PTG [81-132] peptides containing RVxF conserved interaction motif with PP1 were purchased from ProteoGenix SAS.
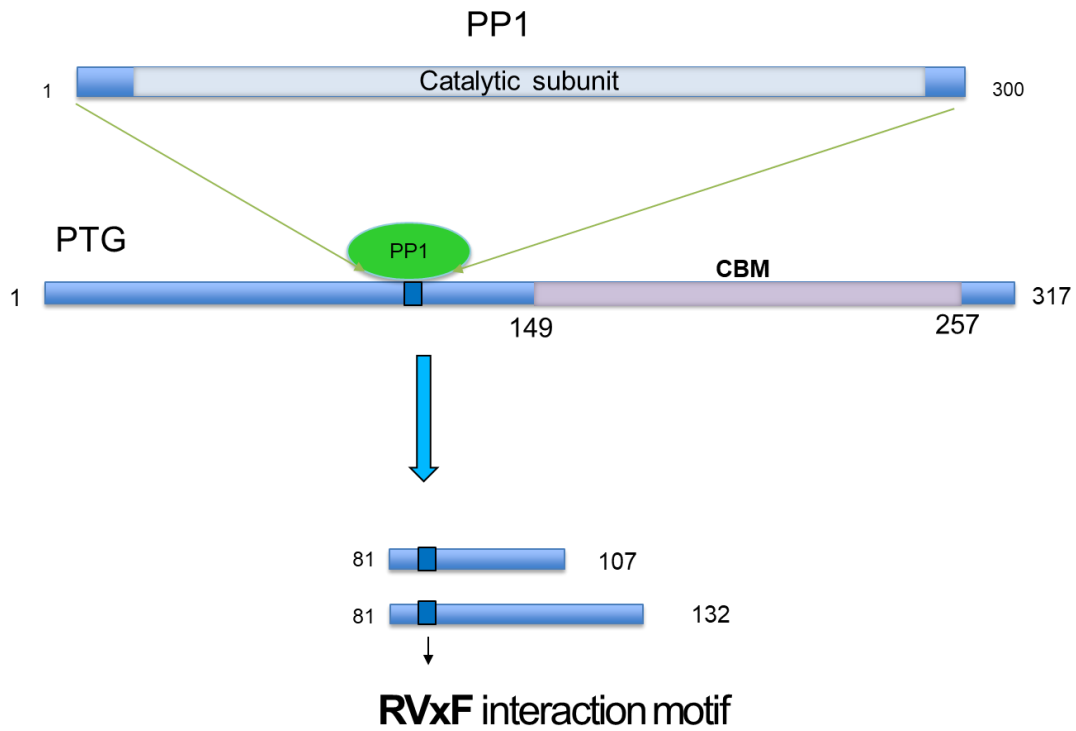
	PTG[81-107]	PTG [81-132]
Sequence	AKKRVVFADSKGLSLTAI HVFSDLPEE	AKKRVVFADSKGLSLTAIH VFSDLPEEPAWDLQFDLLDL NDISSALKHHEEK
Number of amino acids	27	52
Molecular weight	2958.41	5875.63
Theoretical pl	6.81	5.07

4.3. Crystallization

PTG CBM21$^{132\text{-}264}$

The protein was incubated with 3x molar excess of β-cyclodextrin (Calbiochem) and concentrated to 50mg/ml using Vivaspin®2 ultrafiltration centrifugal concentrators (Sartorius). Crystals grew in sitting drops in two different conditions: 3M NaCl, 0.1M BIS-Tris pH5.5 and 0.1M MMT buffer pH4, 25% PEG1500 at 20°C. For cryo-protection a solution of mother liquor supplemented with 20% glycerol was used.

PP1$^{7\text{-}300}$/PTG$^{81\text{-}107}$

PTG$^{81\text{-}107}$ peptide was purchased from ProteoGenix SAS and resuspended in 20mM Tris pH8, 25mM NaCl. PP1$^{7\text{-}300}$ was diluted 10x with buffer 25mM Tris pH8, 50mM NaCl, 1mM MnCl2, 0,5mM TCEP and incubated together with PTG$^{81\text{-}107}$ peptide with 1.1 molar excess for 1 hour at 4°C. The protein/peptide complex was subsequently concentrated to 8mg/ml. Crystals grew in sitting drop in 0.1 M Tris pH 8, 1.6 M Lithium sulfate, 1.4 M Sodium malonate pH 6.0 at 20°C.

PP1/PTG$^{70\text{-}264}$

Extensive crystallisation trials were performed of PP1$^{7\text{-}300}$/PTG$^{70\text{-}264}$ and PP1$^{7\text{-}330}$/PTG$^{70\text{-}264}$ protein complexes resulting in poorly diffracting crystals. β-cyclodextrin powder was directly dissolved in PP1/PTG$^{70\text{-}264}$ protein complex solution at final concentration of 5mM and the whole complex was further concentrated up to 10mg/ml. Small, badly diffracting crystals were obtained in condition 0.1M Na citrate pH 6, 2M NaCl and used for seeding experiments. Big crystals grew in sitting drop in 1.0 M Sodium malonate pH 5.0, 0.1 M Sodium acetate trihydrate pH 4.5 at 4°C, however still diffracting poorly (5-10 Å). The crystal quality improved after dehydration O/N in solution with 10% higher precipitant concentration supplemented with 10% glycerol and 1mM β-cyclodextrin. The crystals were cryo-protected in dehydration solution with 20% glycerol or xylitol, flash frozen in liquid nitrogen.

4.4. Data collection and structure solution

All diffraction data were collected on XRD2 beamline at Elettra Sincrotrone Trieste (Italy), integrated and scaled with XDS (28) and merged with Aimless (29) from the CCP4 Suite (115).

The PTG-CBM21$^{132-264}$ structure was solved by molecular replacement with Phaser using the GL structure 2EEF as search model, following removal of the longest loops. Cycles of manual rebuilding with Coot (114) and refinement with Phenix (30) allowed filling all gaps and reconstituting the correct sequence, while the evident residual density was used to include the cyclodextrin molecule.

The PP1$^{7-300}$/PTG$^{81-107}$ structure was solved by molecular replacement with Phaser using the PP1 structure 6ZEG as search model. The PTG peptide was manually build also taking advantage of the similar GM peptide in the 6DNO structure. Refinement was conducted with Phenix.

The PP1$^{7-300}$/PTG$^{70-264}$ structure was solved by molecular replacement with Phaser using the two above structures as search models. Resolution was cut anisotropically with Staraniso (31). Clear positive electron density was used to build the missing PTG residues. Cycles of manual and automatic refinement were performed by using Coot and Phenix, respectively

4.5. Small-angle X-ray scattering

Size-exclusion chromatography (SEC) coupled to Small-angle X-ray scattering (SAXS) measurements of PP1$^{7-330}$/PTG$^{70-264}$ with and without β-cyclodextrin were performed on BM29 beamline at ESRF (Grenoble, France) via mail-in service. The protein complex at concentration of 4 mg/ml (without β-cyclodextrin) and 6.5 mg/ml (with β-cyclodextrin) was injected onto Superdex Increase 200 3.2/300 in buffer 50mM Tris pH8, 0,5M NaCl, 10% glycerol, 1mM DTT (β-cyclodextrin at the final concentration of 5mM was added for the second run).

Data analysis was conducted by Dr. Gabriele Giachin, Dept. of Chemical Sciences, University of Padua. Briefly, various software pieces from the ATSAS suite (116) were

used for diffraction images selection and buffer subtraction, data reduction (averaging and merging), calculation of MW and Rg, generation of Guinier and Porod plots, ab initio shape determination and dummy atom model definition and finally, modelling of protein complexes with fitting to experimental data. Importantly, given the dynamic nature of the PP1/PTG complex, the data were finally subjected to EOM (Ensemble Optimization Method) for taking into account the multiple reciprocal orientations in the solution complex.

4.6. Characterisation

4.6.1.   Grating-Coupled Interferometry

Grating-coupled interferometry (GCI) experiments were performed by the Creoptix WAVE system (Creoptix®). The GCI is similar to other optical label-free methods, such as surface plasmon resonance (SPR). The target (or ligand) is immobilized to the sensor surface and the analyte is flowed over the surface in what is called the bulk. Light is coupled to the surface creating an evanescent field. Interactions are detected by the resulting change in mass and thus a change in the refractive index.
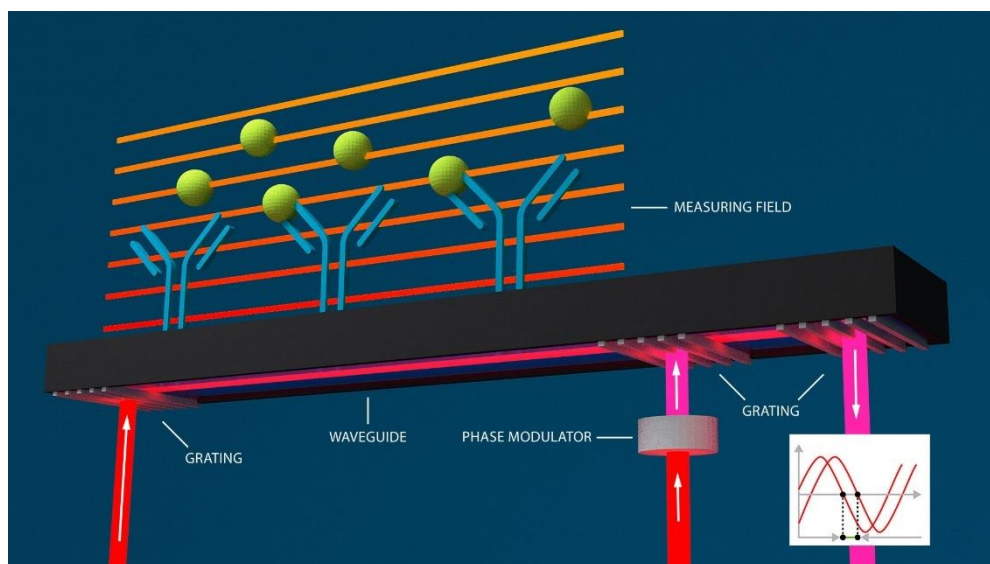

Fig 14. Schematic representation of GCI Creoptix WAVE system (Creoptix ®, www.creoptix.com).

In contrast to traditional SPR, the light in Waveguide Interferometry can travel through the entire length of the sample, which allows more binding events increasing sensitivity for label-free interaction analysis. Another advantage of GCI over SPR is that the

33

evanescent field penetrates less deep into the sample, which results in increased signal-to-noise ratio by minimizing the disturbance caused by bulk refractive index changes (117).

Experimental procedure.

Borate buffer (100 mM sodium borate pH 9.0, 1 M NaCl) was used for chip conditioning. In order to evaluate the kinetics of the interaction of PTG with β-cyclodextrin and PP1, PTG70-264 (20 µg/ml in buffer acetate pH 5.0) was immobilized on 4PCH WAVE chips (quasiplanar polycarboxylate surface; Creoptix) by standard amine coupling according the manufacturer instructions at a density of 1800 pg/mm2 and 1000 pg/mm2 for β-cyclodextrin and PP17-330, respectively. Regeneration free injections of a 1:3 dilution series from 500 µM for β-cyclodextrin in running buffer (10 mM Hepes pH 7.2, 150 mM NaCl, 0.005% Tween 20) were performed at 25 °C, using a flow rate of 100 µl/min. Similarly, regeneration free injections of a 1:2 dilution series from 400 nM for PP1$^{7-330}$ in the same running buffer were performed at 25 °C, using a flow rate of 30 µl/min. Experiments to evaluate the kinetics of the interaction between PP17-330and the two PTG derived peptides (PTG$^{81-107}$and PTG$^{81-132}$, ProteoGenix SAS) were performed reversibly capturing 6xHis-tagged PP1$^{7-330}$ (5 µg/ml in running buffer 10 mM Hepes pH 7.2, 150 mM NaCl, 0.005% Tween 20, 0.2 mg/ml BSA) on 4PCP-NTA WAVE chips according to the manufacturer instructions at a density of 3600 pg/mm2 and 1600 pg/mm2 for peptide PTG$^{81-107}$and PTG$^{81-132}$, respectively. Regeneration free injections of a 1:2 dilution series from 300 nM for the two peptides in running buffer were performed at 25 °C, using a flow rate of 100 µl/min. Blank injections were used for double referencing and a dimethyl sulfoxide (DMSO) calibration curve for bulk correction. Analysis and correction of the obtained data were performed using the Creoptix WAVE control software (correction applied: X and Y offset; DMSO calibration; double referencing). Mass transport binding models with bulk correction were used to fit all experiments. All the experiments have been performed at least in triplicate.

4.6.2. Isothermal Titration Calorimetry (ITC)

ITC experiments were performed using a MicroCal ITC (Malvern Panalytical) at 25°C with a stirring speed of 750 r.p.m. in 25mM Hepes pH 7.4, 0.2M NaCl, 5% glycerol with 180 seconds between the injections. β-cyclodextrin was dissolved in the running buffer and two datasets were collected, one with PTG CBM21 concentration at 100mM and β-cyclodextrin at 3nM with 40 injections of 1µl, and second one at 200mM and 4nM respectively with 80 injections of 1µl. As a negative control a blank run with only the buffer was carried out. Data analysis was done with ORIGIN subtracting the blank titrations fitting 'one site' and 'two sites' model.

4.6.3. Thermal Shift Assay (TSA)

Thermal Shift Assay (TSA) was performed in order to test the stability of the proteins in different pH and in the presence of additives (β-cyclodextrin, maltose, glucose, sucrose, arabinose, xylitol). Reaction mixtures were prepared in the white 96-wells, hard-shell, thin-wall PCR plates (Biorad, #HSP9655) and the assay was performed in a Real Time PCR instrument (CFX96 Touch, Bio-Rad) following the protocol: 2' at 30°C and 1°C increment every 30'' up to 95°C. Data were analyzed in CFX Manager software (BioRad).

Starting Concentration	Reagent	Volume	**Final Concentration**
	H_2O	To reach 20 µl	
5x	Protein buffer	4 µl	1x
0,01-1g/l	Protein	x µl	0,005-0,5
5x	* Additive	4 µl	1x
200x	SYPRO™ Orange (Invitrogen, # S6650)	1 µl	10x

TSA reaction mix composition,   *   optional

4.6.4. Right angle light scattering (OMNISEC)

The absolute molecular weight and oligomeric state of the proteins were measured using OMNISEC (Malvern Panalytical), a Gel Permeation Chromatography (GPC) / Size Exclusion Chromatography (SEC) system connected to low-angle and right-angle light scattering detector. The run was performed in 50mM Tris pH8, 0,5M NaCl, 5% glycerol, at room temperature using Superdex Increase 200 3.2/300 column calibrated with BSA at 2mg/ml (Sigma, #P0634), injecting 50µl of each sample, at flow of 0,08ml/min. The analysis was performed using OMNISEC Software.

5. RESULTS AND DISCUSSION

5.1. Construct design

One of the biggest bottlenecks in protein structural and biochemical characterization is obtaining enough soluble and stable material to start the experimental journey. PTG is a relatively small protein composed of 317 amino acids with a structured central region, containing a carbohydrate binding module 21 (CBM21, residues 133-261), flanked by largely disordered regions. PTG constructs have been designed based on analyses performed with the PSIPRED server (Fig 14), mainly relying on protein secondary structure prediction.

The full-length PTG construct failed to be synthetized and, based on PSIPRED prediction, a shorter sequence encoding for PTG 70-264 was designed and corresponding sequence was purchased from GeneScript and cloned in pGEX-6-b vector as described in methods. The construct encodes for the flexible N-terminal region which contains the PP1-binding signature RVxF ([84]RVVF[87]), and for the C-terminal carbohydrate binding motif (CBM) that is predicted to have an ordered structure. The N-terminal part of the protein is expected to be disordered, which decreases the probability of stable protein production; thus, two further constructs were designed consisting mainly of the CBM domain: PTG CBM 132-264 where a possible β-sheet has been included and PTG CBM 142-264. See methods for cloning details.
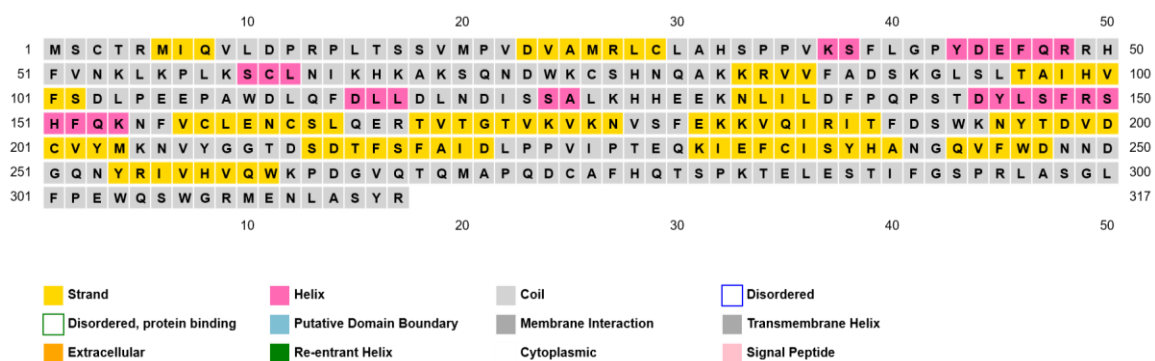

Fig 15. PSIPRED results of secondary sequence prediction of PTG.

5.2. Protein production

<u>PTG 70-264</u>

Expression optimisation

First attempts of protein production were unsatisfactory, yielding poorly soluble protein expression (Fig.16). For further protocol optimisation, the BL21pLysS (DE3) and BL21(DE3) cell strains were chosen, as in total cell extract an expression band was detected, indicating good expression and poor solubility of the protein.
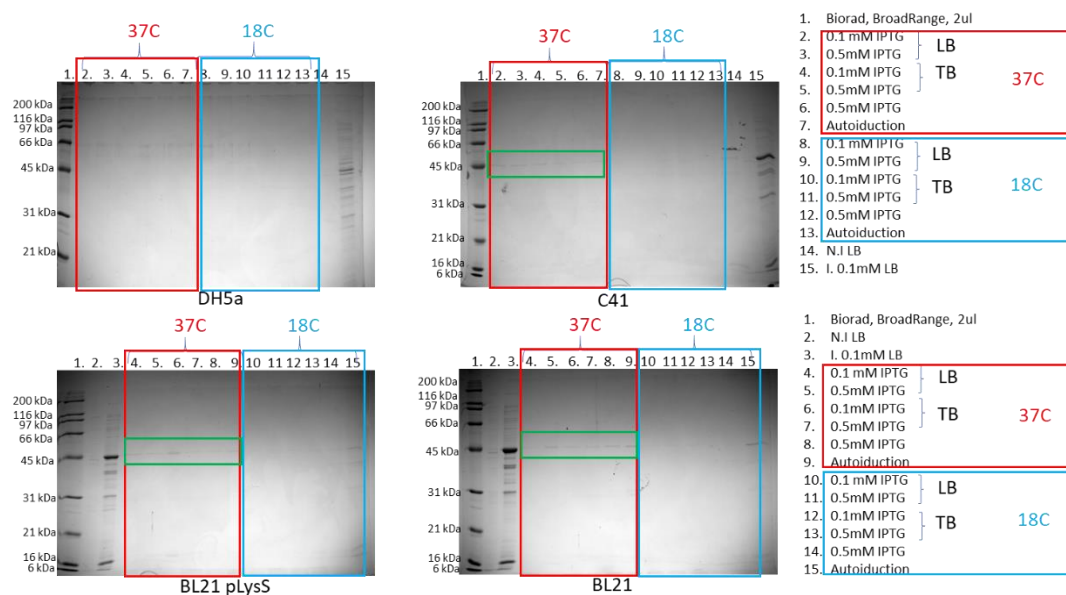

Fig 16. SDS-PAGE gels of PTG 70-264 test expression in different cell strains: DH5alpha, C41(DE3), BL21pLysS (DE3) and BL21(DE3). Elutions from GST affinity purification were loaded and single not induced (N.I.) and induced (I) samples were included in order to estimate total protein expression.

Further expression tests were carried out, narrowing the number of conditions and checking different time points at 37°C and two inducer concentrations. In all conditions, high protein expression was observed (strong band visible in total extract); unfortunately, most of the protein was insoluble, and only a tiny fraction got purified (Fig.17A). Scale-up protein production was performed using BL21(DE3) cells induced with 1mM IPTG for 6h at 37°C. The protein, after affinity binding to the resin, underwent 3C protease treatment for GST-tag removal on column, resulting in very low yield (~100ug/1 liter of culture), confirming that the current expression system is not suitable for obtaining stable protein for further experiments (Fig.17B).
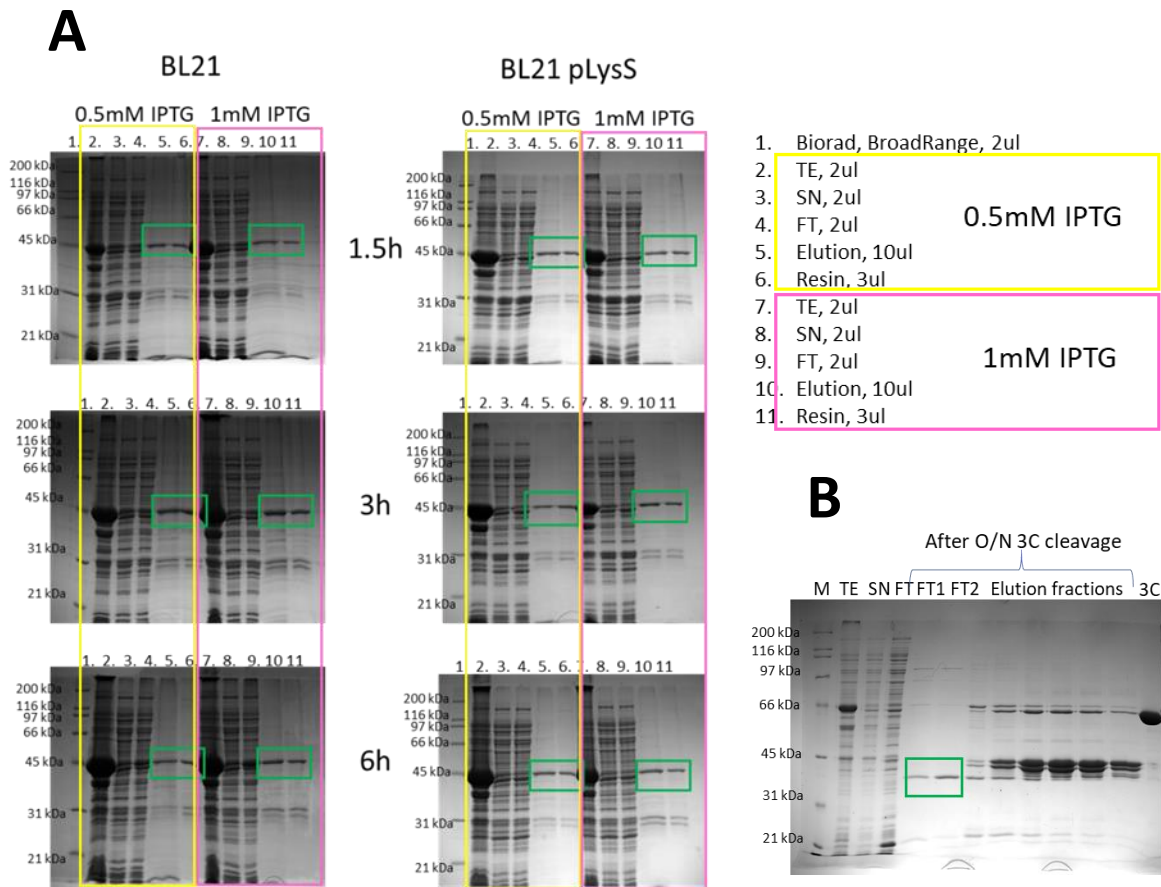
Fig 17. **A** SDS-PAGE gels of PTG 70-264 test expression in different cell strains: DH5alpha, C41(DE3), BL21pLysS(DE3) and BL21(DE3). Elutions from GST affinity purification were loaded and single not induced (N.I.) and induced (I) samples were included in order to estimate total protein expression. **B** SDS-PAGE gels of PTG 70-264 purification and on-column GST-tag cleavage. Total cell extract (TE), soluble fraction-supernatant (SN), flow through (FT), FT fractions containing cleaved PTG after overnight GST-tag removal with 3C protease (FT1 and FT2) and elutions from GST resin containing uncleaved PTG, GST tag and 3C-GST protease were loaded.

To assist in correct protein folding and increase its solubility, PTG was co-expressed with chaperonin GroE complex. Preliminary test expression gave promising results (Fig 18A), with increased soluble PTG, especially in the cultures grown at 18°C. Three best conditions were chosen for mid-scale test expression, which allowed for final expression protocol optimisation (Fig 18B). The soluble GST-tagged protein yield improved significantly (from ~160ug/l to ~ 12mg/l of the cell culture), finally obtaining enough material for binding studies. The scaleup expression was performed in LB medium supplemented with 0,5mg/ml of arabinose for chaperonin expression induction.
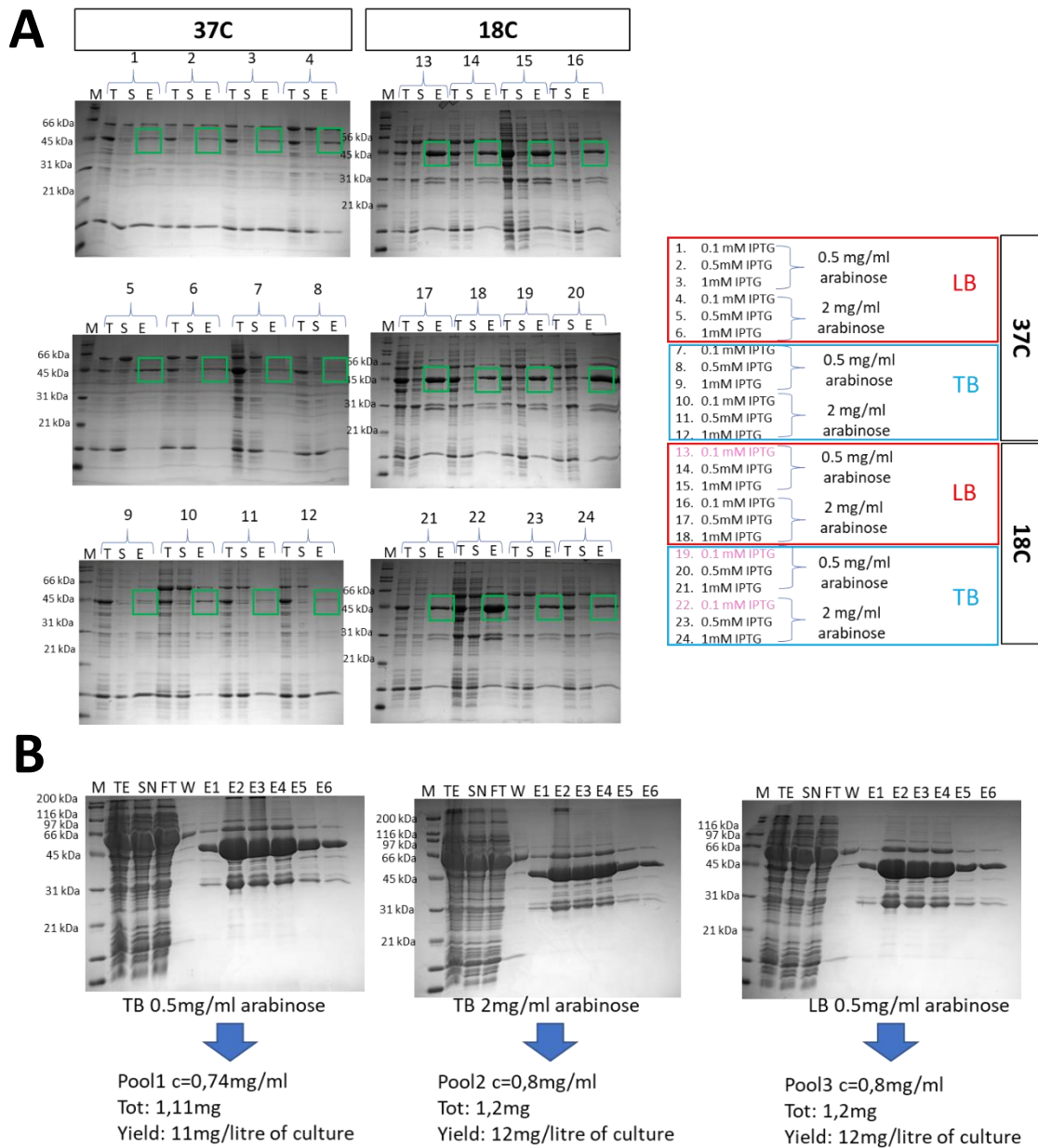
Fig 18. **A.** SDS-PAGE gels of PTG 70-264 test expression in BL21(DE3) with different conditions. Total extract (T), supernatant (S) and elution (E) were loaded. **B** SDS-PAGE gels of PTG 70-264 mid-scale purification. Total cell extract (TE), soluble fraction-supernatant (SN), flow through (FT) and elutions from GST resin.

Purification

The PTG 70-264 purification was performed in two chromatographic steps: glutathione affinity purification and Gel Filtration or Size Exclusion Chromatography (SEC).

During the first affinity step, the GST-tagged PTG binds to the glutathione resin, while unbound sample components are collected in flow-through and subsequent washing steps. The excess of GroEL chaperonin is removed by introducing a wash step with buffer containing 5mM Mg-ATP, which facilitates on-column protein folding and chaperonin dissociation (33). The GST tag cleavage was performed overnight by adding GST-tagged 3C protease. All the GST-containing components remained bound to the column, while the untagged protein was released in the flow through. An important amount of GroEL (1:1 chaperone: PTG ratio) was still present in the protein sample (Fig.19).
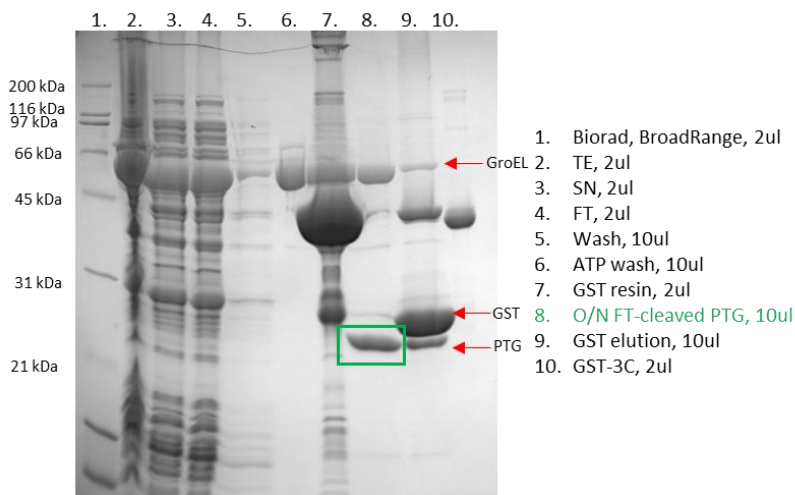

Fig 19. SDS-PAGE gels of PTG 70-264 purification and on-column GST-tag cleavage. Total cell extract (TE), soluble fraction-supernatant (SN), flow through (FT), O/N FT fractions containing cleaved PTG after overnight GST-tag removal with 3C protease and elutions from GST resin containing uncleaved PTG, GST tag and 3C-GST protease were loaded.

The protein was further purified by size exclusion chromatography. In this step, most of the chaperonin fraction was separated from monomeric PTG protein (Fig 20). The protein was used for studies of interaction with PP1.
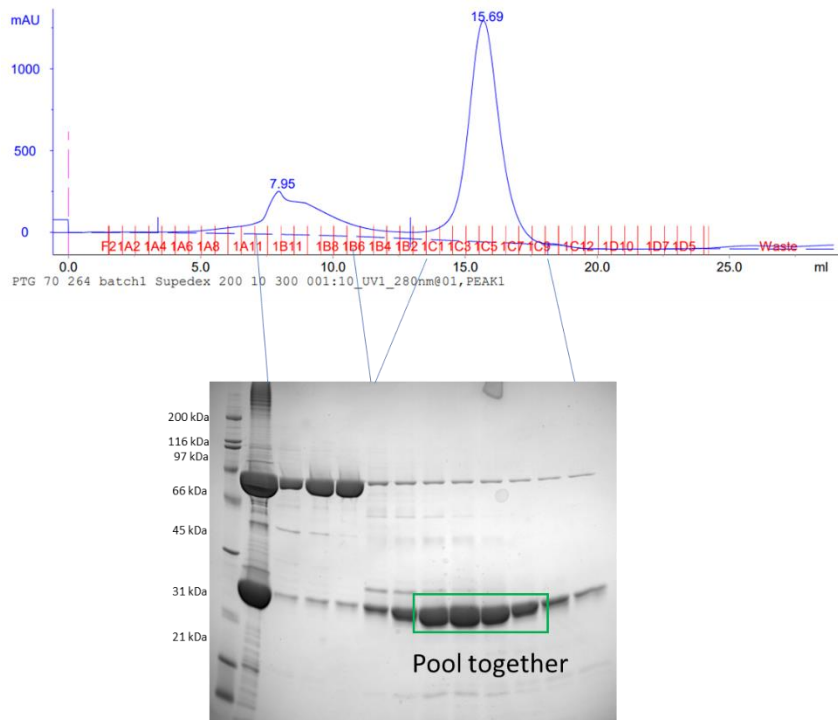
Fig 20. Chromatogram profile of SEC200 purification and SDS-PAGE gels with corresponding purified fractions of PTG 70-264. The first peak corresponds to the column void volume containing protein aggregates and most of the eluted chaperonin. The second peak corresponds to PTG 70-264.

PTG CBM 132-264 and 142-264

Expression optimisation

PTG CBM constructs were cloned into four different expression vectors that bring different tags to the encoded protein in order to facilitate its expression and purification. Extensive small and mid-scale test expression was performed in order to explore the best condition for stable protein production (Fig. 21)

Fig 21. **A.** SDS-PAGE gels of test expression of PTG 132-264 and 142-264 constructs cloned in a series of vectors containing different tags. The red arrow indicates the protein band corresponding to the expected molecular weight. Samples loaded on the gel: total extract (T), supernatant (SN) and elution (E), colour code: 132-264, 142-264. **B** SDS-PAGE gels of His-tagged and GST-tagged PTG 132-264 of mid-scale expression in LB and autoinduction medium. Total cell extract (TE), soluble fraction-supernatant (SN), flow through (FT) and elutions (E) were loaded.

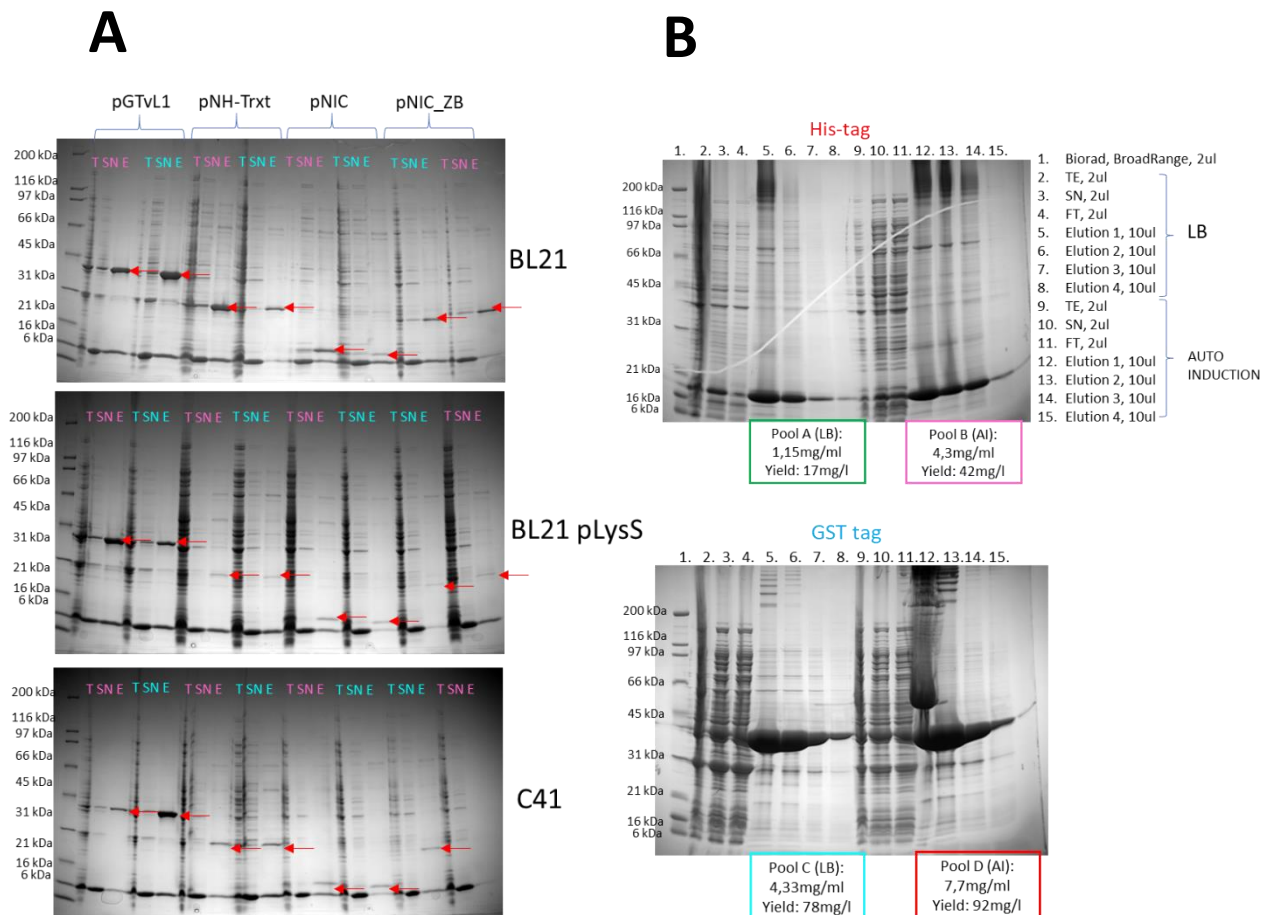Both constructs resulted in good protein yields. The best results were obtained using the pNIC_28_Bse4 cloned constructs expressed in BL21 (DE3) cell strain growing in autoinduction medium. For scale up protein expression protocol, 6xHis-tagged PTG CBM 132-264 construct was chosen.

Purification

The purification of PTG CBM 132-264 was performed in three chromatographic steps: Ni-NTA affinity, anion-exchange chromatography and Size Exclusion Chromatography (SEC).

The His-tagged PTG protein was captured by affinity to Ni-NTA agarose resin and non-bound samples were eliminated by extensive washing. The eluted PTG protein was later dialysed together with His-TEV protease for His-tag removal and lowering the imidazole concentration. The TEV cleavage was partial and the untagged protein together with His-TEV was eliminated form PTG sample by negative affinity (Fig 22).

A

1. Marker, Biorad, Broad range, 2ul
2. Total extract,2 ul
3. SN,2ul
4. FT, 2ul
5. Wash 10ul
6. Elution, 10ul
7. Elution, 5ul

B

1. Marker, Biorad, Broad range, 2ul
2. NiNTA elution (uncleaved),2 ul
3. TEV O/N cleavage mix ,10ul
4. FT from neg affinity (cleaved), 10ul
5. Elution from neg. affinity, 10ul
6. TEV protease, 3ul
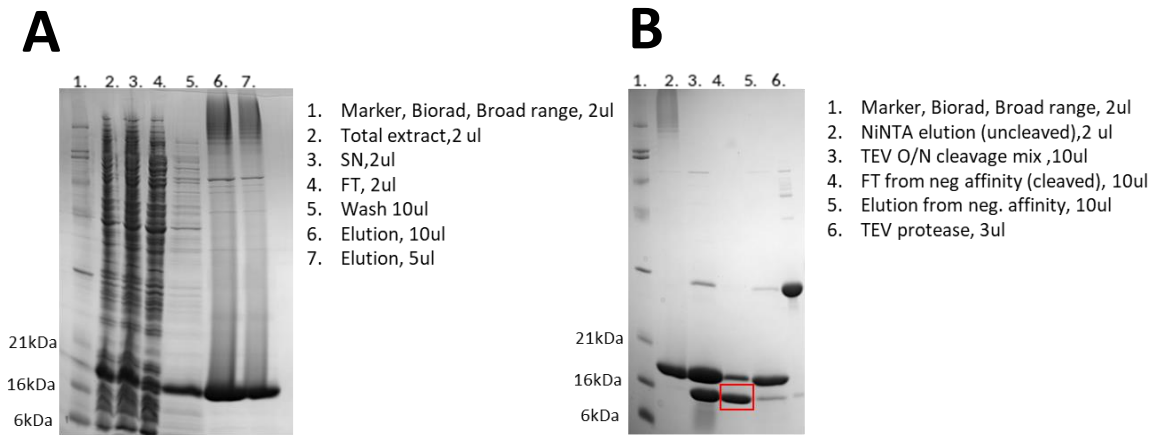
Fig 22. **A.** SDS-PAGE gel of His-tagged PTG 132-264 affinity purification where total cell extract (TE), soluble fraction-supernatant (SN), flow through (FT) and elutions (E) were loaded. **B** SDS-PAGE gel of PTG 132-264 TEV cleavage efficiency and negative Ni-NTA purification step.

Next, PTG underwent an anion exchange chromatographic step, where it remained in the flow-through, with the uncleaved fraction and a minor part of the cleaved protein bound to the resin (Fig. 23). Small-scale trials with both anion and cation exchange resins were performed in order to find the best conditions.

1. Marker, Biorad, Broad range, 2ul
2. HiTrap Q load, 5 ul
3. HiTrap Q FT,10ul
4. A2, 10ul
5. A3, 10ul
6. A4, 10ul
7. A5, 10ul

Fig 23. SDS-PAGE gel with samples from anion exchange chromatography on HiTrap Q HP column (Cytiva). In red PTG 132-264 fraction in the flow through of the column, used for further purification step.

The protein is further concentrated and loaded on Size Exclusion column (SEC75, 16/60, Cytiva) eluting as a single, monomeric peak (Fig. 24). The protein was used for structural and β-cyclodextrin binding studies.
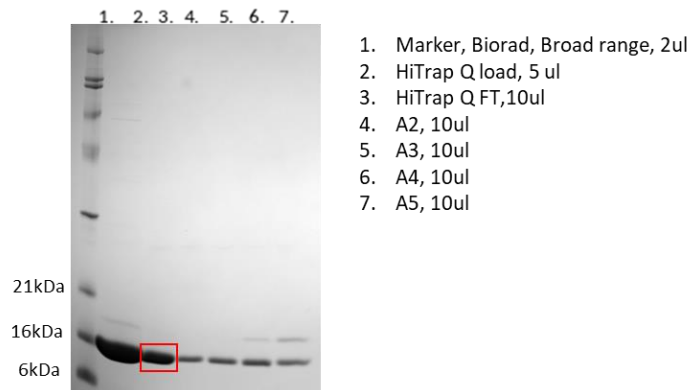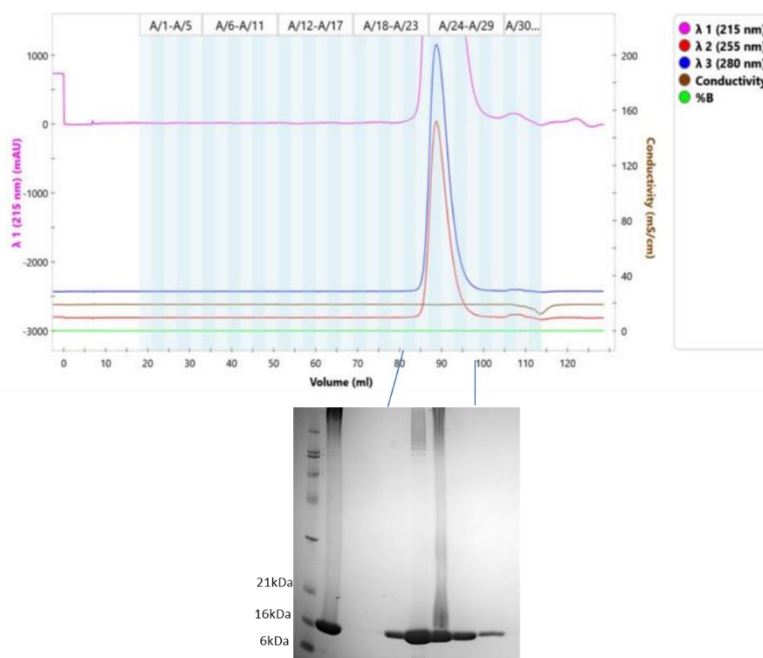

Fig. 24. Chromatogram profile of SEC75 purification and SDS-PAGE gels with corresponding purified fractions of PTG 132-264.

PP1$^{7-300}$ and PP1 $^{7-330}$

Expression and purification

Both constructs were expressed in the BL21 (DE3)/ pGro7 system following the protocol for PTG 70-264 expression with the addition of MnCl$_2$ to the growth medium, as required for its activity and stability.

The purification steps were the same for both proteins. After elution from the Ni-NTA affinity resin, the proteins underwent a 6xHis-tag cleavage step by incubation with His-tag-TEV protease in dialysis to lower the imidazole concentration. The negative affinity step allowed to remove the uncleaved protein and the protease. As the last step, size-exclusion chromatography was introduced and pure, monomeric protein was separated from the remaining impurities (Fig 25). The cleaved PP1$^{7-300}$ protein was used for structural studies and amine-coupling immobilization trials for interaction studies with

PTG N-terminal peptides. Finally, the His-tagged PP1$^{7-330}$ was used for the binding studies, taking advantage of His-tag for a less harsh immobilization protocol.

A

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11.

55 kDa
43 kDa
34 kDa
26 kDa

1. Thermofisher Ladder, 2ul
2. TE, 2ul
3. SN, 2ul
4. FT, 2ul
5. Wash, 10ul
6. Elution, 10ul
7. His-tagged PP1 SEC pool, 10ul
8. PP1 TEV cleavage mix, 10ul
9. FT from negative affinity, 10ul
10. Elution from Ni-NTA after cleavage, 10ul
11. TEV, 10ul
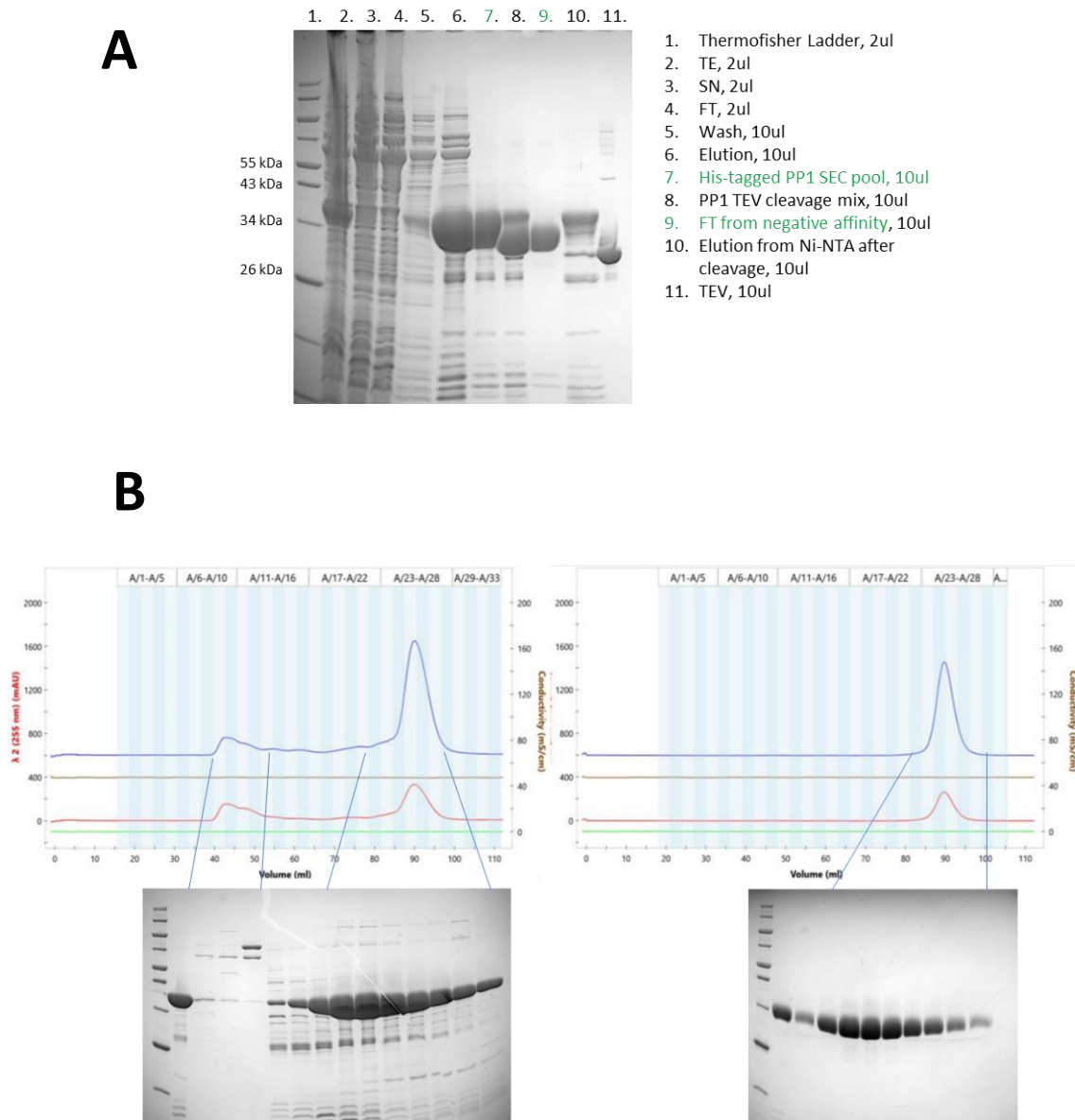
B

Fig 25. **A** SDS-PAGE gel of PP1$^{7-330}$ Ni-NTA affinity purification and overnight negative affinity after TEV cleavage (total cell extract (TE), soluble fraction-supernatant (SN), flow through (FT)). **B** Chromatogram profile of His-tagged PP1$^{7-300}$ SEC200 16/60 (Cytiva) purification. **C.** Chromatogram profile of cleaved PP1$^{7-300}$ SEC200 16/60 (Cytiva) purification.

PTG 70-264/PP1 complex

Expression and purification

The two proteins PTG 70-264 and PP1 were co-expressed in the previously described BL21(DE3)/pGro7 set-up supplementing LB medium with $MnCl_2$ to enhance PP1 stability. Each protein is fused with different N-terminal and cleavable tags: PTG 70-264 with GST and PP1 with His-tag, which was exploited for the complex formation verification. The cell lysate was divided in two samples and one was loaded on glutathione and another on NiNTA affinity resins and, in both cases, the complex was obtained (Fig 26). The glutathione affinity resulted in higher purity of the complex and it was chosen as the first step of purification.
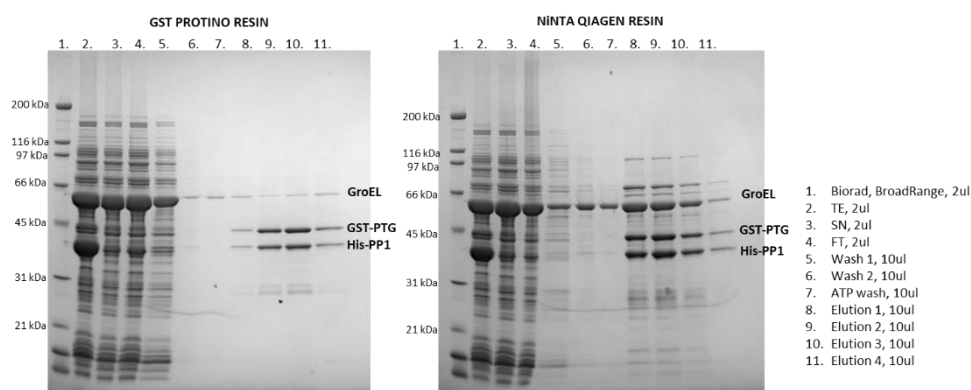
Fig 26. SDS-PAGE gels of PTG/PP1 complex purification by Ni-NTA and glutathione affinity showing in both cases complex formation.

The purification of the complex was performed in two chromatographic steps: glutathione affinity and size exclusion chromatography.

As described for PTG 70-264, the affinity step was carried out using buffers supplemented with $MnCl_2$. After extensive washing, including ATP/$MgCl_2$ for chaperonin removal, two proteases were added on the column: GST-3C for GST tag cleavage from PTG and His-TEV, which cuts the 6xHis-tag from PP1. After O/N incubation, the column flow through was collected, containing the cleaved PTG/PP1 complex and His-TEV protease. A step of negative Ni-NTA affinity removed uncleaved PP1 and TEV. The PTG/PP1 complex was concentrated and loaded on SEC200 10/300 GL (Cytiva) column and eluted as a single monomer (Fig. 27A). The purified complex was used for

crystallization. For SAXS analysis the step of His-tag removal form PP1 protein was excluded (Fig 27B). The average yield of the pure complex is 2-3mg/litre of cell culture.
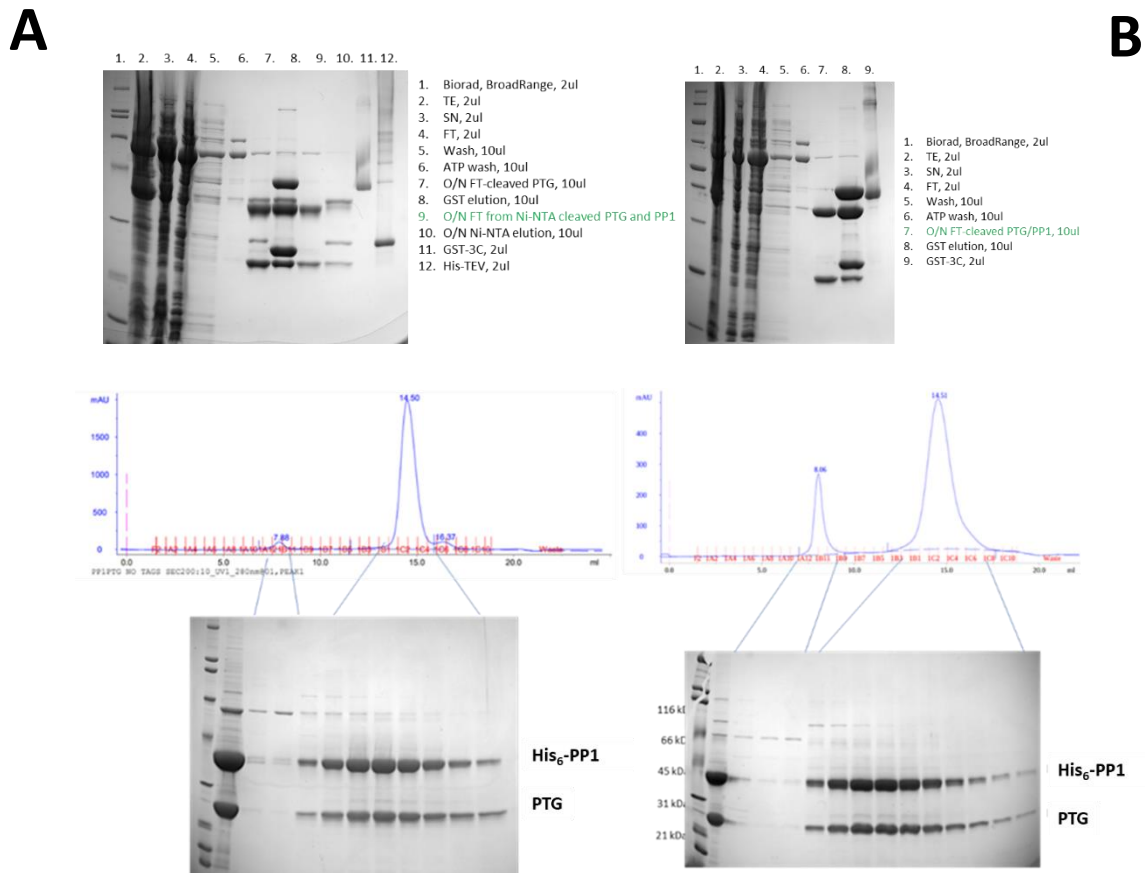

Fig 27. **A** SDS-PAGE gels representing purification steps where tags are cleaved from both proteins. On the first gel samples from affinity purification and overnight GST- and His-tag cleavage were loaded. Second gel contains samples from SEC200 chromatography corresponding to the complex elution peaks **B** SDS-PAGE gels representing purification steps where only GST tag from PTG is cleaved, whereas PP1 remains His-tagged. On the first gel, samples from affinity purification and overnight GST-tag cleavage were loaded. Second gel contains samples from SEC200 chromatography corresponding to the complex elution peaks.

5.3. Thermal shift assay

PTG protein contains CMB21 domain that binds carbohydrates. To understand if addition of common carbohydrates increases the protein stability, which could improve crystallisation outcomes, a thermal shift assay was performed. We tested PTG 132-264 and PTG/PP1 complex stability in presence of a series of sugar additives. In all cases, the TSA resulted in no significant melting temperature shift, suggesting that proteins stability does not increase upon carbohydrate binding (Fig 28).

Protein	additive	Tm
PTG 132-264	-	58.50
	maltose	59.00
	sucrose	59.00
	glucose	58.50
	arabinose	59.50
	xylitol	58.50
	β-cyclodextrin 3x molar excess	58.50
	β-cyclodextrin 10x molar excess	58.50
PTG/PP1 complex	-	50.00
	maltose	50.00
	sucrose	50.00
	glucose	49.00
	arabinose	50.50
	xylitol	50.50
	β-cyclodextrin 3x molar excess	50.00
	β-cyclodextrin 10x molar excess	50.50
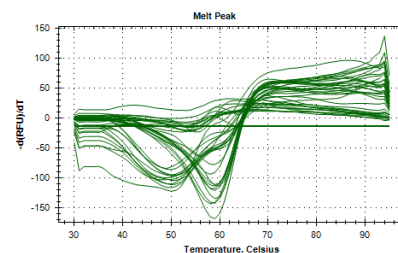
Fig 28. Melting temperature of PTG 132-264 and PTG/PP1 complex alone and in presence of different carbohydrates. Next to the table, melting curves of each sample are shown.

5.4. RALS analysis

To measure absolute molecular weight of the PTG 132-264 and PTG/PP1 complex and the peak dispersity, a static light scattering technique was applied: Right-Angle Light Scattering (RALS) using Omnisec instrument. In both cases, the measured MW corresponded to the expected one, resulting in 16 kDa for PTG 132-264 and 59,9 kDa for the PTG/PP1 complex that elute as monodispersed peaks (Fig 29).

Fig 29. RALS analysis. **A**. A mass distribution in the elution peak corresponding to PTG 132-264 and its absolute molecular weight. **B** A mass distribution in the elution peak corresponding to PTG/PP1 complex and its absolute molecular weight.

5.5. Grating-Coupled Interferometry

For the investigation of binding kinetics of PTG protein and its ligand β-cyclodextrin as well its interactor PP1, grating-coupled interferometry (GCI) experiments were performed. In the first set-up, PTG 132-264 was immobilized on the 4PCH WAVE chip and a series of 1:3 dilutions starting from 500 µM of β-cyclodextrin were injected, resulting in Kd of 38µM. In comparison, *Rhizopus oryzae* fungus- *Ro*GACBM21 binds β-cyclodextrin with a Kd of 5.1 µM (37), while G_M of 8.2µM (34). When titrating PP1 against PTG CBM of 1:3 dilutions starting from 30 µM of PP1, no binding was detected, suggesting that the C-terminal domain of PTG doesn't contribute to the PP1 interaction mechanism. To measure the binding of PTG containing PP1 recognition motif, PTG 70-264 was immobilised on the chip surface and flashed with a series of injections of 1:2 dilution of PP1$^{7-330}$ starting from 400 nM concentration. Kd of around 80nM was evaluated, confirming high affinity between the proteins. At last, the binding kinetics were

measured for two N-terminal PTG peptides containing conserved PP1 RVXF interaction motif. The longer peptide contains as well SALK residues that further contribute to PP1 binding. Both peptides were injected in 1:2 dilution series starting from 300 nM against immobilized PP1 protein. Kd of the shorter peptide 1 was 14 nM and the peptide 2 with both RVXF and SALK sequence was 7 nM, indicating SALK contribution for PP1 binding. The difference between Kd of the peptides and PTG 70-264 towards PP1 is probably due to size difference, small peptides are more available for binding than more bulky long protein. Binding studies confirm that the N-terminal part of PTG is essential for PP1 binding while CBM doesn't contribute to their interaction (Fig. 30, Table 1).
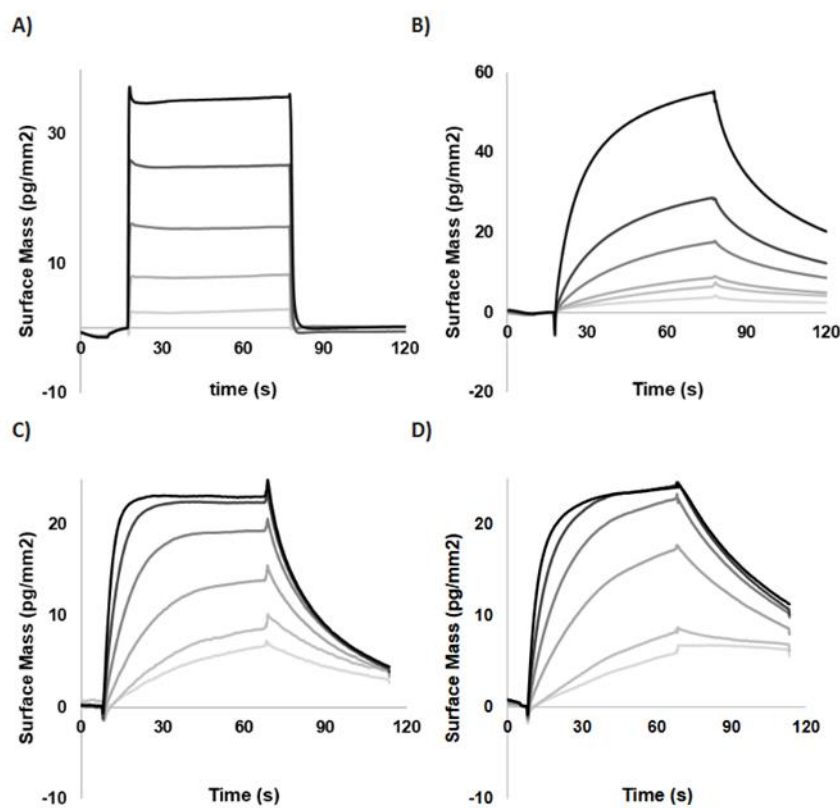
Fig 30: Quantitative binding kinetics by GCI: A) PTG 132-264 versus cyclodextrin. Response curves of 1:3 serial dilutions of cyclodextrin (500uM-6uM range) on a cell coated with 1800 pg/mm2 of PTG 132-264. B) PTG 70-264 versus PP1. Response curves of 1:2 serial dilutions of PP1 (400nM-12.5nM range) on a cell coated with 1000 pg/mm2 of PTG 70-264. C) PP1 versus peptide 1. Response curves of 1:2 serial dilutions of peptide 1 (300nM-9nM range) on a cell coated with 3600 pg/m2 of PP1. D) PP1 versus peptide 2. Response curves of 1:2 serial dilutions of peptide 2 (300nM-9nM range) on a cell coated with 1800 pg/mm2 of PP1. All curves were blank subtracted. Experiments performed in triplicate.

ligand	analyte	k_{on} (M-1s-1)	k_{off} (s-1)	K_D (M)
PTG 132-264	β-cyclodextrin	4.20 ± 0.60 E1	1.55 ± 0.05 E-3	37.3 ± 7.0 E-6
PTG 70-264	PP1	5.00 ± 0.30 E4	4.30 ± 0.01 E-3	87.0 ± 4.1 E-9
PP1 7-330	Peptide1 PTG[81-107]	2.08 ± 0.19 E5	2.95 ± 0.77 E-3	14.1 ± 2.4 E-9
PP1 7-330	Peptide 2 PTG[81-132]	3.57 ± 0.78 E5	2.60 ± 0.66 E-3	7.7 ± 2.0 E-9

Table 1. Kinetic results for each protein.

5.6. Isothermal Titration Calorimetry (ITC)

Grating-coupled interferometry is a robust technique for binding kinetics measurements; however, it cannot reliably estimate the stoichiometry between the target and the analyte. To investigate the number of β-cyclodextrin bindings sites present in PTG, an ITC analysis was carried out. PTG CBM was titrated with β-cyclodextrin. A Kd of 63.18 (58.46 – 68.47) µM was measured. The fit was done using the "One site" model, which means that the existence of only one binding site "type" (with a variable stoichiometry) was supposed. The "Two sites" model was as well tested. In this case, two distinct binding sites are expected with different characteristics. The errors on the Ka1 and Kd1 ("two-site" model) are rather large compared to the errors on Ka2 and Kd2. Considering that the fit of the experimental data does not substantially improve, the "one-site" model is more probable and it fits well the experimental isotherm (Table 2, Fig 31).
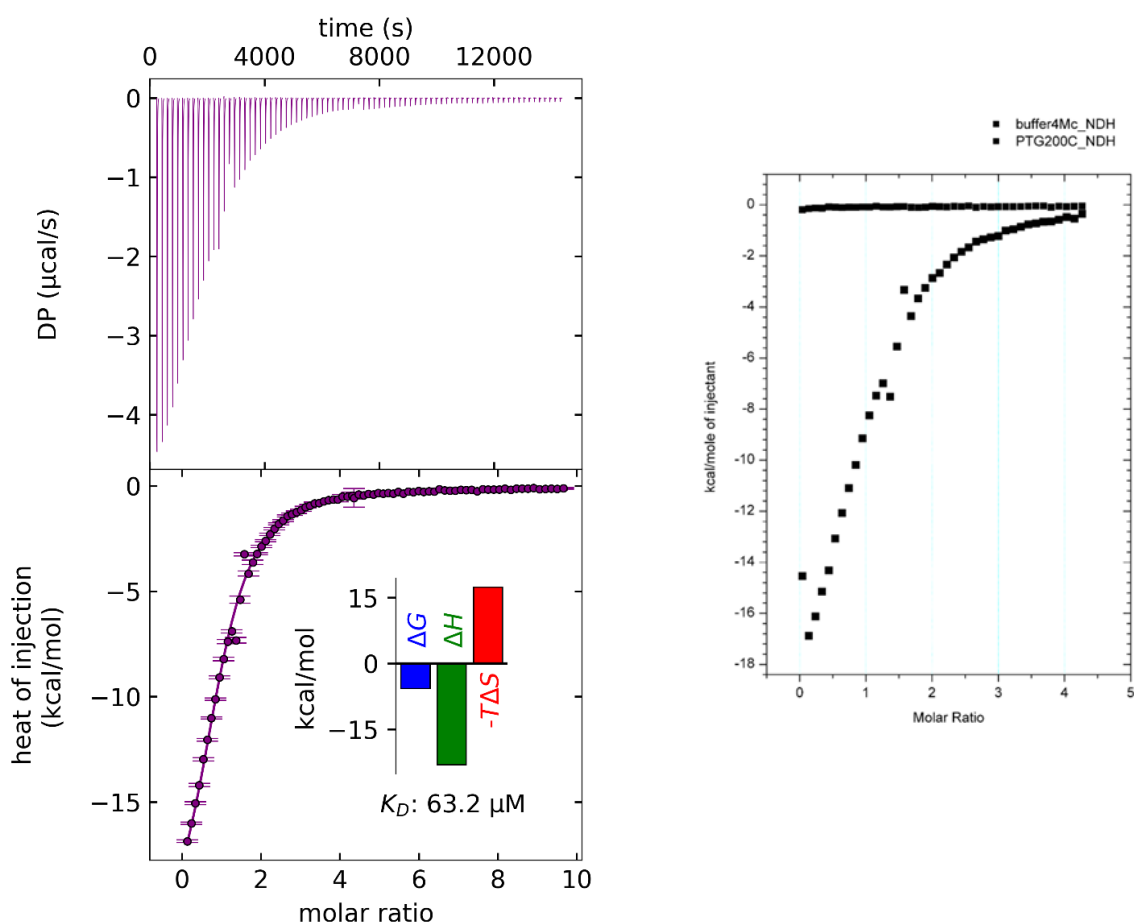
Fig. 31. ITC data from PTG 132-264 titration with β-cyclodextrin. On the left negative control of buffer titration instead of β-cyclodextrin.

	Dataset #2	Dataset #3
'two-sites' model	0.28 ± 0.03	0.40 ± 0.04
Ka1 (M^{-1}) x 10^4	59 ± 38	17 ± 6
ΔH 1(Kcal mol^{-1})	-14.3 ± 0.5	-18.8 ± 0.7
ΔS1 (cal mol^{-1} K^{-1})	-21.6	-39.1
Kd1(1/Ka) µM	1.7 ± 1.1	5.9 ± 2.1
'one-site' model	1.06 ± 0.02	0.94 ± 0.02
Ka2 (M^{-1}) x 10^4	1.64 ± 0.03	1.42 ± 0.05
ΔH 2(Kcal mol^{-1})	-18.6 ± 0.6	-16.4 ± 1.2
ΔS2 (cal mol^{-1} K^{-1})	-42.8	-36.1
Kd2(1/Ka) µM	61 ± 1	70 ± 2

Table 2: A comparison of errors for fitting the data using 'one-site' or 'two-site' model. The errors on the Ka1 and Kd1 are rather large compared to the errors on Ka2 and Kd2

5.7. Crystal structure of PTG CBM21 in complex with cyclodextrin

Extensive crystallization screens of PTG CBM21[132-264] were conducted, testing different protein concentrations, crystallization screens, temperatures, and carbohydrate supplementation. However, only upon β-cyclodextrin addition, crystal growth was observed. Several hits were obtained using JCSG-plus and PACT-plus screens, with two crystals that allowed for crystal structure determination in two different space groups at 1.5 and 2.0 Å resolution (Fig.32). Data collection and refinement statistics are reported in Table 3.
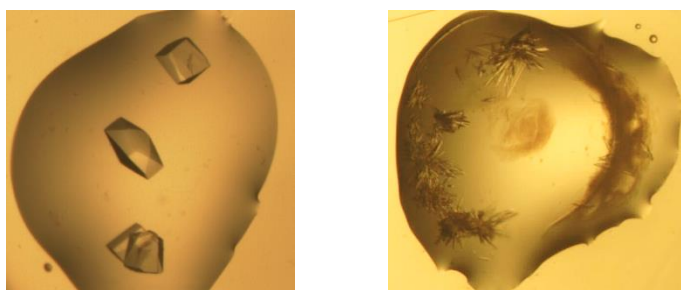
Fig 32: Two crystal forms of PTG 132-264 with β-cyclodextrin. **A** 1.5 Å resolution, grown in 3M NaCl, 0.1M BIS-Tris pH5.5. **B** 2 Å resolution, grown in 0.1M MMT buffer pH4, 25% PEG1500

	PTG-CBM21 orthorhombic	PTG-CBM21 monoclinic
Data Collection		
Space group	P22$_1$2$_1$	I2
Unit-cell parameters (Å, °)	a = 38.59 b = 56.44 c = 70.55	a = 66.83 b = 42.87 c = 168.99 β = 100.67
Wavelength (Å)	0.9718	0.9718
Resolution (Å)	44.07-1.47 (1.50-1.47)	46.81-2.00 (2.05-2.00)
R_{merge} (%)	4.7 (98.1)	10.6 (95.3)
R_{meas} (%)	5.2 (105.8)	11.5 (103.5)
R_{pim} (%)	2.0 (39.1)	4.5 (39.9)

$<I/\sigma(I)>$	17.1 (2.1)	12.0 (2.2)
$CC^{1/2}$	0.999 (0.837)	0.998 (0.778)
Completeness (%)	99.9 (100.0)	99.0 (98.7)
Multiplicity	6.8 (7.1)	6.6 (6.6)
Refinement		
Resolution (Å)	44.07-1.47	41.51-2.00
R_{work}/R_{free} (%)	17.6/19.5	17.3/21.2
R.m.s. deviations		
Bond lengths (Å)	0.005	0.007
Bond angles (°)	0.94	1.06

Table 3. Crystal data collection and refinement statistics.

The PTG 132-264 domain is organized in a β-sandwich of eight antiparallel strands (β2-β9) divided in two β-sheets, with an overall immunoglobulin-like fold (Fig. 33). On one side of the sandwich, the N-terminal region forms a short 3_{10} helix immediately followed by an α-helix and connects to the C-terminal segment through an additional two strands short β-sheet (N-terminal β1 and C-terminal β10). The four chains (one in the orthorhombic crystal and three in the asymmetric unit of the monoclinic system) are identical with an average rmsd of 0.36 Å over Cα atoms.
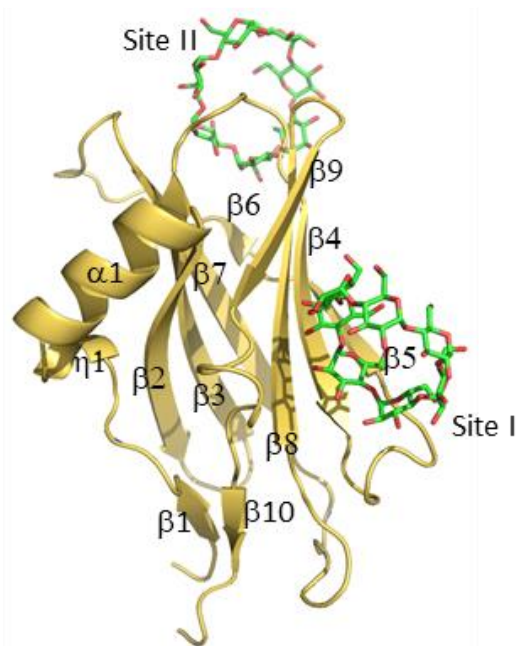
Fig 33. Crystal structure of PTG 132-264 in complex with β-cyclodextrin

Currently, the only reported CBM21 structural information is that of human G_L (PDB: 2EEF, no publication available) and rabbit G_M (PDB: 2M83) (34), both determined by NMR, and crystal structures of starch-binding domain of glucoamylase from *Rhizopus oryzae* fungus- *Ro*GACBM21 (PDB 2VQ4, 2V8L, 2V8M, 4BFO and 4BFN) (35, 36). The eight strands β-sandwich organization is overall conserved. Among the CBM21 structures, the PTG shares the highest grade of similarity with G_L; however, some important differences are observed. The most prominent is the deviation of the β6-β7 position, where the distance between PTG-Y208 Cα atom and the corresponding G_L-Y184 is 5.4 Å. Other changes occur at the β9-β10 loops, the N- and C-termini position, and minor divergences for the β7-β8 loop and the helical region (Fig. 34A). On the other hand, the differences are more marked between the PTG and G_M CBM21, where almost all loops deviate significantly. It lacks the 3_{10} helix, and the position of α1 helix is evidently moved. All those changes contribute to the different protein shape and charge distribution, with most prominent surface shifts at vicinity of the β1-β10 sheet, the 3_{10} helix with its surrounding loops and the β7- β8 loop positions. (Fig. 34B).
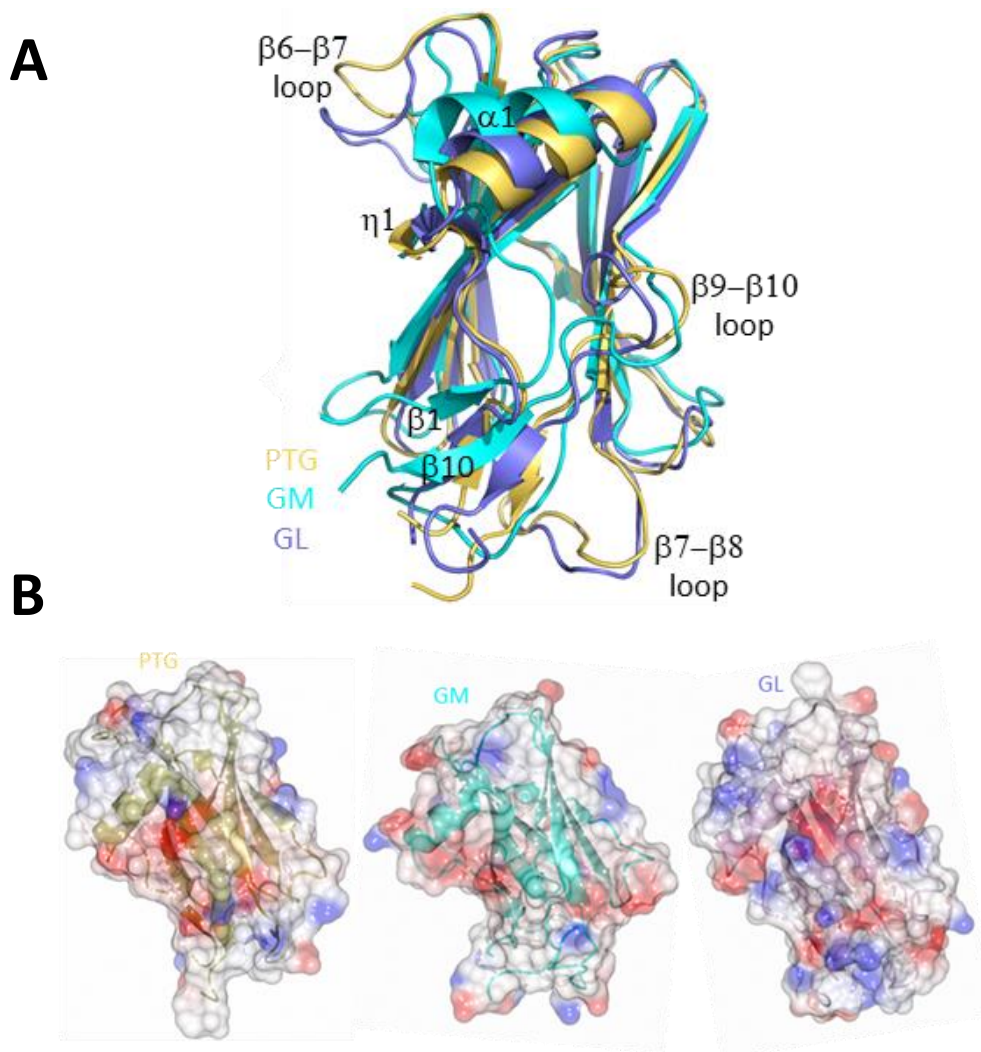
Fig 34. **A** PTG, G$_M$ and G$_L$ CBM21 structure comparison. **B**. PTG, G$_M$ and G$_L$ CBM21 shape and surface distribution.

In the PTG CBM21 crystal structure, we observe two bindings sites for β-cyclodextrin molecule, which accommodates itself at the interface between symmetry-related copies occupying site I in one chain and site II in the other (Fig.35A). Site I is a shallow binding pocket formed by the β4-β5-β8-β9 sheet together with the β4-β5 and the β9-β10 loops, where it accommodates half of the cyclodextrin molecule. The interaction is contributed to the van der Waals contacts with Arg187, Trp193, Tyr196, Cys235 and Trp246 together with H-bonds directed to Arg187, Asn248 and Asn253 side chains. Additional water-bridged H-bonds are formed with Trp193 main chain and Trp246 side chain (Fig. 35B). Site II is generated between the β3-β4 and the β6-β7 loops. The PTG residues Asn177, Phe180, Glu181, Lys182, Asp214 and Tyr203 create a water-mediated close net of H-bonds interacting with β-cyclodextrin. Moreover, extensive van der Waals contacts are

57

present with Lys182, Tyr203 and Phe180 which aromatic residue is inserted in the cyclodextrin ring, resulting in its full immersion (Fig. 35C).
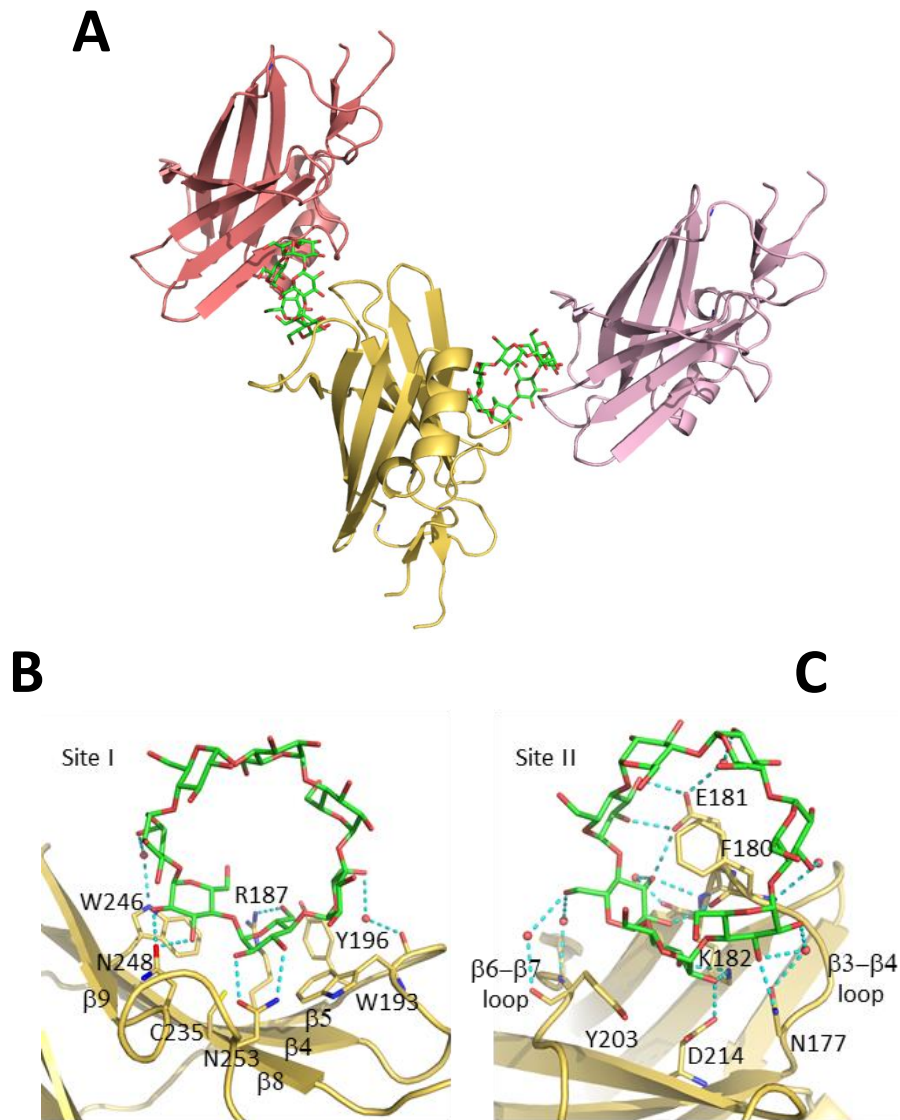
Fig 35. **A** PTG crystal packing contacts with β-cyclodextrin bound at site I at one PTG molecule and site II with another one. **B** β-cyclodextrin bound to site I of PTG. **C** β-cyclodextrin bound to site II of PTG

Site II is highly conserved in sequence among G$_M$ CBM21, G$_L$ CBM21, while site I shows some differences (Fig. 38). Interestingly, both sugar binding sites are present in crystal structures of *Ro*GACBM21 (35, 36). The site II has been identified as the major binding site, allowing accommodation of up to 3 glucose molecules, while site I probably playing role in the recognition of the long-chain polysaccharides, suggesting binding of variety of

sugars with different conformations and length (37, 38). Interestingly, NMR and ITC experiments showed that G_M CBM21 only binds carbohydrates at site II centred on Phe155 (34). No data is currently available for G_L. ITC titration experiments of PTG CBM21 indicate one binding site (section 5.5), which would indicate the site I as crystallographic interface, or as reported for *Ro*GACBM21, an ancillary carbohydrate-interacting site.

5.8. Crystal structure of PP1 in complex with PTG N-terminal peptide

The PTG N-terminal region contains the PP1-binding signature RVxF ([84]RVVF[87]). A number of co-crystallization screens of PP1 with two N-terminal peptides of PTG 81-107 and 81-132 were performed. Crystals grew only with shorter peptide and diffracted to 2.05 Å resolution (Fig. 36), enabling for the reconstruction in the electron density map. Data collection and refinement statistics are reported in Table 4.

Fig 36. Crystal of PP1/PTG[81-107] grown in 0.1 M Tris pH 8, 1.6 M Lithium sulfate, 1.4 M Sodium malonate pH 6.0

	PP1/PTG[8] 1-107
Data Collection	
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 66.38 b = 68.34 c = 119.78
Wavelength (Å)	0.9999
Resolution (Å)	68.34-2.05 (2.11-2.05)
R_{merge} (%)	19.5 (152.2)
R_{meas} (%)	20.2 (157.9)
R_{pim} (%)	5.3 (41.5)
$<I/\sigma(I)>$	10.0 (2.1)
$CC^{1/2}$	0.997 (0.737)
Completeness (%)	100.0 (100.0)
Multiplicity	14.3 (14.3)
Refinement	
Resolution (Å)	59.92-2.05
R_{work}/R_{free} (%)	15.9/19.1
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.92

Table 4. PP1/PTG[81-107] crystal data collection and refinement statistics.

The PTG[81-107] consists of two β-strands, with an intervening 3_{10} helix that accommodates itself on the PP1 C-terminal β-strand (Fig.37A). In the context of the various intermolecular H-bonds involving the main chain atoms responsible for the β-sheet extension, side chains of the PTG peptide also importantly contribute to the binding (Fig. 24B). The aliphatic region of Lys83 is in contact with PP1 Leu289, while its terminal amino group forms a salt bridge with PP1 Asp166 side chain and a H-bond with PP1 Glu287 main chain oxygen (Fig. 36C). PTG Arg84 complement PP1 Asp242 side chain, which also receives a H-bond from main chain nitrogen of PTG Val85. PTG Val85, Phe87 and Leu95 are almost fully buried in the PP1 hydrophobic groove constituted by Ile169, Leu243, Tyr255, Phe257, Arg261, Met283, Leu289 and Phe293, all also becoming barely exposed to the solvent (Fig. 37B and 37C). Following an almost flat region on the PP1 surface, a second groove accommodates PTG His99, stacking to PP1 Tyr78 and hydrogen bonding to Gln294, and PTG Phe101, sandwiched between PP1 Arg74, Leu296 and Pro298 (Fig. 24B and 24D). Additional interactions involve: i) the PTG helical turn that allows Val86, Ala88, Lys91 and Leu93 to embrace the protruding PP1 Met290 and Asp89 to complement the PP1 Arg261 side chain (Fig. 37C) and ii) PTG residues

Thr96, Ala97, Ile98 and Val100 lining with the flattest part of the PP1 surface including Ile295 and the aliphatic regions of Gln294 and Lys297, again becoming largely solvent-excluded (Fig. 37E).
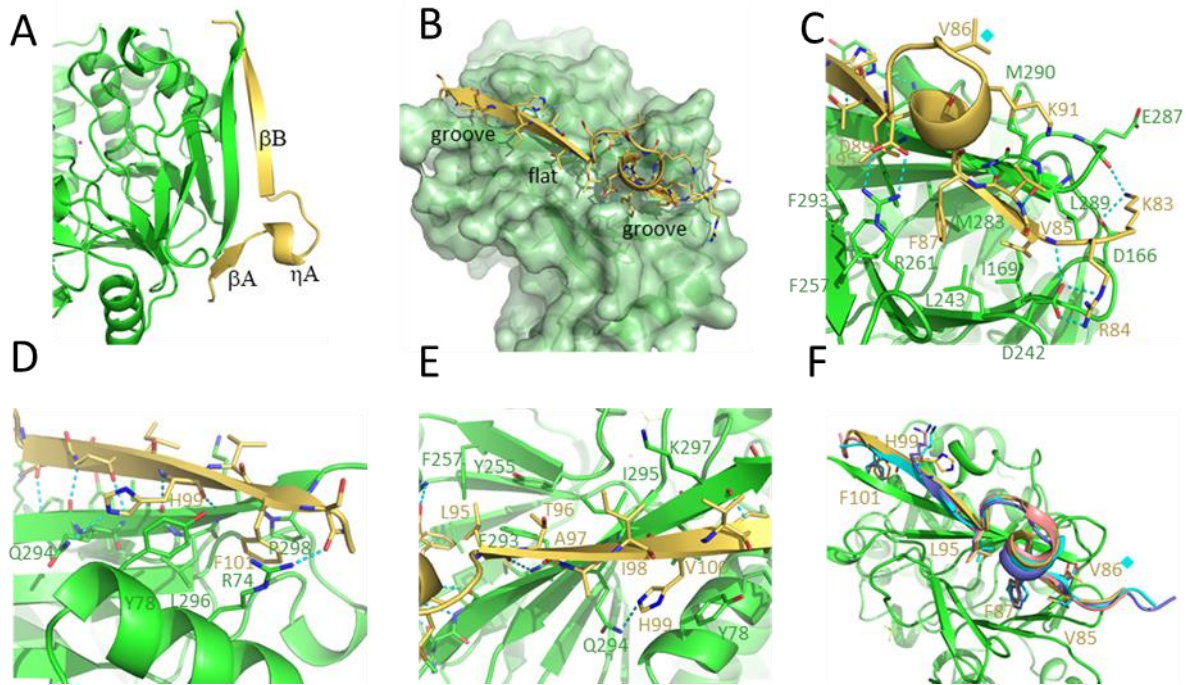
Fig 37. **A.** N-terminal PTG peptide (in yellow) together with C-terminal β-strand of PP1 (in green) **B.** PTG peptide position on the PP1 surface **C.** PTG and PP1 interaction in the conserved RVXF region (Val86 is marked with cyan square, in G$_M$ and G$_L$ is substituted by Ser, that its phosphorylation dictates dissociation from PP1. **D.** βB interaction groove with PP1 **E.** PTG interaction with flat PP1 surface **F.** Comparison of PTG, G$_M$ and G$_L$ binding to PP1 in the conserved RVXF region.
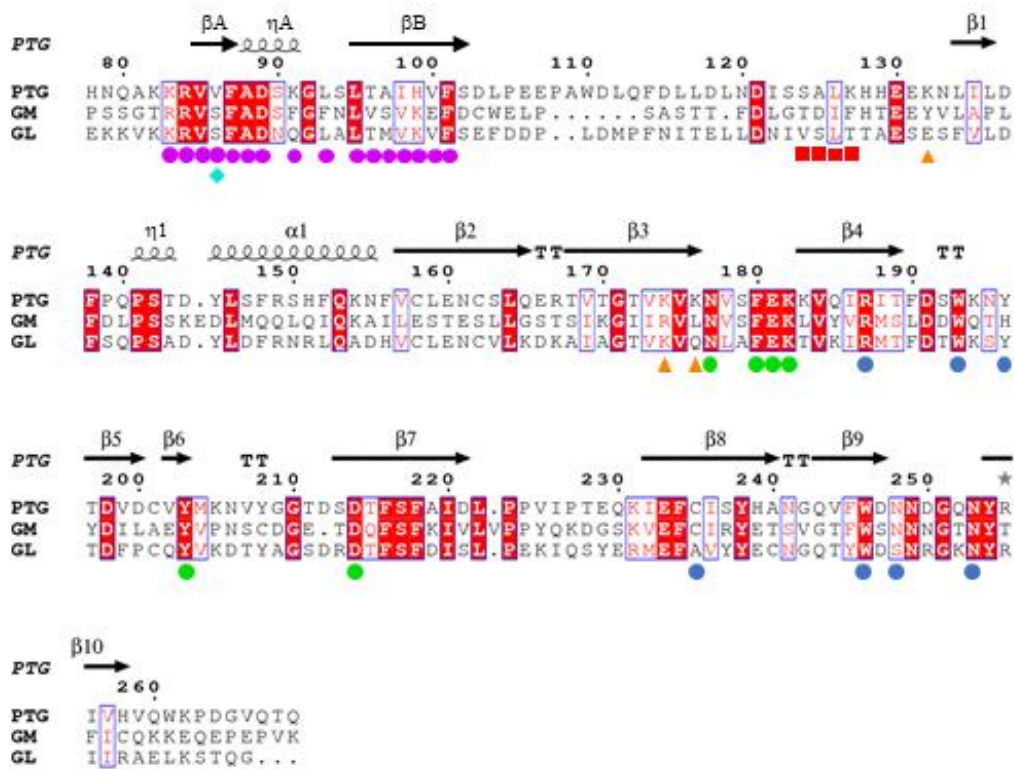
Fig 38. The structure-based sequence alignment of PTG, G$_M$ and G$_L$ CBM21 was generated using ESPript (39). In red are indicated conserved regions among the tree proteins. Purple dots indicate conserved PP1 binding site, green dots highly conserved β-cyclodextrin binding site II, blue dots β-cyclodextrin binding site I, red squares indicate PTG SALK region, triangles PTG ubiquitinated lysines, in cyan Valine substitution in PTG. A secondary structure of PTG is depicted above the alignment box.

Structures of analogous G$_M$ and G$_L$ regions in complex with PP1 have been reported with PDB entries 5ZQV and 5ZT0 representing mouse PP1 and human G$_M$ and G$_L$, respectively, and 6DNO human PP1 and rabbit G$_M$ peptide (34, 51). The superimposition of those structures shows the overall conserved binding mode and secondary structure organization (Fig. 36F). The binding occurs via hydrophobic and stacking interactions of the PTG Val85, Phe87, Leu95, His99 and Phe101, with the difference that, in both G$_M$ and G$_L$, the His99 is substituted by a lysine (Fig. 38). The PP1-G$_L$ structure lacks density from Val78 onwards, impairing the determination of G$_L$ Phe79 side chain position, which corresponds to PTG Phe101 and G$_M$ Phe80, both crucial for the binding interface. It remains to be clarified whether this is a peculiarity of G$_L$ or a crystallographic artifact due to packing restriction and poor resolution of the associated structure (3.3 Å).

Unlike in PTG, the X residue in the RVXF region of G$_M$ and G$_L$ is the serine, that plays crucial role in their regulation. Following the activation of the adrenalin or glucagon signaling cascades, those serines get phosphorylated by protein kinase A, which inhibits

the binding to PP1 and in consequence PP1 become unable to act on glycogen metabolism (40-44) (Fig. 38, 37C and 37F). PTG, instead of serine contains valine, and its regulation mechanism relies on Lys132, Lys174 and Lys176 ubiquitination (Fig. 38), among which none is conserved in G_M and only one in G_L, however, it is not reported as an ubiquitination site (45).

5.9. Crystal structure of the PP1/PTG complex

Extensive crystallization screens of the PTG/PP1 complex were performed, resulting in poorly diffracting crystals (7-10 Å) (Fig. 39). Finally, overnight dehydration of the crystals in solution with 10% higher precipitant concentration supplemented with 10% glycerol and 1mM β-cyclodextrin resulted in anisotropical diffraction to 2.7 Å in two directions, but only to 3.9 Å in the worst direction. Structure was determined by molecular replacement using the above described PTG-CBM and PP1-PTG (81-107) structures. Data collection and refinement statistics are reported in Table 5.
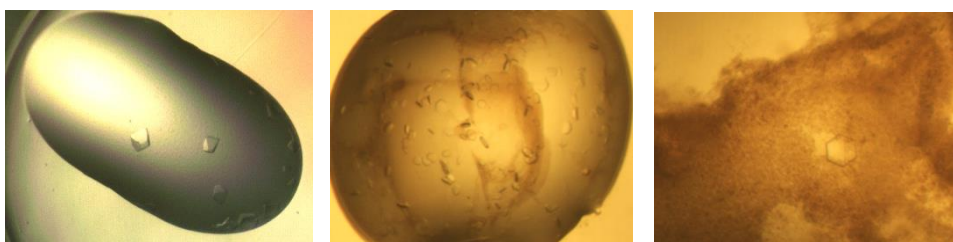
Fig 39. Crystal of PP1/PTG complex grown in 1.0 M Sodium malonate pH 5.0, 0.1 M Sodium acetate trihydrate pH 4.5 at 4°C.

	PP1/PTG[7] 0-264
Data Collection	
Space group	P $6_1$22
Unit-cell parameters (Å, °)	a = 153.60
	b = 153.60
	c = 285.62
Wavelength (Å)	1.0000
Resolution (Å)	120.58-2.69
	(2.88-2.69)
R_{merge} (%)	36.3 (324.7)
R_{meas} (%)	37.4 (334.4)
R_{pim} (%)	9.0 (79.8)
$<I/\sigma(I)>$	13.3 (1.5)
$CC^{1/2}$	0.997 (0.422)
Completeness (%)	Sph. 66.3 (18.9)
	Ell. 95.2 (73.1)
Multiplicity	32.0 (33.9)
Refinement	
Resolution (Å)	52.30-2.69
R_{work}/R_{free} (%)	17.7-21.3
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.23

Table 5. PP1/PTG crystal data collection and refinement statistics

Two PP1/PTG copies/asu were located with additional electron density corresponding to cyclodextrin molecules and part of the PTG sequence intervening between the N-terminal region and the CBM21.

All previously described interaction between the CBM21 (residues 133-259) and β-cyclodextrin are conserved as is the binding mode of the PTG region 83-103 to PP1. Residues 111-128 could be almost identically built in the two copies, crossing PP1 and joining the N-terminal peptide on one PP1 side with the CBM21 on the opposite PP1 surface (Fig. 40A).

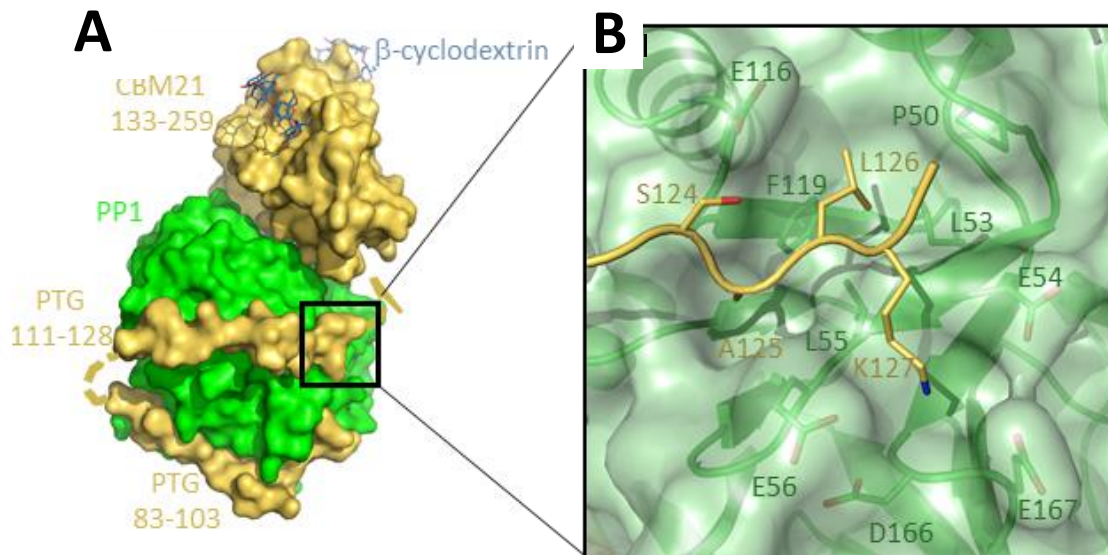Fig 40. **A** PTG/PP1 complex with interaction regions between two proteins **B** PTG N-terminal [124]SALK[127] region

Extensive interactions are only observed for PTG residues [124]SALK[127] (Fig. 40B): Ala125 and Leu126 insert in a small groove defined by PP1 residues Pro50, Leu53, Leu55 and Phe119; the flanking Ser124 contacts PP1 Glu116, while Lys127 inserts in a very acidic region of the PP1 surface defined by Glu54, Glu56, Asp166 and Glu167. The CBM motif sits on an almost flat PP1 surface by using its β1 strand, the following 3_{10} helix and the intervening loop. This poorly interlocked interface is indeed slightly different in the two dimers in the asymmetric unit, then pliable, due to a rotation of the CBMs in the context of identical locations for the other PTG regions (Fig. 41B). It is mainly polar in nature, variably (depending on the orientation) involving PTG residues Ile135, Asp137, Pro139, Gln140, Thr143 and Asp144 and PP1 amino acids Leu40, Lys41, Arg43, Glu44, Leu47 Ser48, Lys147, Lys150 and Asp154.

Interestingly, the PP1/PTG complex organizes in the crystal as a tight pseudo-knotted dimer of dimers (Fig. 41A). PTG region 111-117 pairs with its symmetric mate forming an intertwined antiparallel β-sheet (Fig. 41C). The driving force in this assembly relies however in a very hydrophobic core, formed by residues in all four chains, namely PTG Leu104, Phe106, Leu116, Leu117 and Leu119 and PP1 Ile13, Gly14, Leu17, Leu73, Phe81, Pro82, Pro83, Tyr110 and Tyr114, which become largely solvent-excluded. In order for the tetramer to form,

PTG should detach its CBM21 (and SALK region) from the PP1 surface, thus confirming the dynamic nature of the interaction in this region, while remaining complexed to PP1 through its N-terminal region.
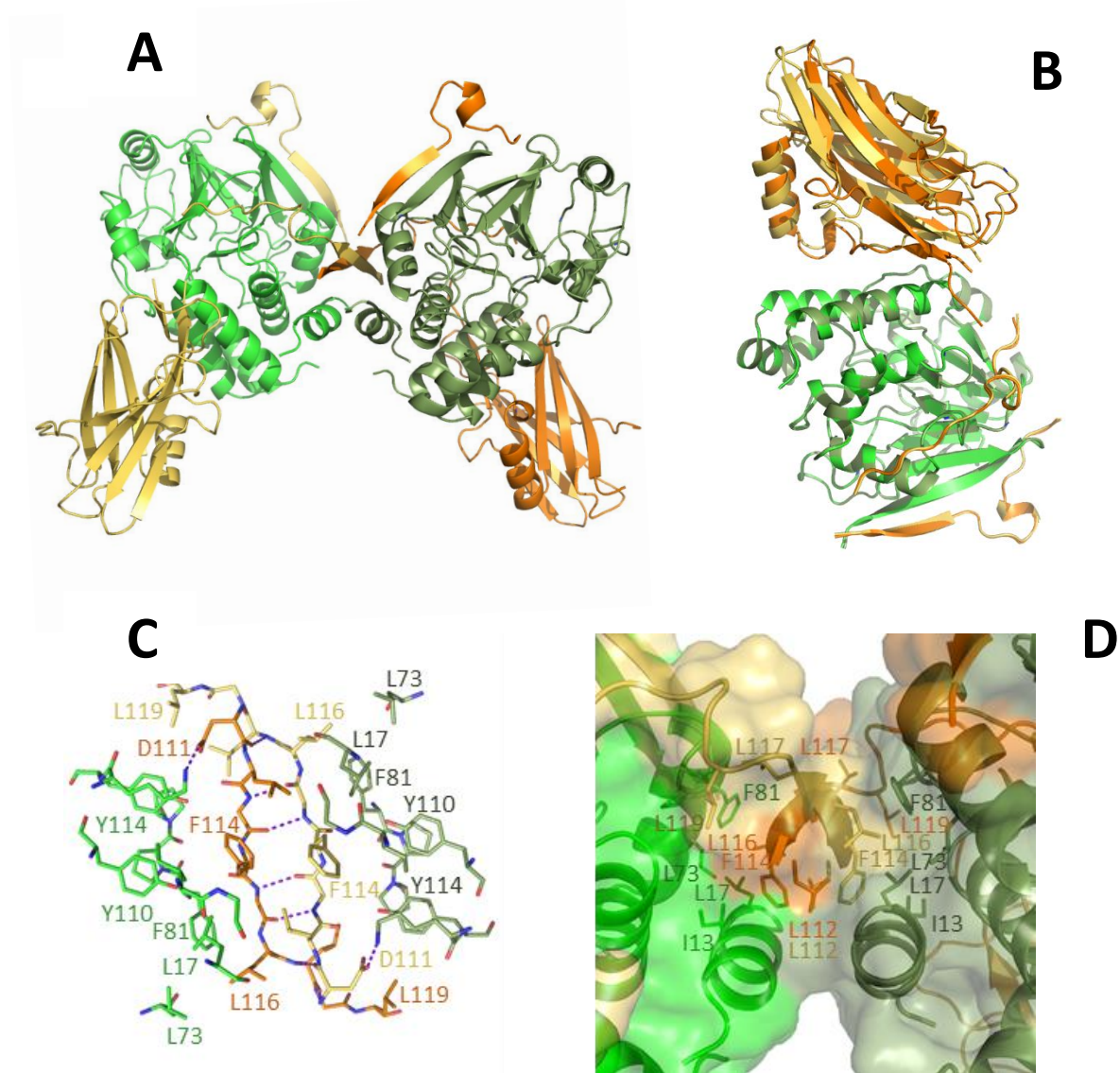
Fig 41. **A** PTG/PP1 representation in crystal as a tight pseudo-knotted dimer of dimers **B** Two PTG/PP1 dimer overlay **C.** Detailed representation of the PTG interplay with its symmetric dimer.

Comparison with other PP1 complexes

A limited number of PP1 structures in complex with large portions of the partner proteins are available. The identical PP1 holoenzymes with Neurabin-1 and Neurabin-2 (49) (PDB 3EGG and 3HVQ) significantly diverge from the PP1/PTG complex (Fig. 42). The flexible region connecting the N-terminal PP1-recognition portion of Neurabin and its PDZ domain also embraces PP1 but on the opposite side with respect to the PTG linker

66

region; the PP1/PDZ interface is adjacent but different from the PP1/CBM21. Mutational studies and the absence of electron density for one PDZ domain out of the two in the asymmetric unit indicate that recognition of PP1 and Neurabin relies on the RVXL motif and surrounding residues, while the other interacting surfaces configure as transient. The path followed by the linker region of Neurabin is similarly retraced in three other PP1 holoenzymes, namely those with PHACTR1 (phosphatase and actin regulator 1), PPP1R15B and PNUTS (PP1 nuclear targeting subunit), directed by the so called Arg motif (46-48).
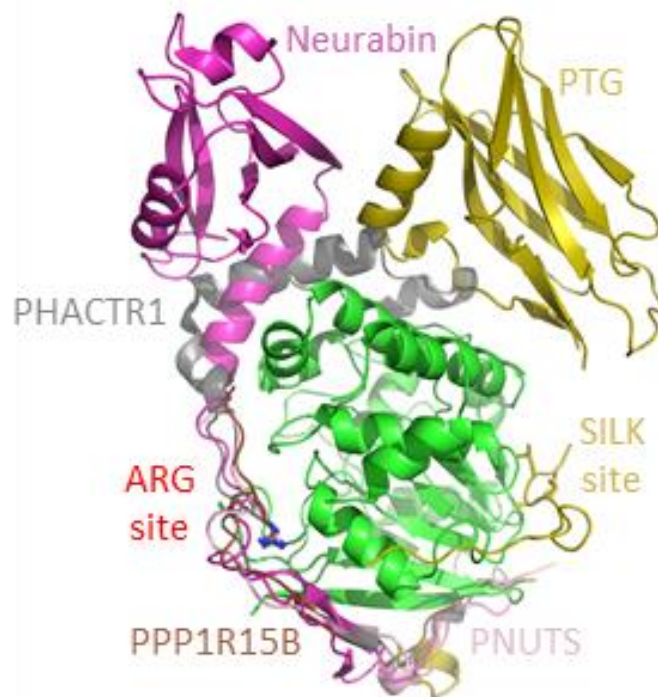
Fig 42. PP1/PTG complex comparison with PP1/Neurabin, PHACTR1, PPP1R15B and PNUTS.

PTG linker instead perfectly superposes, using its ^{124}SALK^{127} sequence, to similar stretches observed in KNL1 (kinetochore scaffold 1 protein) and PPP1R2 (protein phosphatase inhibitor 2) (50-51) (Fig. 43). This region, known as the SILK PP1-binding motif was reported, hypothesized and searched always at the N-terminus of the RVXF motif (52-55). To the best of our knowledge, this is the first evidence of a SILK motif downstream of the RVXF binding region highlighting, that previous computational studies may have missing combinations of PP1-binding motifs in its targeting subunits, as was the case for PTG. Notably, the SALK sequence is not conserved in G$_M$ and G$_L$, with poor homology among the three proteins in the linker region (Fig. 38). It is worth noticing instead that the SSILK region of KNL1 (corresponding to PTG SSALK) is phosphorylated

by Aurora B on both serines, thus impairing its binding to PP1, which in turn is removed from the kinetochore (50, 56).
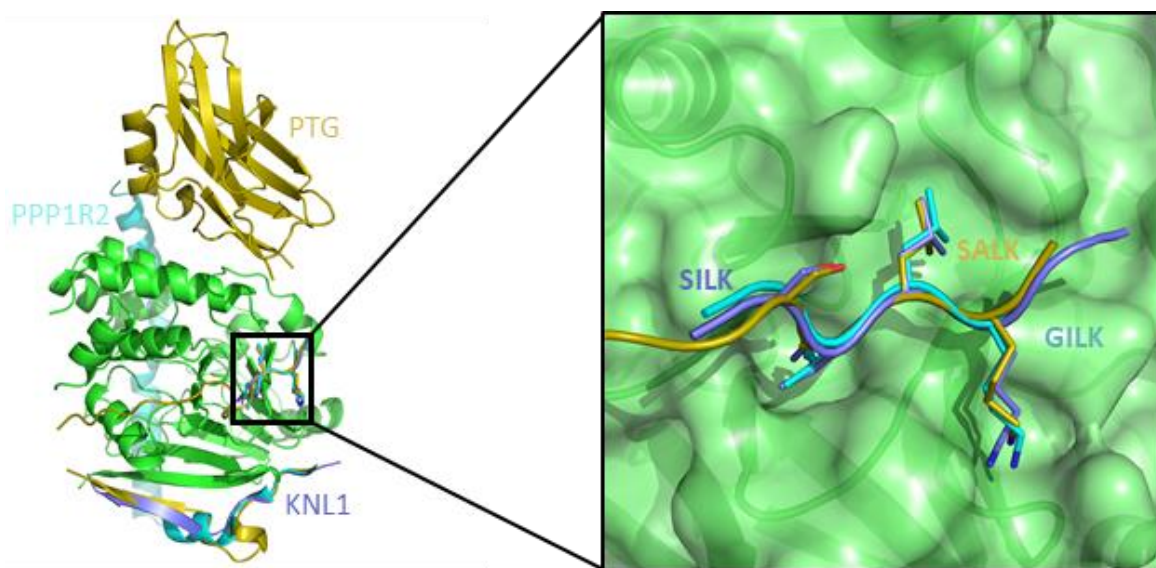
Fig 43. PTG/PP1 SALK interaction region in comparison with PP1R2 GILK and KNL1 SILK regions.

PTG binding to PP1 is dictated by the RVXF motif with the SALK sequence, but not by the CBM21, as demonstrated by grating-coupled interferometry (section 5.5).

5.10.   PP1/PTG complex SEC-SAXS analysis

SAXS provides three-dimensional low-resolution structures and is a powerful tool for structure validation and analysis of flexible systems, which complements high-resolution techniques as X-ray crystallography. We performed two SEC-SAXS experiments investigating PP1/PTG complex in presence and absence of β-cyclodextrin, resulting in similar density envelope in both cases (Fig. 44 B and C). Binding of cyclodextrin is evident by comparing the two elution curves and Rg (Fig. 43 A). SAXS data clearly revealed that in solution the crystallographic structure is only one of the explored orientations; indeed, data are best interpreted as a conformational ensemble, where PTG, while always bound through its N-terminal region to PP1, variously orients its CBM with respect to PP1 (Fig. 44 B and C, Fig. 45). This fully agrees with the affinity data, showing no contribution to the binding affinity by the PTG CBM.
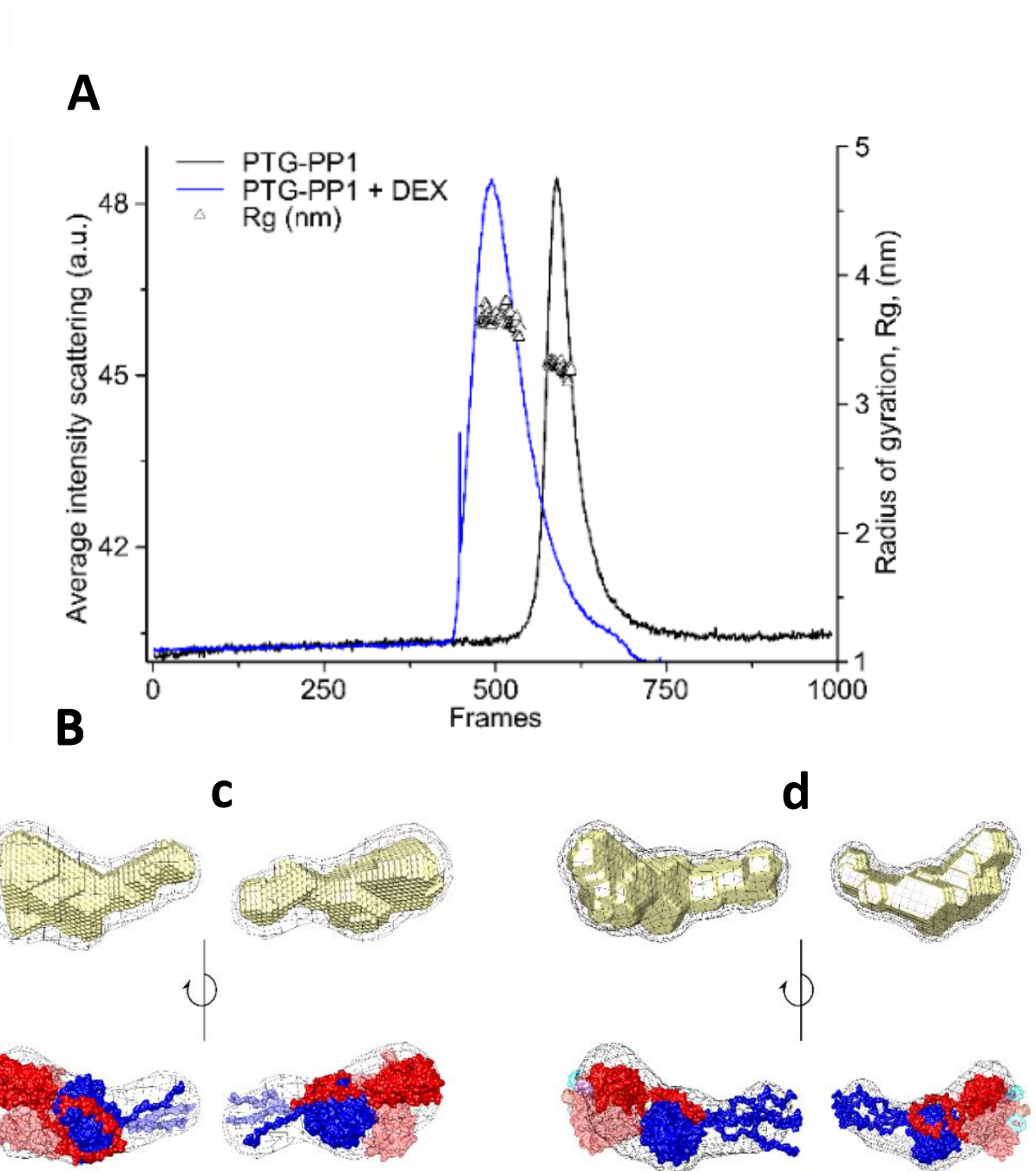
Fig 44. A SEC-SAXS chromatograms of PTG-PP1 complex alone and bound to cyclodextrin (DEX). The black and blue lines represent the total summed scattering intensity for PTG-PP1 and PTG-PP1 + DEX, respectively; triangles represent the calculated radius of gyration, Rg in nm, in the selected frames) B. On the top, ab initio DAMMIF model for PTG-PP1 complex in beads and volume representations; on the bottom, illustrative 3D structures obtained after EOM modelling describing the PTG-PP1 complex, here visualized in surface mode. In blue, a representative ensemble for PP1 and in red for PTG. In faint colour, other ensembles populating the surface envelope of the PTG-PP1 complex. (c) Similarly to panel (d), ab initio DAMMIF model and EOM models for PTG-PP1 complex bound to DEX.

Fig 45. SAXS studies on the PTG-PP1 complex bound to cyclodextrin. (A) Shown are I(q) versus q experimental SAXS profiles for PTG-PP1 complex alone (black dots) and bound to cyclodextrin (DEX, blue dots) with the EOM fit (red line). Chi2 values for the EOM fitting are indicated (χ^2 of 0.98 and 2.3 for PTG-PP1 complex and bound to DEX, respectively). The curves are shifted by an arbitrary offset for better comparison. In the inset, the Guinier fits for the two samples and the calculated radius of gyration (Rg of 33 and 36 Å for PTG-PP1 complex alone and bound to DEX, respectively) are shown. The lower plots show the residuals defined as [Iexp(q) - Icomp(q)]/σexp(q), corresponding to the difference between the experimental and the computed intensities weighted by the experimental uncertainty. (B) Superposition of illustrative 3D EOM structures describing 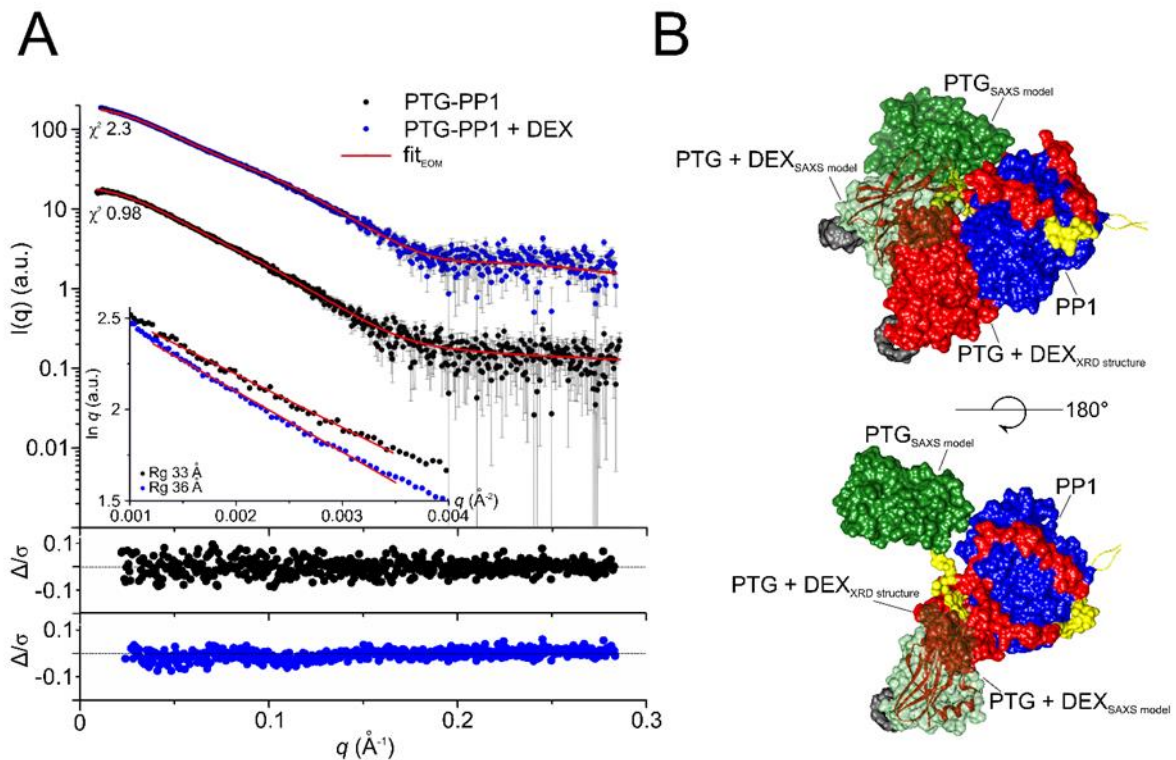the PTG-PP1 complex alone and bound to DEX together with the X-ray crystal structure (XRD) of the PTG-PP1 complex bound to DEX. All the structures are represented in surface mode (except for the PP1 segment from residue 299 to 323, here not shown). The PTG and PP1 structures were solved by X-ray crystallography in this study and used for EOM modelling as rigid bodies. PP1 protein (residues 6-300) is shown in blue; the PTG segments from residue 83 to 102 and from 111 to 128 are shown in red; missing PTG residues were modelled with EOM (residues 103-110 and 129-132) and shown in yellow. The PTG domain from residue 133 to 259 was modelled by EOM as free rigid body connected by the 4 amino acids linker (129-HEEK-132). In red, the PTG domain bound to DEX (in dark grey) in the position interacting with PP1 as observed in the XRD structure (PTG+DEX$_{XRD structure}$). In green and in light green with transparency the PTG domain alone and bound to DEX (in dark grey), respectively, in representative positions where the PTG domain is not interacting with PP1, as postulated in the EOM analysis (in the figure indicated as PTG$_{SAXS model}$ and PTG+DEX $_{SAXS model}$).

6. CONCLUSIONS AND FUTURE PERSPECTIVES

Our research focuses on protein targeting to glycogen, a PP1 regulatory subunit that emerges as a potential druggable target for Lafora Disease treatment. Up to date, no structural information of PTG protein was available. In the presented work we describe a quest for elucidating the basis of mechanism and regulation of PTG/PP1 complex starting from construct design, through protein production and finally arriving to the atomic structures. We were able to obtain a crystal structure of human PTG CBM21 in complex with β-cyclodextrin, which has not been reported before. It reveals a typical immunoglobulin-like fold, and comparing with other regulatory subunits, it shares more similarities with G_L than with G_M. Two carbohydrate binding sites have been identified, among which site II being the main determinant for this interaction and the one conserved among other CBM21 proteins. Site I could play a role as an additional support for binding to long-chain polysaccharides, as hypothesized for *Ro*GACBM21. Furthermore, we analyzed crystal structure of PP1 bound to N-terminal PTG peptide containing RVXF signature motif. It organizes itself on the PP1 surface in two β-strands connected by a 3_{10} helix. The binding mode of this region relies on hydrophobic and stacking interactions with two main PP1 grooves and is well conserved among PTG, G_M and G_L, with the exception of His99 in the second pocket being substituted by a lysine in both G_M and G_L. Finally, we were able to obtain crystallographic data of the PTG/PP1 complex resulting in being the first PP1 holoenzyme reported that includes a large portion of PTG regulatory subunit. It confirms that binding to PP1 occurs mainly due to N-terminal RVXF conserved motif supported by additional SALK sequence, while poor contribution is observed from CBM21. This was confirmed by in-solution SAXS analyses showing that the CBM can assume multiple orientations with respect to PP1 in the context of a tight complex determined by the interaction between PP1 and the PTG N-terminal region. We hypothesize that the SALK sequence separates from PP1 allowing CBM21 domain to change its position, which could facilitate substrate recognition and recruitment. This would also explain the peculiar tetramer complex observed in the PTG/PP1 crystal structure.

The crystallization optimization and structural determination of PTG and PTG/PP1 complex lay the ground for further investigations towards rational drug design targeting

PTG. First proposed experimental steps would be PTG structure analysis for the presence of druggable pockets by using a set of bioinformatic software. Next, an extensive virtual screening for hit identification would be performed. Selected compounds would be evaluated by measuring their binding kinetics. Additionally, co-crystallization trials with PTG protein would be carried out. In the case of obtaining well-diffracting crystals, the data will be used to guide further selection of compound derivatives. Moreover, further structural characterization of PTG in complexes with its different partners, such as PTG/GS, PTG/PP1/GS, PTG/laforin/malin would be attempted.

7. REFERENCES

1. Glycogen and its metabolism: some new developments and old themes Peter J. Roach1, Anna A. Depaoli-Roach, Thomas D. Hurley, and Vincent S. Tagliabracci2 Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, U.S.

2. Enerkin, Glycogen, DOI: 10.1016/B978-0-12-391909-0.50023-2

3. Glycogen metabolism in humans☆, ☆☆ María M. Adeva-Andany ∗, Manuel González-Lucán, Cristóbal Donapetry-García, Carlos Fernández-Fernández, Eva Ameneiros-Rodríguez

4. Glycogen and its Metabolism Peter J. Roach*

5. The dynamic life of the glycogen granule Published, Papers in Press, February 26, 2018, DOI 10.1074/jbc.R117.802843 X Clara Prats‡§1, Terry E. Graham¶ , and X Jane Shearer

6. Meléndez, R., Meléndez-Hevia, E., and Cascante, M. (1997) How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. J. Mol. Evol. 45, 446–455 CrossRef Medline

7. The regulation of muscle glycogen: the granule and its
proteinsT. E. Graham, Z. Yuan, A. K. Hill and R. J. Wilson
Human Health and Nutritional Sciences, University of Guelph, Guelph, ON, Canada

8. Prokaryotes to Eukaryotes: Insights Into the Molecular Structure of Glycogen ParticlesQing-Hua Liu1,2, Jia-Wei Tang3, Peng-Bo Wen3, Meng-Meng Wang4,Xiao Zhang3 and and Liang Wang3,4 Front. Mol. Biosci., 29 April 2021 | https://doi.org/10.3389/fmolb.2021.673315

9. Brewer MK, Gentry MS. Brain Glycogen Structure and Its Associated Proteins: Past, Present and Future. Advances in Neurobiology. 2019;23:17-81. DOI: 10.1007/978-3-030-27480-1_2. PMID: 31667805; PMCID: PMC7239500.

10. Is liver glycogen fragility a possible drug target for diabetes? Cheng Li,Zhenxia Hu, First published: 21 November 2019, https://doi.org/10.1096/fj.201901463RR

11. Rybicka K. K. (1996) Glycosomes–the organelles of glycogen metabolism. Tissue Cell 28, 253–265 10.1016/S0040-8166(96)80013-9

12. Brain Glycogen Structure and Its Associated Proteins: Past, Present and Future.

Brewer MK1,Gentry MS1 Advances in Neurobiology, 01 Jan 2019, 23:17-81 DOI: 10.1007/978-3-030-27480-1_2 PMID: 31667805 PMCID: PMC7239500

13. Sullivan M. A., Aroney S. T., Li S., Warren F. J., Joo J. S., Mak K. S15. Stapleton D. I., Bell-Anderson K. S., and Gilbert R. G. (2014) Changes in glycogen structure over feeding cycle sheds new light on blood-glucose control. Biomacromolecules 15, 660–665 10.1021/bm401714v

14. Quantification of subcellular glycogen in resting human muscle: granule size, number, and location I. Marchand, K. Chorneyko, M. Tarnopolsky, S. Hamilton, J. Shearer NOV 2002https://doi.org/10.1152/japplphysiol.00585.2001

15. Molecular Structure of Human-Liver Glycogen Bin Deng, Mitchell A. Sullivan, Cheng Chen, Jialun Li,Prudence O. Powell, Zhenxia Hu, Robert G. Gilbert Published: March 2, 2016 https://doi.org/10.1371/journal.pone.0150540

16. Goldsmith E, Sprang S, Fletterick R. Structure of maltoheptaose by difference Fourier methods and a model for glycogen. J Mol Biol. 1982;156:411–427.

17. Graham TE. Glycogen: an overview of possible regulatory roles of the proteins associated with the granule. Appl Physiol Nutr Metab. 2009;34:488–492.

18. Shearer J, Graham TE. Novel aspects of skeletal muscle glycogen and its regulation during rest and exercise. Exercise Sport Sci Rev. 2004;32:120–126.

19. Meyer F, Heilmeyer LM, Jr, Haschke RH, Fischer EH. Control of phosphorylase activity in a muscle glycogen particle. I. Isolation and characterization of the protein–glycogen complex. J Biol Chem. 1970;245:6642–6648.[PubMed]

20. Christiansen C, Abou Hachem M, Janecek S, Vikso-Nielsen A, Blennow A, Svensson B. The carbohydrate-binding module family 20: diversity, structure, and function. FEBS J. 2009;276:5006–5029.

21. Machovic M, Janecek S. The evolution of putative starch-binding domains. FEBS Lett. 2006;580:6349–6356

22. Polekhina G, Gupta A, Michell BJ, van Denderen B, Murthy S, Feil SC, Jennings IG, Campbell DJ, Witters LA, Parker MW, et al. AMPKβ subunit targets metabolic stress sensing to glycogen. Curr Biol. 2003;13:867–871.

23. Ganesh S, Amano K, Delgado-Escueta AV, Yamakawa K. Isolation and characterization of mouse homologue for the human epilepsy gene, EPM2A. Biochem Biophys Res Commun. 1999;257:24–28.

24. Jiang S, Heller B, Tagliabracci VS, Zhai L, Irimia JM, DePaoli-Roach AA, Wells CD, Skurat AV, Roach PJ. Starch binding domain-containing protein 1/genethonin 1 is a novel participant in glycogen metabolism. J Biol Chem. 2010;285:34960–34971

25. Animal Models of Glycogen Adithya Raghavan, William J. Craigen

26. Turnbull J, DePaoli-Roach AA, Zhao X, Cortez MA, Pencea N, Tiberia E, Piliguian M, Roach PJ, Wang P, Ackerley CA, Minassian BA. (2011) PTG depletion removes Lafora bodies and rescues the fatal epilepsy of Lafora disease. PLoS Genet. 7: e1002037.

27. Turnbull J, Epp JR, Goldsmith D, Zhao X, Pencea N, Wang P, Frankland PW, Ackerley CA, Minassian BA. (2014) PTG protein depletion rescues malin-deficient Lafora disease in mouse. Ann. Neurol. 75: 442–446.

28. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-132. doi:10.1107/S0907444909047337

29. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr. 2011 Apr 1;67(4):282–92

30. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser: crystallographic software. J Appl Crystallogr. 2007 Aug 1;40(4):658–74.

31. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G. (2018). STARANISO (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi). Cambridge, United Kingdom: Global Phasing Ltd.

32. Buchan DWA, Jones DT (2019). The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz297

33. Katayama H., McGill M., Kearns A., Brzozowski M., Degner N., Harnett B., Kornilayev B., Matkovic-Calogovic D., Holyoak T., Calvet J.P., Gogol E.P., Seed J., and Fisher MT., (2009) "Strategies for folding of affinity tagged proteins using GroEL and osmolytes" J. Structural and Functional Genomics. 10, 57-66. epub Dec 2008

34. Identification of the substrate recruitment mechanism of the muscle glycogen protein phosphatase 1 holoenzyme.Kumar GS, Choy MS, Koveal DM, Lorinsky MK, Lyons SP, Kettenbach AN, Page R, Peti W.Sci Adv. 2018 Nov 14;4(11):eaau6044

35. Crystal structures of the starch-binding domain from Rhizopus oryzaeglucoamylase reveal a polysaccharide-binding path.Tung JY, Chang MD, Chou WI, Liu YY, Yeh YH, Chang FY, Lin SC, Qiu ZL, Sun YJ.Biochem J. 2008 Nov 15;416(1):27-36.

36. Crystal structures of starch binding domain from Rhizopus oryzaeglucoamylase in complex with isomaltooligosaccharide: insights into polysaccharide binding mechanism of CBM21 family.Chu CH, Li KM, Lin SW, Chang MD, Jiang TY, Sun YJ.Proteins. 2014 Jun;82(6):1079-85.

37. The family 21 carbohydrate-binding module of glucoamylase from Rhizopusoryzae consists of two sites playing distinct roles in ligand binding.Chou WI, Pai TW, Liu SH, Hsiung BK, Chang MD.Biochem J. 2006 Jun 15;396(3):469-77.

38. Two unique ligand-binding clamps of Rhizopus oryzae starch binding domain for helical structure disruption of amylose.Jiang TY, Ci YP, Chou WI, Lee YC, Sun YJ, Chou WY, Li KM, Chang MD.PLoS One. 2012;7(7):e41131.

39. Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". Nucl. Acids Res. 42(W1), W320-W324. doi: 10.1093/nar/gku316

40. Phosphorylation of the skeletal muscle glycogen-targetting subunit of protein phosphatase 1 in response to adrenaline in vivo.Walker KS, Watt PW, Cohen P.FEBS Lett. 2000 Jan 21;466(1):121-4

41. Overexpression of the glycogen targeting (G(M)) subunit of protein phosphatase-1.Yamamoto-Honda R, Honda Z, Kaburagi Y, Ueki K, Kimura S, Akanuma Y, Kadowaki T.Biochem Biophys Res Commun. 2000 Sep 7;275(3):859-64.

42. Mutations of the serine phosphorylated in the protein phosphatase-1-binding motif in the skeletal muscle glycogen-targeting subunit.Liu J, Wu J, Oliver C, Shenolikar S, Brautigan DL.Biochem J. 2000 Feb 15;346 Pt 1(Pt 1):77-82.

43. Multisite phosphorylation of the glycogen-binding subunit of protein phosphatase-1G by cyclic AMP-dependent protein kinase and glycogen synthase kinase-3.Dent P, Campbell DG, Hubbard MJ, Cohen P.FEBS Lett. 1989 May 8;248(1-2):67-72.

44. Distinctive regulatory and metabolic properties of glycogen-targeting subunits of protein phosphatase-1 (PTG, GL, GM/RGl) expressed in hepatocytes.Gasa R, Jensen PB, Berman HK, Brady MJ, DePaoli-Roach AA, Newgard CB.J Biol Chem. 2000 Aug 25;275(34):26396-403.

45. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites.Akimov V, Barrio-Hernandez I, Hansen SVF, Hallenborg P, Pedersen AK, Bekker-Jensen DB, Puglia M, Christensen SDK, Vanselow JT, Nielsen MM,

Kratchmarova I, Kelstrup CD, Olsen JV, Blagoev B.Nat Struct Mol Biol. 2018 Jul;25(7):631-640.

46. Molecular basis for substrate specificity of the Phactr1/PP1 phosphatase holoenzyme.Fedoryshchak RO, Přechová M, Butler AM, Lee R, O'Reilly N, Flynn HR, Snijders AP, Eder N, Ultanir S, Mouilleron S, Treisman R.Elife. 2020 Sep 25;9:e61509.

47. G-actin provides substrate-specificity to eukaryotic initiation factor 2αholophosphatases.Chen R, Rato C, Yan Y, Crespillo-Casado A, Clarke HJ, Harding HP, Marciniak SJ, Read RJ, Ron D.Elife. 2015 Mar 16;4:e04871.

48. Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code.Choy MS, Hieke M, Kumar GS, Lewis GR, Gonzalez-DeWhitt KR, Kessler RP, Stein BJ, Hessenberger M, Nairn AC, Peti W, Page R.Proc Natl Acad Sci U S A. 2014 Mar 18;111(11):4097-102.

49. Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites.Ragusa MJ, Dancheck B, Critton DA, Nairn AC, Page R, Peti W.Nat Struct Mol Biol. 2010 Apr;17(4):459-64.

50. KNL1 Binding to PP1 and Microtubules Is Mutually Exclusive.Bajaj R, Bollen M, Peti W, Page R.Structure. 2018 Oct 2;26(10):1327-1336.e4.

51. Structural basis for regulation of protein phosphatase 1 by inhibitor-2.Hurley TD, Yang J, Zhang L, Goodwin KD, Zou Q, Cortese M, Dunker AK, DePaoli-Roach AA.J Biol Chem. 2007 Sep 28;282(39):28874-28883.

52. Cracking the phosphatase code: docking interactions determine substrate specificity.Roy J, Cyert MS.Sci Signal. 2009 Dec 8;2(100):re9.

53. The extended PP1 toolkit: designed to create specificity.Bollen M, Peti W, Ragusa MJ, Beullens M.Trends Biochem Sci. 2010 Aug;35(8):450-8.

54. Docking motif-guided mapping of the interactome of protein phosphatase-1.Hendrickx A, Beullens M, Ceulemans H, Den Abt T, Van Eynde A, Nicolaescu E, Lesage B, Bollen M.Chem Biol. 2009 Apr 24;16(4):365-71.

55. The PP1 binding code: a molecular-lego strategy that governs specificity.Heroes E, Lesage B, Görnemann J, Beullens M, Van Meervelt L, Bollen M.FEBS J. 2013 Jan;280(2):584-95.

56. Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase.Liu D, Vleugel M, Backer CB, Hori T, Fukagawa T, Cheeseman IM, Lampson MA.J Cell Biol. 2010 Mar 22;188(6):809-20.

57. Brown AM. Brain glycogen re-awakened. J Neurochem. 2004 May;89(3):537-52. doi: 10.1111/j.1471- 4159.2004.02421.x. PMID: 15086511.

58. Pellegri G., Rossier C., Magistretti P. J. and Martin J. L. (1996) Cloning, localization and induction of mouse brain glycogen synthase. Brain Res. Mol. Brain Res.38, 191– 199.

59. Glial Glycogen Metabolism☆ I Allaman, E´cole Polytechnique Fe´de´rale de Lausanne, Lausanne, Switzerland PJ Magistretti, KAUST, Thuwal, Saudi Arabia; Ecole Polytechnique Fe´de´rale de Lausanne, Lausanne, Switzerland

60. Brown, A. M., Sickmann, H. M., Fosgerau, K., Lund, T. M., Schousboe, A., Waagepetersen, H. S., and Ransom, B. R. (2005). Astrocyte glycogen metabolism is required for neural activity during aglycemia or intense stimulation in mouse white matter. J. Neurosci. Res. 79, 74–80.

61. Roach, P. J. (2002). Glycogen and its metabolism. Curr. Mol. Med. 2, 101–120.

62. Bollen M, Peti W, Ragusa MJ, Beullens M. The extended PP1 toolkit: designed to create specificity. Trends Biochem Sci. 2010; 35:450–8. [PubMed: 20399103]

63. Structural Basis for Protein Phosphatase 1 Regulation and Specificity Wolfgang Peti1,2,* , Angus C. Nairn3, and Rebecca Page4 1Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI 02912, USA 2Department of Chemistry, Brown University, Providence, RI 02912, USA 3Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06508, USA 4Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA

64. A novel glycogen-targeting subunit of protein phosphatase 1 that is regulated by insulin and shows differential tissue distribution in humans and rodentsShonagh Munro, Hugo Ceulemans, Mathieu Bollen, Julie Diplexcito, Patricia T.W. CohenFirst published: 03 March 2005 https://doi.org/10.1111/j.1742-4658.2005.04585.x

65. Doherty MJ, Young PR & Cohen PTW (1996) Amino acid sequence of a novel protein phosphatase 1 binding protein (R5) which is related to the liver and muscle specific glycogen binding subunits of protein phosphatase 1. FEBS Lett399, 339– 343.

66. Armstrong CG, Browne GJ, Cohen P & Cohen PTW (1997) PPP1R6, a novel member of the family of glycogen-targeting subunits of protein phosphatase 1. FEBS Lett418, 210– 214.DOI: 10.1016/S0014-5793(97)01385-9

67. Identification of Binding Sites on Protein Targeting to Glycogen for Enzymes of Glycogen Metabolism .FongTimothy C.Jensen Ami S.Shah§Nita N.Parekh§Alan R.Saltiel‡§Matthew J.Brady Received 23 June 2000, Revised 3 August 2000, Available online 4 January 2021

68. Protein phosphatase 1--targeted in many directions Patricia T W Cohen

69. Brain glycogen structure and its associated proteins: past, present and future M. Kathryn Brewer, Matthew S. Gentry Department of Molecular and Cellular Biochemistry, Epilepsy and Brain Metabolism Center, Lafora Epilepsy Cure Initiative, and Center for Structural Biology, University of Kentucky College of Medicine, Lexington, KY, 40503 USA

70. Duran J, Gruart A, López-Ramos JC, Delgado-García JM, Guinovart JJ. Glycogen in Astrocytes and Neurons: Physiological and Pathological Aspects. Adv Neurobiol. 2019;23:311-329. doi: 10.1007/978-3-030-27480-1_10. PMID: 31667813; PMCID: PMC7315007.

71. Saez I, Duran J, Sinadinos C, Beltran A, Yanes O, Tevy MF, Martínez-Pons C, Milán M, Guinovart JJ. Neurons have an active glycogen metabolism that contributes to tolerance to hypoxia. J Cereb Blood Flow Metab. 2014 Jun;34(6):945-55. doi: 10.1038/jcbfm.2014.33. Epub 2014 Feb 26. PMID: 24569689; PMCID: PMC4050236.

72. Vilchez D, Ros S, Cifuentes D, Pujadas L, Valles J, Garcia-Fojeda B et al. Mechanism suppressing glycogen synthesis in neurons and its demise in progressive myo clonus epilepsy. Nat Neurosci 2007; 10: 1407–1413

73. The Role of Brain Glycogen in Supporting Physiological Function Laura R. Rich,1William Harris,1 and Angus M. Brown1,2,*DOI=10.3389/fnins.2019.01176 ISSN=1662-453X

74. Pfeiffer-Guglielmi B, Fleckenstein B, Jung G, Hamprecht B. Immunocytochemical localization of glycogen phosphorylase isozymes in rat nervous tissues by using isozyme-specific antibodies. J Neurochem 2003; 85: 73–81.

75. Harris, J.J., Attwell, D. (2012). The Energetics of CNS White Matter. J. Neurosci. 32, 356–371.

76. Reevaluating Metabolism in Alzheimer's Disease from the Perspective of the Astrocyte-Neuron Lactate Shuttle Model April 2013Journal of Neurodegenerative Diseases 2013(5) DOI:10.1155/2013/234572

77. Duran J, Gruart A et al. (2019) Lack of Neuronal Glycogen Impairs Memory Formation and Learning Dependent Synaptic Plasticity in Mice. Frontiers in Cellular Neuroscience

78. Bak, L. K., Walls, A. B., Schousboe, A., and Waagepetersen, H. S. (2018). Astrocytic glycogen metabolism in the healthy and diseased brain. J. Biol. Chem. 293, 7108–7116. doi: 10.1074/jbc.R117.803239

79. PTG Depletion Removes Lafora Bodies and Rescues the Fatal Epilepsy of Lafora Disease Julie Turnbull,1 ,2Anna A. DePaoli-Roach,3 ,*Xiaochu Zhao,1Miguel A. Cortez,4Nela Pencea,1 ,5Erica Tiberia,1Mark Piliguian,1Peter J. Roach,3Peixiang Wang,1Cameron A. Ackerley,5 and Berge A. Minassian

80. LAFORIN: A DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES COMPLEX CARBOHYDRATES Carolyn A. Worby, Matthew S. Gentry, and Jack E. Dixon Departments of Pharmacology, Cellular and Molecular Medicine and Chemistry and Biochemistry

81. The Laforin-Malin Complex Negatively Regulates Glycogen Synthesis by Modulating Cellular Glucose Uptake via Glucose Transporters Pankaj Kumar Singh, Sweta Singh, and Subramaniam Ganesh

82. Malin Decreases Glycogen Accumulation by Promoting the Degradation of Protein Targeting to Glycogen (PTG)* Received for publication,October 22, 2007, and in revised form, December 3, 2007 Published, JBC Papers in Press, December 10, 2007, DOI 10.1074/jbc.M708712200

83. Lafora disease: insights into neurodegeneration from plant metabolism Matthew S. Gentry, Jack E. Dixon, Carolyn A. Worby Trends in Biochemical Sciences Volume 34 Issue 12 Pages 628-639 (December 2009) DOI: 10.1016/j.tibs.2009.08.002

84. Sun et al., Brain glycogen serves as a critical glucosamine cache required for protein glycosylation, Cell Metabolism (2021), https://doi.org/10.1016/j.cmet.2021.05.003

85. Pathogenesis of Lafora Disease: Transition of Soluble Glycogen to Insoluble Polyglucosan August 2017 International Journal of Molecular Sciences 18(8):1743 DOI:10.3390/ijms18081743

86. Tan, J.M.M.; Wong, E.S.P.; Kirkpatrick, D.S.; Pletnikova, O.; Ko, H.S.; Tay, S.P.; Ho, M.W.L.; Troncoso, J.;Gygi, S.P.; Lee, M.K.; et al. Lysine 63-linked ubiquitination promotes the formation and autophagic clearanceof protein inclusions associated with neurodegenerative diseases. Hum. Mol. Genet.2008,17, 431–439

87. Worby, C.A.; Gentry, M.S.; Dixon, J.E. Malin decreases glycogen accumulation by promoting the degradationof protein targeting to glycogen (PTG). J. Biol. Chem. 2008,283, 4069–4076

88. Sharma, J.; Mulherkar, S.; Mukherjee, D.; Jana, N.R. Malin regulates wnt signaling pathway throughdegradation of dishevelled2. J. Biol. Chem. 2012,287, 6830–6839.

89. Vilchez, D.; Ros, S.; Cifuentes, D.; Pujadas, L.; Valles, J.; Garcia-Fojeda, B.; Criado-Garcia, O.;Fernandez-Sanchez, E.; Medrano-Fernandez, I.; Dominguez, J.; et al. Mechanism suppressing glycogensynthesis in neurons and its demise in progressive myoclonus epilepsy. Nat. Neurosci.2007,10, 1407–1413

90. Gentry, M.S.; Worby, C.A.; Dixon, J.E. Insights into lafora disease: Malin is an E3 ubiquitin ligase thatubiquitinates and promotes the degradation of laforin. Proc. Natl. Acad. Sci. USA2005,102, 8501–8506

91. Cheng, A.; Zhang, M.; Gentry, M.S.; Worby, C.A.; Dixon, J.E.; Saltiel, A.R. A role for AGL ubiquitinationin the glycogen storage disorders of lafora and Cori's disease. Genes Dev.2007,21, 2399–2409

92. Lohi, H.; Ianzano, L.; Zhao, X.C.; Chan, E.M.; Turnbull, J.; Scherer, S.W.; Ackerley, C.A.; Minassian, B.A.Novel glycogen synthase kinase 3 and ubiquitination pathways in progressive myoclonus epilepsy.Hum. Mol. Genet. 2005,14, 2727–2736

93. Fernandez-Sanchez, M.E.; Criado-Garcia, O.; Heath, K.E.; Garcia-Fojeda, B.; Medrano-Fernandez, I.;Gomez-Garre, P.; Sanz, P.; Serratosa, J.M.; de Cordoba, S.R. Laforin, the dual-phosphatase responsiblefor lafora disease, interacts with R5 (PTG), a regulatory subunit of protein phosphatase-1 that enhancesglycogen accumulation. Hum. Mol. Genet. 2003,12, 3161–3171

94. Worby, C.A.; Gentry, M.S.; Dixon, J.E. Laforin, a dual specificity phosphatase that dephosphorylates complexcarbohydrates. J. Biol. Chem. 2006,281, 30412–30418

95. Deciphering the role of malin in the lafora progressive myoclonus epilepsy Carlos Romá-Mateo,Pascual Sanz,Matthew S. Gentry First published: 20 July 2012 https://doi.org/10.1002/iub.1072

96. Lafora disease: insights into neurodegeneration from plant metabolism Matthew S. Gentry, Jack E. Dixon, Carolyn A. Worby Trends in Biochemical Sciences Volume 34 Issue 12 Pages 628-639 (December 2009) DOI: 10.1016/j.tibs.2009.08.002

97. Chan EM, Young EJ, Ianzano L, Munteanu I, Zhao X, Christopoulos CC, Avanzini G, Elia M, Ackerley CA, Jovic NJ, Bohlega S, Andermann E, Rouleau GA, Delgado-Escueta AV, Minassian BA, Scherer SW. Mutations in NHLRC1 cause progressive myoclonus epilepsy. Nat Genet. 2003 Oct;35(2):125-7. doi: 10.1038/ng1238. Epub 2003 Sep 7. PMID: 12958597.

98. Ianzano L, Zhang J, Chan EM, Zhao XC, Lohi H, Scherer SW, Minassian BA. Lafora progressive Myoclonus Epilepsy mutation database-EPM2A and NHLRC1 (EPM2B) genes. Hum Mutat. 2005 Oct;26(4):397. doi: 10.1002/humu.9376. PMID: 16134145.

99. Lafora disease in miniature Wirehaired Dachshunds Lindsay Swain,Gill Key,Anna Tauro,Saija Ahonen,Peixiang Wang,Cameron Ackerley,Berge A. Minassian,Clare Rusbridge Published: August 2, 2017 https://doi.org/10.1371/journal.pone.0182024

100. Turnbull, J., Tiberia, E., Striano, P., Genton, P., Carpenter, S., Ackerley, C.A. and Minassian, B.A. (2016), Lafora disease. Epileptic Disorders, 18: S38-S62. https://doi.org/10.1684/epd.2016.0842

101. Tagliabracci VS, et al. 2011. Phosphate incorporation during glycogen synthesis and Lafora disease. Cell. Metab. 13:274–282 [PMC free article] [PubMed]

102. Gayarre, J.; Duran-Trio, L.; Garcia, O.C.; Aguado, C.; Juana-Lopez, L.; Crespo, I.; Knecht, E.; Bovolenta, P.; de Cordoba, S.R. The phosphatase activity of laforin is dispensable to rescue Epm2a−/− mice from lafora disease. Brain 2014, 137, 806–818. [CrossRef] [PubMed]

103. Solaz-Fuster M.D. et al.Regulation of glycogen synthesis by the laforin-malin complex is modulated the AMP-activated protein kinase pathway.Hum. Mol. Genet.2008; 17: 667-678

104. Vernia S. et al. AMP-activated protein kinase phosphorylates R5/PTG, the glycogen targeting subunt of the R5/PTG-PP1 holoenzyme and accelerates its downregulation by the laforin-malin complex. J. Biol. Chem.2009; 284: 8247-8255

105. Yamanami S et al. Comparative study of intraneuronal polyglucosan bodies in brains from patients with Lafora disease and aged dogs. Acta Pathol. Jpn.1992; 42: 787-792

106. Mittal S et al. Lafora disease proteins malin and laforin are recruited to aggresomes in response to proteasomal impairment Hum. Mol.Genet.2007; 16: 753-762

107. Garyali P. et al. The malin-laforin complex suppresses the cellular toxicity of misfolded proteins by promoting their degradation through the ubiquitin-proteasome system. Hum. Mol. Genet.2009; 18: 688-700

108. Ganesh S, Puri R, Singh S, Mittal S, Dubey D. Recent advances in the molecular basis of Lafora's progressive myoclonus epilepsy. J Hum Genet. 2006;51(1):1-8. doi: 10.1007/s10038-005-0321-1. Epub 2005 Nov 26. PMID: 16311711.

109. Autophagy: cellular and molecular mechanisms, Danielle Glick,1,2 Sandra Barth,1 and Kay F. Macleod, J Pathol. 2010 May; 221(1): 3–12. doi: 10.1002/path.2697

110. Carmen Aguado, Sovan Sarkar, Viktor I. Korolchuk, Olga Criado, Santiago Vernia, Patricia Boya, Pascual Sanz, Santiago Rodríguez de Córdoba, Erwin Knecht, David C. Rubinsztein, Laforin, the most common protein mutated in Lafora disease, regulates autophagy, Human Molecular Genetics, Volume 19, Issue 14, 15 July 2010, Pages 2867–2876, https://doi.org/10.1093/hmg/ddq190

111. PTG, a Protein Phosphatase 1-Binding Protein with a Role in Glycogen Metabolism John A. PrintenMatthew J. Bradyand Alan R. SaltielAuthors Info & Affiliations Science • 7 Mar 1997 • Vol 275, Issue 5305 • pp. 1475-1478 •DOI: 10.1126/science.275.5305.1475

112. Protein targeting to glycogen is a master regulator of glycogen synthesis in astrocytes Author links open overlay panelE.Ruchtiab P.J.RoachcA.A.DePaoli-RoachcP.J.MagistrettidabI.Allamana

113. Guerrero Lopez, Rosa & Vernia, Santiago & Sanz, Raul & Abreu-Rodríguez, Irene & Almaraz, Carmen & Garcia-Hoyos, Maria & Michelucci, Roberto & Tassinari, Carlo & Riguzzi, Patrizia & Nobile, Carlo & Sanz, Pascual & Serratosa, Jose & Gómez-Garre, Pilar. (2011). A PTG Variant Contributes to a Milder Phenotype in Lafora Disease. PloS one. 6. e21294. 10.1371/journal.pone.0021294.

114. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501. doi:10.1107/S0907444910007493

115.   MD Winn — CCP4: Collaborative Computational Project Number 4 (1994). "The CCP4 suite: programs for protein crystallography", Acta Cryst. D50, 760-763.

116.   Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., Jeffries, C.M., Svergun, D.I., Franke, D. (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysisJ. Appl. Cryst. 54, 343-355 DOI

117.   Daniel Patko, Kaspar Cottier, Andras Hamori, and Robert Horvath, "Single beam grating coupled interferometry: high resolution miniaturized label-free sensor for plate based parallel screening," Opt. Express 20, 23162-23173 (2012)