Université d'Orléans – Université de Zagreb

UFR Sciences et techniques – Faculté de Nutrition et biotechnologie – Faculté des Sciences

Master Sciences du Vivant

Spécialité : Biotechnologies, Biologie Moléculaire et Cellulaire

INTERNSHIP REPORT

New insights into the processivity and effect of HomoGalacturonan Modifying enzymes (HGME)

by

Filip Košec

(February, 2022 – June, 2022)

Internship institution and address: UMR INRAE 1158 BioEcoAgro, Biologie des Plantes et Innovation, Université de Picardie Jules Verne, 33 rue Saint Leu 80039 Amiens Cedex 1

Mentor: Jérôme Pelloux, professor

Supervisor : Valérie Lefebvre, assistant professor

Acknowledgements

Firstly, I would like to sincerely thank my mentors - professors *Jérôme Pelloux* and *Valérie Lefevbre* for giving me the opportunity to do my internship in BioEcoAgro on this very interesting subject. Thank you for the help with settling in Amiens, for all the good advice, for your support and most of all for the positive attitude and environment I felt over these last few months.

A huge thank you goes out to *Josip Šafran* for showing me the ropes of the lab work, for his continued help during the internship, for all the lunchtime talks and for introducing me to all the people in the lab.

I am grateful to *Adrien Lemaire* for always being there to help no matter how busy he was, to the people of the "Best office" – *Pauline Trezel, Camille Carton* and *Wafae Tabi* for the great working and social atmosphere and to everyone else in the lab.

I would also like to thank *Nely Rodriguez Moraga* and *Francisco Bernardo Ramos Martin* for being great friends, for all the invitations and for filling me in on all the details of the lab.

It is my privilege to say thank you to professors Višnja Besendorfer, Vladimir Mrša and Chantal Pichon for giving me the opportunity to take part in this unique and amazing master's program.

An especially warm thank you goes to my "Grupica" and all the other friends who have been with me these past few years and beyond.

Lastly, my most sincere gratitude goes to my family – my parents *Vesna* and *Berislav*, my aunt *Katica* and uncle *Nikica*, my cousin *Dubravka* and my little nephew *Erol*. Thank you for the support and belief you have always given me.

Volim vas!

Presentation of the laboratory

The BioEcoAgro Joint Cross-Border Research Unit is a new structure which brings together around 400 researchers, technicians, and PhD students from both sides of the Franco-Belgian border. Its ambition is to develop an international center of excellence in the field of biological engineering applied to agriculture, biotechnology, agri-food and the environment. Designed like a laboratory without walls bringing together researchers from INRAE, the University of Liège, the University of Lille and the University of Picardie Jules Verne, it focuses on the combination of (eco) systemic approaches and molecular for:

- understanding the functioning of plants and ecosystems in natural or controlled environments and in the context of climate change
- decryption and control of the synthesis or the bioproduction of active biomolecules (specialized metabolites and polymers of plant origin, enzymes and secondary metabolites of microbial origin, active peptides from the hydrolysis of food proteins)
- biopreservation and food formulation.

The complementary multidisciplinary skills of the BioEcoAgro Transboundary UMR also allow it to develop cross-disciplinary research such as the development of new biocontrol agents for plant diseases: from high throughput screening to their evaluation in the field through the understanding of their mode of action and the development of bioprocesses to obtain them.

The UMR is closely involved in the development of platforms of excellence such as the Realcat equipex, the FoodIsLife, AgricultureIsLife and EnvironmentIsLife (Ecotron) platforms and the national long-term observation and experimentation system on the ACBB environment (Agroecosystems, Biogeochemical Cycles and Biodiversity) in its field crops component. It manages two measurement stations labeled with the European infrastructure ICOS (Integrated Carbon Observatory System).

Invested in upstream research on an international scale, the cross-border BioEcoAgro UMR is also distinguished by its numerous collaborations with companies and actors from the agricultural world of the cross-border region.

Contents

1. - INTRODUCTION

1.1. - Cell wall structure

The plant cell wall is a structure particular to plants which consists of a network of polysaccharide polymers such as cellulose, hemicelluloses and pectins along with various glycoproteins and organic polymers such as lignin (Holland et al., 2020) (Figure 1.).

The plant cell wall can be divided into three distinct layers: the middle lamella, the primary cell wall and the secondary cell wall (Figure 1.). The middle lamella is the layer which forms the interface between two adjacent cells and glues them together to maintain tissue and organ integrity. It is largely composed of pectins and a small amount of proteins in young and growing plants, as well as calcium and magnesium ions (Behboudian et al., 2017; Zamil and Geitmann, 2017). The secondary cell wall is a type of cell wall which usually provides structural support in lignified cells but can also functions as the site of energy storage. Secondary cell walls are mainly composed of cellulose, xylans, glucomannans and lignin and are often found in specialized cells, especially the ones that form the tracheary system (Buchanan et al., 2000; Zhong and Ye, 2015). The primary cell wall is an extensible layer formed during cell growth with the ability to increase its surface as much as 1000-fold. It is composed of cellulose, hemicellulose and pectic polysaccharides (Scheller and Ulvskov, 2010). There are two types of primary cell walls in flowering plants. Type I primary walls are most common in plant species and are composed of cellulose microfibrils interconnected with mostly xyloglucans which can make up to 50% of its mass as well as pectins. Type II primary walls are found only in *Poaceae*. They are interconnected with arabinoxylans instead of xyloglucans and contain very small amounts of pectin (Carpita and Gibeaut, 1993). Despite this already high complexity and variability between the layers of the cell wall, its composition has further evolved and diversified between various plant species and plant tissues due to the need for plant adaptability (Burton et al., 2010; Somerville et al., 2004).

Cellulose, consisting of unbranched β-(1-4)-linked glucan chains, is the primary constituent of plant cell walls making up to 50% of their dry mass (Zeng et al., 2017) and is the primary structural polymer of cell walls with great importance in response to stressful conditions (Zhang et al., 2021). Hemicellulosic polysaccharides include xylans, xyloglucans, mannans, glucomannans and β-(1,3;1,4)-glucans which interconnect the cellulose fibers and provide the cell wall with flexibility, among other functions (Scheller and Ulvskov, 2010). Xylans, made up of β-(1,4)-linked xylose residues with side branches of α-arabinofuranose and/or α-glucuronic acids, are the most abundant of these polymers, and are the primary component responsible for interconnecting cellulose fibers (Scheller and Ulvskov, 2010). Xyloglucans are present in the primary cell walls of all plants except grasses and are considered to be involved in the loosening and tightening of cellulose microfibrils. (Hayashi and Kaida, 2011). Mannans serve both for cellulose binding and as a reserve of carbohydrates in the endosperm and vacuoles (Moreira and Filho, 2008). β-(1,3;1,4)-glucans are unbranched β-(1,4)-linked glucans with occasional β-(1,3)-glucosyl linkages. Their function has not yet been discovered though there are indications that they may function as an attachment surface for other matrix polysaccharides

during rapid cell expansion (Kiemle et al., 2014). Of particular note are hemicellulosic and pectic polymers, which will be covered in more detail later. These polymers have various specific acetyesterification and methylesterification patterns that differ depending on tissues and developmental stages, which points to their respective roles in plant development (Du et al., 2020; Zhong et al., 2018). Lastly, lignin is an amorphic phenolic polymer characteristic of vascular plants which often functions as a barrier against pathogen attack and serves to further reinforce secondary cell wall structure (Vanholme et al., 2019).

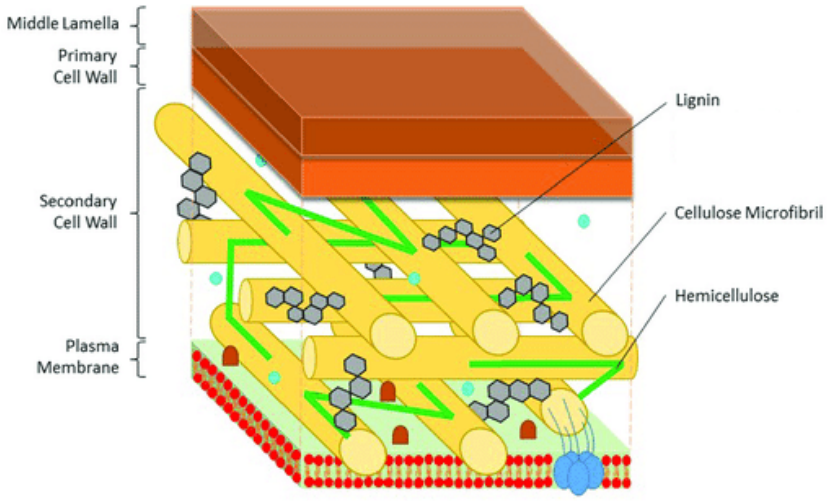

Figure 1. Simplified representation of cell wall structure (Loix et al., 2017).

1.2. - Pectin structure

Pectin is a macromolecular polysaccharide that exists in cell walls of various fruits, vegetables, and other parts of the plant. It can be primarily found in the cell walls of eudicotyledons and monocotyledons except for *Poaceae* which contain low amounts of it (Carpita and Gibeaut, 1993; Vasco-Correa and Zapata Zapata, 2017). Pectin is a heterogenous molecule organized into distinct domains whose abundance and characteristics vary depending on species, tissue, developmental stage, and even location within the cell wall (Voragen et al., 2009). Pectins can be distinguished from hemicellulosic polysaccharides by the fact that the former are easily extracted using hot acid while the latter require alkaline extraction (Scheller and Ulvskov, 2010). Pectins (except Rhamnogalacturonan I) are made from α-(1,4)-linked galacturonic acid (GalA) chains that function as a backbone and are then classified according to the type and presence of branching sugars (Figure 2.).

Homogalacturonan (“smooth” pectin) is a linear polymer of α-(1–4)-linked d-GalA with no substituted side chains. (Sénéchal et al., 2014). This polymer accounts for approximately 60% of the total pectin in most plants, with a few exceptions such as sugar beet and soybean (Ropartz and Ralet, 2020). Understanding the synthesis, transport and dynamics of HG is of utmost importance for unravelling the regulation of plant development. The GalA residues in HG are up to 80% methylesterified at the C6 carboxyl group, and up to 10% acetylesterified at positions O-2 and O-3 (Atmodjo et al., 2013; Voragen et al., 1995). Demethylesterified HGs can form gel like structures with the help of Ca^{2+} ions creating a so called “egg-box” structure (Cabrera et al., 2008). Recently, using super resolution microscopy, homogalacturonan has been shown to organize into

filaments oriented perpendicularly to the surface of cotyledons. These filaments have a different structure depending on their methylesterification state, where methylesterifieded HGs pack more tightly compared to demethylesterified HGs (Figure 3.). The demethylesterification of these filaments then leads to swelling of radial nanofilaments, cell expansion, and wall thickening (Haas et al., 2020). HG-containing structures recovered from cell walls have shown HG connected to RG-I, RG-II and potentially the APAP1 proteoglycan. In vitro, HG fragments of length of up to 500 have been synthesized though such fragments haven't been shown to exist in vivo (Engle et al., 2022; Tan et al., 2013).

Apiogalacturonans are pectins where a β-D-apiofuranosyl monomer or dimer is attached to the O-2 or O-3 of GalA. So far they have only been found in aquatic plants (Ropartz and Ralet, 2020). Xylogalacturonans are pectins where HG is substituted with a xylose monomer or dimer at the O-3 position. They have been identified in various plants such as duckweeds, watermelons, and *A. thaliana*. Methylesterification of xylogalacturonans has been reported in apples (Schols et al., 1995).

A further major component of pectins is Rhamnogalacturonan II (RG-II). It is much more complex than the previous three types of pectins owing to the fact that its side chains are composed of thirteen different sugars and more than twenty different types of glycosidic linkages (Ishii and Kaneko, 1998). There are five types of RG-II side chains. Side-chains A and B are large and branched side chains containing many different residues, while side-chains C, D and E are simple and composed of up to two residues (Ropartz and Ralet, 2020). RG-II can also dimerize via boron cross-links which contribute to the assembly and biophysical properties of the cell wall (Chormova and Fry, 2016).

Rhamnogalacturonan I (RG-I) sometimes also called the "hairy" region of pectin is composed of a backbone containing alternating GalA and Rhamnose residues. Rhamnose is connected to GalA by an α(1→4) linkage and the GalA is then bound to rhamnose by an α(1→2) linkage (Yapo, 2011). The GalA residues are very rarely substituted while the Rhamnose residues can present various modifications. According to the substitutions they are divided into arabinans - substituted by arabinofuranose chains of various length, type I Arabinogalactans substituted by chains containing arabinose and galactose, and type II arabinogalactans consisting of highly branched chains containing a backbone of β-(1,3)-linked D-Galp residues. Lastly, RG-I can rarely be substituted by ester-linked ferulic acid residues in certain species from the *Amaranthaceae* family (Ropartz and Ralet, 2020).

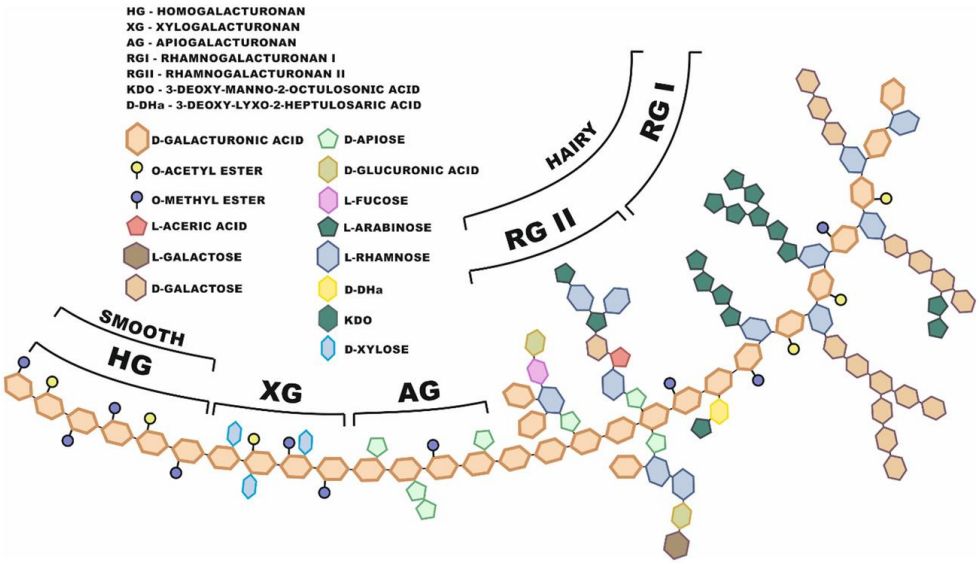

Figure 2. Schematic representation of pectin structure (Zdunek et al., 2021).

1.3. - Homogalacturonan synthesis and homogalacturonan modifying enzymes

HG are synthesized by members of the GALACTURONOSYLTRANSFERASE (GAUT) family. It has been shown that the synthesis of HG by GAUT is dependent on monomers of UDP-GalA, which are synthesized in the cytosol, and transported into the Golgi apparatus where the GAUT1:GAUT7 complex is localized (Reyes and Orellana, 2008). GAUT1 and GAUT7 form a complex which initiates de novo HG synthesis, firstly in a slow phase which changes to a fast synthesis phase when the chain reaches the length of 11 monomers (Zhang et al., 2021). Apart from de novo synthesis which serves to create new HG molecules, many GAUTs are involved in HG elongation because of their ability to further polymerize already existing HG acceptor chains (Engle et al., 2022). HGs are synthesized in a highly methylesterified form and are later demethylesterified by various pectin methylesterases (PME) after their incorporation into the cell wall (Zhang et al., 2021).

As the above-mentioned changes in HGs structure cause modifications in plant development, the understanding of Homogalacturonan modifying enzymes (HGME) is crucial to our comprehension of plant cell wall dynamics. HGMEs include enzymes that modify the structure of HG like pectin methylesterases and pectin acetylesterases (PAE) which remove methyl and acetyl group from HG, respectively. The latter two enzymes are further regulated by Pectin methyl esterase inhibitors (PMEI) and pectin acetyl esterase inhibitors (PAEI). Furthermore, HGMEs also include hydrolysing enzymes like polygalacturonases (PG) which degrade HG using a hydrolytic cleavage mechanism and Pectate lyases (PLLs) which cleave HG through β-elimination mechanism (Jia et al., 2009).

1.4. - Polygalacturonases

Polygalacturonases are a family of enzymes found primarily in plants, bacteria and fungi, with a few known examples among insects (Girard and Jouanin, 1999; Shen et al., 2003; Strong and Kruitwagen, 1968; Yang et al., 2018). They are members of the glycoside hydrolase 28 (GH28) family whose members contains at least one GH28 (Pfam00295) domain (Markovič and Janeček, 2001).

PGs are divided into two distinct families: endo-polygalacturonases (endo-PG, EC 3.2.1.15) which cleave HG chains in the middle of the molecule, and exo-polygalacturonases (exo-PG, EC 3.2.1.82) which hydrolyze one or two GalA residues from the non-reducing end of HG (Sénéchal et al., 2014). It is known that these enzymes prefer unmethylesterified substrates (Bonnin, 2001). Endo-PGs in particular have an affinity for cleaving in HG regions with more than four unmethylesterified GalA residues (Sénéchal et al., 2014). A subset of Exo-PGs called exo-polygalacturonosidases are known to cleave two residues from the non-reducing end of HG (Markovič and Janeček, 2001). PGs from many fruits such as tomatoes (Sitrit et al., 1999), peaches (Pressey and Avants, 1973) and strawberries (Nogota et al., 1993) are known to be calcium-dependent which could be linked to the previously mentioned role of calcium ions in the formation of pectin gels and could play a role in seed germination and fruit abscission (Sénéchal et al., 2014). Plants can also regulate activity of pathogenic PGs from fungi, bacteria, and insects, which degrade the adhesive pectins of the middle lamella and contribute to pathogen penetration, through specific interactions with PGIPs (Nakamura and Iwai, 2019). So far no plant PGIPs have been shown to inhibit plant PGs. PGs are also required for plant cell separation as they hydrolyze the HG backbone of the middle lamella after its demethylesterification, leading to reduction of the structural rigidity normally mediated by Ca^{2+} thus influencing plant development (Daher and Braybrook, 2015). They are also known to play a role in the separation of *A. thaliana* pollen tetrads. (Francis et al., 2006).

1.5. - Roles of pectins in plant development

The plant cell wall plays many roles in the development of plants as it provides the plants' bodies with mechanical support, is involved in plant morphogenesis, transfer of water and nutrients, defense against biotic and abiotic stresses and defense responses against pathogens and predators (Bacic et al., 1988; Carpita and Gibeaut, 1993; Zhang et al., 2021). Plants possess more than 40 different cell types that span various developmental phases, each of which has different cell wall compositions as well as forms (Burton et al., 2010). There has been mounting evidence that the plant cell wall can control many mechanical processes such as cell extension, flexion, loosening and strengthening. Because of that, understanding cell wall dynamics is hugely important for our understanding of plant development.

Cell wall growth and extension is currently thought to be primarily caused by changes in turgor pressure which cause directed strain on cell walls that then expand in an asymmetrical manner (Anisotropically) (Haas et al., 2020). According to current models, this anisotropic expansion is caused by orientation of cellulose microfibrils which restricts cell growth according to their orientation (Baskin, 2005). There is however some evidence that points to the involvement of other cell walls components, such as pectins, in this process. *rsw4* (*radially swollen 4*) and *rsw7 A. thaliana* mutants for example show changes in growth despite there being no change in the amount or structure of cellulose microfibrils (Wiedemeier et al., 2002). The most common view of pectin structure is that it is simply a collection of amorphous polymers. Many have however proposed a more crystalline structure for HG. In one such model, Ca^{2+} ions bond to partially demethylesterified HGs and promote the strengthening of the pectin network (Sénéchal et al., 2014; Zhang et al., 2021). Another model called the "expanding beam" model proposes that pectic nanofilaments possess an expansion capacity and

compact themselves differently based on their degree of methylesterification which causes them to strengthen or loosen (Engle et al., 2022). The degree of pectin methylesterification is controlled by actions of Pectin methyl esterase (PME)(EC 3.1.1.11, CAZY Carbohydrate Esterase Family 8) and Pectin methyl esterase inhibitor (PMEI) pairs which work to integrate the signals from various signalling pathways during plant development, such as the leaf dorsoventral polarity signals (Qi et al., 2017). The control of pectins' degree of methylesterification is particularly crucial for cell and organ shape in the meristematic zone (Peaucelle et al., 2011). An example of this can be seen in pollen tubes of *A. thaliana* where highly methylesterified pectins are deposited at the tip causing a high degree of flexibility. These pectins are then demethylesterified by VANGUARD1, a pectin methylesterase which causes rigidity in the subapical region (Chebli et al., 2012; Jiang et al., 2005). Lastly, pectins can indirectly influence cell development through release of oligogalacturonides (OGs) in response to damage associated patterns. The Wall-Associated Kinase1 (WAK1) has been shown to be a receptor for oligogalacturonides, and is able to activate further enzymes in the defense response signaling pathways (Brutus et al., 2010).

The importance of PGs for plant development was also confirmed using *A. thaliana* mutants with PG overexpression. It was found that these mutants regulate hypocotyl elongation, floral organ patterning (PGX1 and PGX2), timing of flowering, stem lignification and stem thickness (PGX2) (Xiao et al., 2017, 2014).

1.6. - Phylogeny of polygalacturonases

According to some estimations, almost 15% of *Arabidopsis thaliana*'s 27000 genes are dedicated to biogenesis and modification of cell walls (Carpita et al., 2001). The genome of *A. thaliana* has experienced at least four genome duplications about 100-200 million years ago, which have caused a great increase in the number and diversity of genes encoding PG (Prince and Pickett, 2002; Vision et al., 2000). Gene duplications serve to provide organisms with a diversification of the functionality of proteins. Furthermore, long-term selective pressure can induce quick evolution of important amino acid residues (Cao, 2012; Prince and Pickett, 2002). The abundance of PG-encoded genes in plants (68 in *A. thaliana*) points to their diversified functions and specificities which are yet to be determined (Sénéchal et al., 2014). PG genes went through various segmental duplications, the earliest of which happened between the AT3G61490 and AT4G23500 gene around 64 million years ago (Cao, 2012). About 60% of PGs possess a signal peptide and about 20% possess a transmembrane domain (Sénéchal et al., 2014). Research so far has shown that PGs appear in fungi, plants and bacteria (Markovič and Janeček, 2001) with three representatives from insects (Girard and Jouanin, 1999; Shen et al., 2003; Strong and Kruitwagen, 1968), probably due to horizontal gene transfer events. When looking at plant PGs, Rhamno-PGs are considered to be the earliest type and are present in both algae and land plants, while Endo-PGs and Exo-PGs are present only in land plants (Park et al., 2010b). There have been various classifications of plant PGs. Hadfield et al. (1998) first grouped PG genes into three clades from clade A to clade C. Further developments of this system added two additional clades (D and E). While clades A, B and C showed several invariant conserved residues, no such residues were found in clades D and E (Markovič and Janeček, 2001). Another system, based on the work of Kim et al. (2006) maintains the original number of

groups. Each of these systems has its advantages with the first system being better for analyzing the emergence of new PG genes, while the second system works better for analysis within each of the three groups. (Yang et al., 2018). To date sixty-eight PG genes have been identified in *A. thaliana*. This rapid expansion is primarily attributed to gene duplications happening in the clade A of the system proposed by Hadfield et al. (1998) (Ke et al., 2018). It is therefore evident that this high plant PG diversity combined with the plant cell wall's crucial role in plant development requires further study of the enzymes and genes in question.

1.7. - Structure of polygalacturonases

Polygalacturonases all contain a right-hanged parallel β-helix structural domain that consists of 7-12 beta strands which form three to four β-sheets (Figure 4.) that was first described in *Erwinia crysanthemi* (Markovič and Janeček, 2001; Yoder et al., 1993). A key difference between the structures of Endo-PGs and Exo-PGs is the arch that encloses the active sites and is made from two loops. In Endo-PGs these loops are far apart while in Exo-PGs they are very close to each other. This causes the active site in Exo-PGs to be partially or fully closed from one side while it is open on both sides in Endo-PGs such as ADPG2 and PGLR (Abbott and Boraston, 2007; Šafran et al., 2022). Four conserved motifs are known in the active site of PGs. Motif I (NTD) and motif II (DD) contain aspartate residues that form the catalytic triad (Lu et al., 2021). Motif III (GHG) contains a histidine residue which is known to be involved in the catalytic reaction, possibly by maintaining the proper ionization state of some other residue (Armand et al., 2000; Narsimha Rao et al., 1996; Palanivelu, 2006). Motif IV (RIK) is related to ion interactions near the carboxyl end of the substrate and may have some influence on the catalytic activity (Lu et al., 2021; Rexová-Benková, 1990). Finally, a tyrosine residue has been identified which is positioned near the entrance to the catalytic site which assists in substrate binding (Narsimha Rao et al., 1996; Pages et al., 2000). To date only three crystal structures of Exo-PG have been published: one from *Yersinia pestis,* one from the hyperthermophilic *Thermotoga maritima,* and one from *Bacteroides tehtaiotaomicron* (Abbott and Boraston, 2007; Luis et al., 2018; Pijning et al., 2009). The Exo-PG from *Yersinia pestis* has the same right-handed parallel β-helix structure and conserved domains, except for domain I which is NGD in this case. Furthermore, this structure showed the role of two additional residues in the protein-substrate interaction: an arginine residue (R240) and an asparagine residue (N406) (Abbott and Boraston, 2007). The substrate binding sites of PGs are usually made up of predominantly aromatic residues and are surrounded by negative electrostatic potential (Cao, 2012). Apart from conserved residues involved in catalysis, PGs also have a number of conserved cysteine residues which are part of disulfide bridges that stabilize the protein (Markovič and Janeček, 2001). Lastly, changes in pH are known to cause conformational changes in the binding cleft of PGs (Mayans et al., 1997) which could be connected to the various pI of Arabidopsis' PGs that range from 4.8 to 9.8 (Sénéchal et al., 2014).

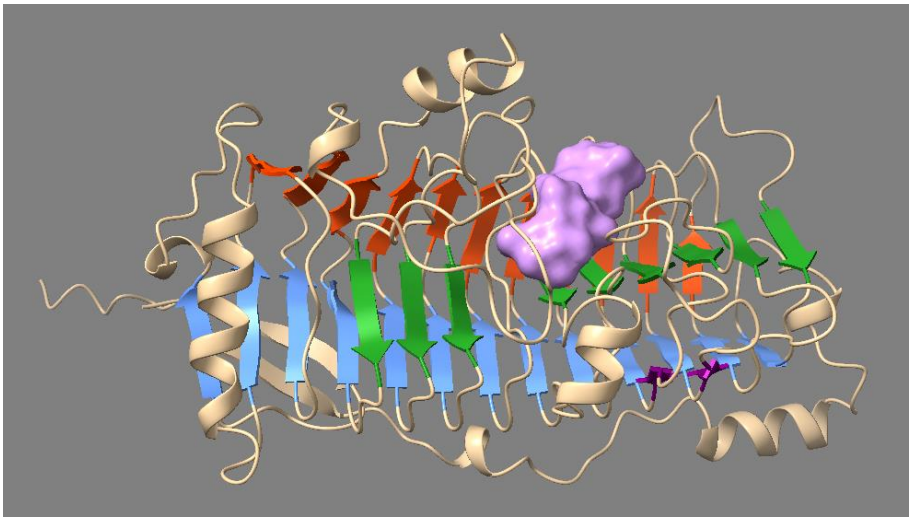

Figure 4. Right-hanged parallel β-helix structural domain in AtPGF9. β-sheets making up the structure colored in blue, green, orange, and purple. Surface of tetragalacturonic acid substrate docked in the active site is shown in light purple.

1.8. - Objectives

PG-produced oligosaccharides have shown potential as prebiotic, antibacterial, anticancer and antioxidant agents in the food industry (Zhu et al., 2019) while several PGs have been shown to cause allergies (Chardin et al., 2003; Ibarrola et al., 2004; Sarkar et al., 2018). It is therefore apparent that understanding polygalacturonases isn't important only in the domain of plant biology, but that it also has many everyday implications.

The genome of *A. thaliana* contains 68 different polygalacturonase-coding genes (Ke et al., 2018). This large diversity of enzymes may point to subtle differences in their biochemical properties such as their mode of action and substrate specificity. Understanding these specificities will allow us to rationalize the abundance of polygalacturonases and what role it plays in plant development. Despite that, the mode of action of most of these enzymes has not yet been characterized. While exploring the roles of PGs in Arabidopsis is attractive because it is a model organism in plant biology, the overwhelming amount of protein characterization that needs to be done may prove to be inefficient for deepening our understanding of the role PGs play in plant development, due to gene compensation. To circumvent this, it may be better to characterize PGs of an organism such as *Marchantia polymorpha* which has only 10 putative PG genes and use the insight gained from those characterizations to expand our understanding of the PGs in *A. thaliana* and other plants.

The goal of this internship is to produce and study four previously uncharacterized putative Exo-polygalacturonases: AT3G61490 (AtPGF9), AT4G23500 (AtPGF12) and AT3G48950 (AtPGF7) from *Arabidopsis thaliana* and Mapoly0009s0072 (MpPG2) from *Marchantia polymorpha*. Previous phylogenetic studies (Figure 5.) have shown that the three *A. thaliana* proteins are very closely related and that MpPG2 is the only PG from *M. polymorpha* which belongs to the same clade as the three Arabidopsis PG. The main objective will be to confirm the exo-acting activity of the four enzymes and further determine how small structural differences between these four closely related enzymes may be of importance to fine-tuning of their activities.

After their expression and purification, the enzymes will be characterized regarding their biochemical properties such as their optimal pH, temperature, melting temperature effects of substrate inhibition and calcium dependence as well as profiling of catalysis products on various substrates. Furthermore, the comparisons of their structural models will allow to relate their structures to the above-mentioned parameters.

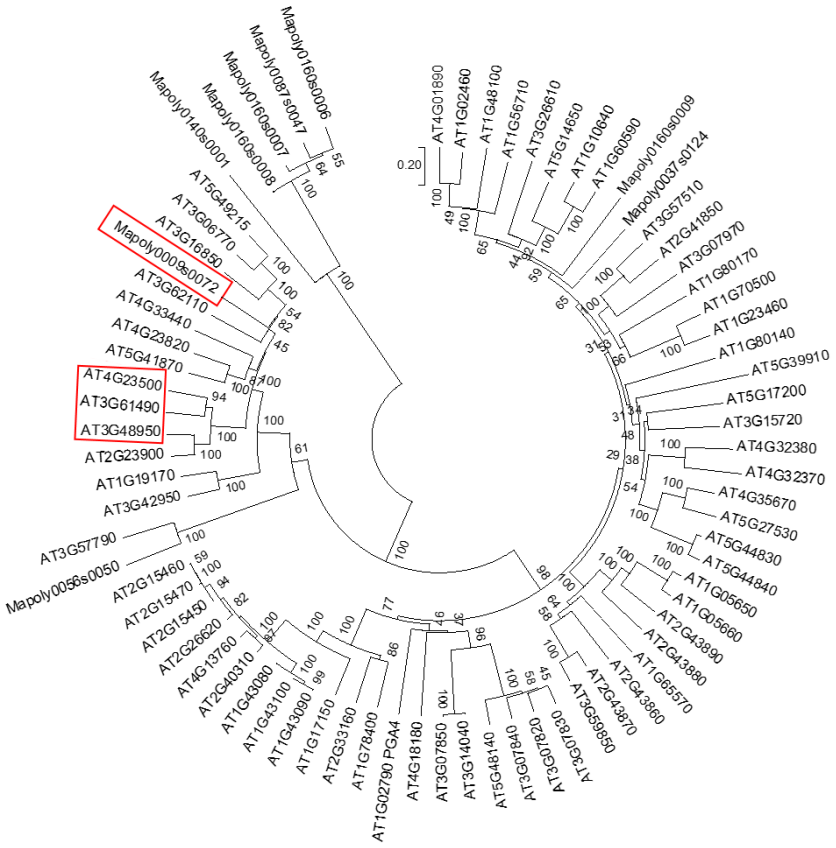

Figure 5. Phylogeny of *A. thaliana* and *M. polymorpha* PGs. PGs that will be the subject of this internship shown in red.

2. - MATERIALS & METHODS

2.1. - Phylogenetic analyses

cDNA sequences of annotated proteins were retrieved from UniProt following query with the "polygalacturonase" keyword. In parallel all *A. thaliana*, and *M.polymorpha* polygalacturonase cDNAs sequences were retrieved from NCBI. The sequences were aligned using MUSCLE in the MEGA11 software package (Tamura et al., 2021) with gap open penalty set to -2.9, gap extend penalty to 0 and the hydrophobicity multiplier to 1.2 . A maximum likelihood tree was created using the IQ-TREE 2.2.0 software package (Minh et al., 2020) with the previously aligned sequences. The options were set to perform 1000 bootstraps, to automatically find the best substitution model and to use the fast SH-aLRT test to check for branch support. (Anisimova et al., 2011). An additional alignment was performed using only the AtPGF9, AtPGF12, AtPGF7 and MpPG2 sequences using MUSCLE with the gap opening penalty increased to -5.0.

2.2 - Structural analysis

Structures of proteins were predicted using the neural network-based Alphafold 2.2.0. model (Jumper et al., 2021) with the amber relaxation option turned on. (link) The protein structures were aligned with the structure of the *Yersinia enterocolitica* Exo-PG containing a digalacturonic acid substrate (Abbot et Boratson (2007), PDB: 2UVF) using the UCSF ChimeraX 1.4 software package (Pettersen et al., 2021). Tetrapolygalacturonic acid structure was obtained from pubchem (CID: 5459352) and aligned onto digalacturonic acid present in the above-mentioned structure using the same program. The structure was then saved and the lowest energy state of the structure was found using the "minimize structure" option in the UCSF Chimera 1.6 software package. (Pettersen et al., 2004)

Analysis of hydrogen bonds in the active site was performed by using the "H-bonds" tool in ChimeraX with the substrate selected.

The proteins surfaces were evaluated for hydrophobicity and electrostatic potential using the respective tools in ChimeraX 1.4. Furthermore, the pI of proteins was calculated using the "Compute pI/Mw" tool on Expasy (Bjellqvist et al., 1993) and potential glycosylation sites were predicted using the NetNGlyc 1.0 web service with default settings (Gupta and Brunak, 2002), signal peptides were checked for using SignalP 6.0 (Teufel et al., 2022), transmembrane domains were predicted using DeepTMHMM 1.0.10 (Hallgren et al., 2022), potential cysteine bridges were assessed by manual visualization of generated protein structures, and lastly tyrosine sulfation was checked using Sulfinator (Monigatti et al., 2002).

2.3. - Protein production and purification

2.3.1. - Production system

AT3G61490 (AtPGF9), AT4G23500 (AtPGF12) and AT3G48950 (AtPGF7) and Mapoly0009s0072 (MpPG2) recombinant proteins were heterologously expressed in *Pichia pastoris* X-33 strain after cloning the coding sequences into the methanol inducible pPICZαB vector (Invitrogen). Coding sequences were cloned in frame with the α-factor secretion signal, at the N-terminus of the proteins and the c-myc and His tags at the C-terminus of the proteins.

2.3.2. - Production of recombinant proteins

A single colony of transformed yeasts was placed into a 125mL baffled plastic Erlenmeyer flask along with 12.5 mL of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) yeast nitrogen base, 0.00004% (w/v) biotin, 1% (v/v) glycerol) with 100 µg / mL of Zeocin (Invitrogen). The cultures were incubated at 30°C and 250 rpm. The next day, the cells' optical density (OD_{600}) was measured and enough of the yeasts to reach an optical density of OD_{600} = 1 was transferred to 100 mL of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) yeast nitrogen base, 0.00004% (w/v) biotin, 0.5% (v/v) methanol) in a baffled erlenmeyer flask. After 24 hours protein production was induced by adding methanol to a final concentration of 0.5% (v/v) (all methanol was assumed to have evaporated). The following day the secreted proteins were collected by

transferring the culture to 50mL falcon tubes, centrifuging them at 1500 g for 5 minutes at 6°C and recovering the supernatant.

2.3.3. - Purification of recombinant proteins

Purification of proteins was carried out immediately after collecting the supernatants using metal ion affinity chromatography with 1mL Histrap™ columns (GE healthcare). Before purification, the supernatant was filtered using sterile GD/XPES filters with a 0.45 µm pore size. The column was placed on a peristaltic pump and the flow was set to 1 mL / min. The column was first rinsed with 10 mL of MilliQ water after which 100 mL of supernatant was applied onto the column. Columns were then washed with 10 mL washing buffer solution (50 mM NaP, 250 mM NaCl, 5 mM Imidazole, pH 7.5) after which the bound proteins were eluted with 10 mL of elution buffer (50 mM NaP, 250 mM NaCl, 500 mM Imidazole, pH 7.5). After elution the column was rinsed with an additional 5 mL of elution buffer followed by 10 mL of MilliQ water and 10 mL of 20% (v/v) ethanol and stored at 4°C. During the purification process all the solutions and eluates were kept on ice.

2.3.4. - Ammonium sulfate precipitation of proteins

For some experiments, in which affinity chromatography was unsuccessful, proteins were precipitated using ammonium sulfate. 150 mL of culture supernatant was placed into a beaker and ammonium sulfate was added to reach a concentration of 50% (w/v) according to Doung-Ly et Gabelli (2014). The proteins were precipitated at 4°C with constant mixing for 2 hours. They were then transferred to 30 mL COREX tubes and centrifuged at 13000 g for 30 minutes. The pellets were resuspended in 50 mM sodium acetate pH 4.5 buffer and ammonium sulfate concentration in the supernatant was then increased to 90% (w/v). The procedure was repeated as previously mentioned.

2.3.5. - Concentration and buffer exchange of purified proteins

Eluates of around 10 mL obtained after purification, which contained proteins of interest, were concentrated using Amicon Ultra-4 Ultracel 10 kDa centrifugation filters. The filters were first rinsed by adding 4 mL of distilled water and centrifugated at 7500 g for 7 minutes after which the proteins were concentrated by repeatedly adding the supernatant to the filter and centrifuging it at 7500 g, 4°C for 15 minutes until the final volume reached ~140 µL. Buffer exchange to 50 mM sodium acetate pH 4.5 buffer was performed using PD SpinTrap™ G25 columns (GE Healthcare) according to the manufacturer's instructions except that the columns were equilibrated with buffer nine times instead of five times.

2.3.6. - SDS-PAGE of purified proteins

To check for successful purification, proteins of concentrated eluates were separated and visualized using SDS-PAGE before performing western blot analysis. SDS PAGE was realized using a resolving gel (0.375 M Tris-HCl pH 8.8, 0.08% (v/v) TEMED, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulfate, 0.3% (w/v) Acrylamide/Bis-acrylamide) and a stacking layer (0.125 M Tris-HCl, pH 6.8, 0.1% (v/v) TEMED, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulfate, 0.05% (w/v) Acrylamide/Bis-acrylamide) with a thickness of 0.75

mm. 15 µL of sample was mixed with 3.5 µL of 5x protein loading buffer (313 mM Tris-HCl pH 6.8, 50% (v/v) glycerol, 10% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 0.02% (w/v) bromophenol blue) and denatured for 20 minutes at 95°C. Samples were loaded onto the gel in TGS buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS) and the migration was performed first at 25 mA for 15 minutes to stack the sample and then at 45 mA for 30 minutes to separate the proteins. The gel was then stained with Quick Coomasie stain (Generon) with the following procedure: the gels were transferred into distilled water and microwaved at maximum power for 1 minute. They were then transferred to the staining solution and left on gentle shaking overnight.

2.3.7. - Western blot of purified proteins

After SDS-PAGE, the gel was transferred in a cathode buffer (25 mM TRIS pH 9.4, 40 mM glycine, 10% (v/v) ethanol) for 15 minutes. In parallel, a polyvinylidene difluoride (PVDF) membrane (Amersham Hybond P 0.45 µM, GE Healthcare Life Sciences**)** was placed in 100% ethanol for 15 seconds, then in distilled water for 2 minutes and finally in anode II buffer (25 mM TRIS pH 10.4, 10% (v/v) ethanol). Transfer of proteins onto the membrane was performed using a semi-dry system (Trans-blot Turbo, BIO-RAD). Four Whatman 3M papers were soaked in cathode buffer, one in anode II buffer and two in anode I buffer (3000 mM TRIS pH 10.4, 10% (v/v) ethanol). The soaked papers, membrane and SDS-page gel were stacked according to Figure 6. and transfer was performed at 100 mA, 25 V for 30 minutes. The membrane was then placed in a 96% EtOH solution for 1 minute and left to dry for one hour. Immunodetection was performed using anti-His antibodies coupled to peroxidase (Sigma). Membranes were blocked at room temperature for 30 minutes in 25 mL of 4% milk (w/v) TBS Tween 20 solution (50 mM Tris, 150 mM NaCl, 0.5 % (v/v) Tween, pH 7.6). They were then incubated at room temperature for 90 minutes in 20 mL of 0.5% milk TBS Tween 20 solution with 1/4000 antibody dilution. The membranes were then washed three times in TBS Tween 20 solution (2 times for 15 minutes and once for 5 minutes) and once in dH_2O for 5 minutes. Detection of peroxidase activity was performed using 3 mL of 1X DAB substrate (Thermo scientific).

2.3.8. - Determination of protein concentration

Protein concentration was determined using the method described in Bradford 1976. Solutions of 0, 0.5, 1, 10, 15, 22.5 and 25 µg / mL of bovine serum albumine (Sigma) in 50 mM sodium acetate pH 4.5 buffer were used as standards. Absorbance was measured on 96-well microplates (Greiner Bio one) at 595 nm (Powerwave XS2, BioTek). Protein concentration was determined by extrapolation from the standard curve created by measuring the absorbance of standard samples.

2.4. - Biochemical and biophysical characterization

2.4.1. - Activity assay

Enzyme activity was determined using the DNS colorimetric method where the yellow 3,5-dinitrosalicylic acid (DNS) reacts with the sugars' reducing ends and turns into the red 3-amino-5-nitrosalicylic acid. To perform the assay, purified proteins were dissolved in 24 µL of 50 mM sodium acetate buffer with 1% (w/v) of polygalacturonic acid (Sigma life sciences, 81235-50G). The samples were then placed in a thermocycler and incubated for 17h at specific temperatures. 150 µL of DNS solution (1% w/v DNS, 1% w/v NaOH, 0.05%

w/v Sodium sulfite, 30% w/v Sodium potassium tartrate) was then added to the samples, which were incubated at 95°C for 10 minutes. The samples were transferred into a 96-well microplate (Greiner Bio one) and the absorbance was read at 550 nm. For quantification of activities, standards of 0, 192.5, 385, 962.5, 1925, 2887.5 and 3850 µg / mL of D-galacturonic acid (Fluka analytical, 48280) were prepared. To perform relative quantification and to correct the absorbance, additional samples were prepared by boiling the proteins at 95°C for 15 minutes. All experiments were realized in triplicates.

2.4.2. - Temperature-dependence of the activity

To assess the optimal temperature for activity of enzymes, an assay was performed as described in section 2.2.1. with 0.5 µg of purified protein in sodium acetate buffer (pH 4.5) and using a range of temperatures (30, 40, 50, 60 and 70 °C).

2.4.3. - pH-dependence of the activity

To assess the optimal pH for enzymes' activities, assays were performed as described in section 2.2.1., at the optimal temperature determined for each enzyme using 50 mM sodium citrate buffer (pH 3, 4 ,5, 6, 7 and 8) with 6 µL of the ammonium sulfate precipitated proteins**.**

2.4.4. - Effects of substrate concentrations on enzyme activity

To assess the activities of enzymes at various substrate concentrations, an assay was performed as described in section 2.2., at optimal pH and temperature, using 10, 20, 30, 40, 50, 60, 70 and 80 mg / mL of PGA and 0.5 µg (AtPGF12) or 0.25 µg (MpPG2) of purified proteins**.**

2.4.5 Determination of melting temperature

To assess protein stability, melting temperatures were determined using the differential scanning fluorimetry (DSF) method described in Niesen et al. (2007.). 19 µL of 3 µg μl^{-1} protein sample was mixed with 1 µL of SYPRO orange dye (Sigma, S5692-50UL) in a microplate. The microplate was then placed into a RT-PCR machine (LC480, Roche) where temperature was increased from 30°C to 70°C at a rate of 0.2°C/10 seconds. The fluorescence was measured every time using an excitation wavelength of 492 nm and an emission wavelength of 610 nm.

2.4.6 Influence of Ca^{2+} on enzyme activity

To check for Ca^{2+}-dependency of enzymes' activities, proteins were incubated as described in section 2.2.1. at optimal pH and temperature, with the addition of Ca^{2+} (400 µM $CaCl_2$) and/or EDTA (400 µM EDTA). The activities were compared with that of control samples in absence of Ca^{2+} and EDTA.

2.4.7 Determination of digestion products using oligoprofiling

To determine the digestion products, reactions were performed using 3 µg of protein in 200 µL of 100 mM ammonium acetate pH 4.5 buffer containing 1% (w/v) PGA (Sigma life sciences, 81235-50G) or 35% methylesterified PGA (Cargill, OF 600 C ND). Samples were incubated overnight at their optimum temperature. 200 µL of 100% EtOH was added to precipitate undigested pectins, centrifuged at 5000 g for 5 minutes and the supernatant was transferred to Eppendorf tubes. The samples were then dried in a speedvac

(Concentrator plus, eppendorf) at 45°C for 3 hours. 200 µL of LC/MS water was then added to redissolve the dried samples and they were transferred to capped liquid chromatography vials and stored at -20°C. Data was collected using a Waters Synapt G2-Si LC/MS machine and was analysed using MassLynx 4.1 (Waters) software. MS-detection was performed in negative mode with end plate offset voltage set to 500 V, capillary voltage to 4000 V, the nebulizer set to 40 psi, dry gas to 8 L / min and dry temperature to 180°C.

For DIONEX (Liquid chromatography) samples were prepared by incubating 15 µL of a mixture of 10 µL wash (25 mM Imidazole) and 25 µL elution (250 mM Imidazole) solution of proteins in 50 mM ph 4.5 sodium acetate buffer overnight at 45°C with PGA (Sigma life sciences, 81235-50G). The sample were then concentrated in 0.5 mL 3kDa Amicon concentration tubes, transferred to capped liquid chromatography vials and stored at -20°C. Data was collected using a Thermofisher DIONEX liquid chromatography system.

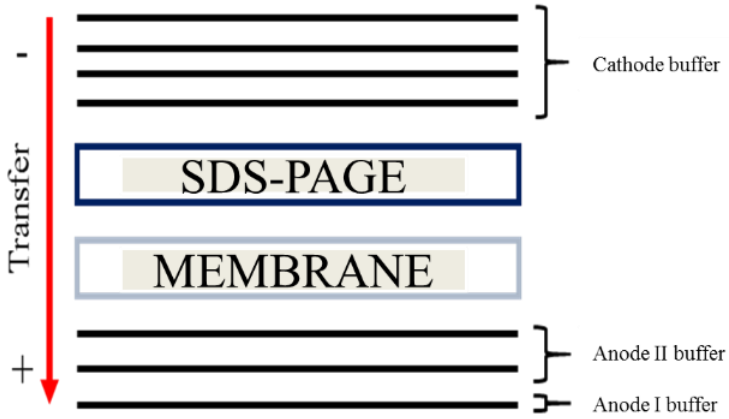

Figure 6. Schematic representation of protein transfer on membrane

3. – RESULTS

3.1. - Phylogenetic analysis of AtPGF9, AtPGF12, AtPGF7 and MpPG2

A maximum likelihood phylogenetic tree was constructed using the sequences of AtPGF9, AtPGF12, AtPGF7 and MpPG2, all the annotated polygalacturonase genes from Swiss-Prot as well as the non-annotated sequences from *A. thaliana* and *M. polymorpha*. In total the tree contains 179 sequences. (Figure 7.) The results showed that the sequences formed 3 distinct clades each containing sub clades. Clade II contains all the sequences from fungal polygalacturonases, except for one sequence from *M. polymorpha* (Mapoly0140s0001). This clade is further neatly divided into three subclades each containing the sequences for either ExoPGs, EndoPGs or Rhamnopolygalacturonases. Clade I contains exclusively sequences from plants. Because most of the polygalacturonases from this group have not been characterized it is not possible to infer their mode of action (ExoPG, EndoPG, RhamnoPG) from their position on the tree. However, we can see from the tree that all seven sequences from this clade characterized as ExoPGs (A. thaliana PGLR1 and PGLR2; Z. mays PGLR1, PGLR2, PGLR3; O. organensis PGLR; P. acerifolia PGLR2) are found on the right side, therefore there are some indications that this tree is neatly divided into subclades which may correspond to the mode of action of the enzymes, just like for Clade II. Clade III contains the sequences for AtPGF9, AtPGF12, AtPGF7 and MpPG2. Unlike Clade I and Clade II which are composed exclusively of either plant

or fungal sequences, Clade III contains a mix of both plant and bacterial sequences. Furthermore, apart from MpPG2 and 3, all the plant sequences in this clade belong to *A. thaliana*. Lastly, it is interesting to note the heterogeneity among sequences from *M. polymorpha* despite coming from the same organism. MpPG2 and MpPG3 are found in Clade III and show short branch length indicating less genetic change, which is also true for MpPG1 found in Clade I while other sequences from *M. polymorpha* have very large branch lengths indicating a large amount of genetic change in some *M. polymorpha* PG genes.

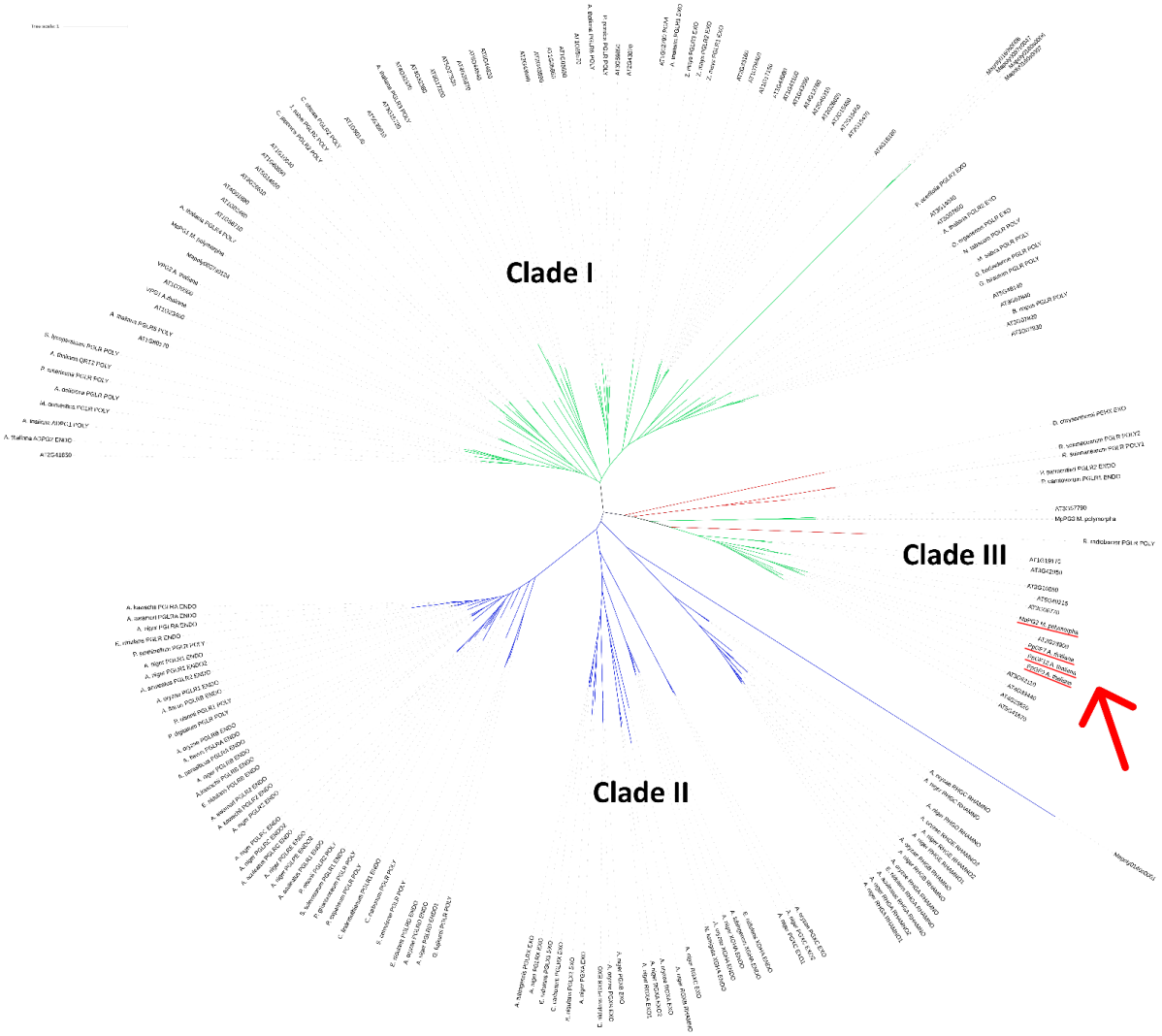

Figure 7. Phylogenetic tree of annotated polygalacturonases from Swiss-Prot and *A. thaliana* and *M. polymorpha*. Green branches represent plant sequences, blue represent fungal sequences and red represent bacterial sequences. AtPGF9, AtPGF12, AtPGF7 and MpPG2 are underlined, and their locations are shown with a red arrow.

3.2. - Structural analysis of AtPGF9, AtPGF12, AtPGF7 and MpPG2

3.2.1. - Protein structure prediction and substrate docking results

The structures of AtPGF9, AtPGF12, AtPGF7 and MpPG2 were predicted using the neural network based Alphafold 2.0. model. Tetragalacturonic acid obtained from pubchem (CID: 5459352) was then docked into the active site using a combination of Chimera and ChimeraX on the basis of structure similarity with a published crystalline exo-polygalacturonase structure from *Yersinia enterocolitica* (Abbott and Boraston, 2007 PDB:2UVF). All four studied proteins possess the right handed parallel ß-helix structural domain first

observed in pectate lyase (Yoder et al., 1993) (Figure 8., left). Furthermore, all 4 proteins possess two loops that arch over the binding cleft and close it from the N-terminal side, just like in previously crystalized exo-PGs (Abbott and Boraston, 2007; Luis et al., 2018; Pijning et al., 2009), with distances between the two closest points ranging from 1.736 Å (AtPGF7) to 1.965 Å (AtPGF12). The tetragalacturonic acid substrate is situated between these two loops and the right-handed parallel ß-helix structural motif. Lastly, the proteins' surface shape is incredibly similar, especially between the *A. thaliana* ExoPGs, with MpPG2 showing a slight difference because of the presence of a longer C-terminal end. (Figure 8, right).

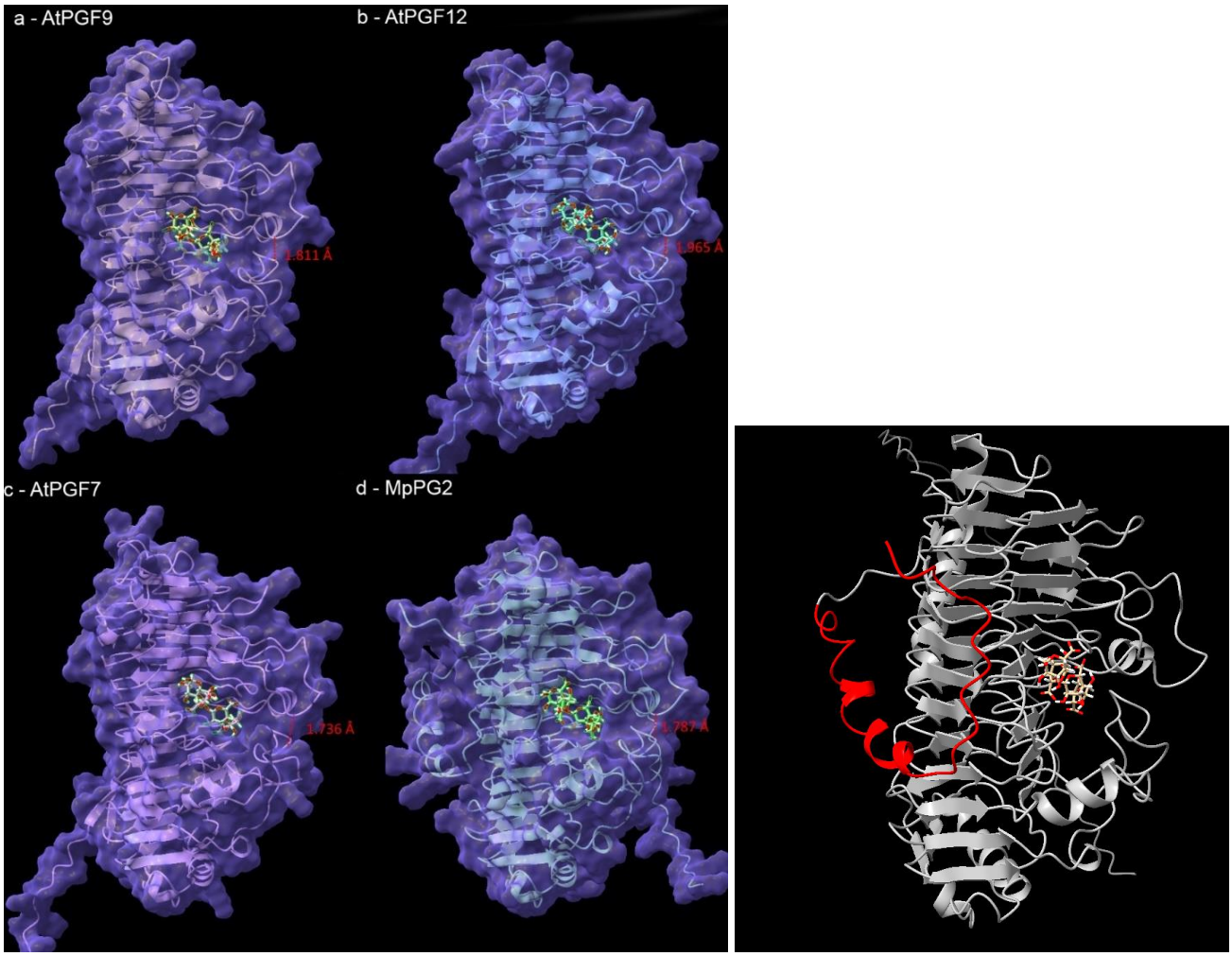

Figure 8. Left - Structure and surface of AtPGF9, AtPGF12, AtPGF7 and MpPG2 with docked tetragalacturonic acid substrate. Distances between nearest points of binding groove loops are shown. Right – Zoomed in view of MpPG2. C-terminal end of the protein is shown in red.

3.2.2. - Alignment & Structural features

An alignment of AtPGF9, AtPGF12, AtPGF7 and MpPG2 was created to gain deeper insight into the similarities and differences between their sequences. All 4 sequences were then annotated with secondary structural features from the predicted structures and with signal peptides, glycosylation sites, and sulfated tyrosines predicted bioinformatically. It was further analyzed for conserved motifs and residues known or inferred from the literature to take part in the catalytic mechanism (Figure 9.). We can see from the alignment that the four proteins possess an incredibly similar secondary structure with few differences between them. The three proteins from *A. thaliana* possess a signal peptide for cellular export on their N-terminal ends.

MpPG2 has a predicted N-terminal alpha helix (α0, Figure 9.) which is not present in other proteins, while the three *A. thaliana* proteins possess an N-terminal beta strand exclusive to them (ß0) which forms a specific beta sheet with the beta strand ß36. After this, the four proteins possess an almost identical secondary structure in the segment found between beta strands ß1 and ß35 with the sole difference being that AtPGF7 does not have the beta strand ß17. After beta strand ß35, the differences between the secondary structures are more specific to each sequence. MpPG2 and AtPGF9 are the only ones to possess alpha helix α7, while MpPG2 additionally possesses two unique alpha helices α9 and α10. Lastly, the *A. thaliana* proteins possess a common α6 alpha helix found at the C-terminal end of the parallel ß-helix structural motif.

The four proteins possess many disulfide bridges with two of them being conserved in all four proteins. Disulfide bridge number 6 is found at the C-terminal end of the parallel ß-helix structural motif. Disulfide bridge number 5 connects the ß-strand ß32 and the α7 α-helix (MpPG2) or a conserved Cysteine at position 489 in the alignment (AtPGF9, AtPGF12, AtPGF7). Disulfide bridge 1 is unique to MpPG2 and connects the alpha helix α0 with the beta strand ß15. Disulfide bridges 2 and 4 are unique to *A. thaliana* proteins with disulfide bridge 2 connecting a conserved Cys residue near the N-terminal end of the protein with beta strand ß36 and disulfide bride 4 connecting beta strand 17 to beta strand 19. Disulfide bridge 5 connects the beta strand ß32 with the alpha helix α7 in MpPG2 or a conserved cysteine residue in AtPGF9, AtPGF12 and AtPGF7 to the same beta strand. (Figure 9.)

The surface electrostatic potential was calculated, highlighting some major differences between the four proteins. When looking at the front side, we see that AtPGF12 and MpPG2 are more positively charged in the vicinity of the active site as compared to AtPGF9 and especially AtPGF7 (Figure 10., upper panel). AtPGF12 has a very large positive patch right next to the active site entrance, AtPGF9 and MpPG2 have a smaller patch and AtPGF7 has no positive patch at all. (Figure 10., circled) The back side of the proteins shows even greater differences, with AtPGF9 and AtPGF12 being mostly positively charged, MpPG2 being a mix of positive and negative charges and AtPGF7 being mostly neutral (Figure 10., lower panel). These differences are also reflected in the calculated pI of the proteins, where AtPGF9, AtPGF12 and MpPG2 have pI of around 5.5 while AtPGF7 has a pI of 8.9.

Furthermore, the proteins also possess many putative glycosylation sites with three of them being fully conserved among the 4 proteins (Figure 11., right). Lastly, MpPG2, AtPGF9 and AtPGF12 also contain a single fully conserved tyrosine sulfation site which is structurally positioned right next to the conserved alpha helix α5 situated between the two loops arching over the active site. (Figure 11., left)

Altogether, these four enzymes have very similar structures, but each of them has its particularities which may contribute to their differences in activity.

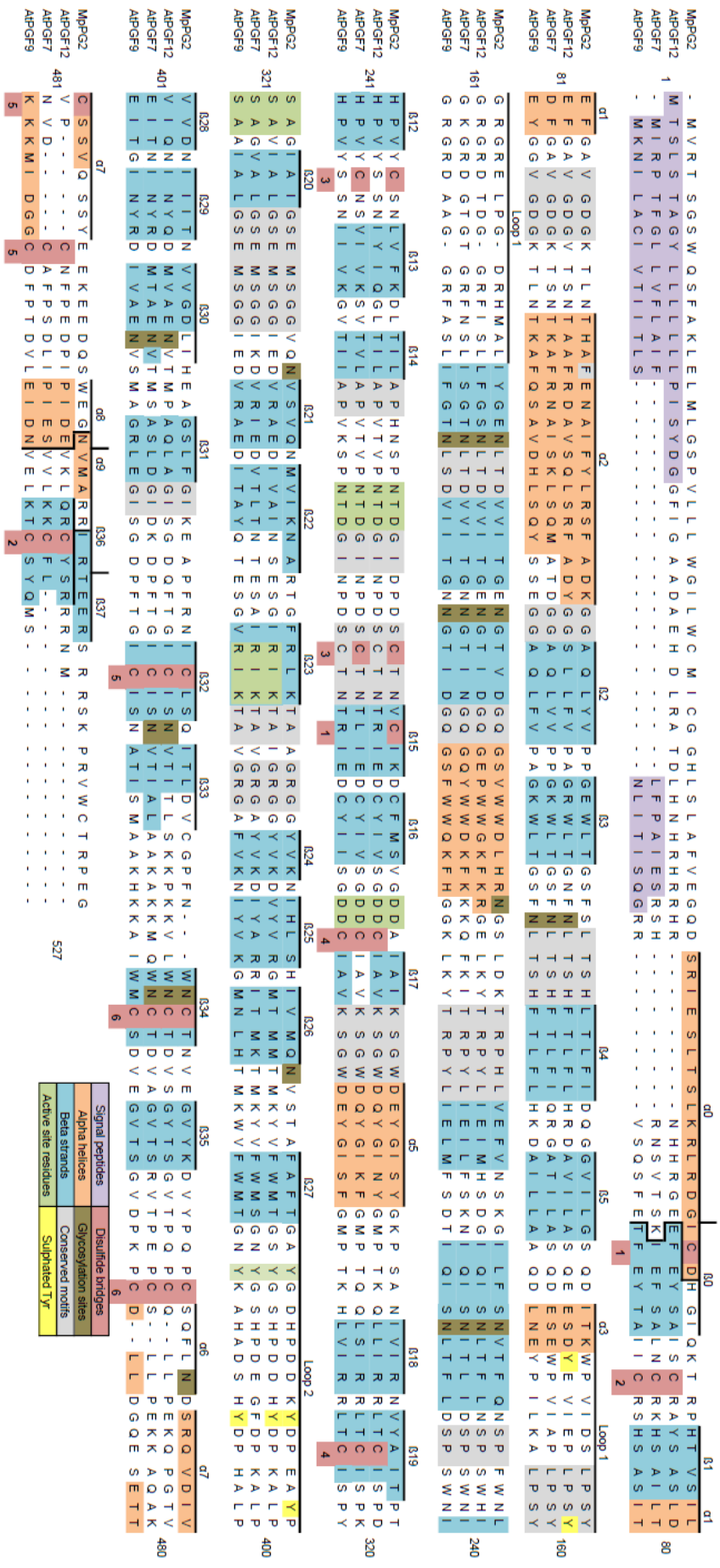

Figure 9. Alignment of AtPGF9, AtPGF12, AtPGF7 and MpPG2 annotated with secondary structures (orange and blue), signal peptides (purple), Active site residues (green), disulfide bridges (red), glycosylation sites (brown), sulfated tyrosines (yellow) and other conserved motifs (grey)

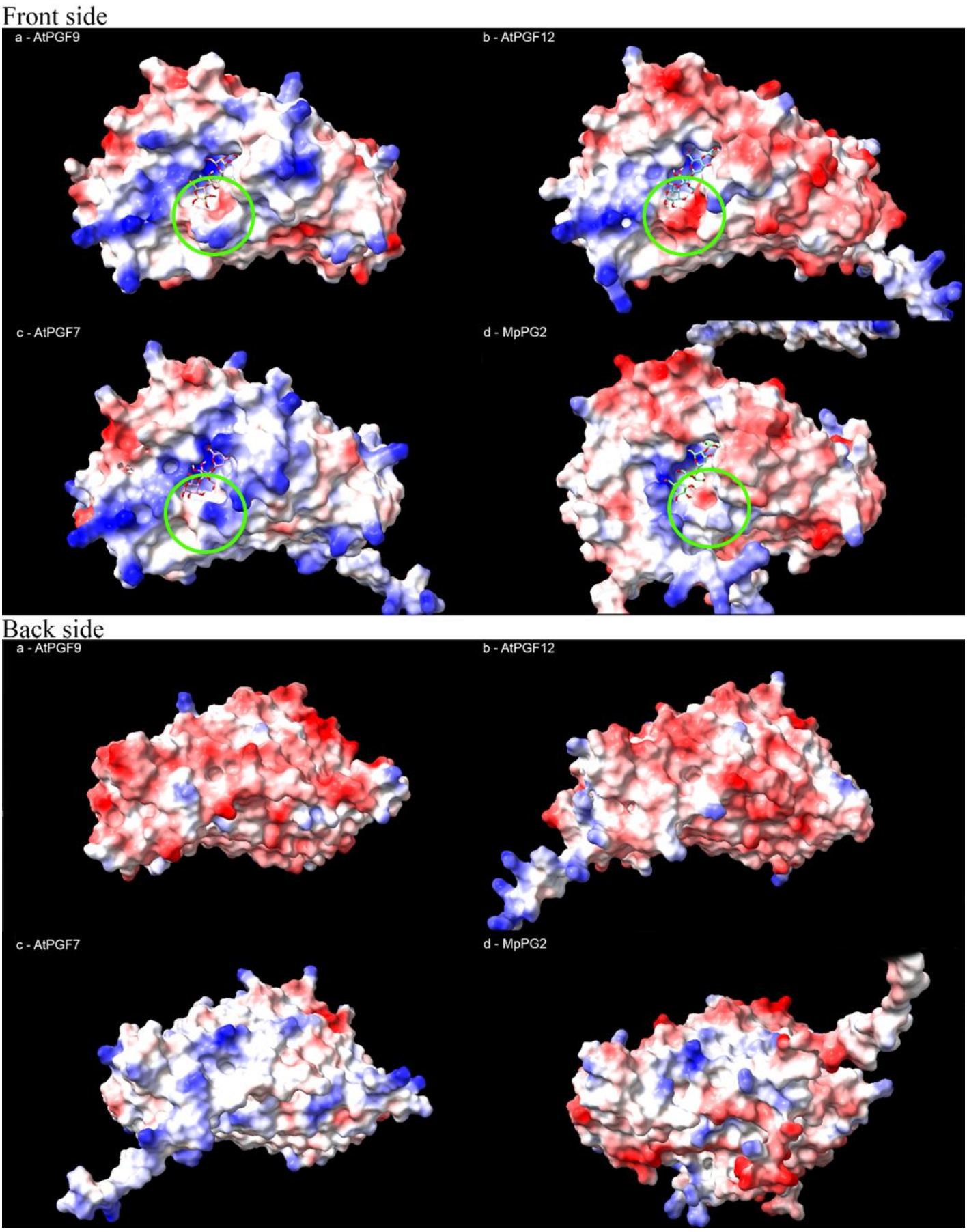

Figure 10. Electrostatic surface potential of front (top) and back (bottom) side of AtPGF9, AtPGF12, AtPGF7 and MpPG2. Positively charged patch near entrance to the active site circled in green. Red indicates positive charge, white neutral charge, and blue negative charge.

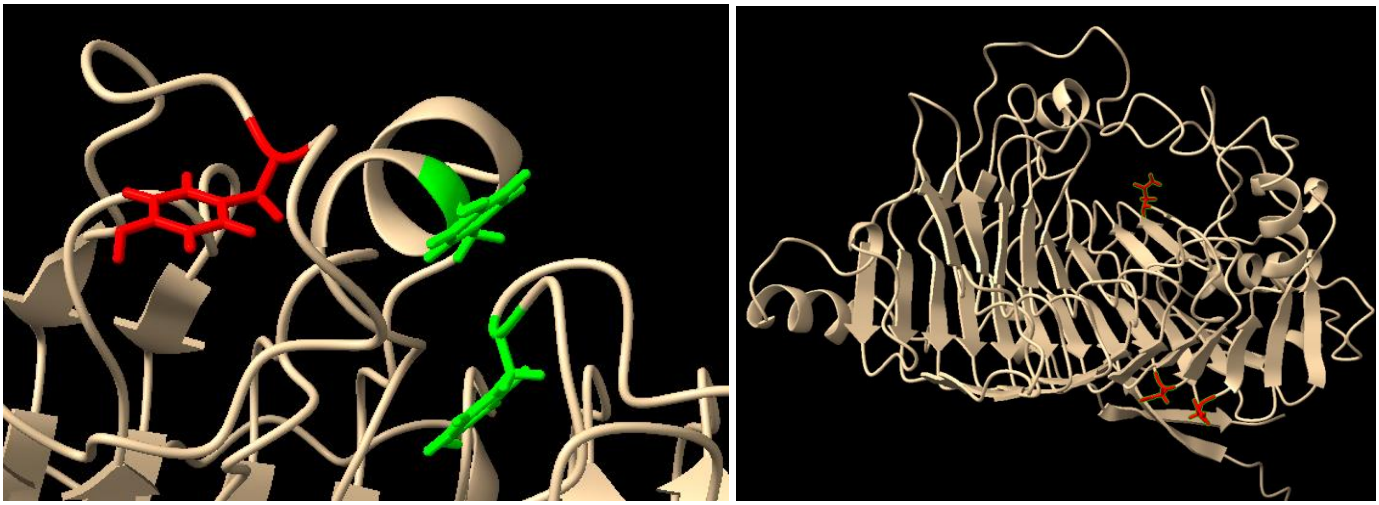

Figure 11. Left - Tyrosine residue near the α5 helix that plugs the gap between loop 1 and 2 of the active site (green). Conserved sulfated tyrosine residue in red. Right - Conserved asparagine residues (red) in turns between ß-sheets near the N-terminal end of the ß-helix structural motif. Both shown on AtPGF9

3.2.3. - Active site architecture

As inferred from the alpha-fold models, the active site of the four enzymes is composed of parts of the parallel ß-helix structural motives enclosed by two loops and an alpha helix (Loop1, Loop 2 and α5, Figure 9.). The active site is completely closed from the N-terminal side. The proteins possess three conserved motifs shown to take part in the reaction mechanism: the NTD motif and the DD motif whose aspartates form the catalytic triad (Figure 12., right) and the RIK motif in which the arginine and lysine take part in substrate binding. However, they do not possess the GHG motif which is conserved in other endo- and exo-PGs. This motif is replaced by the conserved SA motif which is structurally found in the same place as the GHG motif is in other polygalacturonases. (Figure 12., left)

To see which residues may affect substrate binding, an analysis of the possible hydrogen bonds between the docked substrate and each of the enzymes was performed. The differences are shown in Figure 13. We can see that MpPG2 has many more proposed hydrogen bonds than the other three proteins. Furthermore, we can see that for all the proteins the two aspartates from the DD motif have an H-bond with the same COOH residue on the substrate. Lastly, there is a discrepancy between the number of possible H-bonds between the enzymes at subsites +2 and +3. For AtPGF9 and AtPGF12 the number of H-bonds is lower than for AtPGF7 and MpPG2.

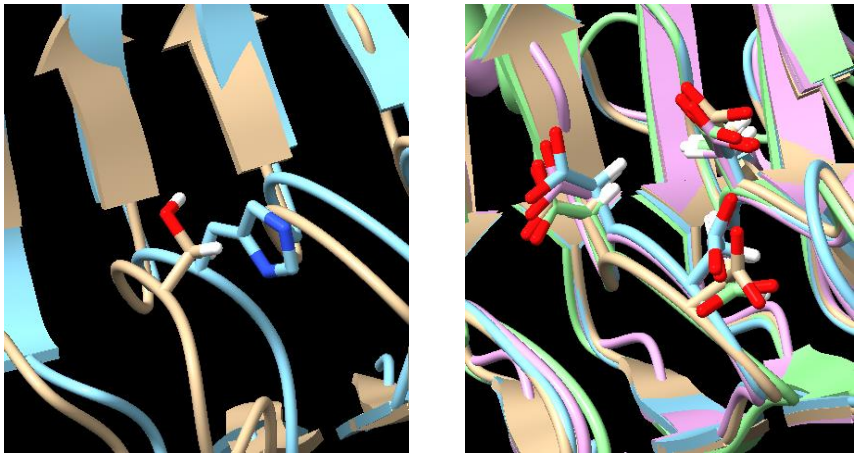

Figure 12. Left – Aligned structures of AtPGF9 (brown) and the Exo-PG from Abbot et Boratson (2008.) (blue, uniport: 2UVF). The serine of the conserved SA motif found in AtPGF9, AtPGF12, AtPGF7 and MpPG2 is shown to be in the same place structurally as the histidine of the GHG motif found in all other characterized polygalacturonaes. Right – Aligned structures of AtPGF9, AtPGF12, AtPGF7 and MpPG2. Aspartates from the conserved NTD and DD motifs which form the catalytic triad are shown.

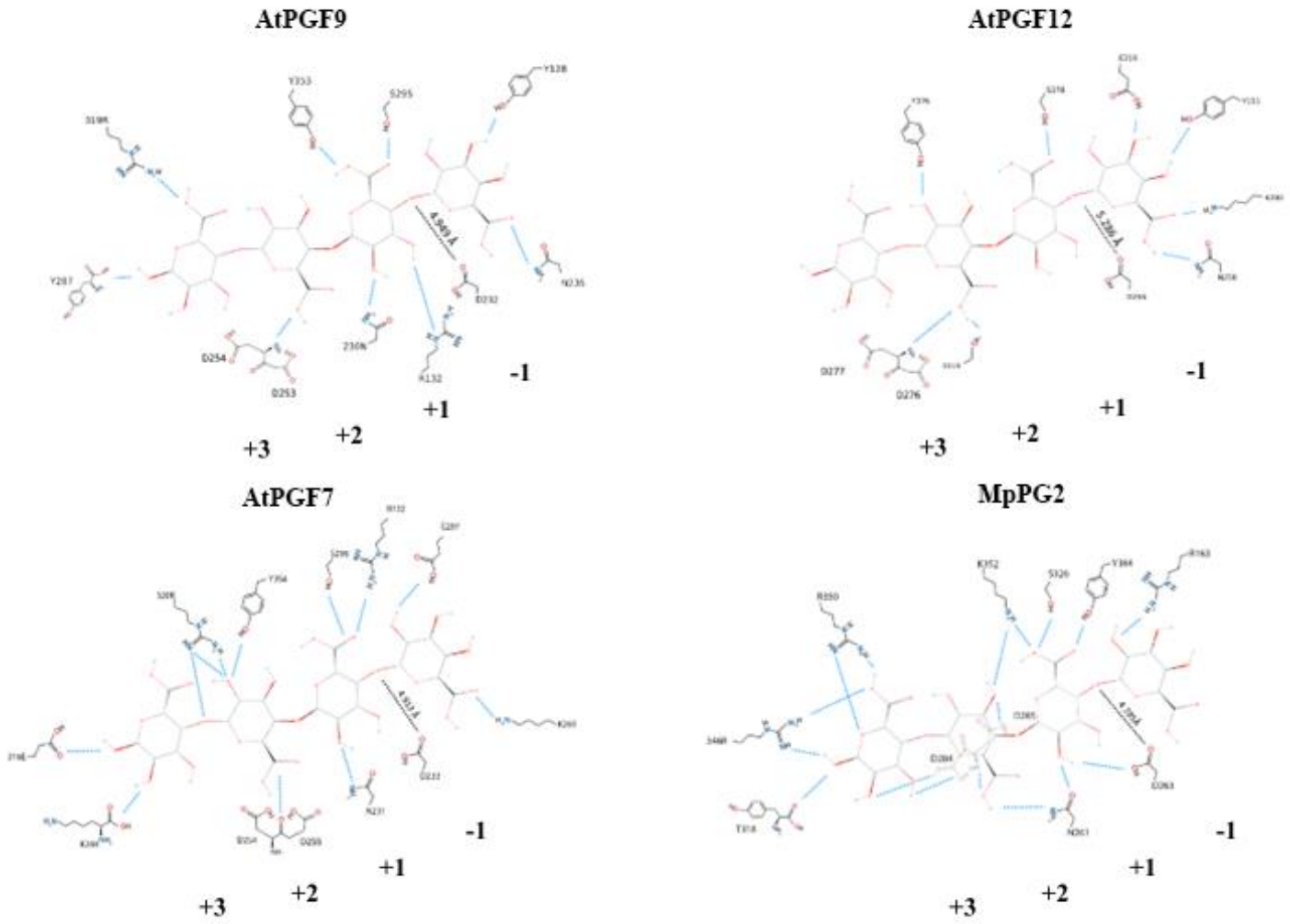

Figure 13. Possible hydrogen bonds (dashed blue lines) between tetragalacturonic acid substrate and active site residues predicted with ChimeraX . Dashed black line – distance between Asp residue of the NTD motif and the glycosidic linkage being broken in catalysis. Subsites (+3, +2, +1, -1) shown beneath substrate-enzyme complex.

3.3. - Production and purification of AtPGF9, AtPGF12, AtPGF7 and MpPG2

Production of AtPGF9 (AT3G61490), AtPGF12 (AT4G23500), AtPGF7 (AT3G48950) and MpPG2 (Mapoly0009s0072) was carried out in the heterologous *Pichia pastoris*, strain X-33 system after cloning into the methanol-inducible pPICZαB expression vector (Invitrogen). Transformants were already available in the laboratory. These proteins were purified by metal ion affinity chromatography thanks to the His-tag. Purification was done immediately after production as we noticed that effectiveness of purification diminishes quickly after expression. We suggest that this is due to protease activity and not problems with detection as western blots made using anti c-myc antibodies gave the same negative results. Average yields per mL of culture were ~0.4 µg for AtPGF9 and MpPG2, 0.3 µg for AtPGF7 and 0.7 µg for PGF12. However, these amounts varied quite a bit from run to run and could be as high as 1.51 µg per mL of culture. SDS PAGE with Coomasie blue staining was used to check for the presence of these proteins at the proper molecular mass in the purified samples, as well as to check the quality of the purification. The predicted size of these proteins together with the added N-terminal signaling peptide and C-terminal c-myc and His-tag was around 60 kDa, which is in line with their sizes observed on gel (Figure 14). Additional faint bands were observed above the corresponding bands which is probably due to the various glycosylation sites predicted on the sequences. The quality of purification was satisfactory as no additional bands could be observed. A further western blot analysis was performed to confirm the presence and purification success of our proteins. We could only observe the presence of AtPGF12 and MpPG2 using anti his-tag antibodies with bound peroxidase (Figure 15.). This is probably due to the activity of various proteases which degraded the His-tag as mentioned before. As it is unlikely that the purification yielded none of our proteins while they can be seen in the SDS PAGE gel and are active, this indicated that protease activity was occurring.

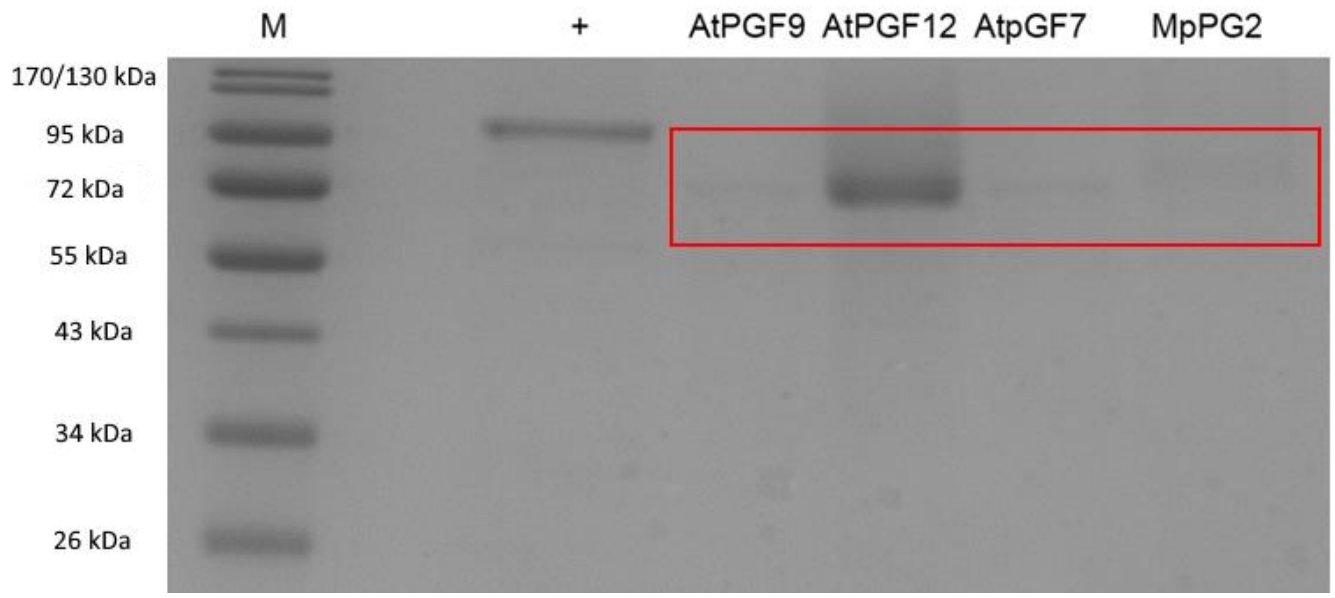

Figure 14. 12% SDS PAGE with Coomassie blue staining of AtPGF9, AtPGF12, AtPGF7 and MpPG2. Corresponding bands highlighted in red.

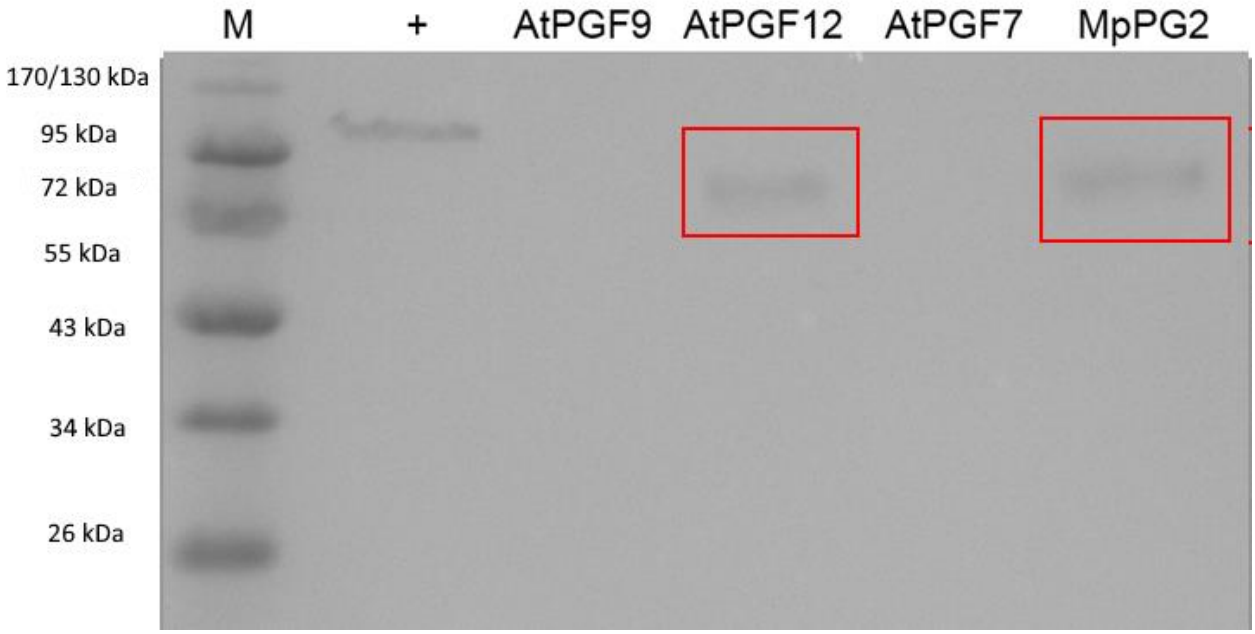

Figure 15. Western blot of AtPGF9, AtPGF12, AtPGF7 and MpPG2 using anti-His antibodies with bound peroxidase. Corresponding bands highlighted in red.

3.4. - Biochemical and biophysical characterization

To confirm that purified proteins were active, we first tested for their activity using the DNS method, where dinitrosalycilic acid reacts with the reducing ends of the substrate and turns into the red 3-amino-5-nitrosalicylic acid, which can be quantified spectrophotometrically. Dependence of temperature on protein activity was measured for all four enzymes at 30, 40, 50, 60 and 70°C. The results showed that AtPGF12 and AtPGF9 present the highest activity at 40°C with activities of 1.24 nmol μg^{-1} min^{-1} and 0.31 nmol μg^{-1} min^{-1} respectively. MpPG2 showed the highest activity of 1.06 nmol μg^{-1} min^{-1} at 50°C while AtPGF7 is the least active with a maximum of only 0.07 nmol μg^{-1} min^{-1} at 60°C. These activities were relatively stable in the range of 10°C around the optimal temperature and dropped sharply afterwards (Figure 16., top).

The dependence of enzyme activity on pH was measured using the same DNS method from pH 2 to pH 8 with increments of 0.5 degrees pH. The activity of these enzymes was measured using protein samples prepared by ammonium sulfate precipitation therefore any comparisons between their absolute activities cannot be made as the amount of active enzyme was unknown. MpPG2 and AtPGF12 showed the highest activity at pH 4 with 4.59 nmol μL^{-1} h^{-1} and 3.98 nmol μL^{-1} h^{-1}, respectively. AtPGF9 showed the highest activity of 3.26 nmol μL^{-1} h^{-1} at pH 4.5 while it is of 2.84 nmol μL^{-1} h^{-1} at pH 5 for AtPGF7. The activities of these proteins drop sharply at pH lower than their ideal pH and gradually at higher pH. The only enzyme still active at pH 6.5 is AtPGF7 (Figure 16., bottom).

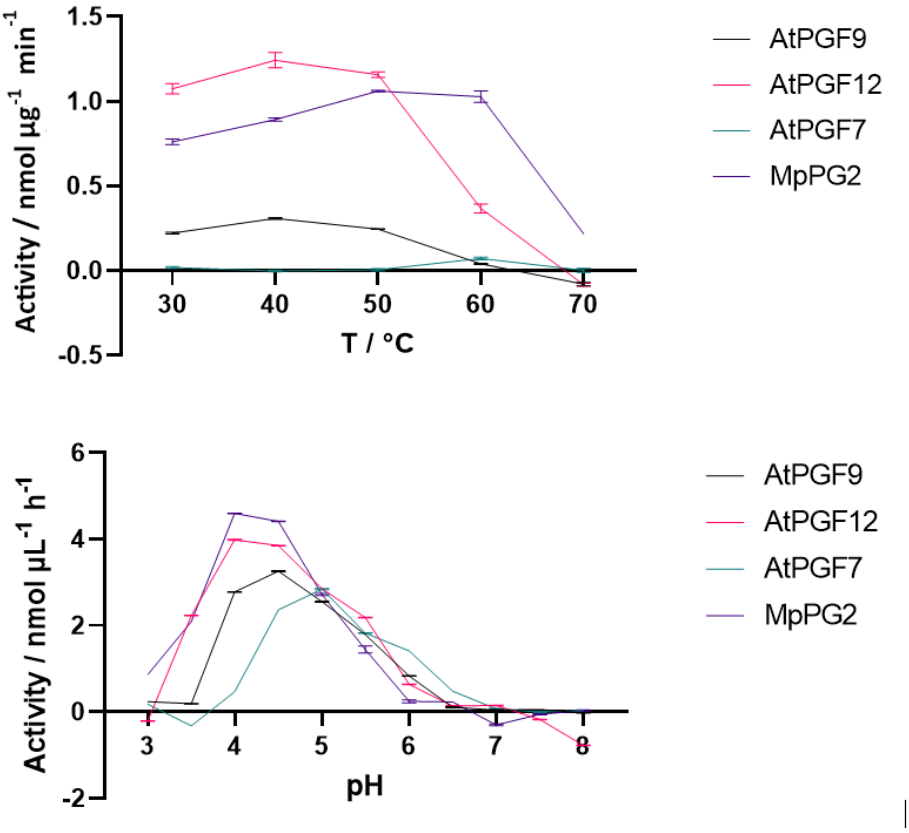

Figure 16. Top – Dependence of enzyme activity on temperature. Bottom – Dependence of enzyme activity on pH.

Activity of proteins was also measured in the presence of Ca^{2+} ions to check for their potential regulatory or catalytic activity. The results for AtPGF12, AtPGF7 and MpPG2 showed no statistically significant difference between the control samples and the Ca^{2+}-treated samples. AtpGF9 however showed a 15.95% increase in activity at the very low Ca^{2+} concentration of 400 µM. The addition of 400 µM EDTA, a divalent cation chelator, to the mixture reduced the amount of activity to 13.23% below the control's activity, which indicates that there are additional divalent ions involved in regulation or catalysis. (Figure 17.)

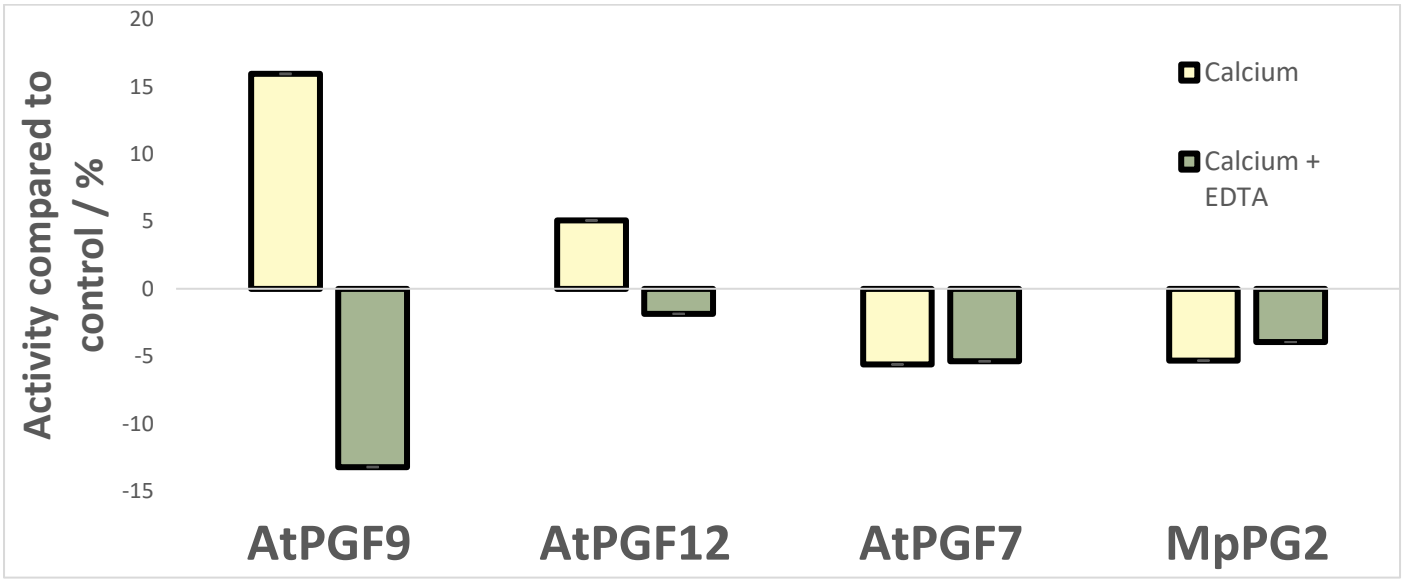

Figure 17. Activity of proteins in the presence of 400 µM $CaCl_2$ or 400 µM $CaCl_2$ and 400 µM EDTA, relative to control (base line).

The melting temperature of AtPGF9 was also measured using differential scanning fluorimetry, that measures the fluorescence of the SYPRO orange die, when bound to hydrophobic parts of the denaturing protein. The melting curve showed that the process of denaturation starts around 50°C. To determine the exact melting point, the curves were derived, and the melting temperature was determined to be 59.95°C (Figure 18.).

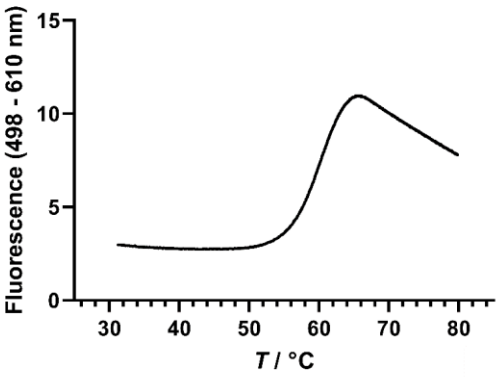

Figure 18. Melting curve of AtPGF9 obtained using differential scanning fluorimetry measuring for fluorescence of SYPRO orange.

The effect of substrate concentration on AtPGF12 and MpPG2 activities were also determined using a range of concentrations previously used in the lab for endo-polygalacturonases (10, 20, 30, 40, 50, 60, 70, 80 mg mL^{-1} PGA). As presented in Figure 19., AtPGF12 and MpPG2 showed the highest activities (2.08 and 2.01 nmol μg^{-1} min^{-1} respectively) at the lowest concentration of 10 mg mL^{-1} of PGA and were then strongly inhibited with higher concentrations of substrate. However, while AtPGF12 completely lost activity with concentrations higher than 40 mg mL^{-1}, MpPG2 managed to retain activity of 0.5 nmol μg^{-1} min^{-1} even at 80 mg mL^{-1} of PGA.

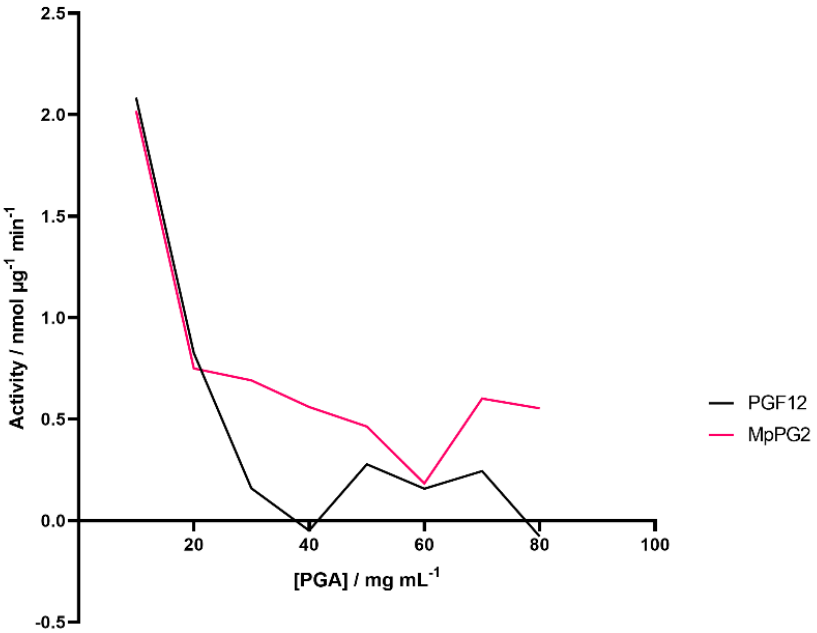

Figure 19. Influence of substrate concentration on activity of AtPGF12 and MpPG2

3.5. - Mode of action of AtPGF7, AtPGF12 and MpPG2

We then assessed the mode of action of the four candidates. Indeed, from structural modelling, AtPGF12, AtPGF9, AtPGF7 and MpPG2 showed characteristics of exo-polygalacturonases. To this aim, the determination of the products from PGA digestion was performed for each enzyme. This technique, named oligoprofiling of enzymes, allows to deduce their activity from the set of oligogalacturonides identified. The results of PGA digestion showed that both AtPGF7 and AtPGF12 were exo-polygalacturonases producing exclusively GalA with AtPGF12 being able to produce almost four times as much in the same conditions. (Figure 20. a, b) The LC-MS analysis of MpPG2 on the other hand showed different results. While MpPG2 does primarily produce GalA, both GalA3 and GalA4 could also be detected in significant proportions when using PGA as a substrate. Intriguingly, no statistically significant increase in the amount of GalA2 was shown. (Figure 20. c).

Furthermore, the enzymatic effects of AtPGF12 and MpPG2 on HG DM35 (35% methylated homogalacturonan) substrates were analyzed using LC-MS. Results showed that there was a significant increase in abundance of unmethylated GalA after digestion with both enzymes, while other OGs, methylated or not were absent (Figure 21.).

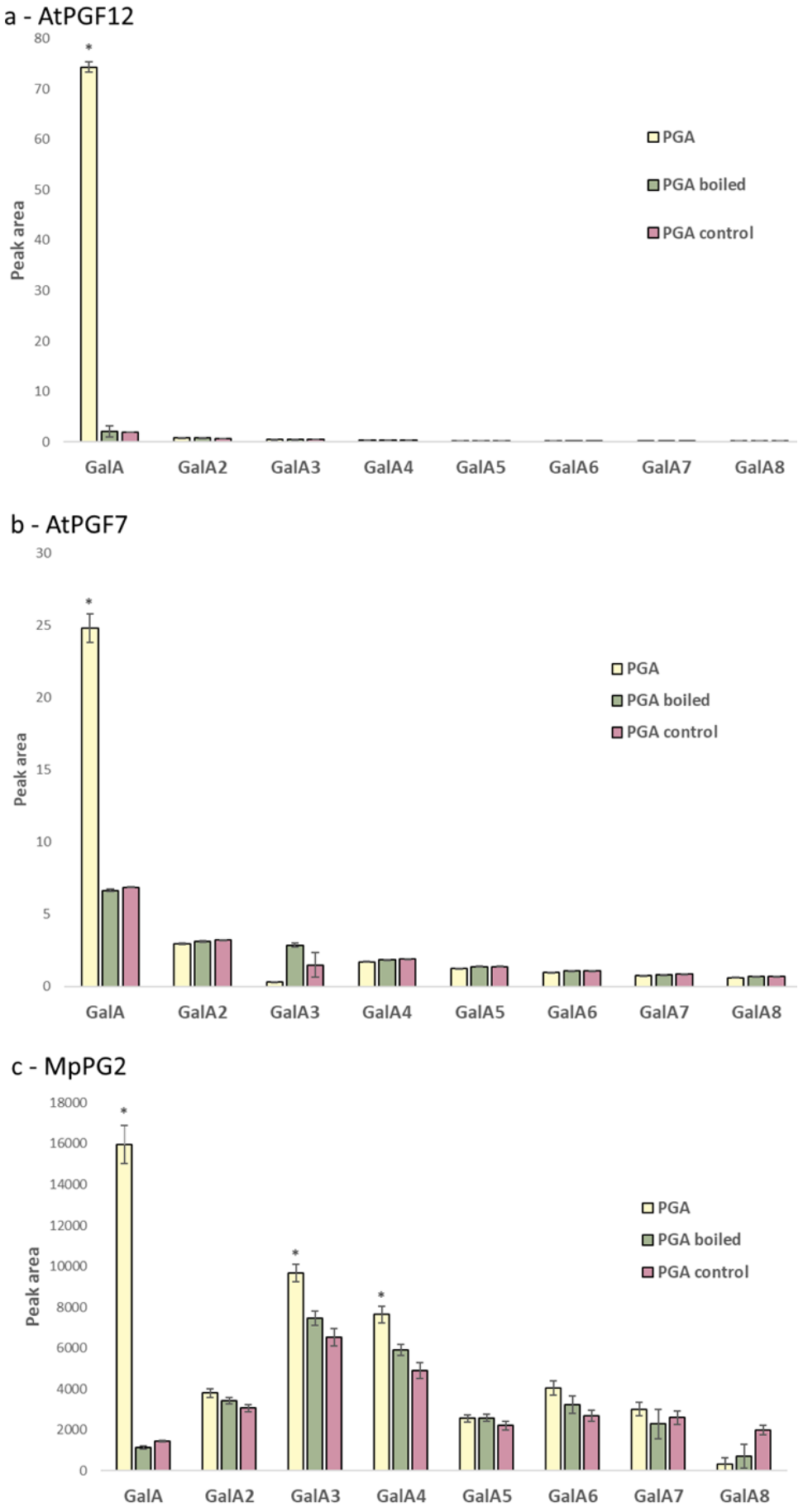

Figure 20. a – LC oligoprofiling of AtPGF12 on PGA substrate, b – LC oligoprofiling of AtPGF7 on PGA substrate, c – LC/MS oligoprofiling of MpPG2 on PGA substrate. Boiled signifies that digestion was performed with boiled proteins and control signifies that substrate was not incubated with any enzyme. Statistically significant results compared to control shown with an asterisk (*).

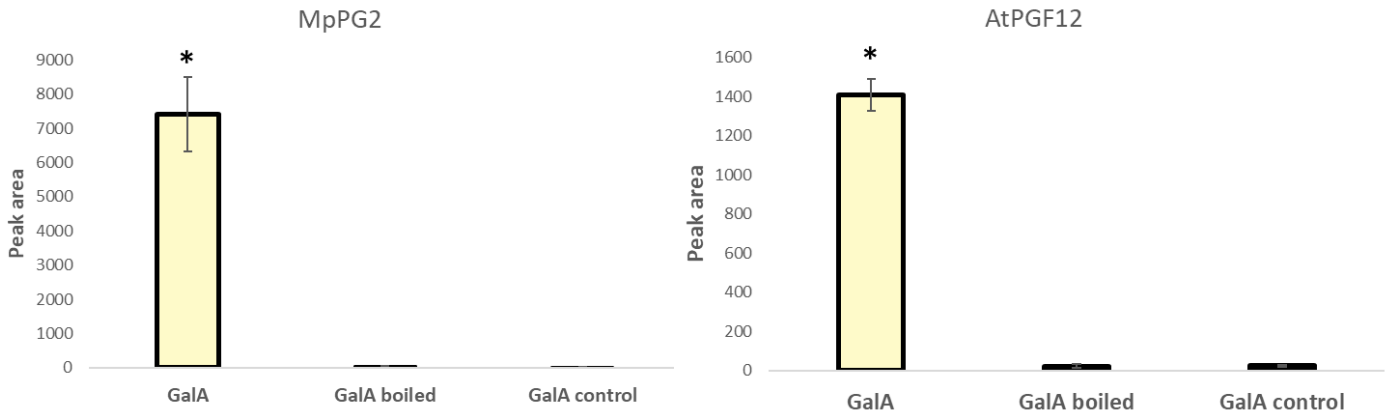

Figure 21. GalA production of MpPG2 (left) and AtPGF12 (right) on DM35 substrate. Boiled signifies that digestion was performed with boiled proteins and control signifies that substrate was not incubated with any enzyme. Statistically significant results compared to control shown with an asterisk (*).

4. – DISCUSSION & CONCLUSIONS

The focus of this study was to gain insights into the phylogeny, structure, biochemical properties, and activity of three putative exopolygalacturonases from *Arabidopsis thaliana* – AtPGF9 (AT3G61490), AtPGF12 (AT4G23500) and ATPGF7 (AT3G48950) and one from *Marchantia polymorpha* – MpPG2 (Mapoly0009s0072). The four proteins were chosen as they possess, based on structural homology modelling, a similar structure and were found to be closely related in previous phylogenetic studies realized in the laboratory. (Figure 5.) We first performed an extensive phylogenetic analysis by creating a large phylogenetic tree that contains the amino acid sequences of many characterized polygalacturonases as well as all the putative polygalacturonases from *A. thaliana* and *M. polymorpha*. Furthermore, thanks to homology modelling using AlphaFold, we studied the overall structure and fold of the four proteins of interest, as well as the architecture of the proteins' active sites. We next performed biochemical analyses to determine the dependency of the enzymes' activities with regards to pH, temperature, Ca^{2+} and substrate concentrations. Lastly, we characterized the digestion products for each of the enzymes using HPLC or LC-MS in order to determine their mode of action (endo or exo).

The phylogenetic tree constructed using the amino-acid sequences of the characterized polygalacturonases as well as all the four putative PG from *A. thaliana* and *M. polymorpha* (Figure 7.) revealed that the three proteins from *A. thaliana* are found in a small clade of highly conserved proteins, including AT2G23900, a putative membrane bound polygalacturonase. Interestingly, MpPG2 from *M. polymorpha* was shown to be the most closely related protein to this clade showing how similar these proteins are. Furthermore, several proteins from *M. polymorpha* including Mapoly0140s001, as well as the clade containing Mapoly0160s008, Mapoly0087s0047, Mapoly0160s0006 and Mapol0160s007 have abnormally long branch lengths. This is likely to indicate that these proteins went through a much higher degree of genetic change compared to the proteins from other groups. These changes could be associated with a diversification of function in development of *M. polymorpha*, confirming it as an attractive model for further research on

polygalacturonases, in particular considering the number of genes encoding PG (10 genes). Lastly, when looking at the overall structure of the phylogenetic tree, three distinct clades can be identified: Clades I and II are composed exclusively of either plant (Clade I) or fungal (Clade II) proteins and are then subdivided into subclades that mostly correspond to the proteins' mode of action (endo versus exo). Clade III, however, contains sequences from both plant (*A. thaliana* and *M. polymorpha*) and bacterial species (*D. crysanthemi, R. solanacearum, R. parmentieri, R. carotovorum, R. radiobacter*). These findings suggest that a horizontal gene transfer might have happened between the common ancestor of *A. thaliana* and *M. polymorpha*, and the common ancestor of the bacterial species. Furthermore, another possibility is that the three clades represent three different lineages of PGs with clade III tracing its origin to a common ancestor of plants and bacteria. However, this is in contradiction with previous reports that grouped plant PGs (Hadfield et al. (1998) and Kim et al. (2006)) and does not explain the rationale for the absence of fungal PG in clade III. Further research is needed to confirm these findings and to confirm the positioning of bacterial PGs in this tree.

The structures of the four proteins were modelled using the neural network-based Alphafold algorithm (Jumper et al., 2021). Analysis of the four models showed that the proteins' structure is very similar to that of other published PG structures (Šafran et al., 2022) with a few key differences. The proteins possess the parallel ß-helix structural domain first discovered in pectate lyase (Yoder et al., 1993). In all proteins this structure is composed of 4 different ß-sheets all with the same number of ß-strands except for AtPGF7 where one of the ß-sheets lacks one ß-strand (ß17, Figure 10.). One of the key differences between the known endo-PG structures and the structure of the four proteins of interest concerns the structure of the substrate binding site. In particular, when looking at the loops that arch over the active site and the ß-helix structure, the distance between them is on average 6.8 times higher in PGLR and ADPG2 (10.29Å for PGLR and 14.46Å for ADPG2, (Šafran et al., 2022)) as compared to AtPGF9, AtPGF12, AtPGF7 and MpPG2 (Figure 8.). Furthermore, all four of these proteins have an α-helix (α5, figure 10.) that is localized between the two loops, closing the binding groove site, and which is also present in the structure of the Exo-PGs from *T. maritima* (Pijning et al., 2009), *B. thetaiotaomicron* (Luis et al., 2018) and *Y. enterocolitica* (Abbott and Boraston, 2007). This α-helix is found right next to a conserved sulfated tyrosine residues in Loop 2 of the active site for AtPGF9, AtPGF12 and MpPG2 (Figure 11., left). It is possible that this sulfatation is of importance in maintaining the structure of these proteins, as tyrosine sulfatation has been shown in flexible parts of proteins, where it enhances intermolecular interactions (Miyanabe et al., 2018). Because of these differences the active site of these four proteins is closed from the N-terminal side which is in line with the enzymes having an exo-acting mechanism, where hydrolysis occurs at the non-reducing end of the HG chain, releasing GalA or GalA2.

LC-MS analyses of the degradation products generated by AtPGF12, AtPGF7 and MpPG2 on various pectic substrates were performed. All three enzymes were tested on polygalacturonic acid substrates while AtPGF12 and MpPG2 were also tested on 35% methylesterified HG substrate. The results for AtPGF12 and AtPGF7 on PGA (Figure 19.) clearly showed the exo-PG activity, as suggested by the structural analysis, as only GalA

was significantly produced following digestion. On the other hand, the results for MpPG2 are not so clear. We noticed that MpPG2 produces GalA as well as smaller amount of GalA3 and GalA4, but no GalA2. This may suggest that MpPG2 has a different mechanism than AtPGF12 and AtPGF7. For example, it may be possible that the substrate positionings that result in the production of GalA3 and GalA4 are less favorable than the one for GalA or that this increase is due to the enzymes completely cutting down the chains with GalA3 and GalA4 being what remains of them. Further research is definitely needed to elucidate MpPG2's mechanism. Lastly, both AtPGF12 and MpPG2 showed increases in the amount of unmethylesterified GalA and no other OG when digesting 35% methyl esterified substrate. It is probable that this increase is due to the enzymes performing HG digestion on various chains until they come across a methylated galacturonic acid residue. However, it is interesting that an increase of concentration was noted only for GalA unlike the digestion in PGA. This means that the appearance of GalA3 and GalA4 with PGA digestion may be associated with prolonged HG digestion which is in line with the results of PGA degradation by MpPG2.

We next calculated the electrostatic potential of the proteins' surfaces (Figure 9.) and observed differences between the four proteins. For instance, when considering the back sides of the proteins, AtPGF7 is much more neutrally charged compared to the three other proteins which are either have a mixed charge (MpPG2) or are positively charged (AtPGF9 and AtPGF12). We can also observe that the isoelectric point of AtPGF9 (pI = 8.8) is about three units higher than the pIs of other proteins. This large difference in the pI may be of importance for protein stability (Alexov, 2004) and could influence the handling of purified proteins. The differences are even more pronounced when looking at the front side and around the active site of the protein (Figure 9.). AtPGF12 is the most positively charged of the four proteins and has a highly positively charged patch in the vicinity of the active site. The patch in AtPGF9 and MpPG2 is less positively charged while in AtPGF7 is mostly neutrally charged (Figure 10., upper panel). If we consider that HG is negatively charged this would indicate that these patches on AtPGF12, AtPGF9 and MpPG2 attract the HG molecule (with attraction being the strongest for AtPGF12), while the patch on AtPGF7 does not. These results coincide with some experiments where AtPGF12 had the highest activity on polygalacturonic acid substrate, followed by AtPGF9 and MpPG2 while AtPGF7 usually showed little to no activity (Figure 16., top). Furthermore, it is possible that this positive patch contributes to the difference in activity we showed earlier on methylated substrate. Considering that methylation increases HG's charge this may reduce the effect this positively charged patch has on substrate binding. We also observed three fully conserved N-glycosylation sites (Figure 11.). These site are found in the loops between ß-sheets of the ß-helix structural motif near the N-terminal ends of the proteins (Figure 11., right) where they may play a role in protein folding because of the sharp turns that are present (Shental-Bechor and Levy, 2008). Lastly, the three proteins from *A. thaliana* have an N-terminal signaling peptide which suggests these proteins are exported from the cytosol and act on pectin as free-floating enzymes in the cell wall. MpPG2, however does not have an N-terminal signaling peptide. Older versions of DeepTMHMM like TMHMM 2.0 and 1.0 showed that MpPG2 has a transmembrane domain upstream of the α0 alpha helix (Figure 10.), which suggests that this protein remains bound to the cell membrane. However, this data is inconclusive and requires further research.

The active sites of most characterized polygalacturonases comprise several residues found in four highly conserved motifs: NTD, DD, RIK and GHG (Palanivelu, 2006). One aspartate residue from the DD motif and the aspartate from the NTD motif are considered to activate an H_2O molecule for nucleophilic attack, while the other aspartate acts as a general acid and is a proton donor to the glycosidic oxygen. (Federici et al., 2001) The four enzymes characterized in this study have the NTD, DD and RIK motifs, while the GHG motif is replaced with an SA motif. One explanation is that the lack of a GHG motif could be required for proper functioning of the protein. As mentioned before, the active sites of AtPGF9, AtPGF12, AtPGF7 and MpPG2 are much narrower than those of endo-PGs and are even narrower than the ones from other crystallized exo-PGs (Abbott and Boraston, 2007; Luis et al., 2018; Pijning et al., 2009), with the pocket containing subsite -1 being especially narrow. (Figure 22.) Considering this, it would be impossible for the cleaved galacturonic acid residue at the reducing end to exit the active site without detachment of the rest of the HG molecule. A cycle of HG binding, catalysis, substrate release, GalA release, and novel substrate binding (Figure 23.) is therefore needed. It is possible that the absence of the GHG motif destabilizes the binding of the HG substrate, improving the substrate release phase of the cycle shown in Figure 23., as the binding of the HG substrate is mostly influenced by -1 and +1 subsites (Mertens et al., 2012). These different hypotheses could be checked for instance by mutating the sequences of these proteins into the GHG motif to see if it has any influence on the proteins' activity and enzyme-substrate interaction. Secondly, it would be of interest to explore whether the serine present in the SA motif plays any role in substrate binding as AT3G62110, a closely related protein from Clade III (Figure 7.), contains three structurally-conserved serine residues. Lastly, it would be of importance to understand how the products of the substrates' hydrolyses can be released from the narrow -1 subsite. An analysis of hydrogen bonding between the substrate and the enzyme at subsites -1 to +3 was performed. (Figure 12.) While these results give insights into the putative mechanisms involved, the crystal structures of proteins will be needed to confirm which of these hydrogen bonds really occur in these proteins while accommodating the substrate.

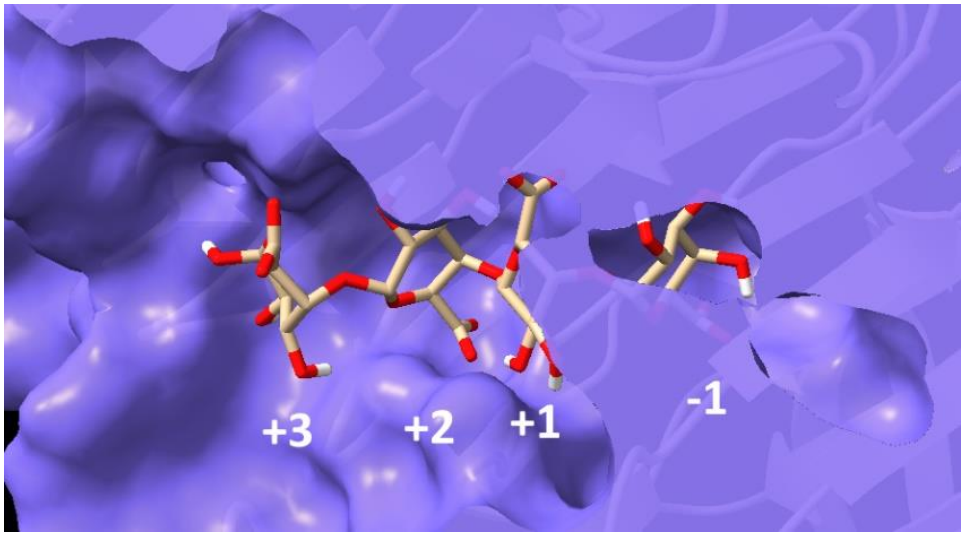

Figure 22. Active site of AtPGF9 with docked tetragalacturonic acid substrate (brown). Subsites are noted

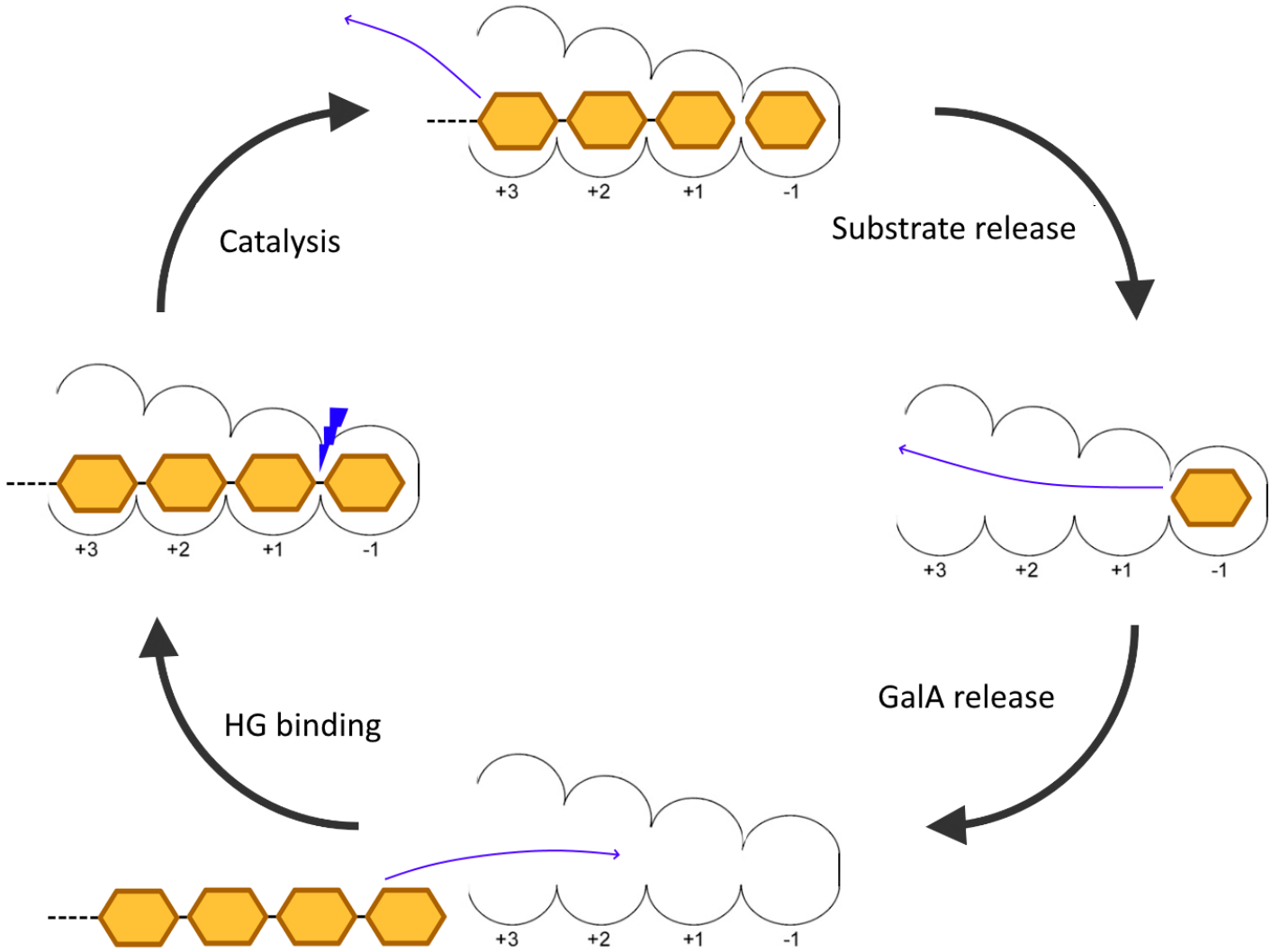

Figure 23. Proposed reaction cycle for AtPGF9, AtPGF12, AtPGF7 and MpPG2. Subsites are noted.

Analysis of the dependance of activity with regards to pH, temperature, substrate and Ca^{2+} concentrations as well as an analysis of the melting temperature were performed to biochemically and biophysically characterize AtPGF9, AtPGF12, AtPGF7 and MpPG2. Despite their structural similarities it was found that these proteins vary substantially in all these parameters. AtPGF9 and AtPGF12 showed the best activity at lower temperatures (40°C) compared to AtPGF7 and MpPG2 (50 and 60°C respectively) (Figure 16., top), which is in agreement with the optimum temperature of most other PGs (Niture, 2008). In line with this, we were able to obtain the melting curve for AtPGF9 showing that protein denaturation starts ~50°C (Figure 18.) Furthermore, it was found that all four proteins have maximum activity for pH of 4.0 to 5.0 (Figure 16., bottom) which is within the range found for other PGs (Šafran, 2021). AtPGF7 is most active at pH 5 which may be due to its unusually high isoelectric point. We also observed that substrate concentration can have an inhibiting effect on protein activity like was already shown for ADPG2 (Pongrac, 2019) and VPG1 (Boras, 2018). Lastly, we showed that the activity of AtPGF9 is slightly improved with the addition of Ca^{2+} in the form of $CaCl_2$. (Figure 17.) This result appears to be in contradiction with previous results showing the inhibiting effects of Ca^{2+} on PGs (Biggs et al., 1997; Buescher and Hobson, 1982; Cabanne and Donèche, 2002; Chun-bin et al., 1990; Conway, 1988), and therefore needs to be confirmed.

In conclusion, this study sought to characterize the activities and structures of three putative exo-polygalacturonases from *A. thaliana* – AtPGF9, AtPGF12, AtPGF7 and one PG from *M. polymorpha* – MpPG2. The results of this study clearly showed that these proteins belong to a closely related clade of polygalacturonases containing sequences from both plants and bacteria. Furthermore, the structures of these proteins showed that they have a narrower active site compared to Endo-PGs and even compared to other Exo-PGs. These proteins also have a conserved structural SA motif which replaces the GHG motif found in other PGs and may have an impact on substrate-enzyme interaction. Biochemical and biophysical characterizations of these proteins demonstrated the the four enzymes are bona fide PG, and confirmed that AtPGF12 and AtPGF7 are indeed exo-acting enzymes. In contrast, MpPG2 may work through a new, yet unidentified mechanism. As *M.* polymorpha has only 10 putative PG genes, it is possible that MpPG2 may act in multiple ways. In this way MpPG2 may compensate by fulfilling the roles of multiple PGs from *A. thaliana* and other plants. Further research should be focused on deepening our understanding on how the small structural differences between these proteins influence their activity, and on extensive additional biochemical characterization of these proteins.

5. - REFERENCES

Abbott, D.W., Boraston, A.B., 2007. The Structural Basis for Exopolygalacturonase Activity in a Family 28 Glycoside Hydrolase. J. Mol. Biol. 368, 1215–1222. https://doi.org/10.1016/j.jmb.2007.02.083

Alexov, E., 2004. Numerical calculations of the pH of maximal protein stability. Eur. J. Biochem. 271, 173–185. https://doi.org/10.1046/j.1432-1033.2003.03917.x

Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., Gascuel, O., 2011. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. Syst. Biol. 60, 685–699. https://doi.org/10.1093/sysbio/syr041

Armand, S., Wagemaker, M.J.M., Sánchez-Torres, P., Kester, H.C.M., van Santen, Y., Dijkstra, B.W., Visser, J., Benen, J.A.E., 2000. The Active Site Topology of Aspergillus nigerEndopolygalacturonase II as Studied by Site-directed Mutagenesis. J. Biol. Chem. 275, 691–696. https://doi.org/10.1074/jbc.275.1.691

Atmodjo, M.A., Hao, Z., Mohnen, D., 2013. Evolving Views of Pectin Biosynthesis. Annu. Rev. Plant Biol. 64, 747–779. https://doi.org/10.1146/annurev-arplant-042811-105534

Bacic, A., Harris, P.J., Stone, B.A., 1988. Structure and function of plant cell walls. Biochem. Plants Compr. Treatise Vol. 14 Carbohydr. 297–372.

Baskin, T.I., 2005. Anisotropic expansion of the plant cell wall. Annu. Rev. Cell Dev. Biol. 21, 203–222. https://doi.org/10.1146/annurev.cellbio.20.082503.103053

Behboudian, M.H., Pickering, A.H., Dayan, E., 2017. Deficiency Diseases, Principles, in: Thomas, B., Murray, B.G., Murphy, D.J. (Eds.), Encyclopedia of Applied Plant Sciences (Second Edition). Academic Press, Oxford, pp. 219–224. https://doi.org/10.1016/B978-0-12-394807-6.00121-0

Biggs, A.R., El-Kholi, M.M., El-Neshawy, S., Nickerson, R., 1997. Effects of Calcium Salts on Growth, Polygalacturonase Activity, and Infection of Peach Fruit by Monilinia fructicola. Plant Dis. 81, 399–403. https://doi.org/10.1094/PDIS.1997.81.4.399

Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S., Hochstrasser, D., 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis 14, 1023–1031. https://doi.org/10.1002/elps.11501401163

Bonnin, E., 2001. Study of the mode of action of endopolygalacturonase from Fusarium moniliforme. Biochim. Biophys. Acta BBA - Gen. Subj. 1526, 301–309. https://doi.org/10.1016/S0304-4165(01)00141-6

Boras, M., 2018. Biochemical characterization of plant polygalacturonases and effects on development.
Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254. https://doi.org/10.1006/abio.1976.9999
Brutus, A., Sicilia, F., Macone, A., Cervone, F., De Lorenzo, G., 2010. A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proc. Natl. Acad. Sci. 107, 9452–9457. https://doi.org/10.1073/pnas.1000675107
Buchanan, B.B., Gruissem, W., Jones, R.L., 2000. Biochemistry & molecular biology of plants. Rockville, Md. : American Society of Plant Physiologists.
Buescher, R.W., Hobson, G.E., 1982. Role of Calcium and Chelating Agents in Regulating the Degradation of Tomato Fruit Tissue by Polygalacturonase. J. Food Biochem. 6, 147–160. https://doi.org/10.1111/j.1745-4514.1982.tb00682.x
Burton, R.A., Gidley, M.J., Fincher, G.B., 2010. Heterogeneity in the chemistry, structure and function of plant cell walls. Nat. Chem. Biol. 6, 724–732. https://doi.org/10.1038/nchembio.439
Cabanne, C., Donèche, B., 2002. Purification and characterization of two isozymes of polygalacturonase from Botrytis cinerea. Effect of calcium ions on polygalacturonase activity. Microbiol. Res. 157, 183–189. https://doi.org/10.1078/0944-5013-00147
Cabrera, J.C., Boland, A., Messiaen, J., Cambier, P., Van Cutsem, P., 2008. Egg box conformation of oligogalacturonides: The time-dependent stabilization of the elicitor-active conformation increases its biological activity. Glycobiology 18, 473–482. https://doi.org/10.1093/glycob/cwn027
Cao, J., 2012. The Pectin Lyases in Arabidopsis thaliana: Evolution, Selection and Expression Profiles. PLoS ONE 7, e46944. https://doi.org/10.1371/journal.pone.0046944
Carpita, N., Tierney, Mary, Campbell, Malcolm, 2001. Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics, in: Carpita, N.C., Campbell, M., Tierney, M. (Eds.), Plant Cell Walls. Springer Netherlands, Dordrecht, pp. 1–5. https://doi.org/10.1007/978-94-010-0668-2_1
Carpita, N.C., Gibeaut, D.M., 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3, 1–30. https://doi.org/10.1111/j.1365-313X.1993.tb00007.x
Chardin, H., Mayer, C., Senechal, H., Poncet, P., Clement, G., Wal, J.M., Desvaux, F.X., Peltre, G., 2003. Polygalacturonase (pectinase), a new oilseed rape allergen. Allergy 58, 407–411. https://doi.org/10.1034/j.1398-9995.2003.00094.x
Chebli, Y., Kaneda, M., Zerzour, R., Geitmann, A., 2012. The Cell Wall of the Arabidopsis Pollen Tube—Spatial Distribution, Recycling, and Network Formation of Polysaccharides. Plant Physiol. 160, 1940–1955. https://doi.org/10.1104/pp.112.199729
Chormova, D., Fry, S.C., 2016. Boron bridging of rhamnogalacturonan-II is promoted in vitro by cationic chaperones, including polyhistidine and wall glycoproteins. New Phytol. 209, 241–251. https://doi.org/10.1111/nph.13596
Chun-bin, L., Cun-de, L., Hou-guo, S.Q. and L., 1990. Effect of the Calcium on Polygalacturonase (PG) Activity and Synthesis at Different Ripening Stages of Tomato Fruits. J. Integr. Plant Biol. 32.
Conway, W.S., 1988. Inhibition of *Penicillium expansum* Polygalacturonase Activity by Increased Apple Cell Wall Calcium. Phytopathology 78, 1052. https://doi.org/10.1094/Phyto-78-1052
Daher, F.B., Braybrook, S.A., 2015. How to let go: pectin and plant cell adhesion. Front. Plant Sci. 6.
Du, J., Kirui, A., Huang, S., Wang, L., Barnes, W.J., Kiemle, S.N., Zheng, Y., Rui, Y., Ruan, M., Qi, S., Kim, S.H., Wang, T., Cosgrove, D.J., Anderson, C.T., Xiao, C., 2020. Mutations in the Pectin Methyltransferase QUASIMODO2 Influence Cellulose Biosynthesis and Wall Integrity in Arabidopsis. Plant Cell 32, 3576–3597. https://doi.org/10.1105/tpc.20.00252
Duong-Ly, K.C., Gabelli, S.B., 2014. Salting out of proteins using ammonium sulfate precipitation. Methods Enzymol. 541, 85–94. https://doi.org/10.1016/B978-0-12-420119-4.00007-0
Engle, K.A., Amos, R.A., Yang, J.-Y., Glushka, J., Atmodjo, M.A., Tan, L., Huang, C., Moremen, K.W., Mohnen, D., 2022. Multiple Arabidopsis galacturonosyltransferases synthesize polymeric homogalacturonan by oligosaccharide acceptor-dependent or de novo synthesis. Plant J. 109, 1441–1456. https://doi.org/10.1111/tpj.15640
Federici, L., Caprari, C., Mattei, B., Savino, C., Di Matteo, A., De Lorenzo, G., Cervone, F., Tsernoglou, D., 2001. Structural requirements of endopolygalacturonase for the interaction with PGIP

(polygalacturonase-inhibiting protein). Proc. Natl. Acad. Sci. U. S. A. 98, 13425–13430. https://doi.org/10.1073/pnas.231473698
Francis, K.E., Lam, S.Y., Copenhaver, G.P., 2006. Separation of Arabidopsis pollen tetrads is regulated by QUARTET1, a pectin methylesterase gene. Plant Physiol. 142, 1004–1013. https://doi.org/10.1104/pp.106.085274
Girard, C., Jouanin, L., 1999. Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, Phaedon cochleariae. Insect Biochem. Mol. Biol. 29, 1129–1142. https://doi.org/10.1016/S0965-1748(99)00104-6
Gupta, R., Brunak, S., 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput. Pac. Symp. Biocomput. 310–322.
Haas, K.T., Wightman, R., Meyerowitz, E.M., Peaucelle, A., 2020. Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells. Science 367, 1003–1007. https://doi.org/10.1126/science.aaz5103
Hadfield, K.A., Bennett, A.B., 1998. Polygalacturonases: Many Genes in Search of a Function. Plant Physiol. 117, 337–343.
Hallgren, J., Tsirigos, K.D., Pedersen, M.D., Armenteros, J.J.A., Marcatili, P., Nielsen, H., Krogh, A., Winther, O., 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. https://doi.org/10.1101/2022.04.08.487609
Hayashi, T., Kaida, R., 2011. Functions of Xyloglucan in Plant Cells. Mol. Plant 4, 17–24. https://doi.org/10.1093/mp/ssq063
Holland, C., Ryden, P., Edwards, C.H., Grundy, M.M.-L., 2020. Plant Cell Walls: Impact on Nutrient Bioaccessibility and Digestibility. Foods 9, 201. https://doi.org/10.3390/foods9020201
Ibarrola, I., Arilla, M.C., Martínez, A., Asturias, J.A., 2004. Identification of a polygalacturonase as a major allergen (Pla a 2) from Platanus acerifolia pollen. J. Allergy Clin. Immunol. 113, 1185–1191. https://doi.org/10.1016/j.jaci.2004.02.031
Ishii, T., Kaneko, S., 1998. Oligosaccharides generated by partial hydrolysis of the borate-rhamnogalacturonan II complex from sugar beet. Phytochemistry 49, 1195–1202. https://doi.org/10.1016/s0031-9422(98)00119-8
Jia, Y.J., Feng, B.Z., Sun, W.X., Zhang, X.G., 2009. Polygalacturonase, Pectate Lyase and Pectin Methylesterase Activity in Pathogenic Strains of *Phytophthora capsici* Incubated under Different Conditions. J. Phytopathol. 157, 585–591. https://doi.org/10.1111/j.1439-0434.2008.01533.x
Jiang, L., Yang, S.-L., Xie, L.-F., Puah, C.S., Zhang, X.-Q., Yang, W.-C., Sundaresan, V., Ye, D., 2005. VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract. Plant Cell 17, 584–596. https://doi.org/10.1105/tpc.104.027631
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2
Ke, X., Wang, H., Li, Y., Zhu, B., Zang, Y., He, Y., Cao, J., Zhu, Z., Yu, Y., 2018. Genome-Wide Identification and Analysis of Polygalacturonase Genes in Solanum lycopersicum. Int. J. Mol. Sci. 19, 2290. https://doi.org/10.3390/ijms19082290
Kiemle, S.N., Zhang, X., Esker, A.R., Toriz, G., Gatenholm, P., Cosgrove, D.J., 2014. Role of (1,3)(1,4)-β-Glucan in Cell Walls: Interaction with Cellulose. Biomacromolecules 15, 1727–1736. https://doi.org/10.1021/bm5001247
Kim, J., Shiu, S.-H., Thoma, S., Li, W.-H., Patterson, S.E., 2006. Patterns of expansion and expression divergence in the plant polygalacturonase gene family. Genome Biol. 7, R87. https://doi.org/10.1186/gb-2006-7-9-r87
Loix, C., Huybrechts, M., Vangronsveld, J., Gielen, M., Keunen, E., Cuypers, A., 2017. Reciprocal Interactions between Cadmium-Induced Cell Wall Responses and Oxidative Stress in Plants. Front. Plant Sci. 8. https://doi.org/10.3389/fpls.2017.01867
Lu, L., Hou, Q., Wang, L., Zhang, T., Zhao, W., Yan, T., Zhao, L., Li, J., Wan, X., 2021. Genome-Wide Identification and Characterization of Polygalacturonase Gene Family in Maize (Zea mays L.). Int. J. Mol. Sci. 22, 10722. https://doi.org/10.3390/ijms221910722

Luis, A.S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé, A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E.C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.-C., Henrissat, B., Martens, E.C., Mosimann, S.C., Abbott, D.W., Gilbert, H.J., 2018. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. Nat. Microbiol. 3, 210–219. https://doi.org/10.1038/s41564-017-0079-1

Markovič, O., Janeček, Š., 2001. Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. Protein Eng. Des. Sel. 14, 615–631. https://doi.org/10.1093/protein/14.9.615

Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., Jenkins, J., 1997. Two crystal structures of pectin lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. Structure 5, 677–689. https://doi.org/10.1016/S0969-2126(97)00222-0

Mertens, J.A., Hector, R.E., Bowman, M.J., 2012. Subsite binding energies of an exo-polygalacturonase using isothermal titration calorimetry. Thermochim. Acta 527, 219–222. https://doi.org/10.1016/j.tca.2011.10.029

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Miyanabe, K., Yamashita, T., Abe, Y., Akiba, H., Takamatsu, Y., Nakakido, M., Hamakubo, T., Ueda, T., Caaveiro, J.M.M., Tsumoto, K., 2018. Tyrosine Sulfation Restricts the Conformational Ensemble of a Flexible Peptide, Strengthening the Binding Affinity for an Antibody. Biochemistry 57, 4177–4185. https://doi.org/10.1021/acs.biochem.8b00592

Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine sulfation sites in protein sequences. Bioinforma. Oxf. Engl. 18, 769–770. https://doi.org/10.1093/bioinformatics/18.5.769

Moreira, L.R.S., Filho, E.X.F., 2008. An overview of mannan structure and mannan-degrading enzyme systems. Appl. Microbiol. Biotechnol. 79, 165–178. https://doi.org/10.1007/s00253-008-1423-4

Nakamura, M., Iwai, H., 2019. Functions and mechanisms: polygalacturonases from plant pathogenic fungi as pathogenicity and virulence factors. J. Gen. Plant Pathol. 85, 243–250. https://doi.org/10.1007/s10327-019-00856-8

Narsimha Rao, M., Kembhavi, A.A., Pant, A., 1996. Implication of tryptophan and histidine in the active site of endo-polygalacturonase from Aspergillus ustus: elucidation of the reaction mechanism. Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol. 1296, 167–173. https://doi.org/10.1016/0167-4838(96)00067-2

Niesen, F.H., Berglund, H., Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. 2, 2212–2221. https://doi.org/10.1038/nprot.2007.321

Niture, S.K., 2008. Comparative biochemical and structural characterizations of fungal polygalacturonases. Biologia (Bratisl.) 63, 1–19. https://doi.org/10.2478/s11756-008-0018-y

Nogota, Y., Ohta, H., Voragen, A.G.J., 1993. Polygalacturonase in strawberry fruit. Phytochemistry 34, 617–620. https://doi.org/10.1016/0031-9422(93)85327-n

Pages, S., Heijne, W.H.M., Kester, H.C.M., Visser, J., Benen, J. a. E., 2000. Subsite Mapping of Aspergillus niger Endopolygalacturonase II by Site-directed Mutagenesis. J. Biol. Chem. 275, 29348–29353. https://doi.org/10.1074/jbc.M910112199

Palanivelu, P., 2006. Polygalacturonases: Active site analyses and mechanism of action. INDIAN J BIOTECHNOL 15.

Park, K.-C., Kwon, S.-J., Kim, N.-S., 2010a. Intron loss mediated structural dynamics and functional differentiation of the polygalacturonase gene family in land plants. Genes Genomics 32, 570–577. https://doi.org/10.1007/s13258-010-0076-8

Park, K.-C., Kwon, S.-J., Kim, N.-S., 2010b. Intron loss mediated structural dynamics and functional differentiation of the polygalacturonase gene family in land plants. Genes Genomics 32, 570–577. https://doi.org/10.1007/s13258-010-0076-8

Peaucelle, A., Braybrook, S.A., Le Guillou, L., Bron, E., Kuhlemeier, C., Höfte, H., 2011. Pectin-Induced Changes in Cell Wall Mechanics Underlie Organ Initiation in Arabidopsis. Curr. Biol. 21, 1720–1726. https://doi.org/10.1016/j.cub.2011.08.057

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. Publ. Protein Soc. 30, 70–82. https://doi.org/10.1002/pro.3943

Pijning, T., van Pouderoyen, G., Kluskens, L., van der Oost, J., Dijkstra, B.W., 2009. The crystal structure of a hyperthermoactive exopolygalacturonase from Thermotoga maritima reveals a unique tetramer. FEBS Lett. 583, 3665–3670. https://doi.org/10.1016/j.febslet.2009.10.047

Pongrac, P., 2019. Impact of polygalacturonases on cell wall remodeling and plant development (info:eu-repo/semantics/masterThesis). University of Zagreb. Faculty of Food Technology and Biotechnology. Department of Chemistry and Biochemistry. Laboratory for Biochemistry.

Pressey, R., Avants, J.K., 1973. Separation and Characterization of Endopolygalacturonase and Exopolygalacturonase from Peaches. Plant Physiol. 52, 252–256.

Prince, V.E., Pickett, F.B., 2002. Splitting pairs: the diverging fates of duplicated genes. Nat. Rev. Genet. 3, 827–837. https://doi.org/10.1038/nrg928

Qi, J., Wu, B., Feng, S., Lü, S., Guan, C., Zhang, X., Qiu, D., Hu, Y., Zhou, Y., Li, C., Long, M., Jiao, Y., 2017. Mechanical regulation of organ asymmetry in leaves. Nat. Plants 3, 724–733. https://doi.org/10.1038/s41477-017-0008-6

Rexová-Benková, L., 1990. Evidence for the role of carboxyl groups in activity of endopolygalacturonase of *Aspergillus niger*. Chemical modification by carbodiimide reagent. Collect. Czechoslov. Chem. Commun. 55, 1389–1395. https://doi.org/10.1135/cccc19901389

Reyes, F., Orellana, A., 2008. Golgi transporters: opening the gate to cell wall polysaccharide biosynthesis. Curr. Opin. Plant Biol. 11, 244–251. https://doi.org/10.1016/j.pbi.2008.03.008

Ropartz, D., Ralet, M.-C., 2020. Pectin Structure, in: Kontogiorgos, V. (Ed.), Pectin: Technological and Physiological Properties. Springer International Publishing, Cham, pp. 17–36. https://doi.org/10.1007/978-3-030-53421-9_2

Šafran, J., 2021. Characterization of pectin remodelling enzymes from Arabidopsis thaliana and Verticillium dahliae: from protein structure to processivity. Université de Picardie Jules Verne.

Šafran, J., Tabi, W., Ung, V., Lemaire, A., Habrylo, O., Bouckaert, J., Rouffle, M., Voxeur, A., Pongrac, P., Bassard, S., Molinié, R., Fontaine, J.-X., Pilard, S., Pau-Roblot, C., Bonnin, E., Larsen, D.S., Morel-Rouhier, M., Girardet, J.-M., Lefebvre, V., Sénéchal, F., Mercadante, D., Pelloux, J., 2022. Differences in the structure of plant polygalacturonases specify enzymes' dynamics and processivities to fine-tune pectins and root development. https://doi.org/10.1101/2022.06.22.497136

Sarkar, M.B., Sircar, G., Ghosh, N., Das, A.K., Jana, K., Dasgupta, A., Bhattacharya, S.G., 2018. Cari p 1, a Novel Polygalacturonase Allergen From Papaya Acting as Respiratory and Food Sensitizer. Front. Plant Sci. 9.

Scheller, H.V., Ulvskov, P., 2010. Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315

Schols, H.A., Bakx, E.J., Schipper, D., Voragen, A.G.J., 1995. A xylogalacturonan subunit present in the modified hairy regions of apple pectin. Carbohydr. Res. 279, 265–279. https://doi.org/10.1016/0008-6215(95)00287-1

Sénéchal, F., Wattier, C., Rustérucci, C., Pelloux, J., 2014. Homogalacturonan-modifying enzymes: structure, expression, and roles in plants. J. Exp. Bot. 65, 5125–5160. https://doi.org/10.1093/jxb/eru272

Shen, Z., Denton, M., Mutti, N., Pappan, K., Kanost, M.R., Reese, J.C., Reeck, G.R., 2003. Polygalacturonase from Sitophilus oryzae: Possible horizontal transfer of a pectinase gene from fungi to weevils. J. Insect Sci. 3, 1–9. https://doi.org/10.1673/031.003.2401

Shental-Bechor, D., Levy, Y., 2008. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. Proc. Natl. Acad. Sci. 105, 8256–8261. https://doi.org/10.1073/pnas.0801340105

Sitrit, Y., Hadfield, K.A., Bennett, A.B., Bradford, K.J., Downie, A.B., 1999. Expression of a polygalacturonase associated with tomato seed germination. Plant Physiol. 121, 419–428. https://doi.org/10.1104/pp.121.2.419

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., Vorwerk, S., Youngs, H., 2004. Toward a Systems Approach to Understanding Plant Cell Walls. Science 306, 2206–2211. https://doi.org/10.1126/science.1102765

Strong, F.E., Kruitwagen, E.C., 1968. Polygalacturonase in the salivary apparatus of Lygus hesperus (Hemiptera). J. Insect Physiol. 14, 1113–1119. https://doi.org/10.1016/0022-1910(68)90050-4

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. 38, 3022–3027. https://doi.org/10.1093/molbev/msab120

Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., Hao, Z., Zhu, X., Avci, U., Miller, J.S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M.G., Kieliszewski, M.J., Mohnen, D., 2013. An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. Plant Cell 25, 270–287. https://doi.org/10.1105/tpc.112.107334

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 1–3. https://doi.org/10.1038/s41587-021-01156-3

Vanholme, R., De Meester, B., Ralph, J., Boerjan, W., 2019. Lignin biosynthesis and its integration into metabolism. Curr. Opin. Biotechnol. 56, 230–239. https://doi.org/10.1016/j.copbio.2019.02.018

Vasco-Correa, J., Zapata Zapata, A.D., 2017. Enzymatic extraction of pectin from passion fruit peel (Passiflora edulis f. flavicarpa) at laboratory and bench scale. LWT 80, 280–285. https://doi.org/10.1016/j.lwt.2017.02.024

Vision, T.J., Brown, D.G., Tanksley, S.D., 2000. The Origins of Genomic Duplications in *Arabidopsis*. Science 290, 2114–2117. https://doi.org/10.1126/science.290.5499.2114

Voragen, A.G.J., Coenen, G.-J., Verhoef, R.P., Schols, H.A., 2009. Pectin, a versatile polysaccharide present in plant cell walls. Struct. Chem. 20, 263–275. https://doi.org/10.1007/s11224-009-9442-z

Voragen, A.G.J., Pilnik, W., Thibault, J.F., Axelos, M. a. V., Renard, C.M.G.C., 1995. 10 Pectins. Food Polysacch. Their Appl. 287–339.

Wiedemeier, A.M.D., Judy-March, J.E., Hocart, C.H., Wasteneys, G.O., Williamson, R.E., Baskin, T.I., 2002. Mutant alleles of Arabidopsis RADIALLY SWOLLEN 4 and 7 reduce growth anisotropy without altering the transverse orientation of cortical microtubules or cellulose microfibrils. Dev. Camb. Engl. 129, 4821–4830. https://doi.org/10.1242/dev.129.20.4821

Xiao, C., Barnes, W.J., Zamil, M.S., Yi, H., Puri, V.M., Anderson, C.T., 2017. Activation tagging of Arabidopsis POLYGALACTURONASE INVOLVED IN EXPANSION2 promotes hypocotyl elongation, leaf expansion, stem lignification, mechanical stiffening, and lodging. Plant J. Cell Mol. Biol. 89, 1159–1173. https://doi.org/10.1111/tpj.13453

Xiao, C., Somerville, C., Anderson, C.T., 2014. POLYGALACTURONASE INVOLVED IN EXPANSION1 Functions in Cell Elongation and Flower Development in Arabidopsis[C][W]. Plant Cell 26, 1018–1035. https://doi.org/10.1105/tpc.114.123968

Yang, Y., Yu, Y., Liang, Y., Anderson, C.T., Cao, J., 2018. A Profusion of Molecular Scissors for Pectins: Classification, Expression, and Functions of Plant Polygalacturonases. Front. Plant Sci. 9, 1208. https://doi.org/10.3389/fpls.2018.01208

Yapo, B.M., 2011. Rhamnogalacturonan-I: A Structurally Puzzling and Functionally Versatile Polysaccharide from Plant Cell Walls and Mucilages. Polym. Rev. 51, 391–413. https://doi.org/10.1080/15583724.2011.615962

Yoder, M.D., Keen, N.T., Jurnak, F., 1993. New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. Science 260, 1503–1507. https://doi.org/10.1126/science.8502994

Zamil, M.S., Geitmann, A., 2017. The middle lamella-more than a glue. Phys. Biol. 14, 015004. https://doi.org/10.1088/1478-3975/aa5ba5

Zdunek, A., Pieczywek, P.M., Cybulska, J., 2021. The primary, secondary, and structures of higher levels of pectin polysaccharides. Compr. Rev. Food Sci. Food Saf. 20, 1101–1117. https://doi.org/10.1111/1541-4337.12689

Zeng, Y., Himmel, M.E., Ding, S.-Y., 2017. Visualizing chemical functionality in plant cell walls. Biotechnol. Biofuels 10, 263. https://doi.org/10.1186/s13068-017-0953-3

Zhang, B., Gao, Y., Zhang, L., Zhou, Y., 2021. The plant cell wall: Biosynthesis, construction, and functions. J. Integr. Plant Biol. 63, 251–272. https://doi.org/10.1111/jipb.13055

Zhong, R., Cui, D., Ye, Z.-H., 2018. Members of the DUF231 Family are O-Acetyltransferases Catalyzing 2-O- and 3-O-Acetylation of Mannan. Plant Cell Physiol. https://doi.org/10.1093/pcp/pcy159

Zhong, R., Ye, Z.-H., 2015. Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation. Plant Cell Physiol. 56, 195–214. https://doi.org/10.1093/pcp/pcu140
Zhu, R., Wang, C., Zhang, L., Wang, Y., Chen, G., Fan, J., Jia, Y., Yan, F., Ning, C., 2019. Pectin oligosaccharides from fruit of Actinidia arguta: Structure-activity relationship of prebiotic and antiglycation potentials. Carbohydr. Polym. 217, 90–97. https://doi.org/10.1016/j.carbpol.2019.04.032

ABSTRACT

Pectins are one of the main polysaccharide components of plants' primary cell walls. They are made of several highly heterogenous domains that vary in monosaccharide content and have been found to contribute to many developmental processes. Homogalacturonan (HG) is the simplest pectin domain and is made of unbranched linear polymers of α-(1-4) linked d-galacturonic acid (GalA) which can be methylesterifed and acetylesterified. Polygalacturonases are enzymes which cleave the homogalacturonan chains. They are divided into Endo-polygalacturonases which cleave in the middle of the chains and Exo-polygalacturonases (Exo-PG) which cleave single GalA residues from the non-reducing end of HG. In this internship we sought to characterize the activities and structures of four closely related putative Exo-PGs: AtPGF9, AtPGF12 and AtPGF7 form *Arabidopsis thaliana* and MpPG2 from *Marchantia polymorpha*. Firstly, the proteins' structures were modelled using Alphafold 2.0 and the differences within the group and among other crystallized PGs were determined. Secondly, the proteins were expressed in a heterologous *Pichia pastoris* system after which their biochemical properties such as optimal pH and temperature, substrate inhibition, calcium dependency and oligogalacturonide production were examined. Our results showed that the four enzymes possess a conserved SA structural motif which replaces a GHG motif found in all other PGs, an unusually narrow active, and that they differ in the amount of possible hydrogen bonds and the surface charge amongst themselves. Furthermore, the proteins have differences in their biochemical characteristics and OG production. AtPGF12 and AtPGF7 were confirmed to act on polygalacturonic acid as Exo-PGs, while MpPG2 showed production of GalA as well as smaller amount of GalA3 and GalA4. To conclude, these four enzymes showed biochemical differences despite their similar structures and close phylogenetic relationship thereby proving that fine-grained analysis of PGs' features is needed to understand their influence in plant development.

Kewyords: *Arabidopsis thaliana*, *Marchantia polymorpha*, homogalacturonan, oligogalacturonides, pectins, polygalacturonase, Alphafold, protein structures

RÉSUMÉ

Les pectines sont l'un des principaux composants polysaccharidiques des parois cellulaires primaires des plantes. Elles sont constituées de plusieurs domaines très hétérogènes dont la teneur en monosaccharides varie et dont on a découvert qu'ils contribuent à de nombreux processus de développement. L'homogalacturonane (HG) est le domaine pectinique le plus simple et est constitué de polymères linéaires non ramifiés de α-(1-4) lié acide d-galacturonique (GalA) qui peut être méthylestérifié et acétylestérifié. Les polygalacturonases sont des enzymes qui clivent les chaînes d'homogalacturonane. Elles se divisent en Endo-polygalacturonases qui clivent au milieu des chaînes et en Exo-polygalacturonases (Exo-PG) qui clivent les résidus GalA uniques de l'extrémité non réductrice de l'HG. Dans ce stage, nous avons cherché à caractériser les activités et les structures de quatre Exo-PGs putatifs étroitement liés : AtPGF9, AtPGF12 et AtPGF7 d'*Arabidopsis thaliana* et MpPG2 de *Marchantia polymorpha*. Premièrement, les structures des protéines ont été modélisées à l'aide d'Alphafold 2.0 et les différences au sein du groupe et parmi d'autres PGs cristallisées ont été déterminées. Ensuite, les protéines ont été exprimées dans un système hétérologue *Pichia pastoris*, après quoi leurs propriétés biochimiques telles que le pH et la température optimaux, l'inhibition du substrat, la dépendance au calcium et la production d'oligogalacturonides ont été examinées. Nos résultats ont montré que les quatre enzymes possèdent un motif structural SA conservé qui remplace un motif GHG trouvé dans toutes les autres PGs, un site actif inhabituellement étroit, et qu'elles diffèrent dans la quantité de liaisons hydrogène possibles et la charge de surface entre elles. En outre, les protéines présentent des différences dans leurs caractéristiques biochimiques et dans la production d'OG. Il a été confirmé qu'AtPGF12 et AtPGF7 agissent sur l'acide polygalacturonique en tant qu'Exo-PGs, tandis que MpPG2 a montré la production de GalA ainsi qu'une plus petite quantité de GalA3 et GalA4 . En conclusion, ces quatre enzymes ont montré des différences biochimiques malgré leurs structures similaires et leur étroite relation phylogénétique, ce qui prouve qu'une analyse fine des caractéristiques des PGs est nécessaire pour comprendre leur influence sur le développement des plantes.

Mots clés : *Arabidopsis thaliana*, *Marchantia polymorpha*, homogalacturonane, oligogalacturonides, pectines, polygalacturonase, Alphafold, structures des protéines.