

SHORT REPORT

Open Access

In silico analysis of the polygalacturonase inhibiting protein 1 from apple, *Malus domestica*

Lerato BT Matsaunyane^{1,2*}, Dean Oelofse¹ and Ian A Dubery^{2*}

Abstract

Background: The *Malus domestica* polygalacturonase inhibiting protein 1 (*MdPGIP1*) gene, encoding the *M. domestica* polygalacturonase inhibiting protein 1 (*MdPGIP1*), was isolated from the Granny Smith apple cultivar (GenBank accession no. DQ185063). The gene was used to transform tobacco and potato for enhanced resistance against fungal diseases.

Findings: Analysis of the *MdPGIP1* nucleotide sequence revealed that the gene comprises 993 nucleotides that encode a 330 amino acid polypeptide. *In silico* characterization of the *MdPGIP1* polypeptide revealed domains typical of PGIP proteins, which include a 24 amino acid putative signal peptide, a potential cleavage site [Alanine-Leucine-Serine (ALS)] for the signal peptide, a 238 amino acid leucine-rich repeat (LRR) domain, a 46 amino acid N-terminal domain and a 22 amino acid C-terminal domain. The hydropathic evaluation of *MdPGIP1* indicated a repetitive hydrophobic motif in the LRR domain and a hydrophilic surface area consistent with a globular protein. The typical consensus glycosylation sequence of Asn-X-Ser/Thr was identified in *MdPGIP1*, indicating potential N-linked glycosylation of *MdPGIP1*. The molecular mass of non-glycosylated *MdPGIP1* was calculated as 36.615 kDa and the theoretical isoelectric point as 6.98. Furthermore, the secondary and tertiary structure of *MdPGIP1* was modelled, and revealed that *MdPGIP1* is a curved and elongated molecule that contains sheet B1, sheet B2 and 3₁₀-helices on its LRR domain.

Conclusion: The overall properties of the *MdPGIP1* protein is similar to that of the prototypical *Phaseolus vulgaris* PGIP 2 (*PvPGIP2*), and the detected differences supported its use in biotechnological applications as an inhibitor of targeted fungal polygalacturonases (PGs).

Keywords: Disease resistance, PGIP, Polygalacturonase, Inhibitor, Structure

Background

Polygalacturonase inhibiting proteins (PGIPs) are part of the innate immune system of plants. PGIPs may specifically inhibit fungal polygalacturonases (PGs) [1-3]. The action of PGIPs on PGs during fungal attack slows down the infection rate and facilitates the prolonged existence of mid-sized oligogalacturonides (damage-associated molecular pattern molecules, DAMPs), which in turn can elicit a general defence response from the plant [4-6]. PGIPs are not specialized inhibitors of a single PG, but rather versatile proteins that capable of recognising different surface motifs of structurally variable PGs [6]. Knowledge about the structural properties of PGIPs can

provide valuable insight into the nature and dynamics of these interactions with fungal PGs, and also helps to identify promising candidate PGIPs for biotechnological approaches to improve plant disease resistance.

Yao *et al.* [7] isolated mRNA containing the complete coding sequence region of the *MdPGIP* gene from cDNA of *M. domestica* cv Golden Delicious [GenBank: MDU77041]. Similarly, Arendse *et al.* [8] isolated the complete *MdPGIP1* gene from gDNA of *M. domestica* cv Granny Smith [GenBank:DQ185063]. The sequence of the *MdPGIP1* accession DQ185063 was compared to the *MdPGIP* accession MDU77041 and the results showed that the two gene sequences share a 100% identity. The *MdPGIP1* gene sequence elicited interest in its potential use as an anti-fungal agent and was subsequently used to transfer into potato [9] and tobacco [10]. *MdPGIP1* inhibits PGs from *Botryosphaeria obtuse* and *Diaporthe ambigua*, which are both apple pathogens. In

* Correspondence: lmatsaunyane@arc.agric.za; idubery@arc.agric.za

¹Agricultural Research Council - Vegetable and Ornamental Plant Institute (ARC-VOPI), Roodeplaat, Pretoria, South Africa

²Department of Biochemistry, University of Johannesburg, P.O. Box 524, Auckland Park 2006, South Africa

addition, further studies performed to date (Matsaunyane and Oelofse, unpublished) indicate that the protein also inhibits PGs from *Verticillium dahlia*, *Botrytis cinerea*, *Colletotrichum acutatum* and *Colletotrichum coccodes*, but not that of *Fusarium verticillioides*. To further explore the biochemical characteristics of *MdPGIP1* as a potential tool in improving disease resistance of food crops, *in silico* analyses were performed to compare the properties of *MdPGIP1* to other characterised PGIPs. This further characterization forms part of new information on the *MdPGIP1* encoded protein.

Methods

The *MdPGIP1* gene sequence with the GenBank accession [DQ185063] was used during *in silico* analysis in this study. The nucleotide sequence was translated into the encoding polypeptide using the <http://web.expasy.org/translate/> database. The amino acid composition of the *MdPGIP1* protein was calculated using the <http://www.biology.arizona.edu/biochemistry/biochemistry.html> database. In addition, the polarity, functional group side chains and their respective charge, and the amino acids' water propensity, were also calculated on this database. The hydrophobicity plot of the *MdPGIP1* contiguous amino acid residues was determined by constructing the Kyte-Doolittle hydropathy graph (<http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm>) [11]. The molecular weight of *MdPGIP1* was determined by compiling a ProtScale of the polypeptide using its respective constituting residues (<http://web.expasy.org/cgi-bin/protscale/protscale.pl>). Software from the NetNGlyc 1.0 Server (<http://genome.cbs.dtu.dk/services/>) was used to analyse the possible N-linked glycosylation sites of *MdPGIP1*. The putative crystal structure of *MdPGIP1* was modelled using SWISS-MODEL [12], a protein structure

homology-modeling server, accessible via the ExpASy web server (www.Swissmodel.expasy.org).

Findings

Amino acid composition and primary structure of *MdPGIP1* supports its folded structure and function

Although subject to modification, the linear sequence of amino acids, as represented by the primary structure of a protein, holds the required information for protein folding, the biological and cellular processes and activities of proteins. The amino acid residues that comprise the *MdPGIP1* polypeptide were therefore analysed, based on the translated nucleotide sequence [13]. The primary structure of *MdPGIP1* with its identified domains is shown in Figure 1. Features include a putative 24 amino acid signal peptide (Figure 1A), a potential cleavage site [Alanine-Leucine-Serine (ALS)] for the signal peptide (Figure 1, ALS in pink), a 46 amino acid N-terminal domain (Figure 1B), a 238 amino acid leucine-rich repeat (LRR) domain (Figure 1C) and a 22 amino acid C-terminal domain (Figure 1D).

The amino acid composition of the polypeptide, the polarity of the amino acids, type of side chain found in their respective functional (R) group, charge of the amino acids, as well as their water propensity was also calculated (<http://www.biology.arizona.edu/biochemistry/biochemistry.html>). The water propensity of amino acids can be used to indicate, *in silico*, the location of the respective amino acid in the final structure of a protein during folding [11]. The interior of a globular protein normally houses hydrophobic residues, whereas the outer side is a location for hydrophilic residues. The Kyte-Doolittle hydropathy graph was subsequently constructed (Figure 2) to obtain further insights into the effect of the different water propensities of the *MdPGIP1* residues on

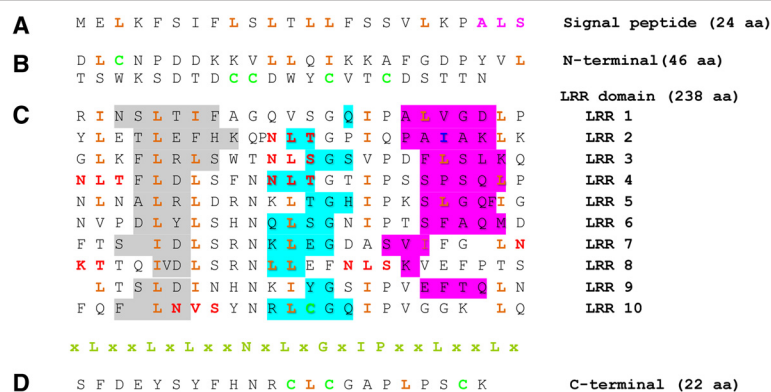


Figure 1 The translated primary structure of the *Malus domestica* polygalacturonase inhibiting protein 1 (*MdPGIP1*). Single letter codes are used to present amino acids. **A:** signal peptide, **B:** N-terminal domain, **C:** leucine-rich repeat LRR domain, and **D:** C-terminal domain. Hydrophobic amino acids leucine (L) and isoleucine (I) are highlighted in orange and the LRR consensus sequence is highlighted in lime. The N-glycosylation sites (N-X-S/T) are highlighted in red, with the cleavage site (A-L-S) highlighted in pink. N- and C-terminal cysteine residues are highlighted in sea green. Sheet-B1, sheet-B2 and 3_{10} -helix are highlighted with grey, blue and pink background colours, respectively.

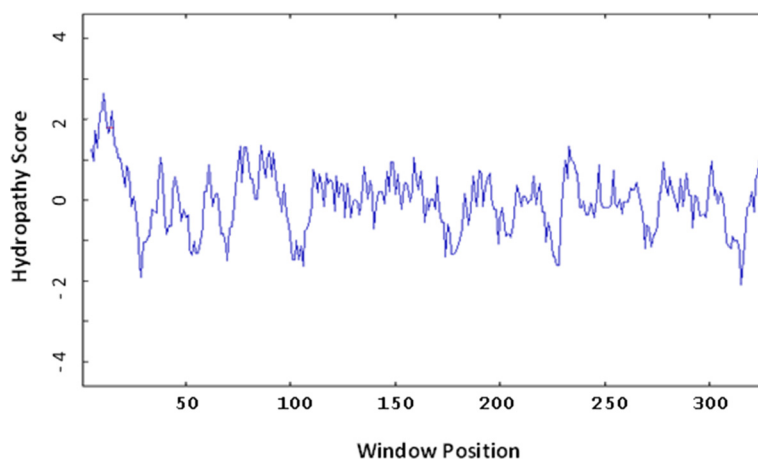


Figure 2 A hydrophathy plot constructed for the *Malus domestica* polygalacturonase inhibiting protein 1 (*MdPGIP1*) polypeptide. The number of amino acids, of the total 330, that were effective in the construction of the hydrophathy plot was 322 with a window size of 9.

the structure of the protein (<http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm>, [11]). The total number of amino acids, of the total 330, that were effective in the construction of the hydrophathy plot was 322. The window size of the *MdPGIP1* hydrophathy plot was 9 and strong negative peaks (indicative of hydrophilic areas) were observed on the plot. This is an indication of a possible surface area of a globular protein. The hydrophathy plot also indicated the existence of a repetitive region between residues 71 and 300 on the window number (x-axis), and by the strong positive (hydrophobic) peaks on the hydrophobicity score (y-axis) (Figure 2).

PGIPs are known to be peripheral membrane-associated proteins secreted to the apoplast [14]. The hydrophathy plot was used to identify portions of *MdPGIP1* that could be associated with the membrane. The most hydrophobic residues served as a basis for this identification and these residues were found from residue 9 to 21 on the protein [11]. The residues found at this location on *MdPGIP1* are Met, Glu, Leu, Lys, Phe, Ser, Ile, Phe, Leu, Ser, Leu, Thr, Leu, Leu, Phe, Ser, Ser, Val, Leu, Lys, Pro, Ala, Leu and Ser (Figure 1). The relative hydrophobicity of this portion is confirmed in Figure 2, where the hydrophathy score is the highest throughout the span of the *MdPGIP1* polypeptide length at a value between one and two. This hydrophathy score is thought to be a deciding factor for this portion to be membrane associated, thus confirming this aspect for *MdPGIP1*, similar to other PGIPs [15].

***MdPGIP1* belongs to the Leucine Rich Repeat superfamily**

The structural features of *MdPGIP1* were found to be consistent with typical PGIP features described by other authors [16-19]. Structural studies of PGIP proteins are important for it is known that the change of one or a few residues may confer new PGs recognition

specificities to a PGIP and may improve its inhibitory strength [6,20]. The recognition specificity is determined by variation in the amino acids comprising the LRR domain [6,20]. This data may support the planning of mutational strategies towards improving the properties of natural PGIPs and the versatility of their recognition capabilities against the many diverse microbial PGs [21].

As mentioned, evaluation of the primary structure of *MdPGIP1* indicates a LRR region of 238 amino acids containing 123 hydrophobic amino acids (Figure 1). The LRR domain spans from residues 71 to 308. The *MdPGIP1* polypeptide contains 10 LRRs as indicated in Figure 1, hence the protein belongs to the LRR family [16,18,22]. All the PGIPs isolated to date comprise 10 LRRs matching the extracytoplasmic LRR consensus LxxLxxLxxLxLxxNxLxGxIPx, features that also apply to *MdPGIP1*.

LRR motifs play an important role in the cellular functions of several proteins [23]. A typical motif contains 20 to 29 residues and these motifs have been identified in plants, animals, as well as in microorganisms [20,23]. In the case of PGIP, these leucine residues are important in the binding of PGIP to the cell wall through the interaction of the residues with the acidic pectin within the cell wall matrix [14,15].

Physicochemical properties and glycosylation of the *MdPGIP1* pre-protein

Following the analysis of how hydrophobicity and hydrophilicity affect the *MdPGIP1* structure, further analyses were performed to determine the molecular weight of *MdPGIP1*. ProtScale was used to compute the profile of *MdPGIP1* and produced its constituent amino acids (<http://web.expasy.org/cgi-bin/protscale/protscale.pl>). The molecular weight of all the residues that make up *MdPGIP1* was determined per residue and mapped on the

polypeptide to assist with determining the molecular weight of the protein (graph not shown).

Purified *MdPGIP1* was found to have a molecular mass of between 44 to 54 kDa (cv Golden Delicious) [24] and 46 kDa (cv Granny Smith) [10]. Using the residues' molecular mass and their mapping throughout the *MdPGIP1* polypeptides, the molecular mass of the non-glycosylated apo-protein was calculated and determined to be 36.615 kDa. The bulk of the *MdPGIP1* polypeptide consists of Leu and Ser at 16.1% and 10.6%, respectively (Figure 1), a combined contribution of 29% to the molecular weight of *MdPGIP1*.

Lastly, the theoretical isoelectric point value (pI) of *MdPGIP1* was calculated to be 6.98. The relatively high pI of the mature *MdPGIP1* polypeptide is attributed to the presence of the positively charged Lys and Arg residues. These residues are believed to interact with the acidic pectin in the cell wall matrix, supporting the cell wall association of PGIP [16]. Addition of amino sugars during glycosylation of the pre-protein can further increase the pI to 8.0 [10], generating a basic functional protein.

PGIPs have been reported to be glycoproteins [4,22,25], undergoing post-translational glycosylation [26,27]. Glycosylation enables proteins to participate in biological processes, such as attaching to the extracellular matrix, as well as protein-ligand interactions, and has been shown to contribute to protein stability and increase resistance to protease digestion [27]. These are important features related to the function of PGIP in an extracellular environment.

N-glycosylation initially occurs in the endoplasmic reticulum and the target residue is Asp that is found in the sequence Asn-X-Ser/ Thr, where X can be any residue except Pro [28,29]. The *N*-linked glycosylation sites of *MdPGIP1* were analysed using the NetNGlyc 1.0 Server (<http://genome.cbs.dtu.dk/services/>) and the analysis is shown in Table 1. The typical consensus glycosylation

Table 1 The calculated N-linked glycosylation sites found on the *Malus domestica* polygalacturonase inhibiting protein 1 (*MdPGIP1*) polypeptide

Position	Possibility	N-Glycosylation
106 NLTG	9/9	++
130 NLSG	8/9	+
144 NLTF	5/9	-
154 NLTG	8/9	+
238 NKTT	8/9	+
254 NLSK	9/9	++
291 NVSY	9/9	++

The position of the receptor sites for *N*-glycosylation is included. The possibility of glycosylation occurring at that position is indicated with a + if it is positive, and - for highly unlikely.

sequence required for this modification (Asn-X-Ser/Thr) was identified in *MdPGIP1*, supporting its characterization as a glycoprotein. Aspects of the glycosylation process that can be modified and thus affect biological function, include glycan composition, glycan structure and glycan length [30].

Sequence comparison of the *MdPGIP1* encoded protein

PGIPs have been identified in many dicot and monocot plants. *PGIP* genes are often found as small gene families that encode PGIP isoforms with different specificities and affinities towards secreted fungal PGs [6]. The phylogenetic relationship between *MdPGIP1* was compared with PGIPs from other plants as shown in Figure 3. Amino acid comparisons between the different PGIPs showed that PGIPs from fruit bearing trees share a high percentage of similarity with each other when compared to other plants. *MdPGIP1* shares a 99.4%, 98%, and 98% amino acid identity with the PGIPs from *Malus pumila*, *Malus hupehensis* and *Pyrus communis* (members of the Rosaceae family), respectively.

The *MdPGIP1* amino acid sequence was also compared PGIPs from non-fruit bearing *Eucalyptus* tree species (members of the Myrtaceae family), namely, *E. grandis*, *E. saligna*, *E. nitens*, *E. urophylla* and *E. camaldulensis*. Interestingly, *MdPGIP1* was found to share a 97% amino acid identity with *E. grandis*, and a 96% amino acid identity with *E. saligna*, and *E. nitens*. *E. urophylla* and *E. camaldulensis* share a 95% amino acid identity with *MdPGIP1*. In contrast, *MdPGIP1* shares only a 55%, 48% and 53% identity with *PvPGIP1*, *PvPGIP2* and *GmPGIP* (members of the Fabaceae family) respectively.

An analysis of PGIP sequences from different eudicotyledonous species (Fabaceae, Brassicaceae, Rosaceae and Rutaceae) indicated that diversification of *PGIP* genes during evolution has been driven by positive selection [31], limited to a small number of PGIP residues that are mostly solvent exposed and located in the β -sheet B1 corresponding to the concave surface of the protein (below).

Structural modeling: *MdPGIP1* shares a similar structure with *PvPGIP2*

Two types of repeats, types A and B, are found in plant LRR domains [23]. Eight repeats of 28 amino acids were found in type A and 29 amino acids in type B. Short β -strand, $\beta\alpha$ loop, α helix and $\beta\alpha$ loop regions are formed by the type A repeats. This leads into repeats that form a parallel β -strand. The type A formation is repeated until the C-terminal, ending with a β -strand. β -Sheets are formed parallel to the α helix on the face of the protein. These sheets are formed by stabilising residues found on adjacent repeats. The formation of β -sheets creates curvature to the protein and gives it a horseshoe shape.

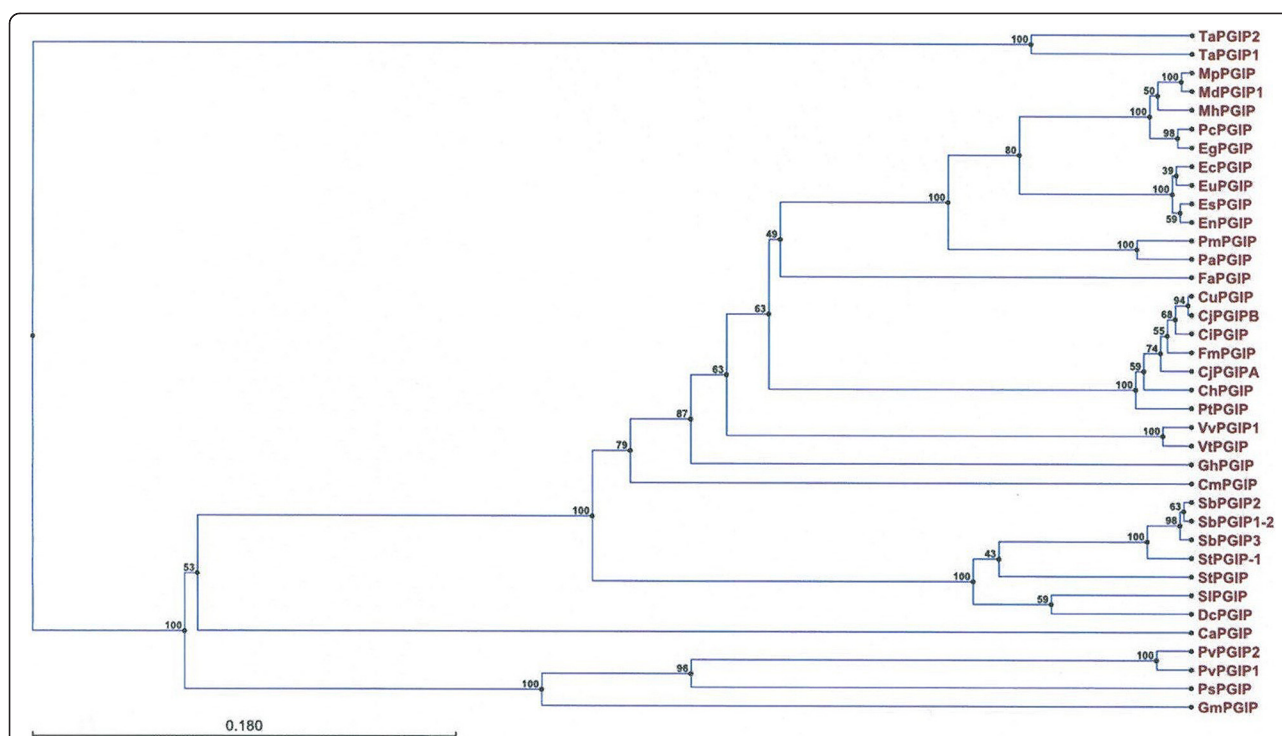


Figure 3 A phylogenetic tree representing the hierarchical clustering of the pairwise similarities between polygalacturonase inhibiting proteins (PGIPs).

The tree was constructed using the unweighted pair group method with arithmetic mean (UPGMA) and bootstrapping was also performed to verify the accuracy of the pairwise similarities identified. *Triticum aestivum*2 (TaPGIP2; AM180657), *T. aestivum*1 (TaPGIP1; AM180656), *Malus pumila* (MpPGIP; JQ001783), *M. domestica* PGIP (MdPGIP; U77041), *M. hupehensis* (MhPGIP; FJ449708), *Pyrus communis* PGIP (PcPGIP; L09264), *Eucalyptus grandis* (EgPGIP; AY445043), *E. camaldulensis* (EcPGIP; AF159168), *E. urophylla* (EuPGIP; AF159169), *E. saligna* (EsPGIP; AF159170), *E. nitens* (EnPGIP; AF159171), *Prunus mahaleb* (PmPGIP; AF263465), *P. americana* (PaPGIP; AY883418), *Fragaria ananassa* (FaPGIP; EU117215), *Citrus unshiu* (CuPGIP; AB016204), *C. jambhiri*B (CjPGIPB, AB015198), *C. iyo* (CiPGIP, AB016205), *Fortunella margarita* (FmPGIP; AB020529), *C. jambhiri*A (CjPGIPA, AB013397), *C. hystrix* (ChPGIP; AB071018), *Poncirus trifoliata* (PtPGIP, AB050528), *Vitis vinifera*1 (VvPGIP1; AF499451), *V. thunbergii* (VtPGIP; EU037367), *Gossypium hirsutum* (GhPGIP; EU711352), *Cucumis melo* (CmPGIP; KC215471), *Solanum brevidens*2 (SbPGIP2; DQ185391), *S. brevidens*1-2 (SbPGIP1-2; DQ185393), *S. brevidens*3 (SbPGIP3; DQ185392), *S. tuberosum*1 (StPGIP1; AY662681), *S. torvum* (StPGIP, FJ943498), *S. lycopersicum* (SlPGIP; L26529), *Daucus carota* (DcPGIP; AF055480), *Capsicum annum* (CaPGIP; EU560898), *Phaseolus vulgaris*2 (PvPGIP2, AJ864507), *P. vulgaris*1 (PvPGIP1; X64769), *Pisum sativum* (PsPGIP; AJ749705) and *Glycine max* (GmPGIP; X78274).

The molecular structure of PGIP2 from *Phaseolus vulgaris* (PvPGIP2) was elucidated using X-ray crystallography (Figure 4A) [20,32]. Although the presence of a single β -sheet in PvPGIP2 was predicted, it was shown that two β -sheets (sheet B1 and B2) were present in PvPGIP2. In addition to the two β -sheets, helices were also found on the LRR domain of the PvPGIP2 molecule. The protein structure was found to be curved and elongated which is typical of other PGIPs [23]. The residues found in the β -strand/ β -turn motif of PGIP were reported to be critical in the protein's affinity and specificity towards PG ligands [20,32].

The crystal structure of PvPGIP2 served as a template to which the MdPGIP1 was modelled. The putative structure of MdPGIP1 was modelled using SWISS-MODEL [12] and the modeling results are shown in Figure 4B. The structure of MdPGIP1 was also found to be curved and elongated. In addition, sheet B1, sheet B2 and 3_{10} -helices were also identified on the LRR domain of the MdPGIP1 molecule.

Sheet B1 of the MdPGIP1 LRR domain is located near the N-terminal in the concave inner side of the LRR region (Figure 4). The MdPGIP1 sheet B1 comprises 38 residues, of which 19 are hydrophobic, and these are located at residues 75, 77–78, 99, 101, 123–124, 126, 147–148, 171–172, 197, 220, 243–244, 267, 289 and 290 on the MdPGIP1 polypeptide (Figure 5). The hydropathy plot confirmed the observation where the hydrophobicity scores at these residue positions are relatively high. Sheet B2 is located on the convex outer side of the LRR region and comprises 29 residues, with 16 of those being hydrophilic (Figure 5). These hydrophilic residues are found at position 85, 108, 132, 134, 154, 156, 180, 182, 203, 205, 226, 228, 275, 296, 298 and 300 on the MdPGIP1 polypeptide. This water propensity of the sheet B2 residues is confirmed on the hydropathy plot. Sheet B2 determines the folding of PGIPs by connecting β -sheet B1 and 3_{10} -helices [32]. In addition, it is thought to form an additional surface on the PGIP for interaction with PGs [33].

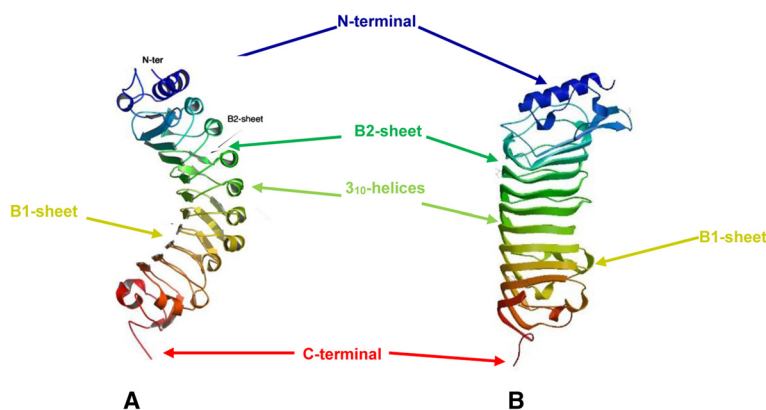


Figure 4 The ribbon representation of the folded structure of the *Malus domestica* polygalacturonase inhibiting protein 1 (*MdPGIP1*) in comparison to *Phaseolus vulgaris* PGIP2 (*PvPGIP2*). **A:** *PvPGIP2* structure, and **B:** *MdPGIP1* structure. PGIPs have evolved a wide interacting surface within the concave face of the LRR domain that is subject to evolutionary pressure for diversification [2,31,33].

Glycine residues found in sheet B2 of PGIPs are thought to be responsible for the bending of the sheet [32,33] and these residues were identified in the analysis of *MdPGIP1* (Figure 5, green highlighted residues). Further analysis of the glycine residues revealed that they are located, in sheet B2 of *MdPGIP1* (Figure 5, green highlighted residues), at positions similar to those identified on *PvPGIP2*. The speculation is that the *MdPGIP1* sheet B2 glycine residues are also responsible for the bending of the sheet, as was described for *PvPGIP2* [20,32]. However, the *MdPGIP1* molecule does have a curve that is more relaxed compared to *PvPGIP2*.

The amino acids within sheet B2 of *PvPGIP2* also comprise several aromatic and hydrophobic residues that

contribute to the bending of the *PvPGIP2* molecule [32,33]. These residues were identified as Phe-133, Phe-156, Phe-172, Phe-176, Phe-194, and Tyr-169. During analysis of the *MdPGIP1* polypeptide, amino acids identified at the same positions were Leu-133, Leu-156, Phe-172, Val-176, Phe-194, and Leu-169. Leucine and valine are hydrophobic aliphatic residues with smaller side chains and these residues may be the cause of the more relaxed curve observed on the *MdPGIP1* molecule.

While sheet B2 is thought to form an additional surface on the PGIP protein for interaction with PGs [32,33], sheet B1 residues determine the affinity and specificity of *PvPGIP2*. Asparagine residues have been found to form an Asn-ladder on *PvPGIP2* and these were found to form

<i>Malus domestica</i> 1	DLCNPDKKVLLQIKKAFGDPYVLTWSKSDTDCDD--WYCVTCDST--N
<i>Phaseolus vulgaris</i> 2	ELCNPQDKQALLQIKKDLGNPTTLSSWLPTTDCNRTWLGVLCDTDTQTY
<i>Malus domestica</i> 1	RINSLTIFAGQVSG--QIPALVGDLPYLETFEHLKQPNLTGPIQPAIAKL
<i>Phaseolus vulgaris</i> 2	RVNNLDLSGLNLPKPYPIPSSLANLPYLNFLYIGGINNLVGPPIPAIAKL
<i>Malus domestica</i> 1	KGLKFLRLSWT NLS GSVPDFLSQL NLT FLD LSF NNLTGTIPSSPSQL PN
<i>Phaseolus vulgaris</i> 2	TQLHYLYITH TNV SGAIPDFLSQIKTLVTL DFS YNALSGLTPPSIS SLPN
<i>Malus domestica</i> 1	L N AL R LDR N KLT G HIPK S L G Q F IG N V P DL Y L S H N Q L S G NI P T S F A Q M D F T
<i>Phaseolus vulgaris</i> 2	LVG I T F D G NR I SG A IP D SY G S F SK L F T SM T IS R N R LT G K I PP T F A N L N L A
<i>Malus domestica</i> 1	SIDLS R N K LE G DASVIFGL N KT Q IVDLS R N L LE F NLSKVE F PT S LT S LD
<i>Phaseolus vulgaris</i> 2	FVDLS R N M LE G DASVLF G SD K N T Q K IHL A K N SL A FD L G V L S K N L N GL D
<i>Malus domestica</i> 1	INH N K I Y G S I P V E F T Q L N -F Q FL N V S Y N R L C G Q I P V G G K L Q S F D E Y S Y F H
<i>Phaseolus vulgaris</i> 2	LR N N R I Y G T L P Q GL T Q L K F L H S L N V S F NN L C G E I P Q GG N L Q R F D V S A Y N
<i>Malus domestica</i> 1	NRCLCGAP L PS C K
<i>Phaseolus vulgaris</i> 2	NKCLCGSP L P A C T

Figure 5 Alignment of the amino acids from *Phaseolus vulgaris* polygalacturonase inhibiting protein 2 (*PvPGIP2*) and *Malus domestica* PGIP1 (*MdPGIP1*) without the signal peptides. Glycine residues and aromatic and hydrophobic residues thought to be responsible for the bending of sheet B2 in *PvPGIP2* are highlighted in green and red, respectively. Asparagine residues are highlighted in blue within the LRR domains.

hydrogen bonds with amide groups and the main-chain carbonyl. This quality also influences the bending of the protein [32,33]. Twelve of the 20 Asn residues observed on the P_vPGIP2 molecule were also observed on MdPGIP1 (Figure 5, blue highlighted residues).

Conclusion

Although phylogenetically distant from the archetypal P_vPGIP2, the overall properties of the MdPGIP1 protein are broadly similar to that of the PGIPs thus far characterised. However, even slight structural differences may confer new or broader recognition specificities to a PGIP or may improve its inhibitory strength. Based on the foundation laid in the present study, future studies of the detected differences will add support to the biotechnological use of MdPGIP1 in recombinant transgenic applications as a targeted inhibitor of fungal PGs. Moreover, it may assist in the identification of promising candidate PGIPs for crop protection, and in improving the properties of natural PGIPs and thus the versatility of their recognition capabilities against the many diverse microbial PGs.

Abbreviations

A-L-S: Signal peptide cleavage site; DAMP: Damage-associated molecular pattern; LRR: Leucine-rich repeat; N-X-S/T: N glycosylation site; PG: Polygalacturonase; PGIP: Polygalacturonase inhibiting protein; UPGMA: Unweighted pair group method with arithmetic mean.

Competing interests

LBTM and DO are staff members of the South African ARC. The ARC partially financed the publication of this manuscript. The authors declare that they have no competing interest regarding the publication of this article.

Authors' contributions

All authors conceptualised the research. LBTM performed the analyses. DO and IAD supervised the research and contributed to writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The South African Agricultural Research Council (ARC) and the University of Johannesburg are thanked for supporting the research.

Received: 20 November 2014 Accepted: 20 February 2015

Published online: 11 March 2015

References

- De Lorenzo G, Ferrari S. Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi. *Curr Opin Plant Biol.* 2002;5:295–9.
- Federici L, Caprari C, Mattei B, Savino C, Di Matteo A, De Lorenzo G, et al. Structural requirements of endopolygalacturonase for the interaction with PGIP (polygalacturonase inhibiting protein). *Proc Natl Acad Sci U S A.* 2001;98:13425–30.
- King D, Bergmann C, Orlando R, Benen JAE, Kester HCM, Visser J. Use of amide exchange mass spectrometry to study conformational changes within the endopolygalacturonase II - polygalacturonic acid - polygalacturonase inhibiting protein system. *Biochemistry.* 2002;41:10225–33.
- De Lorenzo G, Cervone F, Bellincampi D, Caprari C, Clark AJ, Desiderio A, et al. Polygalacturonase, PGIP and oligogalacturonides in cell-cell communication. *Biochem Soc Trans.* 1994;22:394–7.
- Sanabria NM, Huang J-C, Dubery IA. Self/non-self perception in plants in innate immunity and defense. *Self/Non-Self Imm Recog Signal.* 2010;1:40–5.
- Federici L, Di Matteo A, Fernandez-Recio J, Tsernoglou D, Cervone F. Polygalacturonase inhibiting proteins: players in plant innate immunity. *Trends Plant Sci.* 2006;11:65–9.
- Yao C, Conway WS, Ren R, Smith D, Ross GS, Sams CE. Gene encoding polygalacturonase inhibitor in apple fruit is developmentally regulated and activated by wounding and fungal infection. *Plant Mol Biol.* 1999;39:1231–41.
- Arendse MS, Dubery IA, Berger DK. Isolation by PCR-based methods of a plant antifungal polygalacturonase-inhibiting protein gene. *Electron J Biotechnol.* 1999;2:152–9.
- Gazendam I, Oelofse D, Berger DK. High-level expression of apple PGIP1 is not sufficient to protect transgenic potato against *Verticillium dahliae*. *Physiol Mol Plant Pathol.* 2004;65:145–55.
- Oelofse D, Dubery IA, Meyer R, Arendse MS, Gazendam I, Berger DK. Apple polygalacturonase inhibiting protein1 expressed in transgenic tobacco inhibits polygalacturonases from fungal pathogens of apple and the anthracnose pathogen of lupins. *Phytochemistry.* 2006;67:255–63.
- Kyte J, Doolittle R. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157:105–32.
- Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics.* 2006;22:195–201.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. In: Walker JM editor. *The Proteomics Protocols Handbook, Protein Identification and Analysis Tools on the ExPASy Server.* Humana Press, 2005. p. 571–607.
- De Lorenzo G, D'Ovidio R, Cervone F. The role of polygalacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi. *Annu Rev Phytopathol.* 2001;39:313–35.
- Protsenko MA, Buza NL, Krinitsyna AA, Bulantseva EA, Korableva NP. Polygalacturonase-inhibiting protein is a structural component of plant cell wall. *Biochemistry (Mosc).* 2008;73:1053–62.
- Toubart P, Desiderio A, Salvi G, Cervone F, Daroda L, De Lorenzo G. Cloning and characterization of the gene encoding the endopolygalacturonase-inhibiting protein (PGIP) of *Phaseolus vulgaris* L. *Plant J.* 1992;2:367–73.
- Stotz HU, Powell ALT, Damon SE, Greve LC, Bennett AB, Labavitch JM. Molecular characterization of a polygalacturonase inhibitor from *Pyrus communis* L. cv Bartlett. *Plant Physiol.* 1993;102:133–8.
- Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol.* 2001;11:725–32.
- Jones JGD. Putting knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol.* 2001;4:281–7.
- Leckie F, Mattei B, Capodicasa C, Hemmings A, Nuss L, Aracri B, et al. The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed β -strand/ β -turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. *EMBO J.* 1999;18:2352–63.
- Borras-Hidalgo O, Caprari C, Hernandez-Estevéz I, De Lorenzo G, Cervone F. A gene for plant protection: expression of a bean polygalacturonase inhibitor in tobacco confers a strong resistance against *Rhizoctonia solani* and two oomycetes. *Front Plant Sci.* 2012;3:268.
- Mattei B, Bernalda MS, Federici L, Roepstorff P, Cervone F, Boffi A. Secondary structure and post-translational modifications of the leucine-rich repeat protein PGIP (polygalacturonase-inhibiting protein) from *Phaseolus vulgaris*. *Biochemistry.* 2001;40:569–76.
- Kobe B, Deisenhofer J. Proteins with leucine-rich repeats. *Curr Opin Struct Biol.* 1995;5:409–16.
- Yao C, Conway WS, Sams CE. Purification and characterization of a polygalacturonase-inhibiting protein from apple fruit. *Phytopathology.* 1995;85:1373–7.
- Favaron F, D'Ovidio R, Porceddu E, Alghisi P. Purification and molecular characterization of a soybean polygalacturonase-inhibiting protein. *Planta.* 1994;195:80–7.
- Parekh RB, Dwek RA, Thomas JR, Opdenakker G, Rademacher T, Wittwer AW, et al. Cell-type specific and site specific N-glycosylation of type I and type II human tissue plasminogen activator. *Biochemistry.* 1989;28:7644–61.
- Bloom JW, Madanat MS, Ray MK. Cell line and site specific comparative analysis of the N-linked oligosaccharides on human ICAM-1des454–532 by electrospray ionisation mass spectrometry. *Biochemistry.* 1996;35:1856–64.
- Rudd PM, Joao HC, Coghill E, Fiten P, Saunders MR, Opdenakker G, et al. Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry.* 1994;33:17–22.
- Nakano M, Kakehi K, Tsai M-H, Lee YC. Detailed structural features of glycan chains derived from α 1-acid glycoproteins of several different animals: the

presence of hypersialylated, O-acetylated sialic acids but not disialyl residues. *Glycobiology*. 2004;14:431–41.

30. Walsh C. *Posttranslational modification of proteins: Expanding nature's inventory*. Englewood, Colo: Roberts and Co. Publishers; 2006.
31. Casasolia M, Federici L, Spinellia F, Di Matteo A, Vella N, Scalonina F, et al. Integration of evolutionary and desolvation energy analysis identifies functional sites in a plant immunity protein. *Proc Natl Acad Sci U S A*. 2009;106:7666–71.
32. Di Matteo A, Federici L, Mattei B, Salvi G, Johnson KA, Savino C, et al. The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat involved in plant defense. *Plant Biol*. 2003;100:10124–8.
33. Di Matteo A, Bonivento D, Tsernoglou D, Federici L, Cervone F. Polygalacturonase-inhibiting protein (PGIP) in plant defence: a structural view. *Phytochemistry*. 2006;67:528–33.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

