ORIGINAL PAPER

In-silico analysis of the structure and binding site features of an α-expansin protein from mountain papaya fruit (VpEXPA2), through molecular modeling, docking, and dynamics simulation studies

Carlos Gaete-Eastman[1] · Luis Morales-Quintana[1] · Raúl Herrera[1] · María Alejandra Moya-León[1]

Abstract Fruit softening is associated to cell wall modifications produced by a set of hydrolytic enzymes and proteins. Expansins are proteins with no catalytic activity, which have been associated with several processes during plant growth and development. A role for expansins has been proposed during softening of fruits, and many fruit-specific expansins have been identified in a variety of species. A 3D model for VpEXPA2, an α-expansin involved in softening of *Vasconcellea pubescens* fruit, was built for the first time by comparative modeling strategy. The model was validated and refined by molecular dynamics simulation. The VpEXPA2 model shows a cellulose binding domain with a β-sandwich structure, and a catalytic domain with a similar structure to the catalytic core of endoglucanase V (EGV) from *Humicola insolens*, formed by six β-strands with interconnected loops. VpEXPA2 protein contains essential structural moieties related to the catalytic mechanism of EGV, such as the conserved HFD motif. Nevertheless, changes in the catalytic environment are observed in the protein model, influencing its mode of action. The lack of catalytic activity of this expansin and its preference for cellulose are discussed in light of the structural information obtained from the VpEXPA2 protein model, regarding the distance between critical amino acid residues. Finally, the VpEXPA2 model improves our understanding on the mechanism of action of α-expansins on plant cell walls during softening of *V. pubescens* fruit.

Introduction

Expansins are proteins with no catalytic activity, which have the ability to induce extension and relaxation of plant cell walls in a pH dependent mode. It is believed that during fruit softening they promote relaxation of the cell wall structure allowing the access of different hydrolases to their substrates, through the disruption of hydrogen bonds between cellulose microfibrills and cross-linking glycans in the cell wall [1, 2]. Two families of expansins have been described: α-expansins (EXPA), with sequences phylogenetically related to the originally described expansin from cucumber, and β-expansins (EXPB), with sequences related to group-1 grass pollen allergens [3]. Recent database searches and phylogenetic sequence analysis revealed two additional expansin families, called expansin-like A and B, although no evidence of expansin activity has yet been described for these expansin-like families [4]. Both α- and β-expansins consist of two domains. The N-terminal domain 1 (D1) is distantly related to the catalytic domain of family-45 glycoside hydrolase (GH45) proteins, most of which are fungal β-1,4-D-endoglucanases. Proteins from this family have been crystallized and their mechanism of action determined. Their tertiary structure forms a six-stranded β-barrel with a groove for substrate binding [4, 5].

✉ Carlos Gaete-Eastman
cgaete@utalca.cl

✉ María Alejandra Moya-León
alemoya@utalca.cl

[1] Laboratorio de Fisiología Vegetal y Genética Molecular, Instituto de Ciencias Biológicas, Universidad de Talca, Casilla 747, Talca, Chile

The C-terminal domain 2 (D2) is distantly related to group-2 grass pollen allergens, and it is described as two stacked β sheets with an immunoglobulin-like fold [6, 7]. The first work performed in plant expansin's structure was in the crystal of ZmEXPB1 (Zea m 1), a beta expansin from maize, confirming the two domain structure but lacking a second Asp residue that serves as a catalytic base during the hydrolytic activity of GH45 enzymes [8]. The function of these two domains has not yet been tested experimentally because heterologous expression of plant expansins (especially critic on α-expansin) has proven difficult, limiting detailed structure and function studies. Recently, an exhaustive structural and functional characterization of a bacterial expansin-like protein (BsEXLX1) has been performed, showing structural and functional insights which the authors suggest may be extrapolated to plant expansins [9].

Expansin's functional assays have been previously performed using pure cellulose paper and cellulose/xyloglucan composite [2, 10] that demonstrate that expansins bound to non-crystalline polysaccharides on the cellulose microfibrils surface, allowing turgor-driven polymer creep and leading to cell wall loosening [7]. Nevertheless, so far no evidence for protein-ligand interaction at the molecular level has been provided for α-expansins.

In our lab, a ripening-specific expansin from mountain papaya fruit (VpEXPA2) was isolated, showing the typical sequence structure of an EXPA family [11]. Sequence analysis suggests that VpEXPA2 contains two domains, with eight cysteines, four tryptophans and the HFD motif, which is part of GH45 glycoside hydrolase active site. The aim of the present study was to perform a structural characterization of VpEXPA2 protein and to provide the basis of its interaction with two types of cell wall polysaccharides: cellulose and hemicellulose. These substrates were parameterized by using the octameric form of cellooligosaccharide [12] and two types of xyloglucans [13]. By using bioinformatics tools a 3D model of this α-expansin was obtained and used to determine protein-ligand interactions with cell wall polysaccharides. The results showed different substrate affinities that provides an explanation for the interaction mode of EXPA family protein.

Materials and methods

The 3D structure of VpEXPA2 using homology modeling

The 3D model of VpEXPA2 protein was built by comparative modeling based on high-resolution crystal structures of a homology protein. The protein sequence of VpEXPA2 [10] was retrieved from NCBI protein sequence database: accession number ABF48653 (http://www.ncbi.nlm.nih.gov/protein). Basic local alignment search tool (BLAST) search [14] and

Protein Structure Prediction Server (PSIPRED) [15] were used for selecting the 3D models with the closest homology available in the Brookhaven Protein Data Bank (PDB). The crystal structure of Phlp1, corresponding to the β-allergen from Timothy grass (*Phleum pratense*) (PDB: 1N10) was selected as template, with 34 % of sequence identity (*p-value* $3e^{-11}$) and an alignment score of 682.0 determined by PSIPRED. The first requirement in the construction of the models is the sequence alignment with the templates, focused on the identification of structurally conserved regions common to the different templates and targets. The pair-wise sequence alignment was obtained by using ClustalW [16]. The comparative model of the structure was generated by using MODELLER 9v8 software (http://salilab.org/modeller/) [17]. Ten models were generated for the structure and the model showing the lowest MODELLER objective function and RMS deviation with respect to trace (Cα atoms) of the crystal structure of the template, were saved for further refinement and validation. Protein hydrogen atoms were added according to environment at pH 4.5 using the HBUILD module of CHARMM version c31b1 [18]. Structural refinements were accomplished by molecular dynamic simulations (MDS), and equilibration methods using Nano Molecular Dynamics (NAMD v2.8) software [19] and Chemistry of Harvard Molecular Modeling (CHARMM27) force field for lipids and proteins [20, 21] along with the TIP3P model for water [22]. After an initial minimization, the system was subjected to a short molecular dynamics of 2 ns run in order to remove wrong contacts and to fill empty pockets. All backbone atoms were restrained using a harmonic force constant of 5 kcal mol^{-1} Å^{-2} and the loops were left free during relaxations. All MDS were done using a time step of 1 fs, with a 12 Å spherical cut off for non-bonding interactions and a switching function of 10 Å for the van der Waals term. Finally, to evaluate the accuracy of the model, ProSA [23], PROCHECK [24], and Verify3D [25] were employed.

Docking studies

Docking studies were performed to predict the putative binding of different octasaccharides to the protein: two different hemicellulose octasaccharides [13] and one cellodextrin 8-mer that resembles a water-soluble cellulose molecule [12]. For this purpose, different docking investigations using the Autodock Vina program were carried out [26]. The different PDBQT files (molecular structure file formats used by Autodock Vina) were generated using MGLTools. For calculations a grid size of 54 Å×56 Å×124 Å centered at the interface with 0.375 Å spacing was used, that includes all protein surface. A Lamarckian genetic algorithm was used as a search method, and the rest of the parameters were set with their default values.

Molecular dynamics (MD) simulations

Molecular dynamics of each substrate inside VpEXPA2 active site was studied by using NAMD v2.8 [19] and CHARMM27 [20, 21] along with the TIP3P model for water [22]. Ligands were built using GlyCAM software (http://glycam.ccrc.uga.edu/), and ParamChem [27] was used to provide and check the force field parameters required for the ligands. The initial coordinates for MDS calculations were taken from the docking experiments. TIP3P water molecules were added (the dimensions of each water box were 82.5 Å×68.7 Å×86.7 Å approximately, which ensured the coverage of the whole complex surface), and the systems were neutralized by adding KCl counterions to balance the net charges of the system. After the construction of the solvent environment, each complex system was composed of about 62,100 atoms. The systems were minimized and pre-equilibrated using the relaxation routine implemented to relax the initial models. After that, a first 2 ns long equilibration MD simulation was performed on each complex system, which was followed by a 20 ns long MD simulation. During MD simulations, the equations of motion were integrated with a 1 fs time step in the NVT ensemble. The van der Waals cut-off was set to 12 Å. The temperature was maintained at 300 K°. Long-range electrostatic forces were taken into account by means of the particle-mesh Ewald (PME) approach [28]. Data were collected every 2 ps during the MD runs. Visualization of protein-ligand complexes and MD trajectory analysis were carried out with the VMD v1.9.1 software [29].

Results

Building the 3D structure of VpEXPA2 based on homology modeling

A 3D structure model for VpEXPA2 was built, based on the crystalline structure of Phlp 1, a β-expansin from *Phleum pratense* (PDB: 1N10). Phlp 1 and VpEXPA2 share 28 and 43 % of identity and similarity respectively by BLAST program (considering the sequence of the mature protein), fulfilling the requirements for using Phlp 1 as a template. The alignment between the target sequence and the template was manually optimized, incorporating information of the secondary structure of expansins, especially related to structural motifs (e.g., six cysteines, four tryptophans, and the catalytic HFD motif) which were spatially restricted during the modeling to avoid any distortion. In addition, the insertions present in both sequences, which are conserved in α and β expansin families [30], were also considered; and the last six residues did not show any secondary structure, because the sequence used as template was shorter at the C-terminal.

Considering the moderate sequence identity between VpEXPA2 and Phlp 1 an exhaustive evaluation of geometric and energetic stability of VpEXPA2 model was performed. In this way, ten conformers with the lowest total energy were selected after energy minimization. The best conformer was determined by using different evaluation methods. The stereochemical quality of VpEXPA2 model was analyzed with PROCHECK. The most favored regions corresponded to 85.8 %; 14.2 % corresponded to allowed and generously allowed regions; and most importantly no badly modeled residues were found (Table 1). On the other hand, ProSA analysis showed low ProSA energy scores in most structurally conserved regions, however a small region corresponding to a loop distant from the active site showed unfavorable ProSA energy score (Figure S1). Finally, the analysis using Verify3D program showed favorable scores for the entire 3D protein structure, and no residues showing values under 0 (data not shown). Consequently, the final structure of VpEXPA2 protein was accepted for subsequent analysis.

The model obtained for VpEXPA2 displayed the two domains proposed for α-expansins: the 'catalytic' domain (D1; residues 1–127) and the cellulose-binding domain (CBD) (D2; 137–223), connected by a short linker (residues 128–136) (Fig. 1a). The spatial location of six cysteine residues (in positions 26, 54, 57, 62, 69, and 130) allowed the formation of three disulfide cross-linkages, which were confirmed in the structural model (Fig. 1a, asterisks). The other two cysteine residues (in positions 82 and 96) do not form disulfide cross-linkages. Superposition of VpEXPA2 with ZmEXPB1 (PDB: 2HCZ) gives a good overlap between the two structures, specially the β-strand residues that become part of the β-barrel on D1 and β-sandwich on D2 (Fig. 1b). Both structures have short α-helices and loops but these do not overlap (Fig. 1b). EXPA and EXPB have two main sequence differences reflected in two insertions located at either sides of the conserved HFD motif, located toward the N-terminal from the HFD motif in EXPA and toward the C-terminal in EXPB [4]. This insertion is observed as a long extra loop in VpEXPA2 compared to ZmEXPB1 (Fig. 1b, white arrow), and the lack of a short α-helix in VpEXPA2 compared to ZmEXPB1 (Fig. 1b, red arrow). In addition, a second loop, which is located next to the 'active site', is longer in VpEXPA2 (Fig. 1b, yellow arrow).

Table 1 Validation of VpEXPA2 protein structure using PROCHECK program (Ramachandran plot)

	Core (%) [a]	Allow (%) [b]	Gener (%) [c]	Disall (%) [d]
VpEXPA2	85.8	13.7	0.5	0.0

[a] Most favored region; [b] Additional allowed regions; [c] Generously allowed regions; [d] Disallowed regions
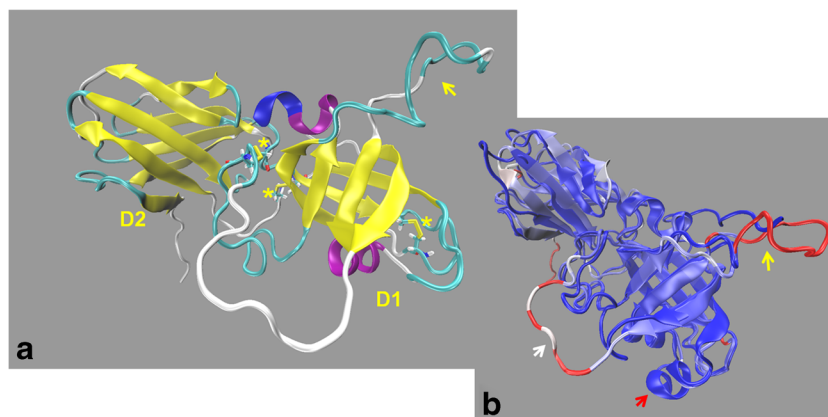
Fig. 1 VpEXPA2 protein model. (**a**) The structures of the two domains are shown: the catalytic domain (D1) on the right, and the β-sandwich of the CBD domain (D2) on the left; the three disulfide bridges are highlighted with *yellow asterisks*, the glycine rich-like loop with a *yellow arrow*, α-helices in *purple* and 3_{10} helix in *blue*. (**b**) Structural alignment between VpEXPA2 and ZmEXPB1, showing the α-loop (*white arrow*), β-helix (*red arrow*), and the glycine rich-like loop (*yellow arrow*)

D1 fold is composed of six β sheets interconnected by several short loops and two α-helices (Fig. 1a), where the 'catalytic' motif (His102, Phe103, and Asp104) is surrounded by a group of residues (Thr11, Tyr13, Asp17, Ala41, and Ala42) highly conserved across the expansin superfamily (Fig. 2a). The 'catalytic' motif is located in a groove between two loops, which corresponds to residues 16–20 and 88–94, respectively (Fig. 2b, white arrows).

D2 fold is composed of eight β-strands assembled in two β-sheets, forming a ß-sandwich structure (Fig. 3a). It contains mainly aromatic (Phe, Tyr, and Trp) residues displayed in line with the groove (that span both domains) and the active site. In addition, other residues from D1 are Gly24, Gly27, Tyr13, and Phe12, which all together could interact with cellulose microfibrils (Fig. 3).

Limited but significant homology has been proposed between D1 of expansins and the catalytic site of

endoglucanases from family 45 (GH45; EC 3.2.1.4) [4]. Taking that into consideration, the structural superposition of VpEXPA2 D1 and endoglucanase V from *Humicola insolens* (EGV) (PDB: 2ENG) was analyzed (Fig. 4a). A good overlap between the two β-barrels was observed, but not in the rest of the structure which was mainly composed by loops and short α-helices. In addition, the spatial orientation of critical residues for the catalytic activity of GH45 was examined in both structures (Fig. 4b). The key residues in EGV, such as His119, Phe120, Asp121, Tyr8, Asp10, Ala73, Ala74, and Thr6 were spatially conserved in our model (Fig. 4b). During the molecular dynamics simulation (MDS) of VpEXPA2 structure all these residues were not disturbed and the 3D architecture of the active site remained unchanged during 2 ns of MD simulation. Additionally, the residues Asp121 and Asp10 in EGV have been highlighted as very important for its catalytic mechanism, the first residue acting as catalytic acid and the second

Fig. 2 The 'active site' of VpEXPA2 (D1) structure. (**a**) The relevant residues are in licorice view. (**b**) A surface view of VpEXPA2 showing two loops (*white arrows*) surrounding the open groove where the active site is located (licorice)
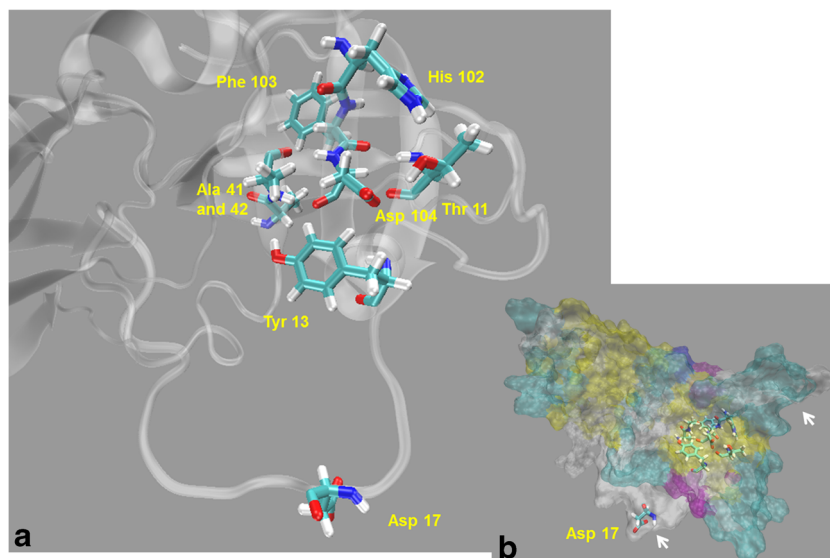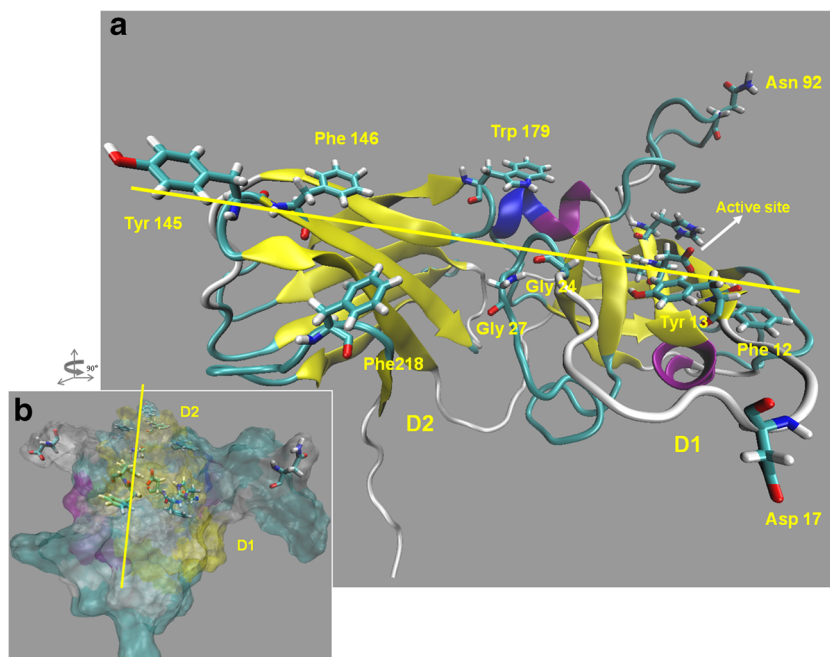
Fig. 3 The 'carbohydrate binding domain' (CBD) (D2) structure of VpEXPA2. (**a**) Non-polar residues are shown in licorice, and the groove orientation is indicated as a *yellow line*. (**b**) VpEXPA2 in surface representation, where the groove location is shown as a *yellow line*

one as catalytic base [5]. Interestingly, when the VpEXPA2 active site is analyzed, the acid Asp104 residue was found in the same position and orientation, and the hydrophobic environment for Asp104 was given by the residues Ala41, Ala42, and Tyr13 (Fig. 4b). This environment should stabilize the protonate state of Asp104, allowing its action as catalytic acid, which could protonate the glycosidic oxygens from polysaccharides such as cellulose, as described in EGV [5]. Nevertheless, the residue corresponding to basic Asp10 is missing in VpEXPA2 (Fig. 4b), and the closest Asp residue (Asp17) is too far to directly participate in the hydrolytic reaction (16.7±2 Å) (Fig. S2). Finally, with the aim to suggest a hypothesis about the role of this missing catalytic Asp, a detailed analysis of the structural alignment between EGV

(PDB: 2ENG) and VpEXPA2 D1 was performed (Fig. 5). The first portion of both proteins (until Tyr8 and Tyr13 residues in EGV and VpEXPA2, respectively) is the same in terms of secondary structure and spatial conformation of Tyr lateral chain, however the distance between the catalytic Asp10 to Tyr8 in EGV is 4.8 Å, and in VpEXPA2 the distance between Asp17 and Tyr13 is 14.8 Å (Fig. 5a). When the sequences of EXPA, EXPB, and EGV were aligned and carefully analyzed it can be observed that the catalytic Asp is two residues away from Tyr residue in EGV, and by contrast, it is four residues away in VpEXPA2 and other EXPA sequences. This spacement between Tyr and Asp of four residues is a conserved pattern in EXPA sequences, nevertheless this Asp residue is missing in EXPB sequences (Fig. 5b).

Fig. 4 **a** Structural superposition of VpEXPA2 model and EGV structure. The cartoon model of EGV is colored in *blue*, and that of VpEXPA2 D1 in *red*. **b** *Red* residues correspond to VpEXPA2 whereas *blue* and *asterisk* labeled residues correspond to EGV
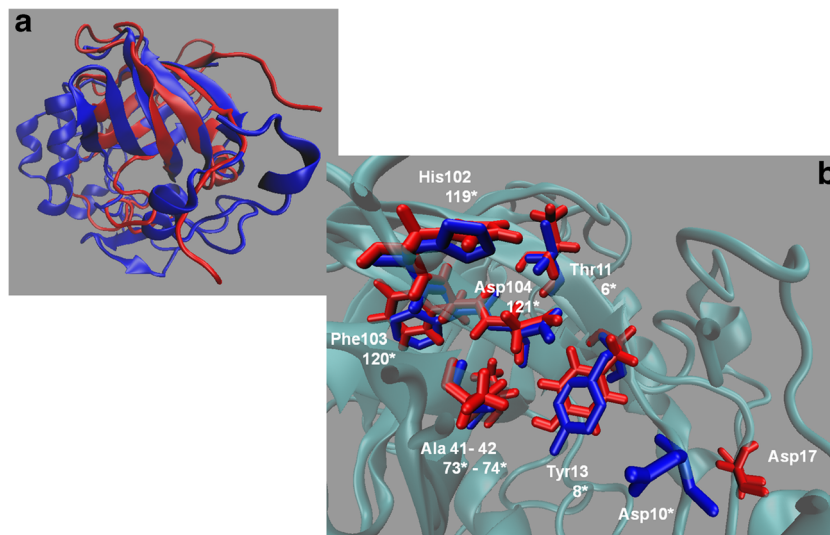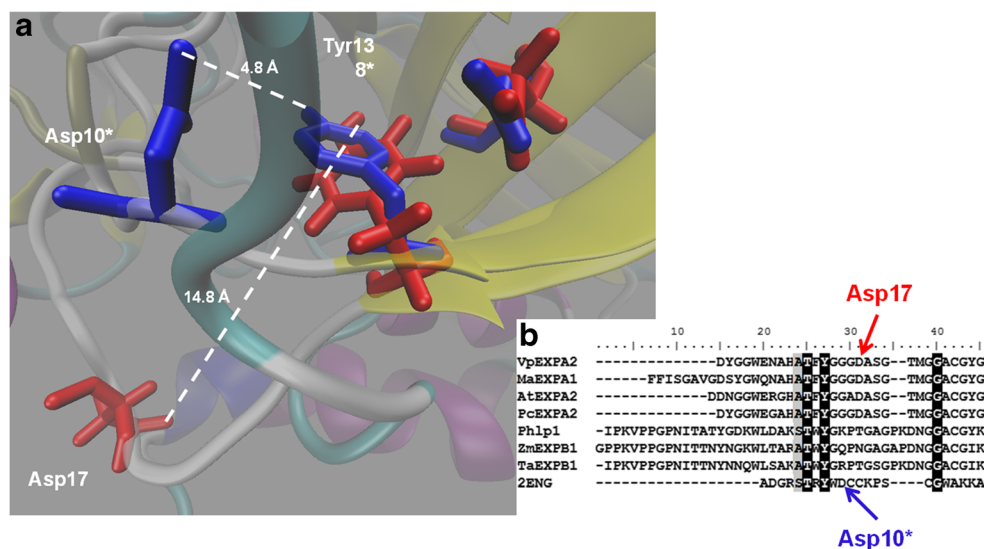
Fig. 5 a Structural superposition of the active sites of VpEXPA2 and EGV. Residues in *blue* and labeled with *asterisks* correspond to EGV, and residues in red and labeled *without asterisks* correspond to VpEXPA2. **b** Partial protein sequence alignment of selected members of α- and β-expansins and EGV. Gaps are indicated by *dashes*, letters with black background are identical amino acids and *letters with gray background* are similar amino acids. Full length protein sequence of four α-expansin (EXPA), three β-expansin (EXPB and Phlp1), and EGV (2ENG) were

aligned using BioEdit Sequence Alignment Editor v7.1 (Hall, 1999). Sequences correspond to the following GeneBank accession numbers: VpEXPA2 (*Vasconcellea pubescens*, DQ499635), MaEXPA1 (*Musa acuminata*, AY083168), AtEXPA2 (*Arabidopsis thaliana*, U30481), PcEXPA2 (*Prunus cerasus*, AF350937), Phlp1 (pollen allergen, *Phleum pratense*, X78813), ZmEXPB1 (Zea mays, AAK56124), TaEXPB1 (*Triticum aestivum*, AAS21276), and 2ENG (endoglucanase V, *Humicola insolens*, P43316)

Ligand binding analysis

The refined VpEXPA2 model was used to study protein-ligand conformations by automatic docking with three substrates. As shown in Table 2, for the three ligands tested, negative energies were obtained, indicating a favorable protein-ligand interaction. The strongest binding interaction was found between VpEXPA2 protein and cellodextrin 8-mer (-7.2 kcal mol^{-1}), while the binding energy for the two-hemicellulose octasaccharides XXXGXXXG and XXFGXXFG was less favorable (-6.0 and -6.1 kcal mol^{-1}, respectively).

Additionally, Fig. 6 shows the docking arrangement of the ligands into the substrate open groove of VpEXPA2 and their interaction with the active site. The VpEXPA2- cellodextrin 8-mer complex indicates that the substrate locates in the open groove, and interacts with the active site and Asn91 from the 'glycine rich-like loop' (Fig. 6a). A detailed view of the active site shows that His102 residue is protonated, while Asp104 residue is not. In addition, the distance between the carboxylic group of Asp104 and O$_4$ from β-1,4 glycosidic bond of the cellodextrin 8-mer polymer is 4.2 Å, at an adequate distance, however the non-protonated state of Asp104 residue avoids its interaction as a catalytic acid with the glycosidic oxygen through hydrogen bonds (Fig. 6b). Interestingly, in the VpEXPA2-XXFGXXFG complex, the conformation of the substrate in the open groove and catalytic site is different compared to cellulose; the substrate covers most of D1 surface but it does not interact with the 'glycine rich-like loop'

(Fig. 6c). The distance between the carboxylic group of Asp104 and O$_4$ from β-1,4 glycosidic bond of the XXFGXXFG polymer is 6.9 Å, which is inadequate in distance, and also the non-protonated state of Asp104 does not allow its interaction with the glycosidic oxygen through hydrogen bonds (Fig. 6d). With the aim to get more detail about the residues which are interacting with the substrate, the residues detected at the interface between cellodextrin 8-mer or XXFGXXFG and VpEXPA2 were painted in yellow surface representation (Fig. 6a and c, respectively), detecting a higher number of residues interacting with XXFGXXFG polymer than with cellodextrin 8-mer (Table 3). Different modes of interaction are observed for VpEXPA2 and the ligands. The interaction of VpEXPA2 with cellodextrin 8-mer is mainly through the β-barrel and the 'glycine rich-like loop', nevertheless when the substrate is XXFGXXFG, the interaction is mainly through the α-loop and β-sandwich (Table 3). In the case of VpEXPA2-XXXGXXXG complex the interaction was also analyzed, obtaining a distance of 8.9 Å between the carboxylic group of Asp104 and O$_4$ from β-1,4 glycosidic

Table 2 Affinity energy of three different octasaccharide substrates with VpEXPA2 protein model

Substrates	VpEXPA2 Affinity energy (kcal mol^{-1})
GGGGGGGG	-7.2
XXXGXXXG	-6.1
XXFGXXFG	-6.0

Fig. 6 Molecular docking simulation of interaction of VpEXPA2 with cellodextrin 8-mer (**a**, **b**) and hemicellulose octasaccharide XXFGXXFG polymer (**c**, **d**). D1 is colored in blue, D2 is colored in green, and residues located up to 3 Å are colored in *yellow* (**a**, **c**). The 'catalytic' residues of VpEXPA2 and the ligand are shown in licorice representation (**b**, **d**). *Dash lines* indicate distance between atoms
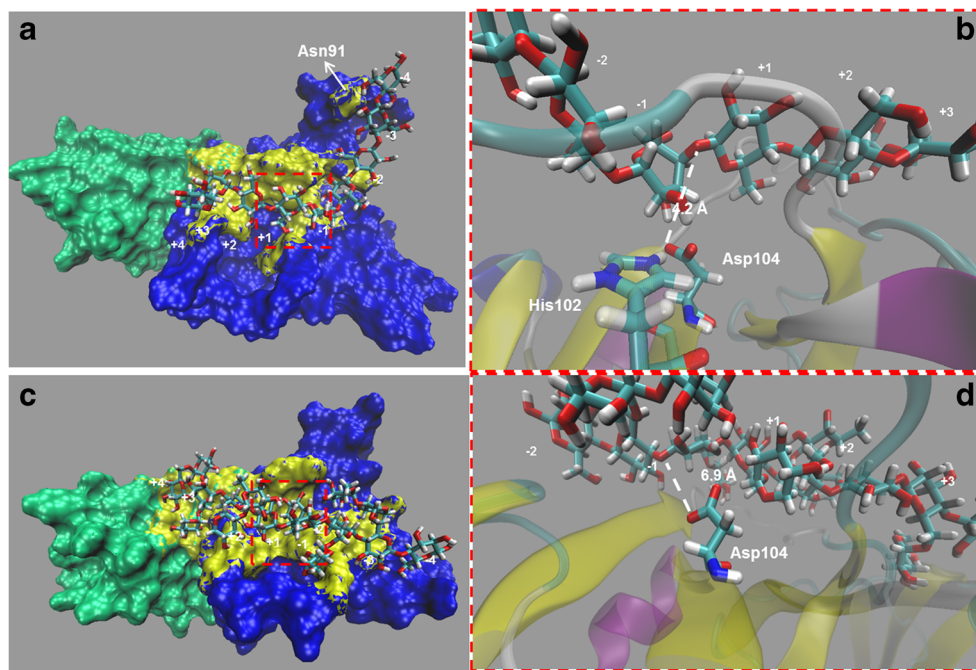
bond of the XXXGXXXG polymer, which indicates an unfavorable interaction (data not shown). In conclusion, both VpEXPA2-hemicellulose octasaccharide complexes present major structural changes in substrate conformation compared to cellodextrin 8-mer, that leads to a higher separation between the carboxylic group of Asp104 and β-1,4 glycosidic bond (distance greater than 6 Å), that does not support a stable interaction between VpEXPA2 and hemicellulose substrates (Table 2).

Molecular dynamics (MD) simulations

From molecular docking simulations, VpEXPA2 protein showed a strong interaction to cellodextrin 8-mer polymer, and in contrast, a low interaction with hemicellulose octasaccharide polymers (XXFGXXFG and XXXGXXXG). The protein-substrate interaction of these substrates with the HFD motif of VpEXPA2 was studied using MD simulations.

In the simulation with cellodextrin 8-mer as substrate, it can be observed that from the beginning of the MD simulation the carboxylic group of Asp104 residue interacts with O_4 from β-1,4 glycosidic bond; this interaction is maintained during the MD simulation at a distance of around 5.1 Å (Fig. 7a, black line). After 5–6 ns a temporal separation of Asp104 and O_4 from β-1,4 glycosidic bond is observed when the substrate reorients in the cavity returning to distances close to 4.8 Å at the end of the MD simulation (Fig. 7a, black line). In the case of hemicellulose octasaccharide polymers, XXXGXXXG and XXFGXXFG, the two complexes showed similar performance during the molecular dynamics simulation, revealing an increasing separation between the carboxylic group of Asp104

and O_4 from β-1,4 glycosidic bond (more than 7.5 Å) (Fig. 7a; blue and red lines). During the MD simulation of XXXGXXXG the ligand moves away from Asp104 (to around 10.5 Å) during the first ns, then the distance temporally decreases to 9 Å and then the ligand reorients into the cavity at a distance of 11 Å from Asp104 at 3.5 ns of MDS (Fig. 7a, blue line). In the case of XXFGXXFG a distance around to 11 Å from Asp104 is shown during the first part of MD simulation, reaching 9 Å after 7 ns of MD simulation and 7.5 at the end of MD simulation (Fig. 7a, red line). In conclusion, the two-hemicellulose octasaccharide polymers are located too far from the HFD motif and are unfavorable oriented during the MD simulation compared to cellodextrin 8-mer.

Another important residue is Asp17, which is located in loop-α and far from Asp104 to allow a hydrolytic mechanism. With the aim to verify if these two residues remain separated, the distance between Asp17 and Asp104 was determined during a 20 ns of MD simulation. At the beginning of MD simulation of VpEXPA2-cellodextrin 8-mer complex, the carboxylic group of Asp17 residue is too far from O_4 from β-1,4 glycosidic bond to directly participate in a hydrolytic reaction (around 22 Å). After the first ns of MD simulation this distance decreases (11–17 Å) as a consequence of the movement of loop-α, but never gets closer during MD simulation (more than 10 Å; Fig. 7b, black line), which confirms the hypothesis that Asp17 has lost its function as critical residue for the catalytic activity of VpEXPA2. The same analysis was performed for the two complexes with hemicellulose octasaccharide polymers, obtaining similar results. At the beginning of MD simulation with XXXGXXXG and XXFGXXFG polymers the distances between Asp17 and Asp104 were higher than 15 Å (Fig. 7b,

Table 3 List of residues which are located at ≤ 3 Å in the complex between VpEXPA2 and GGGGGGGG or XXFGXXFG

Residue	VpEXPA2-GGG GGGGG	VpEXPA2-XXFGXXFG	Protein location
Thr 11	*	*	a
Phe 12		*	a
Tyr 13	*	*	a
Gly 14		*	b
Gly 15		*	b
Gly 16		*	b
Thr 21		*	b
Gly 23		*	b
Gly 24	*		b
Ala 25	*	*	b
Gly 27		*	c
Ala 42		*	a
Asn 80	*	*	a
Phe 81	*		a
Cys 82	*		d
Phe86	*		d
Leu 88	*		d
Asn 91	*		d
His 102	*		e
Asp 104	*	*	e
Gly 119		*	a
Asn 141		*	f
Asn 178		*	f
Trp 179	*	*	f
Asn 182		*	f
Gln 184		*	f

The location of each residue within the VpEXPA2 structure is indicated as: β-barrel (a), loop-α (b), loop D1 (c), 'rich glycine loop' (d), active site (e), β-sandwich (f)

blue and red line). During MD simulation XXXGXXXG moves closer to Asp104, but the distance was still too long (around 12 Å) after 1.5 and 2.5 ns of MDS, and even increases during the last part of MD simulation (around 16 Å after 7.5 ns), reaching the minimum distance at 11 ns (around 12.5 Å) (Fig. 7b, blue line). In the case of XXFGXXFG, very similar dynamics behavior was showed, reaching the minimum distance at 11 ns (around 11 Å), which then quickly increased at 13 ns (15 Å), maintaining this distance relatively constant until the end of MD simulation (Fig. 7b, red line). Finally, the distance between the carboxylic group of Asp17 residue from VpEXPA2 and the O_4 from β-1,4 glycosidic bond of the ligands (cellodextrin 8-mer, XXXGXXXG and XXFGXXFG) (Fig. 7c) was studied during 20 ns of MD simulation. The results indicate distances higher than 9 Å and relatively constant until the end of MD simulation for XXXGXXXG and XXFGXXFG ligands (Fig. 7c, red and blue lines), which

confirms no interaction. The only exception was cellodextrin 8-mer ligand with distance higher than 10 Å at 11 ns of MD simulation, but reaching a minimum distance at 19 ns (around 6 Å) (Fig. 7c, black line).

Additionally, the stability of ligands in the open groove of VpEXPA2 was studied during 20 ns of MD simulation. In the case of hemicellulose octasaccharide polymers significant conformational changes were determined for both molecules tested, with RMSD values of around 10 Å for XXXGXXXG and 25 Å for XXFGXXFG (Fig. 8a, blue and red line, respectively). In the case of cellodextrin 8-mer polymer, high conformational stability was determined, with RMSD values between 2.5 and 4.5 Å (Fig. 8a, black line). Interestingly, the distance between both 'catalytic' Asp residues in all three molecular dynamics trajectories showed low RMSD values (below 2.5 Å) (Fig. 8b). However, the two complexes with hemicellulose octasaccharide polymers (XXXGXXXG and XXFGXXFG) displayed a greater fluctuation during the MD simulation, while in the complex with cellodextrin 8-mer the RMSD remained relatively constant (Fig. 8b).

Discussion

Since the discovery of the first expansin protein 23 years ago [1], a large multigene family has been described in a number of studies in different plant species. Several reports indicate that expansins participate in multiple and diverse developmental processes, including plant growth, pollination, abscission, adaptive responses to different stresses, and fruit softening [31]. Until now, studies on expansin action have been limited to protein crude extracts which provides limited information about its molecular mechanism of action, as the heterologous expression of EXPA from plant species has proven to be difficult [9]. In this context, a 3D model of an EXPA was built for the first time, by using comparative modeling and molecular dynamic simulation. The model obtained fulfills the structural characteristics proposed for EXPA protein family, including the two domains described previously and most probably related to its function [7]. Our approach was to use the structure of group 1 pollen β allergen (Phlp1) from *Phleum pratense* (an EXPB) as template, and carefully took into consideration insertions in the primary sequence that both expansin families differentially have [30], thereby avoiding possible structural distortions when building the model.

For D1, the catalytic domain, our model displays the three disulfide bonds in agreement with other expansins previously reported [4, 32]. However, an additional cysteine pair, highly conserved in α-expansins, suggests the possibility for a fourth disulfide bond. Unfortunately, this bond could not be modeled in our simulation, as a consequence of the long distance between both Cys residues (Cys 82–96; 19 Å) (data not shown), however the same pattern was found in ZmEXPB1 (Cys 58–
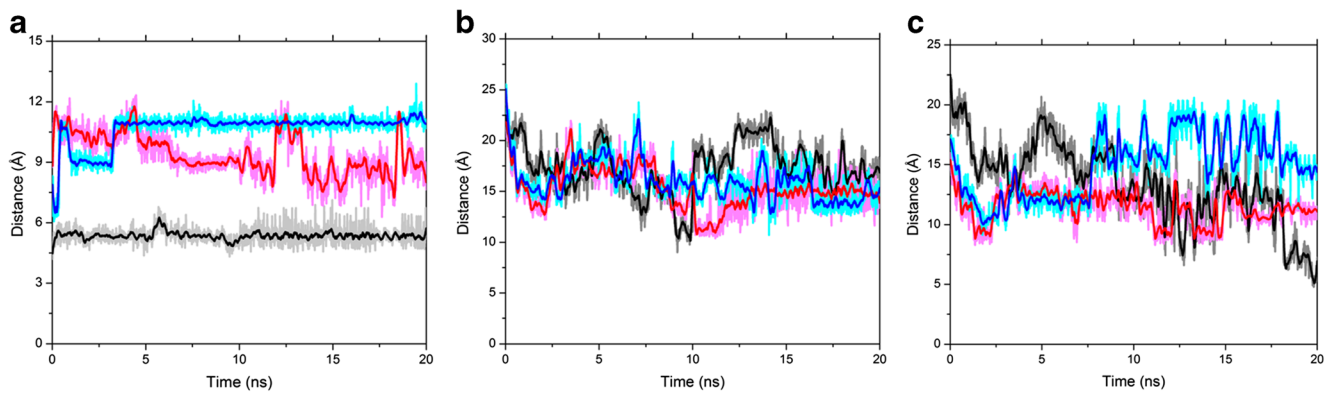
Fig. 7 a Distance between the carboxylic group of Asp104 and O_4 from β-1,4 glycosidic bond during 20 ns of MD simulation of VpEXPA2 in complex with cellodextrin 8-mer (*black line*), XXXGXXXG (*blue line*) and XXFGXXFG (*red line*). **b** Distance between Asp 17 and Asp 104 during 20 ns of MD simulation of VpEXPA2 in complex with cellodextrin 8-mer (*black line*), XXXGXXXG (*blue line*) and XXFGXXFG (*red line*). (**c**) Distance between carboxylic group of Asp17 and O_4 from β-1,4 glycosidic bond during 20 ns of MD simulation of VpEXPA2 in complex with cellodextrin 8-mer (*black line*), XXXGXXXG (*blue line*) and XXFGXXFG (*red line*)

128; 20 Å) [8]. Nevertheless, the strict conservation of three disulfide bonds between EXPA and EXPB families, suggests that they are critical for stabilization of the tertiary structure and for expansin's protein folding maintenance [8, 32]. The model also shows the triad HFD and a group of residues in the active site, which are highly conserved and essential for the catalytic function of GH45 protein family. Our model locates the catalytic site in a long groove, displaying most of the residues involved in the catalytic mechanism of GH45 enzymes with the same lateral chain orientation, albeit the second Asp residue required for the hydrolytic mechanism which is missing (Asp10 in EGV). Interestingly, when the nearest Asp residue was search in VpEXPA2 3D model, Asp17 was identified. In EGV protein the residues involved in catalysis are Tyr8 and Asp10, while in VpEXPA2 they correspond to Tyr13 and Asp17, and therefore they are two residues apart in papaya's expansin increasing the distance between Tyr and Asp from 4.8 to 14.8 Å. This amino acid pattern was

conserved in the several α-expansins analyzed (Fig. 5b) and also in other ripening-related α-expansins [11] (Fig. 3), and therefore, it could be possible that the insertion of two residues lead to α-expansins to loss their hydrolytic activity. No other Asp residue was found near Tyr in several β-expansins analyzed, in agreement with the result previously reported by Yennawar et al. (2006) [8].

Additionally, the existence of the 'glycine rich-like loop' matching with the structure of the 'glycine rich loop' from EGV (from Leu88 to Glu94 in VpEXPA2), suggests a possible function in the expansin's molecular mechanism, through a conformational change during the substrates binding, next to the catalytic cleft (Fig. 1a, yellow arrow). In VpEXPA2 this loop shows a moderate similarity with EGV's loop and shares the same spatial conformation, although it is longer. The most interesting observation is that one of the most important residues in this loop (Asp114 in EGV) has been mutated to Asn85 in VpEXPA2, changing its conformation and moving far from
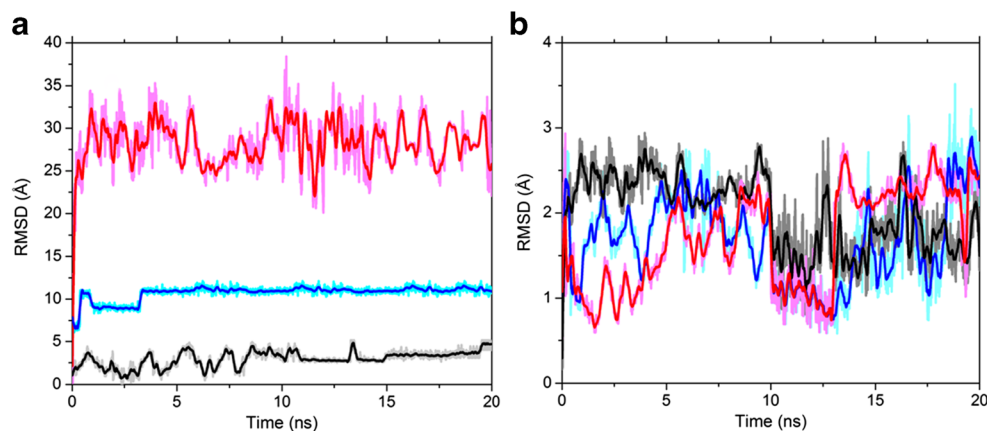

Fig. 8 a RMSD plot for carboxylic group of Asp104 and O_4 from β-1,4 glycosidic bond during 20 ns of MD simulation of VpEXPA2 in complex with cellodextrin 8-mer (*black line*), XXXGXXXG (*blue line*) and XXFGXXFG (*red line*). (**b**) RMSD plot for the distance between Asp17 and Asp104 during 20 ns of MD simulation of VpEXPA2 in complex with cellodextrin 8-mer (*black line*), XXXGXXXG (*blue line*) and XXFGXXFG (*red line*)

Asp104, probably as a consequence of Pro84 next to it (Figure S3). Interestingly, the mutation Asp114Asn designed in EGV by site-directed mutagenesis, reduced Kcat/Km by a factor of 160 and also the activity (20-fold less), suggesting an important role during catalysis [5]. In addition, a change of Asp for Asn92 appears between EGV and VpEXPA2 (Fig. 3a and b), which is located at the edge of the loop and solvent-exposed, plays an important role during MD simulation interacting with the cellodextrin, in the same way as the loop from EGV with cellobiose [5]. Sequence analysis of EXPA family shows Asn92 being highly conserved among all EXPA sequences (data not shown). These differences (in addition to the loss of catalytic Asp residue) could contribute to explain the loss of hydrolytic activity of expansins, if it occurred during the evolutionary process.

For D2, the cellulose-binding domain, the structure of mountain papaya expansin is in agreement with that of the pollen allergen from maize (ZmEXPB1). We also corroborate the existence of a strip along the open groove of polar, non-polar, and aromatic residues, which could interact with polysaccharides that covers both protein domains, as in ZmEXPB1 and carbohydrate-binding module (CBM) proteins [8, 33]. Interestingly, on the surface of the β-sandwich there are three aromatic residues (Tyr145, Phe146, and Trp179) forming a planar surface (Fig. 3a), typical from type A cellulose binding domain (CBM), which could provide a planar hydrophobic contact with crystalline polysaccharides (as cellulose microfibrils) through CH-π interactions [12]. Also this motif could facilitate the movement of the expansin protein on cellulose surface according to the mechanism proposed for cellulases [34]. In fact, our molecular docking simulations results show that VpEXPA2 can bind cellulose and hemicellulose polymers with different affinity energies. Nevertheless, the most significant data is the difference in the mode of interaction described, as the binding of VpEXPA2 with cellodextrin 8-mer polymer takes place through the 'glycine rich-like' loop and not in the case of the hemicellulose polymers tested (Fig. 6a and c). As a consequence of the different mode of interaction the distance between Asp104 and the closest O_4 β-1,4 glycosidic bond is higher in the case of hemicellulose polymers (6.9 Å) compared to cellulose (4.2 Å) (Fig. 6b and d). A second relevant observation is that despite molecular docking simulations were performed using a grid for the entire protein, in all simulations, no matter the substrate type, the substrate was bound primarily to D1, suggesting a lower affinity of D2 for the substrates tested. These results are in agreement with previous reports showing that binding affinities of D2 from the *Bacillus subtilis* (BsEXLX1) expansin and other type-A CBMs are 1000 times lower to soluble cellooligosaccharides than cellulose, mainly as a consequence of high affinity for more ordered cellulose regions (as crystalline microfibrils) [12].

A comparison between the conformations of cellodextrin 8-mer and hemicellulose octasaccharide polymers during 20 ns of MD simulations shows that the first one binds to VpEXPA2 protein in energetic and structurally more stable manner than the second ones (Figs. 7 and 8). However the persistent large distance between the substrate and important residues for catalysis, Asp17 and Asp104, do not allow any polysaccharide hydrolytic activity, in accordance to experimental data previously reported [2]. Additionally, the binding of cellodextrin 8-mer and hemicellulose octasaccharide polymers is not only with D1 of VpEXPA2, but also includes part of D2. This was not notorious in our system due to the short size of the ligands tested compared to the real substrate size in the cell wall. Therefore, we share the hypothesis that the function of domain D1 is highly dependent on the binding of cellulose microfibril to D2 domain. The inclusion of D2 domain provides a long open groove for substrate binding in agreement with the structure of ZmEXPB1 [8]. In this context, recent data shows that a functional D2 from FaEXPA2 has affinity not only for cellulose but also xylan and pectin, although with a significant less affinity for the last two [35], in agreement with previous reports that proposed cellulose-hemicellulose interface as plant expansin's target [7]. Finally, it could be very interesting to test in our homology model of VpEXPA2 the affinity for xylan, pectin, and a microfibril of cellulose, which could be a more realistic framework of real substrates for plant expansin proteins.

Conclusions

In summary, great advances have been made concerning the expression patterns of different members of the expansin family in a number of species. Although the regulation of gene expression by different plant hormones, environmental stimuli and growth factors, and information about the structure of some promoters have helped us to gain insight about the regulation of expansin genes, more effort is needed to understand the mechanism of action at the molecular level of the expansin family protein. With the reports of experimental molecular structure of EXPB from maize, the main structural characteristics of plant expansins were unveiled [8]. However, how much in common is shared within the EXPA family? In this study, the *in silico* prediction of VpEXPA2 structure presented here successfully support most key structural features from other previously reported expansins, like the two-domain structure, with a conserved folding of β-barrel for D1 and β-sandwich for D2, forming an extended open groove that spans both domains, that could bind cellulosic and hemicellulosic substrates, with a higher affinity for the first ones.

One of the intriguing issues about expansin is the limited but significant homology of D1 with GH45 family protein, which has cellulase activity. The analysis of VpEXPA2 D1 structure shows a high structural conservation of composition and conformation of active site, with the exception of catalytic Asp17. Moreover, the amino acidic insertions increase the distance between both catalytic Asp residues leading to the loss of their hydrolytic activity. Interestingly, in EXPB the second catalytic Asp residue was missing [8]. A second difference that could lead to loss of the hydrolytic activity in EXPA was found in a loop which is key to substrate stabilization in EG45, through the changes of key Asp residue.

Moreover, docking simulations determine that VpEXPA2 has a lower binding energy for cellulosic substrate (cellodextrin 8-mer) than for hemicellulosic ones, which was in agreement with the previous studies that indicates a cellulose binding preference, compared to other cell wall polymers [2, 10, 35]. Therefore, VpEXPA2 model constitutes a useful tool to guide experimental hypotheses aimed at improving our understanding on the mechanism of action of α-expansins on plant cell walls.

References

1. McQueen-Mason S, Durachko DM, Cosgrove DJ (1992) Two endogenous proteins that induce cell wall expansion in plants. Plant Cell 4:1425–1433

2. McQueen-Mason S, Cosgrove DJ (1994) Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. Proc Natl Acad Sci U S A 91:6574–6578

3. Kende H, Bradford K, Brummell D, Cho HT, Cosgrove DJ, Fleming A, Gehring C, Lee Y, McQueen-Mason S, Rose JK, Voesenek LA (2004) Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol 55:311–314

4. Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242

5. Davies GJ, Tolley SP, Henrissat B, Hjort C, Schülein M (1995) Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1.9 Å resolution. Biochemistry 34: 16210–16220

6. Fedorov AA, Ball T, Valenta R, Almo SC (1997) X-ray crystal structures of birch pollen profilin and Phl p 2. Int Arch Allergy Immunol 113:109–113

7. Cosgrove DJ (2000) Loosening of plant cell walls by expansins. Nature 407:321–326

8. Yennawar NH, Li L-C, Dudzinski DM, Tabuchi A, Cosgrove DJ (2006) Crystal structure and activities of EXPB1 (Zea m1), a β-expansin and group-1 pollen allergen from maize. Proc Natl Acad Sci U S A 103:14664–14671

9. Georgelis N, Tabuchi A, Nikolaidis N, Cosgrove DJ (2011) Structure-Function analysis of the bacterial expansin EXLX1. J Biol Chem 286:16814–16823

10. Whitney SEC, Gidley MJ, McQueen-Mason S (2000) Probing expansin action using cellulose/hemicellulose composites. Plant J 22:327–334

11. Gaete-Eastman C, Figueroa CR, Balbontín C, Moya M, Atkinson RG, Herrera R, Moya-León MA (2009) Expression of an ethylene-related expansin gene during softening of mountain papaya fruit. Postharvest Biol Technol 53:58–65

12. Georgelis N, Yennawar NH, Cosgrove DJ (2012) Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. Proc Natl Acad Sci U S A 109:14830–14835

13. Hanus J, Mazeau K (2006) The xyloglucan-cellulose assembly at the atomic scale. Biopolymers 81:59–73

14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DF (1990) Basic local alignment search tools. J Mol Biol 215:403–410

15. Jonnes DT (1999) Protein secondary structure prediction based on position specific scoring matrices. J Mol Biol 292:195–202

16. Thompson J, Higgins D, Gibson T (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

17. Šali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234:779–815

18. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculation. J Comput Chem 4:187–217

19. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802

20. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field M, Fischer JS, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1992) Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. FASEB J 6:A143

21. Schlenkrich M, Brickmann J, MacKerell AD Jr, Karplus M (1996) In: Roux B, Merz KM (eds) A molecular perspective from computation and experiment. Birkhauser, Boston, pp 31–81

22. Jorgensen WL, Chandresekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

23. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17:355–362

24. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291

25. Lüthy R, Bowie JU, Eiserberg D (1992) Assessment of protein models with three- dimensional profiles. Nature 356:83–85

26. Trott O, Olson AJ (2010) AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31:455–461

27. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force field. J Comput Chem 31:671–690

28. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593

29. Humphrey W, Dalke A, Schulten K (1996) VMD - visual molecular dynamics. J Mol Graph 14:33–38

30. Li Y, Darley CP, Ongaro V, Fleming A, Schipper O, Baldauf SL, McQueen-Mason SJ (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin. Plant Physiol 128:854–864

31. Cosgrove DJ, Li LC, Cho H-T, Hoffmann-Benning S, Moore RC, Blecker D (2002) The growing world of expansins. Plant Cell Physiol 43:1436–1444

32. Cosgrove DJ, Bendinger P, Durachko D (1997) Group I allergen of grass pollen as cell wall-loosening agents. Proc Natl Acad Sci U S A 94:6559–6564

33. Gilbert HJ, Knox JP, Boraston AB (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr Opin Struct Biol 23:669–677

34. Jervis EJ, Haynes CA, Kilburn DG (1997) Surface diffusion of cellulases and their isolated binding domains on cellulose. J Biol Chem 272:24016–24023

35. Nardi C, Escudero C, Villarreal N, Martínez G, Civello PM (2013) The carbohydrate-binding module of *Fragaria x ananassa* expansin 2 (CBM-FaEXP2) binds to cell wall polysaccharides and decreases cell wall enzymes activities 'in vitro'. J Plant Res 126:151–159