HAL open science

Endopolygalacturonases reveal molecular features for processivity pattern and tolerance towards acetylated pectin

Gwenaëlle André-Leroux, Dominique D. Tessier, Estelle Bonnin

▶ **To cite this version:**

Gwenaëlle André-Leroux, Dominique D. Tessier, Estelle Bonnin. Endopolygalacturonases reveal molecular features for processivity pattern and tolerance towards acetylated pectin. Biochimica et Biophysica Acta Proteins and Proteomics, 2009, 1794 (1), pp.5-13. 10.1016/j.bbapap.2008.09.004 . hal-02660618

HAL Id: hal-02660618

https://hal.inrae.fr/hal-02660618

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Biochimica et Biophysica Acta 1794 (2009) 5–13

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Endopolygalacturonases reveal molecular features for processivity pattern and tolerance towards acetylated pectin

G. André-Leroux [a], D. Tessier [b], E. Bonnin [b,*]

[a] Institut Pasteur, Unité de Biochimie Structurale, 25 rue du Dr Roux, 75724 Paris Cedex 15, France
[b] INRA, UR-1268 Biopolymères Interactions Assemblages, BP 71627, 44316 Nantes Cedex 3, France

ARTICLE INFO

Article history:
Received 30 January 2008
Received in revised form 1 September 2008
Accepted 6 September 2008
Available online 25 September 2008

Keywords:
Acetylated homogalacturonan
Tolerance profile
Processivity pattern
Oligogalacturonates docking
Normal mode analysis

ABSTRACT

Endopolygalacturonases (EndoPGs) hydrolyse the 1–4 linkages between two α-D-galacturonic acids (GalA) of the smooth homogalacturonan regions of pectin. GalA may be methyl-esterified on the carboxylic group and acetyl-esterified on the hydroxylic groups. EndoPG activity most often decreases with such increasing degree of substitution. In this paper, we used bioinformatics and molecular modelling technics to explain the tolerance profile at the molecular scale and processivity scheme of three endoPGs with respect to acetylated pectin substrate; the first two enzymes originate from *Aspergillus niger* (*An*PGI and *An*PGII) and the third from *Fusarium moniliforme* (*Fm*PG). Partly acetylated and methylated homogalacturonan fragments in complex with the three PGs were successively modelled *in silico*. The amino acid residues involved in substrate binding were identified for each enzyme. Similarly, the docking pattern of the differently decorated oligomers in the catalytic groove was individually characterized for each enzyme. This work shows full agreement with our previous extensive mass spectrometry analysis of the hydrolytic products that established distinct tolerance profiles for the three endoPGs and earlier work that ascertained processivity, specifically for *An*PGI. In our previous work, *An*PGI was shown to be the most powerful enzyme among the three enzymes with an enhanced tolerance towards O2- and O3-acetylated substrates. We report here amino acids of *An*PGI that are unique in binding the pectin backbone and that are identified as possibly crucial for its specificity, namely $S191_{AnPGI}/D240_{AnPGI}$. Similarly, topologically equivalent residues in *An*PGII and *Fm*PG were identified that could impede such binding; $S234_{AnPGII}/S91_{AnPGII}$ and $S245_{FmPG}/V89_{FmPG}$. In addition, we report here, from normal mode analysis computed on *An*PG1, a shear bending motion of 15 Å of amplitude that fully accredits the processive action pattern for this enzyme, with $D240_{AnPGI}$ and $R96_{AnPGI}$ working as crampons to favour the sliding of the substrate. Conversely, the same method clearly evidences a hinge binding motion for *An*PGII and *Fm*PG that should only authorize one hydrolytic event per enzyme/substrate encounter.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Pectins are among the most abundant polysaccharides in many plant primary cell walls. Together with hemicelluloses, they define the two major networks in which the cellulose microfibrills are embedded. Pectin molecules are composed of different structural regions, which have been extensively described in previous reviews [1–3]. The homogalacturonan regions (HG) consist of a repetition of α (1, 4)-linked D-galacturonic acid (GalA). The rhamnogalacturonan regions *I* (RG*I*) are constituted by an alternating sequence of α (1, 4)-linked D-galacturonic acid and α (1, 2)-linked rhamnopyranosyl residues. Various side chains (mainly arabinan, galactan or arabinogalactan) can link those rhamnose residues. The nature and lengths of the side chains vary among plant species. GalA in pectin can be partially methyl and/or acetyl-esterified. The degree of substitution (including acetylation) is defined as the number of substituents (acetyl groups) for each 100 GalA. A minor type of rhamnogalacturonan (RG*II*) shows a very complex structure composed of 12 different monomers. The structures of the HG and RG regions are quite well known but their way of combination remains unclear [4].

Basically, homogalacturonan is the best substrate for endopolygalacturonases (poly [1, 4-α-D-galacturonide] glycano-hydrolase, EC 3.2.1.15, endoPGs) as endoPGs hydrolyse α (1, 4) linkages between two GalA residues [5]. They are widely distributed in plants, fungi, yeasts and bacteria [6] and are the subject of a thorough research including their purification, properties and mode of action [for reviews see [7,8]]. Based on their sequences and their structurally related catalytic folding, endoPGs have been classified in glycoside hydrolase family 28, GH28 [9] (http://afmb.cnrs-mrs.fr/CAZY/). It has been shown that family 28 enzymes cleave glycosidic bonds with inversion of the anomeric configuration [10]. Their activity is highly modulated by the presence of non-sugar substituents such as methyl and acetyl groups [11–13]. In a previous paper, we investigated the action pattern of

* Corresponding author.
E-mail address: bonnin@nantes.inra.fr (E. Bonnin).

1570-9639/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2008.09.004

Fusarium moniliforme PG (*Fm*PG) towards poorly to highly methylated pectin fragments using molecular modelling and bioinformatics [14]. The tolerance of the enzyme towards methyl groups was understood in comparison with *An*PGII known to be strictly non-tolerant to methyl-esters [12]. Amino acids that could have an impact on the behaviour of each enzyme towards defined methylated substrates were identified. Afterwards, the study was extended to acetyl-esters. An extensive degradation of highly acetylated pectin from sugar beet was carried out by *An*PGI, *An*PGII and *Fm*PG and their hydrolysis products were carefully analyzed by mass spectrometry [15]. From that study, the most depictive products were selectively docked into each of the PG grooves. Our docking energies all showed accurate correlation with the hydrolytic schemes of the three PGs. In addition to the molecular docking, the electrostatic signature as well as the intrinsic flexibility through normal mode analysis was computed and assessed for each PG. Those molecular features clearly impact the tolerance of *An*PGI, *An*PGII and *Fm*PG towards acetyl groups and amino acids that could guarantee or impede the processive hydrolysis of the PGs have been identified.

2. Materials and methods

2.1. Coordinates of the enzymes

We used the atomic coordinates of the native *An*PGI, *An*PGII and *Fm*PG proteins that were solved at 1.70, 1.68 and 1.73 Å and deposited in the PDB bank under accession number 1nhc, 1czf and 1hg8, respectively (http://www.rcsb.org/pdb/home/home.do) [16–18]. They share the same single stranded right-handed β-helix topology, the same catalytic set of aspartate residues $D191_{FmPG}$, $D212_{FmPG}$ and $D213_{FmPG}$ (*Fm*PG numbering) and thus the same inverting catalytic mechanism.

2.2. Sequence alignment

The sequences of *An*PGI, *An*PGII and *Fm*PG were retrieved from the PDB databank in the PDB format (http://www.rcsb.org/pdb/home/home.do), whereas the precursors of the proteins were extracted from the Uniprot database with P26213, P26214 and Q07181 accession numbers, respectively. Multiple alignments were performed with the ClustalW algorithm with standard parameters.

2.3. Construction of the substrates

The following abbreviations will be used to describe the oligomers produced by the enzymes: G is the shortcut of GalA, M the one of methylated GalA, A_2 the abbreviation of GalA acetylated on *O*-2 and finally A_3, the abbreviation of GalA acetylated on *O*-3. For instance, A_3MG is a trimer in which the non reducing end is acetylated on *O*-3, the central residue is methylated and the reducing end is free of substituent. In addition, there is no way to discriminate from which side of the product, the cleavage occurred. For instance, the fraction GA_2 could either come from the hydrolysis of GG/GA_2 or GA_2/GG, so the two dockings shall be considered. The decorated oligomers were built on the parental scaffold of the linear GalA oligomer and positioned with respect to their position in the complex with *Fm*PG [14]. To position the substrates into *An*PGI and *An*PGII crevices, the two PGs were successively superimposed onto the complex *Fm*PG/oligoGalA, using the pairwise structure tool from the ebi website (http://www.ebi.ac.uk/DaliLite/) [19].

2.4. Molecular modelling

2.4.1. Electrostatic profile

The electrostatic profile was calculated for each of the three PGs using the APBS (Adaptive Poisson–Boltzmann Solver) tool interfaced in Pymol (http://www.pymol.org) [20,21] after a PDB2PQR conversion [22] using most of the default parameters except the pH that was set up at 4, where the degradations were carried out [15]. The former program evaluates the electrostatic properties of nanoscale biomolecular systems using the Poisson-Boltzmann equation (PBE), which is one of the most popular continuum models for describing electrostatic interactions between molecular solutes in salty, aqueous media. The latter program includes several steps as it first adds a limited number of missing heavy atoms to biomolecular structures, determines side-chain pKas, places missing hydrogens, optimizes the protein for favourable hydrogen bonding and finally assigns charge and radius parameters for a given forcefield; here AMBER was chosen.

2.4.2. Normal mode analysis

Normal mode analysis is a classical technique to generate functionally relevant movements in biological macromolecules. The principle is to use an all atom model with a highly simplified and quadratic potential energy between atoms that are considered to be linked by a spring of universal strength provided that they are located less than 10 Å away in the PDB structure [23]. For low-frequency normal modes, which give large-amplitude and collective modes, this Elastic Network Model gives excellent results. We used the corresponding NOMAD program to compute the flexibility of each PG into motion modes of low frequency (http://lorentz.immstr.pasteur.fr/nomad-ref.php) [24]. With the PDB coordinates of the PGs as input files, NOMAD calculates and explores each motion shaping that associates large-amplitude movements and correct stereochemistry.

2.4.3. Docking of substrates

Molecular docking was carried out on an SGI computer with the Accelrys© package (Accelrys. Inc, San Diego, CA, USA). Molecular displays and energy minimisations were performed using the molecular and simulation modelling environment InsightII, combining Biopolymer and Discover modules. When necessary, tools of the recently transferred InsightII version of linux Discovery Studio© have been used. To compare our data with our previous work on *Fm*PG complexed to methylated pectin, similar cycles of docking and optimization were used. The strategy implies the CFF91 force field and steepest descent minimization algorithm during which the backbone of the protein was kept fixed and the side chains and sugar moieties allowed to relax (15,000 iterations) [14,25,26]. We computed similar energetic criteria as previously defined: 1. potential energy of the protein, 2. potential energy of the substrate, 3. potential energy of the complex and 4. binding energy of the complex, which is exclusively the contribution of the residues involved through van der Waals and electrostatic interactions with the oligosaccharide. The binding energy can be directly correlated to the stability.

3. Results

3.1. Docking of oligomers in the catalytic groove

From our previous experimental study, GalA trimer, tetramer and pentamer were found to be the archetypal products resulting from pectin hydrolysis by the three enzymes [15]. They are all methyl- and/or acetyl-esterified and released in appreciable quantities whatever the enzyme used. A_3MG is the most predominant trimer produced. The corresponding substrate A_3MG/GG was thus built for subsequent docking in the catalytic groove of each PG. Conversely, DP_4 and DP_5 are delivered with various substitution patterns where the most abundant configurations are GA_2MG and GA_2MGG. Similarly, GA_2MG/GG and GA_2MGG/GG were built for further docking. The binding energies for all the modelled complexes are registered in Table 1. *An*PGI evidences stronger binding, regardless of the substrate length, as compared to the two others. *An*PGI shows the best binding energy for GA_2MG/GG with −186 kcal/mol, so when GA_2MG spans from subsite −4 to −1. This

Table 1
Binding energies (in kcal/mol) for *An*PGI, *An*PGII and *Fm*PG in complex with oligogalacturonates

Binding energy	*An*PGI	*An*PGII	*Fm*PG
GA_2/GG	−135	−112	−115
GG/ GA_2	−131	−113	−124
GA_3/GG	−114	−94	−119
GG/ GA_3	−135	−106	−108
A_3MG/GG	−141	−113	−140
GA_2MG/GG	**−186**	−129	−159
GA_2MGG/GG	−165	**−150**	**−176**

The most stable complex for each enzyme, suggested by the lowest energy, is highlighted in bold.

is fully relevant with the product quantification [15], where GA_2MG was measured as the oligomer released in the highest quantity by *An*PGI action. *An*PGII and *Fm*PG both display much lower binding energies for those short substrates. Thus, the three enzymes show a binding profile that correlates very well with the experimental hydrolytic pattern. Particularly, the percentage of products with a degree of polymerisation lower than 6 was markedly lower for *An*PGII and *Fm*PG, 64% and 56% of total GalA respectively as compared to 75% for *An*PGI. In addition, the docking study shows that *An*PGII and *Fm*PG, with binding energies of −150 and −176 kcal/mol respectively, give more stable complex with the fraction GA_2MGG/GG. This is fully consistent with the experimental data that assessed this moiety as the most abundant oligomer produced by those enzymes [15].

Our docking study evidences that *An*PGI gives more stable complexes for oligosaccharides with A_2 and A_3 acetylations as compared to the two other PGs. This is again in full accordance with the experimental data that assessed *An*PGI as the most potent enzyme and the most tolerant towards acetylated fragments [15]. In more details, Table 1 shows that the three enzymes display a markedly higher binding energy for A_2 acetylation as compared to A_3, at subsite −3. For instance, for *An*PGII, −129 kcal/mol for GA_2MG/GG has to be compared to −113 kcal/mol for A_3MG/GG, and the difference is in the same order whatever the enzyme. Similarly at the subsite −1, A_2− is better tolerated than A_3− acetylation, as for instance, −135 kcal/mol for GA_2/GG compared to −114 kcal/mol for GA_3/GG in the case of acetylated GalA in subsite −1 of *An*PGI. This is consistent with higher amount of A_2 oligosaccharides compared to A_3 ones that were quantified after enzymatic digestion [15]. The only exception is for subsite −1 in *Fm*PG that shows rather equal binding affinity between GA_2/GG and GA_3/GG with −119 kcal/mol and −114 kcal/mol, respectively. In any case, it is consistent with the tolerance profile of the three PGs that all accept A_2 and A_3 in subsite −1.

3.2. Comparison of the three proteins

GH28 family enzymes are conformed as single stranded right-handed beta-helix fold. The catalytic triad $D191_{FmPG}$, $D212_{FmPG}$, $D213_{FmPG}$, plus $H188_{FmPG}$, $N189_{FmPG}$, $T190_{FmPG}$, $H234_{FmPG}$, $R267_{FmPG}$ and $K269_{FmPG}$ are strictly conserved among the overall family (see alignment of *An*PGI, *An*PGII and *Fm*PG on Fig. 1) [16–18]. These conserved residues are involved in the subsites −1 and +1 having the highest affinity for GalA [14]. Sequence analysis of the three enzymes, through 1D alignment, evidences a high homology degree with 30% strict identity shared. As expected and already published [16], *An*PGI and *An*PGII share 57% strict identity. *Fm*PG sequence is slightly more similar to *An*PGII sequence than to *An*PGI with aligned sequence score of 43% and 40%, respectively. The pairwise structure comparison by 3D alignment using *An*PGI as the reference shows an average rms

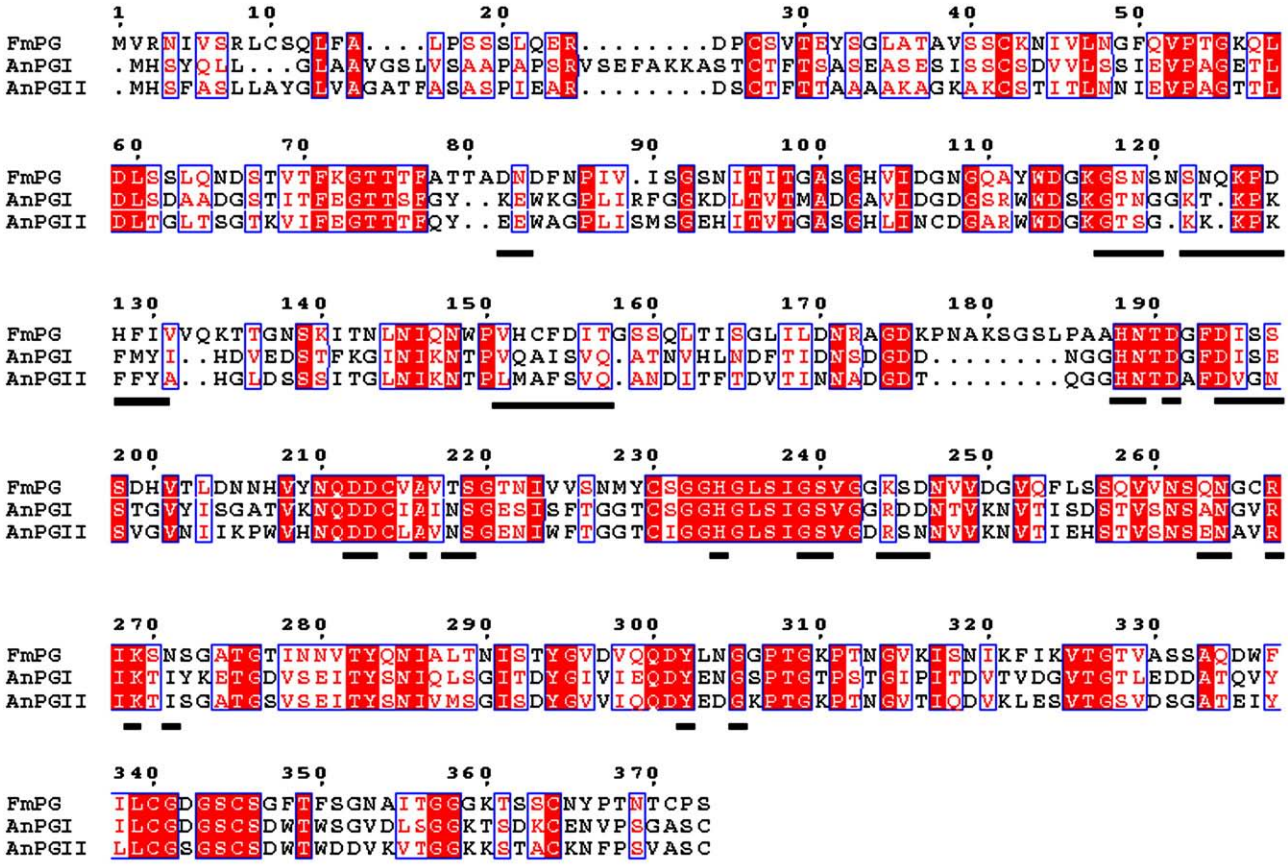

Fig. 1. ClustalW alignment of precursors of *An*PGI, *An*PGII and *Fm*PG proteins from the Uniprot database with P26213, P26214 and Q07181 code respectively. The underlined amino-acids belong to the catalytic groove of at least one of the three enzymes.

deviation of 0.8 Å and 1.5 Å between equivalent Cα for *An*PGII and *Fm*PG, respectively. Basically, the residues that cover the catalytic cavity are very well conserved among the three PGs. Still, some discrepancies on the sequences can be pinpointed as they are visible on the clefts and could be crucial for substrate tolerance or processivity pattern. The regions of the 1D sequence covering the catalytic groove are underlined on Fig. 1 below the alignment. Some notable differences in or next to the catalytic clefts are one amino-acid insertion $N83_{FmPG}$, one amino acid insertion $R96_{AnPGI}/S91_{AnPGII}$ that 3D aligned with $V89_{FmPG}$, a motif of low complexity $S118_{FmPG}NSN123_{FmPG}$, an eight-residue insertion-$P177_{FmPG}$-$L184_{FmPG}$, $D240_{AnPGI}/S234_{AnPGII}$ that 3D aligned with $S245_{FmPG}$ and eventually the variation region

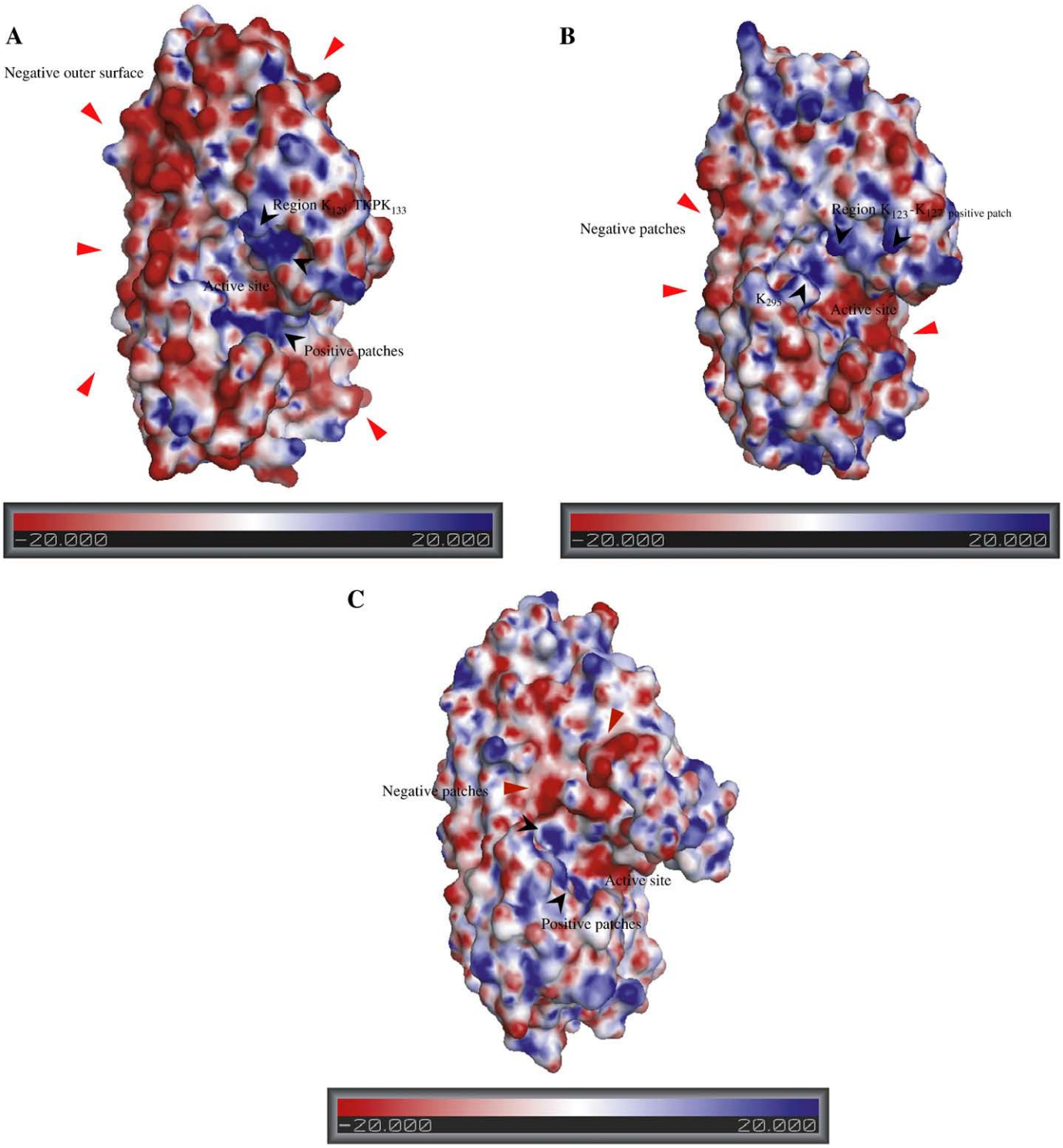

Fig. 2. APBS Electrostatics of *An*PGI (A), *An*PGII (B) and *Fm*PG (C) contoured at pH 4. The residues coloured in red are negative, the ones coloured in blue are positive and the ones in white are neutral. The negative patches are highlighted with red arrows, the positive patches with black arrows. The negative active site, due to the three conserved Aspartate residues, is mentioned. The electrostatics has been homogeneously scaled between −20 and +20 for each.

L303$_{FmPG}$–G306$_{FmPG}$. Only V89$_{FmPG}$ and residues equivalent in *An*PGs (R96$_{AnPGI}$/S91$_{AnPGII}$) belong to a structural β strand, the other residues belong to loops or to T3 turns.

The calculation of the electrostatic surface clearly shows distinct profiles for the three PGs (Fig. 2A–C). At pH 4, where the degradation was carried out [15], *An*PGI, *An*PGII and *Fm*PG have a total charge of −6, +9 and +5, respectively. The distribution of charge in *An*PGI evidences an outer surface highly negative combined to a catalytic groove that contains the positive lysine rich motif K129$_{AnPGI}$TKPK133$_{AnPGI}$, where K133$_{AnPGI}$ is in position to interact with the substrate at subsites −2/−1, consequently highly attractive to a negative demethylated pectate (Fig. 2A). *An*PGII shows an outer surface less negative and a catalytic crevice slightly more negative as compared to *An*PGI. However, the enzyme associates positive patches at the apex of two loops that face each other apart from the fissure, on one side the lysine cluster, K123$_{AnPGII}$–K125$_{AnPGII}$ and K127$_{AnPGII}$, and on the other side K295 (Fig. 2B). Finally, the electrostatic footprint is less pronounced for *Fm*PG that displays a balanced and scattered positive/negative outer surface and the inner fissure the less charged among the three (Fig. 2C).The docking study maps the subsites in *An*PGI and *An*PGII and evidences that both enzymes have a 7 subsites active site that split into five (−5 to −1) and two (+1 to +2) apart from the catalytic site that spans subsites −1 and +1. Similarly, *Fm*PG active site defines up to 7 subsites that cover from −5 to +2, thus confirming an additional subsite when compared to our previous results [14]. From the computational docking, we were able to list precisely the amino acids involved in the binding of oligogalacturonates through apolar, polar and electrostatic contributions (Table 2). Interestingly, remarkable differences in the electrostatic signature of each PG can be zoomed in at subsites −4, −5 and +2 for the three enzymes, the inner subsites being most homogeneous. *An*PGI clearly shows a negatively charged surface at subsites −4 and −5 with E192, D240 and D241 and, reversely, a highly positively charged surface at subsite +2 with H183, H229 and R262 (Fig. 3A). *An*PGII displays a highly polar surface at subsites −4 and −5 with residues Q155, N186, S234 and N235 and a polar charged at subsite +2 with H177, H223, E252 and R256 (Fig. 3B). Finally, *Fm*PG stands in between the two other profiles with a highly polar surface at subsites −4 and −5 with residues T157, S197, S245 and S246 and a highly positively charged surface at subsite +2 with residues H188, H234 and H267 (Fig. 3C).

Table 2
List of amino acids of *An*PGI, *An*PGII and *Fm*PG involved, either through hydrogen bonds, van der Waals interactions or salt bridges, in the binding of decorated oligogalacturonates

	***An*PGI**	***An*PGII**	***Fm*PG**
Subsite - 5	Q161	Q155	T157
	E192	N186	S197
	D240	**S234**	**S245**
Subsite - 4	Q161	Q155	T157
	E192	N186	S197
	D240	**S234**	**S245**
	D241	N235	D246
Subsite - 3	K133	K127	H128
	S159	**S153**	**D155**
	V160	V154	I156
	S191	V184	S196
	N213	S208	S219
	D240	**S234**	**S245**
Subsite - 2	D189	D183	D194
	I190	V184	I195
	A211	A205	A216
	N213	N207	T218
	R239	R233	K244
Subsite - 1	T125	T120	S118
	V155	L149	V151
	Q156	M150	H152
	D186	*D180*	*D191*
	G234	G228	G239
	S235	S229	S240
	K264	K258	K269
Subsite + 1	N126	S121	N119
	N184	N178	N189
	D207	*D201*	*D212*
	D208	*D202*	*D213*
	Y297	Y291	Y302
	S301	K295	G305

The catalytic residues are in italic. Amino acids with non polar side chains are in black, amino acids with uncharged polar side chains are in green, amino acids with positively and negatively charged polar side chains are in blue and red, respectively. Amino acids in bold are proposed for site directed mutation.

3.3. Processivity pattern through NOMAD analysis

The default parameters of NOMAD compute 16 motion modes. The first six are discarded as they refer to the own translations and rotations of the protein. The first motion mode that can be relevant starts at 7. Indeed, the motion mode analysis evidences remarkable flexibility at modes 7 and 8 that are shear bending and hinge bending motions for the three PGs. Shear bending flexibility is characterized by the glissade of the N-terminus portion along the C-terminus one, apart from the catalytic cleft (See arrows in Fig. 4A). The hinge bending motion describes the successive opening or closure of the same domains over the active groove (See arrows in Fig. 4B, C). Interestingly, despite a strictly conserved topology and a high sequence identity, the differences in scaling motions between hinge and shear bending are outstanding between *An*PGI on one hand and *An*PGII and *Fm*PG on the other hand. Clearly, *An*PGI promotes a straightforward shear bending motion with measured amplitude of 15 Å between the Cα of N126$_{AnPGI}$ and the carbonyl oxygen of N299$_{AnPGI}$. Those residues define the two extreme positions along the trajectory (Fig. 4A). Indeed, N126$_{AnPGI}$ and N299$_{AnPGI}$ are located at the apex of loops apart from the catalytic site and submitted to the most substantial amplitude of motions. The distance variations between the two residues, measured during hinge bending motion exhibits a rather weak amplitude of 5 Å in *An*PGI. Reversely, *An*PGII and *Fm*PG definitely favour the opening/closure flexibility of the hinge bending motion as shown by the amplitudes of 24 Å and 18 Å, respectively, measured between Cα and carbonyl oxygen of the homologous residues S121$_{AnPGII}$, D293$_{AnPGII}$ and N118$_{FmPG}$, N304$_{FmPG}$ (Fig. 4B, C). It is noteworthy that the shear bending motion is weaker for *An*PGII and *Fm*PG with amplitudes of 4.5 Å and 9 Å, respectively.

4. Discussion

From previous experimental data and our *in silico* calculations, it is clear that the three PGs, despite sharing a common architecture, a similar subsite distribution with a strictly conserved catalytic mechanism, diverge in substrate binding specificity, in tolerance behaviour towards acetylated or methylated pectins and in processivity pattern. We advance here molecular clues that should explain those macromolecular properties and we propose site-directed mutations to validate our assumptions.

It was previously demonstrated that acetyl groups *O*-2 or *O*-3 on homogalacturonan GalA backbone are energetically favourable and that the presence of acetyl groups did not alter the conformational behaviour of the backbones [28]. It can also be calculated from the length of C–O bond (1.43 Å) and C–H bond (1.07 Å) that the presence of an acetyl group induces a limited protuberance of around 18 Å^3. Compared to the volume of the active site cavity (849 Å^3 on average), it can be assumed that acetyl groups on the homogalacturonan backbone do not modify very much the geometry and the volume of the substrate and instead could favour the binding through hydrogen bonding. Indeed, we report for *An*PGI possible hydrogen bonding between H138, S159, S191, E192 and R239 and acetyl group when positioned as A_2, or between Q161 and R239 and acetyl groups when positioned as A_3. Similarly for *An*PGII, we list possible hydrogen bonding between R233 and N207 and acetyl group A_2. No hydrogen

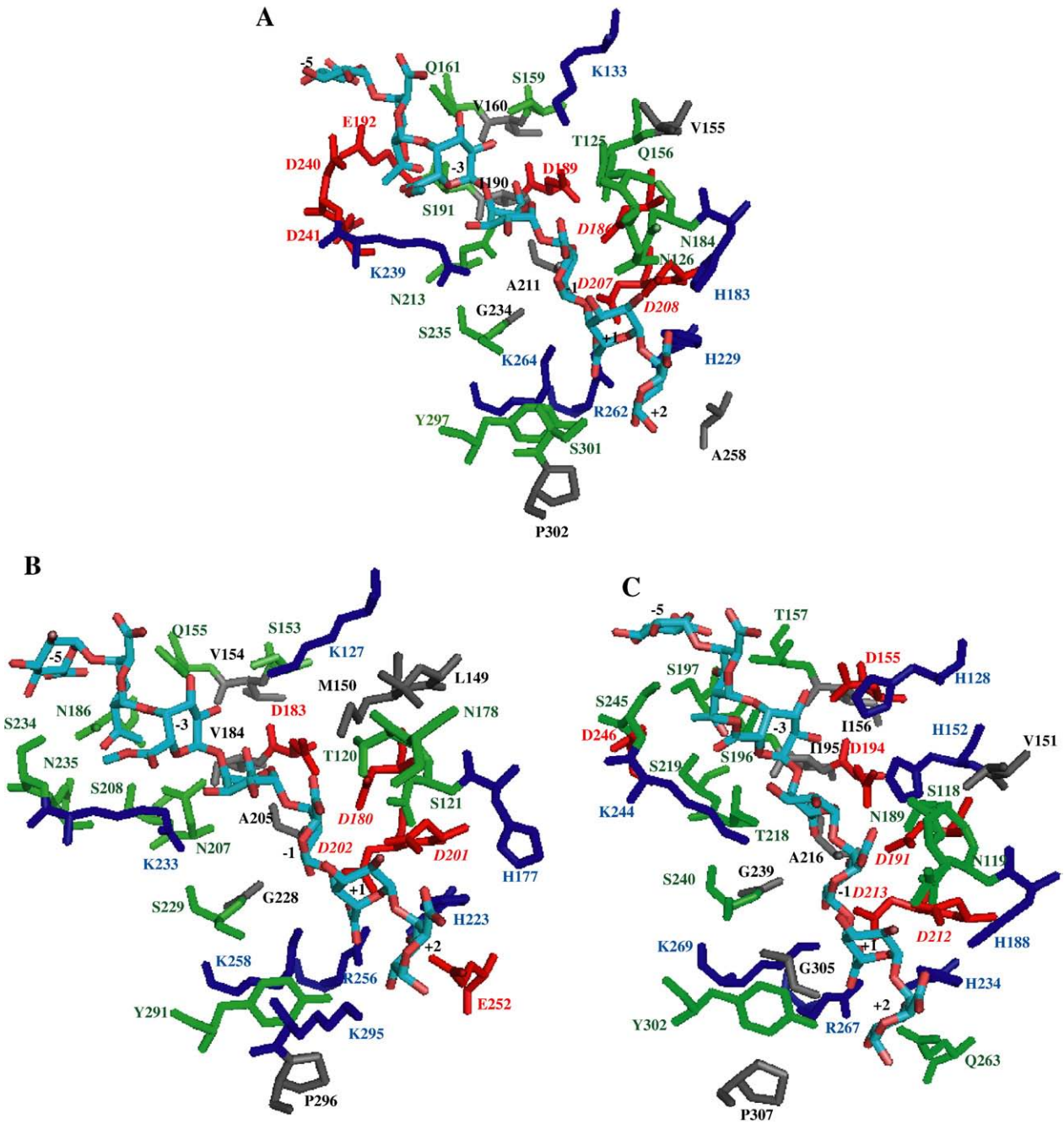

Fig. 3. Close view of the Acetylated/methylated GA_2MGG/GG oligomer docked in the catalytic groove of *An*PGI (A), *An*PGII (B) and *Fm*PG (C), respectively. The substrate GA_2MGG/GG is sticked and atom type coloured in cyan, the subsites are numbered. The catalytic triad is annotated in italic bold. The amino acid residues that are involved in the binding of substrate are sticked and atom type coloured with respect to the colours used in Table 2: amino acids with non polar side chains are coloured in grey, amino acids with uncharged polar side chains are in green, amino acids with positive charge are blue and with negative charge are red.

bonding can be listed between *An*PGII and A_3 or between *Fm*PG and A_2 or A_3. Our molecular modelling study reveals binding profiles that associates very well with the experimental hydrolytic pattern of the three PGs. Indeed, *An*PGI always evidences higher binding energies, as compared to *An*PGII and *Fm*PG, showing that *An*PGI binds more strongly the oligogalacturonates, especially short ones. It is fully relevant as *An*PGI is also the most efficient enzyme among the three PGs [15]. This is particularly true for GA_2MG/GG, which displays the best energy when docked in *An*PGI active site and from which GA_2MG was the main product with DP_4 [15]. *An*PGII and *Fm*PG both display much weaker binding energies for short substrates. This is in line with the experimental data that quantified the total of digested Gal A as markedly lower for the two enzymes as compared to *An*PGI. In addition, the docking study shows that the highest binding energy is obtained for GA_2MGG/GG which is shown experimentally to be also the most abundant oligomer produced by those enzymes [15]. Our study evidences that the three PGs share a similar unequal distribution of subsites apart from the active subsites −1 and +1 with a rather conserved core of residues from subsites −2 to +1. In line with that, we calculate that each PG has a completely different total charge for an identical pH of optimum activity, that the electrostatic contour is distributed contrastingly for the three enzymes, with interesting variations especially at the external subsites. We suggest that those distinct electrostatic properties could impact the binding. *An*PGI

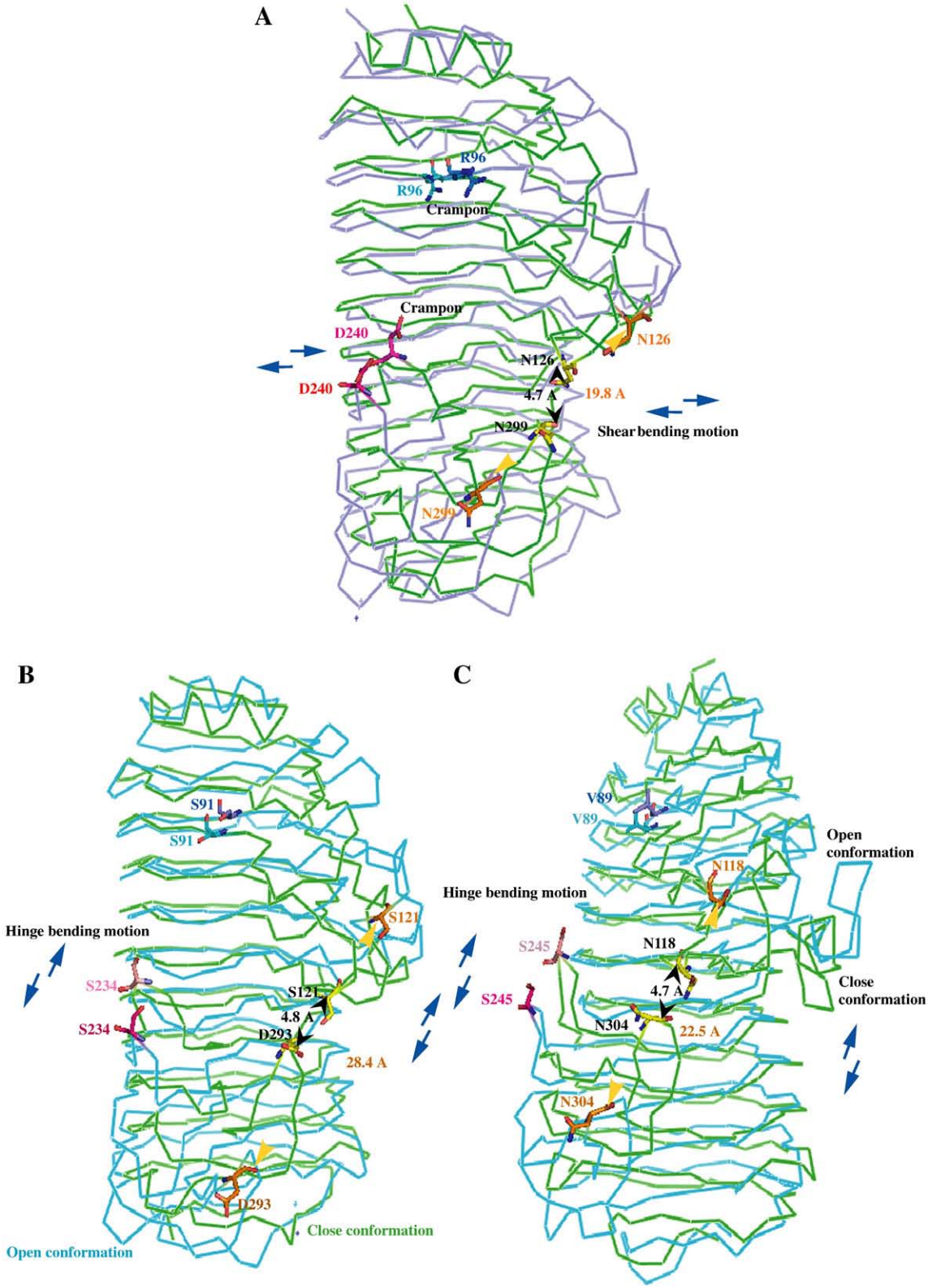

Fig. 4. NOMAD description of motions for *An*PGI (A), *An*PGII (B) and *Fm*PG (C). The two extreme positions are highlighted for the three enzymes where the shear bending motion for *An*PGI (A) and the hinge motions for in *An*PGII (B) and *Fm*PG (C) are evidenced by arrows. The amino acids that trace the distance run along the trajectory are atom sticked and atom type coloured in yellow and orange. The residues $D240_{AnPGI}$ and $R96_{AnPGI}$ that are supposed to act as crampons for the incoming substrate are annotated in (A) and atom sticked in clear and deep pink. The equivalent residues $S234_{AnPGII}$, $S91_{AnPGII}$ and $S245_{FmPG}$, $V89_{FmPG}$ are positioned in (B) and (C) for comparison.

displays a highly negative contour of surface with fissure subsites −5 and −4 clearly negative and subsite +2 markedly positive. The fact that *An*PGI is far more potent could be due to a particularly efficient step of substrate binding as the negative outer surface could repel the negative pectate and drive it straightforwardly to the contrasting positive groove. *An*PGII is the most positively charged, *Fm*PG having

an intermediate positive charge. Our results for *An*PGI and *An*PGII are in line with the very low tolerance of *An*PGI and *An*PGII towards methylated pectin, as the enzymes are clearly not active on this substrate [12]. On the opposite, our result for *Fm*PG is in line with its relative tolerance for less charged methylated pectin [27]. In details, subsites −5 and −4 in *An*PGII and *Fm*PG are highly polar as exclusively composed of Asparagine, Serine, Threonine and Glutamine residues. Subsite +2 of *An*PGII is less positive than subsite +2 of its counterpart *An*PGI, this of *Fm*PG is slightly more polar. Those elements together suggest that the polar patch at subsites −5 and −4 of *An*PGII or *Fm*PG could attract the substrate and be the key to the binding of rather long pectate substrate for which affinity to the protein seems to increase with the DP.

From our recent experimental work [15] and despite a narrower catalytic groove that one could expect to hamper the docking of decorated substrates, *An*PGI is the most potent enzyme and the most tolerant one towards acetylated A_2 and A_3 pectate. *Fm*PG is the most tolerant one to methylated pectins [13,14] and *An*PGII is both less tolerant to A_2, A_3 pectate and to methylated pectins [15]. From the global electrostatic profile of each PG and the previous extensive study of methylated and not methylated substrates that were docked in the active site of *Fm*PG [14], we gained molecular clues about the methylated substrate binding capacity [14] or acetylated substrate binding capacity. Methyl decorations abolish the negative charge of the pectate, rendering it more hydrophobic, more neutral and thus less sensitive to the electrostatic field contouring the protein. A_2 or A_3 are decorations that replace the hydroxyl groups on carbon C_2 and C_3, respectively. They do not screen any negative charge of the GalA moieties and thus do not change the sensitivity of the substrate towards positive or negative charges of the proteins. In that sense, *An*PGI offers the most contrasting negative outer surface *vs* positive groove, so the latter can easily be perceived as highly attractive to a negatively charged pectate. *Fm*PG is the one with the less pronounced electrostatic signature and thus the one most likely to accept the pectin substrate. If we zoom into the active cleft of each PG, we have identified a residue as possibly crucial to modulate such tolerance. Namely, due to its position on a protruding loop, D240$_{AnPGI}$, S234$_{AnPGII}$ and S245$_{FmPG}$ define subsites −3, −4 and −5 in *An*PGI, *An*PGII and *Fm*PG, respectively. Facing the residue and located at the end of a β-sheet on the other side of the cleft, we have S191$_{AnPGI}$, G185$_{AnPGII}$ and S196$_{FmPG}$. Obviously, the resulting tandem is unique for each enzyme. D240$_{AnPGI}$ has a long and charged side chain that is suspected to be responsible for the peculiar positioning of the substrate on the glycone side, as compared to the two others and S191$_{AnPGI}$ hydrogen bonds the A_2 group in subsite −4 (Fig. 3A). S234$_{AnPGII}$ and S245$_{FmPG}$ could be responsible for a comparable global positioning of the substrate that the two enzymes share. However, S196$_{FmPG}$ hydrogen bonds the A_2 group in subsite −4 whereas G185$_{AnPGII}$ does not, so we suggest that this hydrogen bond is critical to boost the tolerance (Fig. 3B, C). We propose S196$_{FmPG}$D for mutation to gain some tolerance or D240$_{AnPGI}$S reversely to decrease it.

The NOMAD motion modes explore the conformational space where we suggest that each PG could be subjected to allosteric conformational changes upon binding of substrate. *An*PGI likely shows an exceptional shear bending motion of 15 Å of amplitude that combines with a 5 Å weak opening. Reversely, *An*PGII and *Fm*PG possibly display hinge bending motion of 24 and 18 Å, respectively, that combines with a weaker shear bending motion of 4.5 and 9 Å, respectively. These motions should not only enable the correct positioning of the oligogalacturonates in the active site but should favour or impair the processive pattern of each enzyme. In *An*PGI, the exceptional shear flexibility combined with a weak opening suggests that the substrate could slide along the crevice after its cleavage with only partial unbounding, for an ensuing hydrolysis. The narrow cleft of the enzyme topologically supports this hypothesis. Indeed, processive enzymes attack the same polymer chain repeatedly during one encounter and in a narrow cleft, the entrapped substrate is more likely to slide along the subsites than to be released. This is actually fully relevant with the work of van Pouderoyen and coworkers who assessed experimentally that *An*PGI excised processively several GalA residues of pectate at one enzyme–glycan encounter [16]. Reversely, we suggest that the groove opening, clearly observed for *An*PGII and *Fm*PG, may be required upon catalysis to release the substrate once cleaved. This is in line with the experimental results that clearly evidence a non-processive scheme for those two enzymes [14,16,19].

In details, R96$_{AnPGI}$ has been reported as crucial for the processivity of *An*PGI [16]. The authors have abolished the positive charge of R96$_{AnPGI}$ (located behind subsite −4) with a mutation into a Serine residue and consequently the enzyme is not longer processive. Reversely, when *An*PGII that possesses S91$_{AnPGII}$ at this position is mutated into an Arginine residue, the enzyme acquires processivity. They suggested that the positive charge of the Arginine residue binds more strongly the negatively charged pectin than the Serine, thus emphasizing the glissade of the substrate along the active site. In line with this, we observed that, along *An*PGI trajectory motion, this residue comes to face the residue D240$_{AnPGI}$, positioned on the C-terminal side and identified as part of subsites −4 and −5. Those two residues with long side chain and opposite charge apart from the crevice could act as crampons onto the substrate to help the sliding and favour the processivity. As was already mentioned earlier in this paper, *An*PGII possesses S234$_{AnPGII}$ at position equivalent to D240$_{AnPGI}$. This residue has a non charged shorter side chain that could not compensate D240$_{AnPGI}$ in acting as crampon for the substrate. *Fm*PG neither displays an Arginine nor a Serine residue at position equivalent to R96$_{AnPGI}$ or S91$_{AnPGII}$. In 3D structure, this position is occupied by the short length and hydrophobic Valine residue V89$_{FmPG}$. In position equivalent to D240$_{AnPGI}$, *Fm*PG has S245$_{FmPG}$ and *Fm*PG is known to be non-processive [29]. The tandem V89$_{FmPG}$/S245$_{FmPG}$, similarly to the couple S91$_{AnPGII}$/S234$_{AnPGII}$ could be insufficient to clamp the substrate for another round of hydrolysis. We propose that V89$_{FmPG}$ could be rationally substituted into an Arginine residue to test the conversion into a processive enzyme.

5. Conclusion

Sequence alignment of three highly homologous enzymes, *An*PGI, *An*PGII and *Fm*PG, calculation of their electrostatic contour, and computation of their intrinsic flexibility have been extensively combined to the docking of poorly to highly decorated substrates to comprehend, at a molecular scale, their differences in the tolerance profiles towards acetylated pectins or in their processive–or not–pattern of hydrolysis. Our work suggests that, besides the topology of the catalytic cleft that goes from narrow with *An*PGI to large with *An*PGII, motion flexibility should be pivotal for processivity property. Indeed, the appreciative shear bending motion of *An*PGI should favour the processive pattern whilst the hinge binding motion observed in *An*PGII and *Fm*PG should conversely favour the release of substrate after one catalytic event. The tandem of residues, D240$_{AnPGI}$/R96$_{AnPGI}$, that is 1D/3D replaced by S234$_{AnPGII}$/S91$_{AnPGII}$, and by S245$_{FmPG}$ / V89$_{FmPG}$ is unique and has large-amplitude deviation during the identified motions. We suggest that the couple D240$_{AnPGI}$/R96$_{AnPGI}$ could dynamically come over the catalytic cleft to clamp the substrate. This could be pivotal to enhance the sliding of the substrate along the fissure. In addition, we suggest that the electrostatic contour outside of and inside the catalytic cleft explains how *An*PGI is not only the most potent but also the most tolerant enzyme to hydrolyse acetylated pectin. We have identified the couple S191$_{AnPGI}$/D240$_{AnPGI}$ as critical for the binding of acetylated substrate. Clearly from this work, D240$_{AnPGI}$ has been identified as crucial for the processivity when in tandem with R96$_{AnPGI}$ but also for the tolerance profile of the enzyme towards acetylated substrate when this time in association to S191$_{AnPGI}$. Combinations of site-directed substitutions according to

the residue variations occurring in the other two should be performed experimentally to confirm our *in silico* procedure as powerful and successful in identifying PGs mutations that could modulate the tolerance as well as the hydrolysis pattern profiles.

Acknowledgement

The authors wish to thank Dr Marie-Christine Ralet for fruitful discussions about the structure of the produced oligomers.

References

[1] A.G.J. Voragen, W. Pilnik, J.-F. Thibault, M.A.V. Axelos, C.M.G.C. Renard, Pectins, in: A.M. Stephen (Ed.), Food Polysaccharides and Their Applications, Marcel Dekker, New-York, 1995, pp. 287–340.
[2] H. Schols, A.G.J. Voragen, Complex pectins: structure elucidation using enzymes, in: J. Visser, A.G.J. Voragen (Eds.), Pectins and Pectinases, Elsevier Science B.V., Amsterdam (The Netherlands), 1996, pp. 3–19.
[3] M.-C. Ralet, E. Bonnin, J.-F. Thibault, Pectins, in: A. Steinbüchel (Ed.), Biopolymers, vol. 8, Wiley-VCH Verlag Gmbh, Weinheim, 2003, pp. 345–380.
[4] J.-P. Vincken, H.A. Schols, R.J.F.J. Oomen, M.C. McCann, P. Ulvskov, A.G.J. Voragen, R.G.F. Visser, If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture, Plant Physiol. 132 (2003) 1781–1789.
[5] H. Neukom, Uber den Abbau von Pektinstoffen, Schweitz. Landwirtsch. Forsch 2 (1963) 112–122.
[6] W.M. Fogarty, C.T. Kelly, Pectic enzymes, Microbial Enzymes and Biotechnology, Applied Science Publishers, London (UK), 1983, pp. 131–182.
[7] T. Sakai, T. Sakamoto, J. Hallaert, E.J. Vandamme, Pectin, pectinase and protopectinase: production, properties and applications, Adv. Appl. Microbiol. 39 (1993) 213–294.
[8] C. Lang, H. Dönenburg, Perspectives in the biological function and the technological application of polygalacturonases, Appl. Microbiol. Biotechnol. 53 (2000) 366–375.
[9] B. Henrissat, G.J. Davies, Structural and sequence-based classification of glycoside hydrolases, Cur Opin. Struct. Biol. 7 (1997) 637–644.
[10] B. Henrissat, Enzymology of cell-wall degradation, Biochem. Soc. Transac. 26 (1998) 153–156.
[11] E.M.W. Chen, A.J. Mort, Nature of sites hydrolyzable by endopolygalacturonase in partially-esterified homogalacturonans, Carbohydr. Polym. 29 (1996) 129–136.
[12] J.A.E. Benen, H.C.M. Kester, J. Visser, Kinetic characterization of *Aspergillus niger* N400 endopolygalacturonases I, II and C, Eur. J. Biochem. 259 (1999) 577–585.
[13] E. Bonnin, A. Le Goff, G.J.W.M. van Alebeek, A.G.J. Voragen, J.-F. Thibault, Mode of action of *Fusarium moniliforme* endopolygalacturonase towards acetylated pectin, Carbohydr. Polym. 52 (2003) 381–388.
[14] G. André-Leroux, D. Tessier, E. Bonnin, Action pattern of *Fusarium moniliforme* endopolygalacturonase towards pectin fragments: comprehension and prediction, Biochim. Biophys. Acta 1749 (2005) 53–64.
[15] M.C. Ralet, M.J. Crépeau, E. Bonnin, Evidence for a blockwise distribution of acetyl groups onto homogalacturonans from a commercial sugar beet (*Beta vulgaris*) pectin, Phytochem. (2008) 1903–1909.
[16] G. van Pouderoyen, H.J. Snijder, J.A.E. Benen, B.W. Dijkstra, Structural insights into the processivity of endopolygalacturonase I from *Aspergillus niger*, FEBS Lett. 554 (2003) 462–466.
[17] Y. van Santen, J.A.E. Benen, K.-H. Schröter, K.H. Kalk, S. Armand, J. Visser, B.W. Dijkstra, 1.68-Å crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis, J. Biol. Chem. 274 (1999) 30474–30480.
[18] L. Federici, C. Caprari, B. Mattei, C. Savino, A. Di Matteo, G. De Lorenzo, F. Cervone, D. Tsernoglou, Structural requirements of endopolygalacturonase for the interaction with PGIP (polygalacturonase-inhibiting protein), PNAS 98 (2001) 13425–13430.
[19] L. Holm, J. Park, DaliLite workbench for protein structure comparison, Bioinformatics 16 (2000) 566–567.
[20] L. DeLano, The case for open-source software in drug discovery, Drug Discov. Today 10 (2005) 213–217.
[21] N.A. Baker, D. Sept, S. JosepS, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041.
[22] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) 665–667.
[23] M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, Phys. Rev. Lett. 77 (1996) 1905–1908.
[24] M. Delarue, Y.-H. Sanejouand, Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model, J. Mol. Biol. 320 (2002) 1011–1024.
[25] G. André, A. Buléon, M. Juy, N. Aghajari, R. Haser, V. Tran, Amylose chain behavior in an interacting context II—Molecular modelling of a maltopentaose fragment in the barley α-amylase catalytic site, Biopolymers 49 (1999) 107–119.
[26] G. André, A. Buléon, R. Haser, V. Tran, Amylose chain behaviour in an interacting context III—Complete occupancy of the AMY2 barley α-amylase cleft and comparison with biochemical data, Biopolymers 50 (1999) 751–762.
[27] E. Bonnin, A. Le Goff, R. Körner, J. Vigouroux, P. Roepstorff, J.-F. Thibault, Hydrolysis of pectins with different degrees and patterns of methylation by the endopolygalacturonase of *Fusarium moniliforme*, Biochim. Biophys. Acta 1596 (2002) 83–94.
[28] M. Kouwijzer, H. Schols, S. Pérez, Acetylation of rhamnogalacturonan I and homogalacturonan: theoretical calculations, in: J. Visser, A.G.J. Voragen (Eds.), Pectins and Pectinases, vol. 14, Elsevier Science, Amsterdam (The Netherlands), 1996, pp. 57–65.
[29] E. Bonnin, A. Le Goff, R. Körner, G.J.W.M. van Alebeek, T.M.I.E. Christensen, A.G.J. Voragen, P. Roepstorff, A. Caprari, J.-F. Thibault, Biochim. Biophys. Acta 1526 (2001) 301–309.