



Exploring molecular determinants of polysaccharide lyase family 6–1 enzyme activity

Sébastien Violot, Frédéric Galisson, Loïc Carrique, Vinesh Jugnarain, Léa Conchou, Xavier Robert, Aurélien Thureau, William Helbert, Nushin Aghajari, Lionel Ballut

► To cite this version:

Sébastien Violot, Frédéric Galisson, Loïc Carrique, Vinesh Jugnarain, Léa Conchou, et al.. Exploring molecular determinants of polysaccharide lyase family 6–1 enzyme activity. *Glycobiology*, 2021, 10.1093/glycob/cwab073 . hal-03379698

HAL Id: hal-03379698

<https://cnrs.hal.science/hal-03379698>

Submitted on 15 Oct 2021

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.

Structural Biology

Exploring molecular determinants of polysaccharide lyase family 6–1 enzyme activity

Sébastien Violot², Frédéric Galisson², Loïc Carrique², Vinesh Jugnarain², Léa Conchou², Xavier Robert², Aurélien Thureau³, William Helbert⁴, Nushin Aghajari^{1,2}, and Lionel Ballut^{1,2}

²Molecular Microbiology and Structural Biochemistry, UMR 5086, CNRS Université de Lyon, 7 passage du Vercors, Lyon 69367, France, ³Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin, BP 48 91192 Gif-sur-Yvette, France, and ⁴Centre de Recherches sur les Macromolécules Végétales (CERMAV), CNRS and Grenoble Alpes Université, BP53, 38000 Grenoble Cedex 9, France

¹To whom correspondence should be addressed: Tel: +33-(0)4-72-72-26-34/+33-(0)4-72-72-26-33; Fax: +33-(0)4-72-72-26-04; e-mails: lionel.ballut@ibcp.fr; nushin.aghajari@ibcp.fr

Received 15 April 2021; Revised 24 June 2021; Editorial Decision 6 July 2021; Accepted 7 July 2021

Abstract

The polysaccharide lyase family 6 (PL6) represents one of the 41 polysaccharide lyase families classified in the CAZy database with the vast majority of its members being alginate lyases grouped into three subfamilies, PL6_1–3. To decipher the mode of recognition and action of the enzymes belonging to subfamily PL6_1, we solved the crystal structures of Pedsa0632, Patl3640, Pedsa3628 and Pedsa3807, which all show different substrate specificities and mode of action (endo-/exolyase). Thorough exploration of the structures of Pedsa0632 and Patl3640 in complex with their substrates as well as docking experiments confirms that the conserved residues in subsites –1 to +3 of the catalytic site form a common platform that can accommodate various types of alginate in a very similar manner but with a series of original adaptations bringing them their specificities of action. From comparative studies with existing structures of PL6_1 alginate lyases, we observe that in the right-handed parallel β -helix fold shared by all these enzymes, the substrate-binding site harbors the same overall conserved structures and organization. Despite this apparent similarity, it appears that members of the PL6_1 subfamily specifically accommodate and catalyze the degradation of different alginates suggesting that this common platform is actually a highly adaptable and specific tool.

Key words: protein–carbohydrates recognition, structure of alginate lyases, surface-binding site

Introduction

Polysaccharide lyases (PLs) are a group of enzymes (EC 4.2.2.-) divided into 41 families, (May 2021) which display a broad diversity of substrate specificities (www.cazy.org; Lombard et al. 2014). Among this variety of polysaccharide-specific enzymes, members of the PL6 family harbor four different activities including mannuronate-specific alginate lyases (M-specific lyases, EC 4.2.2.3), guluronate-specific alginate lyases (G-specific lyases, EC 4.2.2.11), chondroitin B lyases (chondroitin sulfate (CS)- and dermatan sulfate (DS)-specific lyases, EC 4.2.2.19) and oligo-alginate lyases (EC 4.2.2.26) (Garron and Cygler 2010; Mathieu et al. 2016).

Enzymes belonging to this family cleave uronic acid-containing polysaccharides *via* a β -elimination mechanism resulting in the formation of an unsaturated hexuronic acid residue (Zhu and Yin 2015).

The catalytic mechanism underlying breakage of the glycosidic bond can be divided into three steps (Figure 1). In a first step, neutralization of the negatively charged C6 carboxyl group by, in most cases, a Ca^{2+} ion allows the resonance stabilization of the enolate thereby formed. This enolate formation causes the reduction of the pK_a α -proton on the C5, facilitating its abstraction by a general base catalyst (Xu et al. 2017). Interestingly, this first step does not

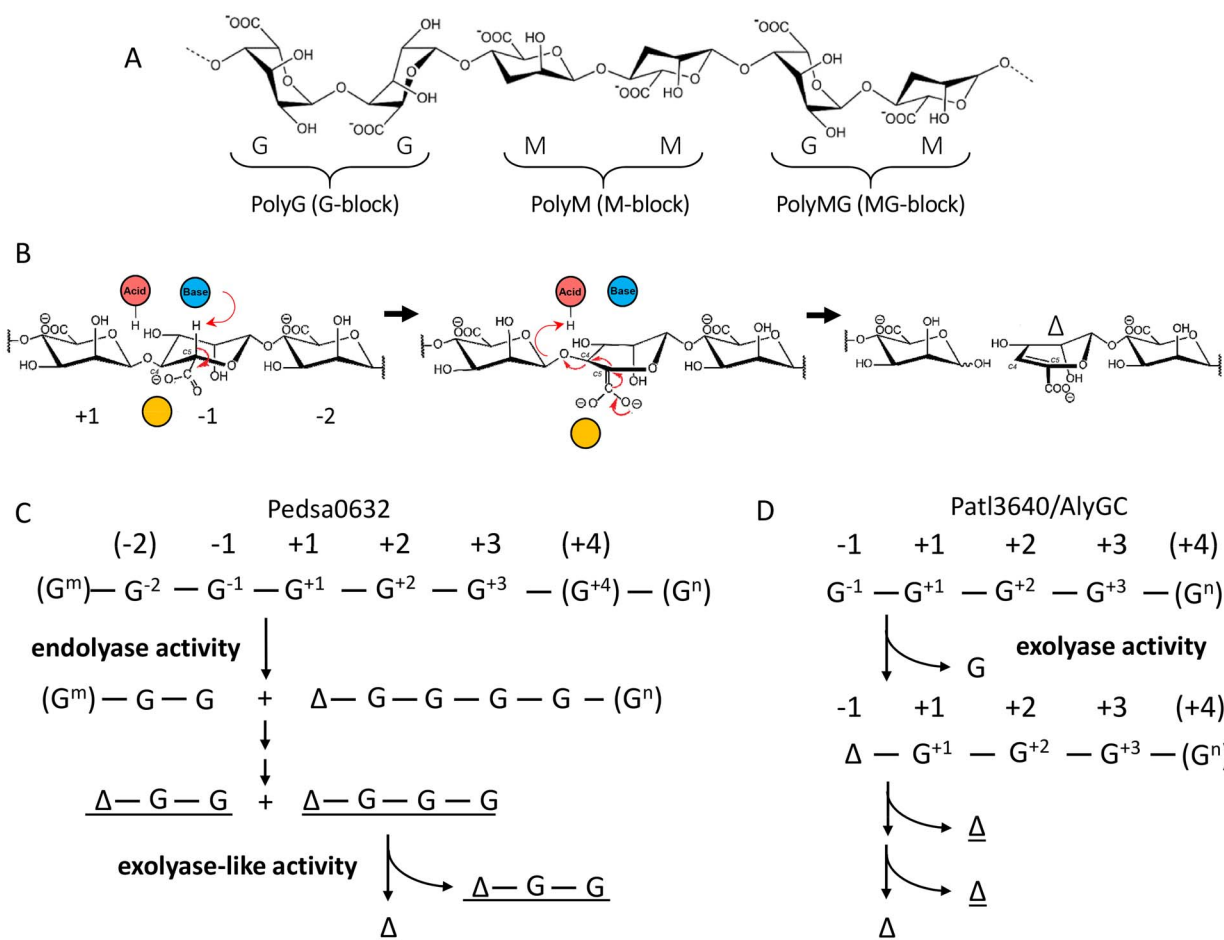


Fig. 1. Schematic illustrating the cleavage modes of endo- and exotype PL6: **(A)** poly α -L-gulonate, poly β -D-mannuronate and mixed poly α -L-gulonate and poly β -D-mannuronate **(B)** β -elimination mechanism on a polyM substrate. The uronate neutralizer (Ca^{2+} , Na^{2+} or water) is depicted as an orange sphere, and acid and base as a red and blue sphere, respectively. **(C)** Endolyase activity observed in Pedsa0632. **(D)** Exolyase activity observed in PatI3640 and AlyGC. Subsites are numbered from -1 to +3 and other potential subsites are shown in parentheses. Final products Δ , ΔGG or ΔGGG for Pedsa0632 and Δ for PatI3640 and AlyGC are underlined.

seem to be strictly Ca^{2+} dependent: In some enzymes, calcium could be replaced by other divalent cations such as Mn^{2+} or by a water molecule (Mathieu et al. 2016; Lyu et al. 2019). In the last step, β -elimination of the 4-O-glycosidic bond results in the concomitant formation of an unsaturated C4-C5 bond within the hexuronic acid when a transfer of electrons from the carboxyl group occurs. This results in the formation of an oligosaccharide with a 4-deoxy-L-erythro-hex-4-eno-pyranosyluronic acid at the nonreducing terminal end. During this very last step, the leaving group must be protonated by a side chain acting as a general acid. Depending on the substrate, proton abstraction can occur either in a *syn* configuration, when the C5 proton and the glycosidic oxygen of the bond are situated on the same side of the sugar ring, like for M/G-M bond-specific lyases, or in an *anticonfiguration* when the groups are placed on opposite sides of the sugar ring, as for G/M-G specific lyases and chondroitin B lyases (Lombard et al. 2010).

Except for chondroitin B lyases, other members of the PL6 family are alginate lyases. Alginate is a linear polysaccharide and the main constituent of the brown algae cell wall. It is composed of mannuronic acid (β -D-mannuronate, M-residues) and its C5 epimer guluronic acid (α -L-gulonate, G-residues); both of which arrange

into different blocks of polyM (M-blocks), polyG (G-blocks) and polyMG heteropolymer (MG-blocks) (Figure 1A) (Haug et al. 1967; Aarstad et al. 2012). In this context, PL6 family members harbor activities toward M-blocks, G-blocks, but also MG-blocks (Garron and Cygler 2010). Alginate lyases frequently have broad substrate specificity (Mathieu et al. 2016) and can cleave more than one type of alginate (e.g. polyG plus polyM or polyG plus polyMG) (Li et al. 2019). The polysaccharide lyase family PL6 is subdivided into three subfamilies, PL6_1, PL6_2 and PL6_3, according to sequence similarity (Mathieu et al. 2016). Chondroitin B lyases, polyG, polyM and polyMG alginate lyases are members of PL6_1. In contrast hereto, PL6_2 and PL6_3 contain essentially polyMG lyases.

Regarding 3D structures, PL6 members adopt the same right-handed parallel β -helix fold formed by three β -sheets. This was first observed for the single-domain (one β -helix) chondroitin B lyase from *Pedobacter heparinus* DSM 2366 (PDB entry 1OFL) and more recently within the single-domain AlyF from *Vibrio splendidus* OU2 (PDB entries 5Z9T, 6ITG, 6A40 and 7BZ0), the single-domain BCELPL6 from *Bacteroides cellulosilyticus* CRE21 (PDB entry 6QPS) and the two-domain enzymes (two β -helices) AlyGC from *Paraglaciicola chatamensis* S18K6T (PDB entry 5GKQ) (Huang

et al. 1999; Xu et al. 2017; Lyu et al. 2019; Stender et al. 2019; Zhang et al., 2021). On one side of the β -helix, a β -sheet forms a groove (chondroitinase B, AlyGC, B*Ce*PL6) or a pocket (AlyF) delimitating the binding site in which at least six subsites can be described (subsites -3 to $+3$). Between subsites -1 and $+1$, two conserved residues, respectively, a lysine and an arginine, are identified as the catalytic residues in PL6 enzymes (Xu et al. 2017; Stender et al. 2019). In addition, the Ca^{2+} ion located between subsites $+1$ and $+2$ is substituted by a water molecule in AlyF, whereas a Ca^{2+} ion is bound in the G6 bound structure. The O-C4 glycosidic bond cleavage always occurs between subsites -1 and $+1$ with two different types of activity: i) endolyase activity (AlyF or B*Ce*PL6) or ii) exolyase activity (AlyGC). Enzymes in PL6_1 display both endo- and exolyase activities as for Pedsa0632 from *Pedobacter saltans* or Mase04135 from *Alteromonas macleodii* (Mathieu et al. 2016).

Altogether, PL6 family members interact with at least four different substrates (M-blocks, G-blocks, MG-blocks and CS/DS) in a specific manner despite sharing common fold and catalytic residues. This fact held together with their ability to cleave substrates, either in an endo- or exolyase mode, raises the questions of how these enzymes discriminate between substrates and how they cleave either in the middle or at the nonreducing end of the polysaccharide chain.

We solved the crystal structures of four different PL6_1 enzymes from marine bacteria, namely Pedsa0632, Pedsa3628 and Pedsa3807 from *Pedobacter saltans*, as well as Patl3640 from *Pseudoalteromonas atlantica* T6c and compared them to the already available structures. Complexes were obtained for Pedsa0632 and Pedsa3628 after soaking with a Δ GGG and in parallel, computer-aided docking of G₇ into Pedsa0632 was carried out to get more insight into the alginate-binding mode. Systematic superposition of the experimentally determined complexes issued from soaking experiments and from docking assays was performed. Upon comparison of residues interacting with the different oligosaccharides, the large conservation of the residues involved in catalysis and substrate binding prompted us to conclude that each member of the PL6 family has developed its own specific binding mode likely issued from a number of very precise modifications and adaptations in order to reach such a level of specificity from a common β -helix fold. These observations could form the bases for future functional prediction of PL6 family members based on sequence analyses.

Results

Conservation of the β -helix fold

The crystal structures of Pedsa0632, Pedsa3628, Patl3640 and Pedsa3807 were determined in their native state, respectively, to 1.99, 1.93, 2.32 and 1.58 Å resolution (Table I). Crystals of Pedsa0632 and Pedsa3807 contain two molecules per asymmetric unit, while crystals of Pedsa3628 and Patl3640 contain one molecule per asymmetric unit. Two additional structures of Pedsa0632 and Patl3640 in complex with Δ GGG obtained by soaking experiments were determined to 2.17 and 2.05 Å resolution, respectively (Table I).

All four 3D structures show the same right-handed parallel β -helix fold, the only difference being that Pedsa0632, Pedsa3628 and Pedsa3807 are composed of a single domain (one β -helix), while Patl3640 is composed of two domains (two β -helices) (Table II). The structures respect the β -helix fold, as can be seen using the canonical nomenclature of chondroitinase B with the three β -sheets designated as PB1–3 and the connecting turns or loops designated as T1–3, with the catalytic residues being always brought by PB1

(Figure 2, Supplementary Figures S1 and S2). For all four enzymes, we always observed 12 β -strands for PB1, in comparison to 13 or 14 for PB2 and PB3 (Figure 2). When compared to the other available structures, a topological representation shows the general conservation between all members of the PL6 family. The main differences observed are found at C-terminal and N-terminal parts where several secondary structure elements can be added, thereby considerably changing their size (Figure 2, Supplementary Figures S1 and S2). Minor modifications are also observed such as secondary structures inserted in between β -helices. (Figure 2, Supplementary Figures S1 and S2). When comparing the primary structure of the eight available crystal structures, we found that CS/DS lyases, chondroitinase B and Pedsa3807 present the highest identity ($\sim 60\%$). Regarding the alginate lyases, AlyGC and Patl3640, both two-domain structures, they present also a high sequence identity ($\sim 48\%$). On the other hand, the one-domain alginate lyases Pedsa0632 and Pedsa3628 present a sequence identity of only 35% (Supplementary Table SI). As expected, alginate lyases and CS/DS lyases show poor sequence conservation with a highest identity of 29% between Pedsa0632 and Pedsa3807.

In order to highlight the most conserved regions in the β -helix, we used the PROMALS-3D program to align both the primary- and secondary structures of alginate and CS/DS lyases. This shows a conserved region from β -sheet PB1₆ to β -sheet PB3₁₀ corresponding to the center of the right-handed parallel β -helix at the very place where the two catalytic residues reside (Supplementary Figure S1).

AlyGC has been shown to be dimeric in solution, while AlyF and B*Ce*PL6 are monomeric in solution (Xu et al. 2017; Lyu et al. 2019; Stender et al. 2019). In order to determine the oligomeric state in solution of the three alginate lyases studied herein (Pedsa0632, Patl3640 and Pedsa3628), a SEC-SAXS experiment was performed. For Pedsa0632, monomeric, dimeric and trimeric forms exist in equilibrium (Supplementary Table SII) with a calculated molecular weight of 48.8 kDa for a monomer. Patl3640 and Pedsa3628 are both dimeric with observed molecular weights of 185 kDa and 107 kDa, respectively (Supplementary Table SII).

Binding mode of Pedsa0632 and Patl3640

Alginate lyases belonging to PL6_1 cleave polysaccharides organized into G-blocks, M-blocks or MG-blocks, which are distributed throughout the alginate. They all have a preferred substrate and are classified as poly- β -D-mannuronate lyases (M-lyases), which cleave glycosidic bonds between M-M moieties; poly- α -L-guluronate lyases (G-lyases), which cleave glycosidic bonds between GG moieties; and MG or GM-lyases (Garron and Cygler 2010). Among the six alginate lyases discussed herein, the two-domain enzymes Patl3640 and AlyGC are exolyases with polyG as preferred substrate. Patl3640 does not seem to have the ability to cleave other types of alginate blocks, whereas AlyGC cleaves polyM as well with a specific activity being 4 times lower than for polyG. The four other alginate lyases, Pedsa0632, Pedsa3628, AlyF and B*Ce*PL6, possess only one domain but present broader substrate specificity (Table II).

In order to bring out the binding mode that could prevail for a given substrate, we performed soaking experiments with two types of compound, a Δ GGG and a Δ MGM with Δ being a 4-deoxy-L-erythro-5-hexoseulose uronic acid. After several assays in which soaking times varied, diffraction data could be collected on a crystal of Pedsa0632 soaked with Δ GGG. The asymmetric unit displayed a monomer in which Pedsa0632 bound Δ GGG and a monomer with a Δ GG bound. Cleavage of the tetra-saccharide indicates that the

Table I. Data collection and refinement statistics

Structure-ID	Pedsa0632			Pedsa0632_Δ GGG/ΔGG			Pedsa3628			Patl3640			Patl3640_Δ			Pedsa3807		
Data collection																		
Beamline	ID30-A3_ESRF			ID29_ESRF			ID30B_ESRF			ID30B_ESRF			ID29_ESRF			ID23-1_ESRF		
Wavelength (Å)	0.9677			1.07227			0.95372			1.07000			0.97625			0.97242		
Space group	<i>P</i> 2 ₁			<i>P</i> 2 ₁			<i>P</i> 2 ₁			<i>P</i> 3 ₁ 2 1			<i>P</i> 3 ₁ 2 1			<i>P</i> 2 ₁ 2 ₁ 2 ₁		
Cell dimensions																		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	81.29	47.29	116.73	81.18	47.58	116.20	58.19	51.75	74.26	79.53	79.53	269.47	78.23	78.23	266.79	78.33	87.45	148.49
α , β , γ (°)	90.00	98.03	90.00	90.00	98.25	90.00	90.00	97.22	90.00	90.00	90.00	120.00	90.00	90.00	120.00	90.00	90.00	90.00
Resolution range (Å)	80.49–1.99			47.58–2.17			48.49–1.93			48.16–2.32			47.53–2.05			75.36–1.58		
Total reflections	399,271			151,609			214,990			1,059,752			789,224			1,758,910		
Unique reflections	58,008			46,036			32,727			42,777			60,589			135,111		
<i>R</i> _{meas} (%)	11.3 (89.1)			19.8 (119.6)			19.6 (134.8)			11.1 (313.1)			7.9 (156.0)			7.4 (127.2)		
<i>CC</i> _{1/2} (%)	99.9 (85.0)			99.0 (43.3)			99.6 (54.0)			100.0 (66.3)			99.9 (77.2)			100.0 (84.4)		
<i>I</i> / σ (<i>I</i>)	11.7 (2.2)			4.8 (1.1)			8.0 (1.4)			18.5 (1.0)			18.02 (1.80)			19.4 (2.1)		
Multiplicity	6.9 (7.4)			3.3 (3.3)			6.6 (6.5)			24.8 (18.6)			13.0 (13.2)			13.0 (12.7)		
Completeness (%)	94.7 (98.7)			97.8 (90.5)			98.5 (91.8)			97.3 (83.9)			99.9 (99.7)			96.0 (99.5)		
No. mol./asymm. unit	2			2			1			1			1			2		
Refinement																		
<i>R</i> _{work} / <i>R</i> _{free} (%)	21.02/24.66			23.12/28.62			17.89/20.94			24.99/28.23			20.91/24.31			18.02/21.02		
No. atoms																		
Protein	6270			6293			3258			5418			5396			7618		
Ligand/ion	/			84			5			6			32			/		
Water	610			372			344			53			261			908		
Average B-factor (Å ²)																		
Protein	32.3			40.7			28.7			79.4			54.5			23.3		
Ligand/ion	/			64.4			43.4			79.9			62.6			/		
Water	42.8			1.4			37.3			71.1			57.1			35.5		
r.m.s.d.																		
Bond lengths (Å)	0.008			0.014			0.008			0.012			0.008			0.016		
Angles (°)	1.07			1.67			1.28			1.69			0.94			1.64		
Ramachandran																		
Favored (%)	97.5			97.3			95.3			94.2			95.7			97.5		
Allowed (%)	2.2			2.6			4.5			5.4			4.2			2.6		
Outliers (%)	0.3			0.1			0.2			0.4			0			0		

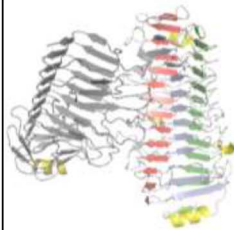
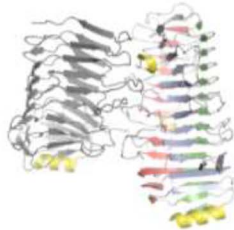
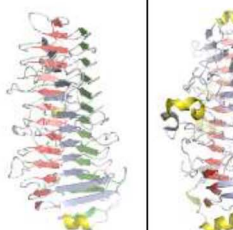
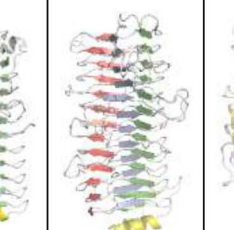
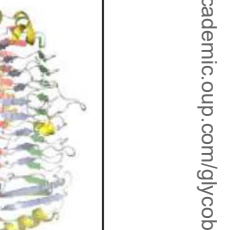
enzyme remains active in its crystalline form. The resulting products are in accordance with a previous study in which the final products observed are Δ, ΔGG and ΔGGG (Mathieu et al. 2016). Indeed, Pedsa0632 has been described as an endolyase that cleaves a range of oligosaccharides of various size and that produces ΔGGG and ΔGG as main end products after extensive incubation times (Mathieu et al. 2016). However, the presence of a Δ suggests that the enzyme is capable, to a certain extent, of cleaving oligosaccharides with an exolytic mode of action (Figure 1C and 1D).

In this structure, ΔGGG and ΔGG bind in subsites −1 to +3 and −1 to +2, respectively, involving β-sheets PB1₅ to PB1₉ (Figure 3A and 3B). Altogether, this tends to indicate that after the cleavage of ΔGGG into Δ and ΔGG, the enzyme is unable to cleave the glycosidic bond any further. When comparing the two oligosaccharides, the three first sugar moieties at positions −1 to +2 superposed very well (all atom RMSD: 0.28 Å) with the only difference being that ΔGGG is further stabilized by an H-bond between the carboxyl group of the fourth residue at the reducing end and the side chain of Arg215 (Figure 3B, 3C and 3D). Another interesting feature is the water molecule replacing the Ca²⁺ ion in Pedsa0632, which we hereafter refer to as the catalytic site water molecule. Indeed, in the monomer with a bound ΔGGG, three acidic residues, Glu212, Glu241 and Glu243, in addition to Asn209 interact with this water molecule. As concerns the monomer with the ΔGG, only Glu212 and Glu241 are involved in such a network, resulting in a drastic displacement of the water molecule (~2.5 Å). It is important to note that in the ΔGG bound monomer, after making a complete

flip, the side chain of Asn209 is turned toward the solvent leaving room for the water molecule indicating that the active site is not in a proper configuration for catalysis and that ΔGG cannot be considered as a substrate (Figure 4A and 4B). Asn209 therefore could play the role of a trigger that engages or disengages the substrate, thereby impairing or enabling the catalytic action of the enzyme.

In addition to Pedsa0632, the 3D structure of a complex between Patl3640 and a linear Δ after soaking with ΔGGG was determined, indicating as for Pedsa0632 that Patl3640 remains active in its crystalline form. The enzyme has been crystallized in the absence of calcium and the electron density indicates the presence of a water molecule at the Ca²⁺ binding site. This observation supports the divalent cation being dispensable for enzymatic activity. Patl3640 is a two-domain enzyme with an exolytic mode of action restricted to polyG (Figure 3, Table II). It shows similarities with AlyGC in terms of structure and mode of action with the exception that AlyGC is also active on polyM. In the structure, the linear Δ forms H-bonds with three basic residues: His249, Arg267 and Arg300 (Figure 5). It is interesting to note that after Patl3640 has catalyzed the cleavage of ΔGGG, the final product is still present in the crystal after a complete hydrolysis of the tetra-saccharide. The linear Δ is situated at subsite −1 as is Δ of the ΔGGG bound to Pedsa0632 (Figure 5). However, in Patl3640, the linear Δ is slightly displaced and interacts via an H-bond with the catalytic site water molecule, which is in turn further stabilized by interactions with the two acidic residues Glu248 and Glu250. Interestingly, Asn216 and Glu219, which normally interact with the water molecule in the binding site, adopt other

Table II. Characteristics of PL6 family members with available crystal structures

	Patl3640	AlyGC	Pedsa0632	AlyF	Pedsa3628	BCeIPL6
Number of domains	2	2	1	1	1	1
PDB code(s)	7o77 7o7t	5gkq 5gkd	7o7a 7o84	6itg 6a40 5z9t 7z70	7o79	6qps
Main activity	Exo PolyG	Exo PolyG	Endo PolyMG	Endo PolyG	Endo PolyMG	Endo PolyM
Activity on M-G-block	No	n.d.	Yes	Yes	Yes	No
Activity on M-block	No	Yes	No	Yes	No	Yes
Activity on G-block	Yes	Yes	Yes	Yes	No	No
Overall structure						

The overall structures showing the right-handed parallel β -helix fold are represented by a ribbon model. β -sheets corresponding to PB1, PB2 and PB3 are colored blue, red and green, respectively. Helices are shown in yellow and the second domain of Patl3640 and AlyGC in gray. Three types of alginate blocks are cleaved by these enzymes, namely polyG by AlyF, polyM by BCeIPL6 and polyMG by Pedsa0632 and Pedsa3628. Other differences exist as BCeIPL6 and Pedsa3628 only recognize one type of substrate and Pedsa0632 recognizes both polyG and polyMG, while AlyF recognizes all three types of substrate.

orientations in this structure (Figure 4C). Nevertheless, the water located in the position is also observed in the Pedsa0632- Δ GGG complex, suggesting that the presence of a water molecule at this position is essential for the enzyme to adopt an active conformation.

Comparison of Pedsa0632-, AlyGC- and AlyF-binding modes

The structure of the Pedsa0632 complex was compared with the structures of the catalytic mutant R241A of AlyGC (Xu et al. 2017), a two-domain enzyme that was co-crystallized with an MMMM (M_4) tetra-saccharide (Lyu et al. 2019) and with AlyF, a single-domain enzyme, in complex with a GGGG (G_4) tetra-saccharide and with a Δ GG (Lyu et al. 2019).

Comparative studies of AlyGC- M_4 and Pedsa0632- Δ GGG 3-D structures show that the two substrates superimpose well, both spanning subsites -1 to $+3$. Despite differing conformations and nature of the sugar moiety, β -D-mannuronate and α -L-guluronate, interactions between conserved residues of the two enzymes and the same chemical functions of the glycosidic residues are entirely conserved (Figure 3D and 4, Supplementary Figures S4A, and S3A, Table III). This could indicate that the conformation of the substrates in the catalytic site of the alginate lyases whatever the chemical structure of these substrates, polyG, polyM and likely polyMG, is largely constrained by the position and conformation of the residues lining the binding site. Eleven residues interact with Δ GGG or M_4 in Pedsa0632 and AlyGC, respectively. The two enzymes have seven interacting residues in common, the only differences being Arg154 and Arg303 in AlyGC interacting with M_4 , respectively, at

subsites $+3$ and -1 , and Arg265 in Pedsa0632 interacting with Δ at subsite -1 . It should be noted that another conserved residue, the catalytic Arg241, was mutated to Ala in AlyGC (Table III). Arg154 and Arg303 in AlyGC are substituted by Trp185 and Glu331, respectively, in Pedsa0632, suggesting that at least these two residues could be specific for the AlyGC-binding site. On the other hand, Arg265 in AlyGC (Arg293 in Pedsa0632) does not interact with M_4 in AlyGC, due to the complete reorientation of the Δ carboxyl group at subsite -1 (Supplementary Figures S3A and S4A). The most striking difference is the lack of interaction between Δ GGG or Δ GG and the catalytic site water molecule (Figure 4A and 4B). Indeed, in AlyGC, the carboxyl group of the second residue of the M_4 at subsite $+1$ interacts with Ca^{2+} (Figure 4D). This lack of interaction in Pedsa0632 is due to a 90° flip of the carboxyl group, which is oriented toward the solvent.

The 11 residues of AlyGC involved in M_4 binding are all conserved in Patl3640. It is important to note that despite this conservation, Patl3640 has no activity on polyM indicating that the substrate specificity is not restricted to these 11 conserved residues (Table III). Interestingly, as earlier reported for AlyGC, we confirmed that Patl3640 is more active on polyG in the presence of calcium. As opposed hereto, Pedsa0632 is not activated by calcium upon degradation of polyG (Supplementary Figure S5).

The binding mode of Pedsa0632 was also compared to that of AlyF, both one-domain endolytic enzymes, with Pedsa0632 acting on PolyG and PolyMG substrates and AlyF on PolyG and alginate substrates. None of them needs calcium for activity and both form the trisaccharide Δ GG as final product upon polyG cleavage (Lyu et al. 2019).

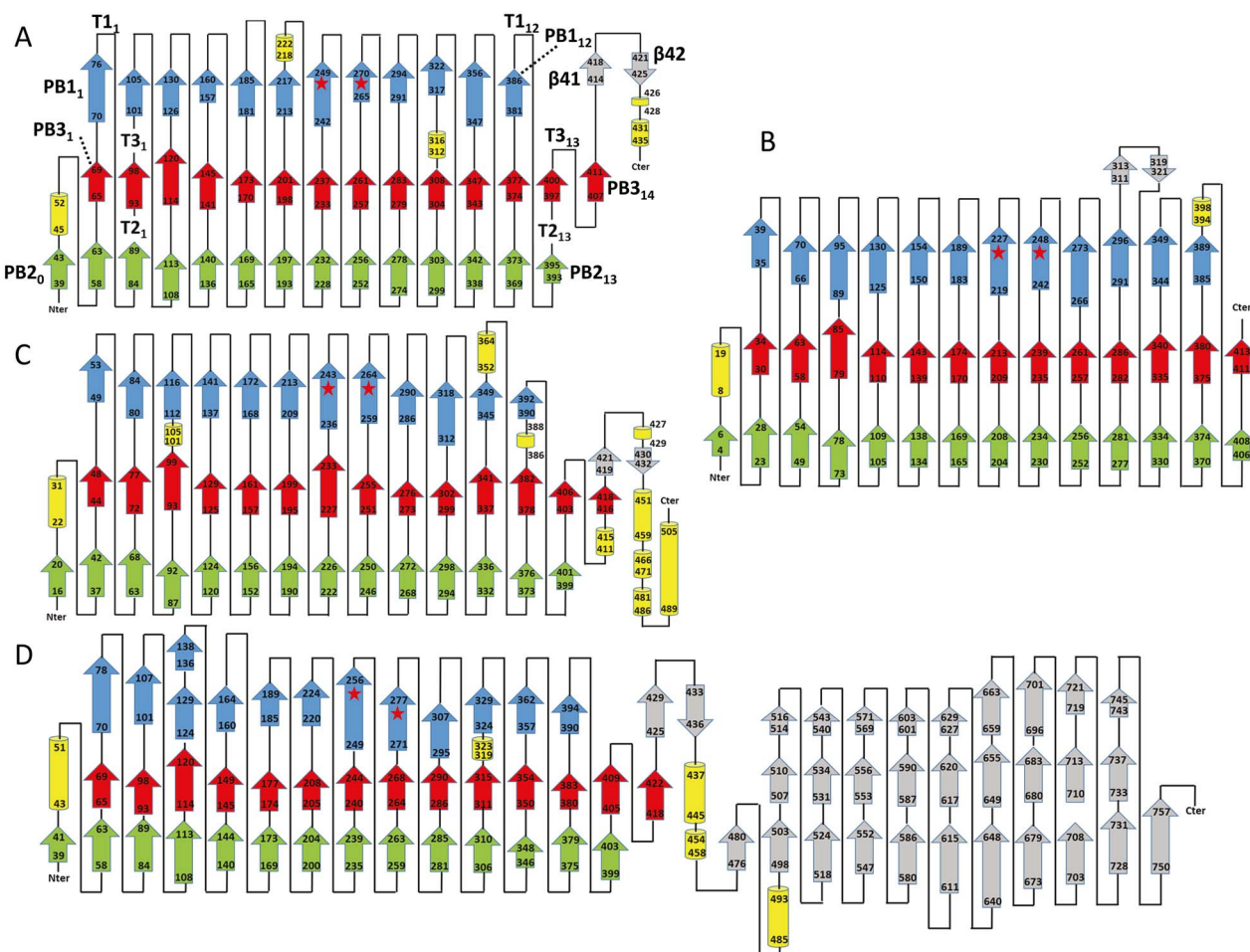


Fig. 2. Topological representation of PL6 structures: relative spatial positions and orientations of the secondary structure elements of PL6 family members show a right-handed parallel β -helix fold made of three β -sheets designated as PB1 (blue), PB2 (red) and PB3 (green). β -strands are numbered from 1 (e.g. PB1₁) to X (e.g. PB1_X), X being the last β -strand, with the exception of β -sheet PB2, which starts at PB2₀. Other β -strands (e.g. β 41) are numbered relative to their position in the primary structure. Sizes and relative positions of secondary structures (β -sheets and helices (yellow)) are respected. Turns between β -sheets designated as T1–3 (black lines), are used as connectors between secondary structures without size consideration. (A) Pedsa0632, (B) Pedsa3628 and (C) Pedsa3807 are single-domain PL6 family members. (D) Patl3640 is a two-domain PL6 family member. The second domain (gray) shows the same β -helix fold. Red stars indicate catalytic residues. Secondary structures insertions: (A) In Pedsa0632, two helices (helix 218–222 and helix 312–316) are present in turn T16 and turn T310. (B) In Pedsa3628, helix 394–398 is introduced in turn T112 and a small β -sheet composed of two β -strands (311–313, 319–321) is inserted in turn T110. (C) In Pedsa3807, two helices (helix 101–105 and helix 352–364) are inserted within turn T33 and turn T111.

When superposing Δ GGG from Pedsa0632 on G₄ from AlyF it can be observed that both substrates span subsites –1 to +3, with G₄ being slightly translated by ~ 1 Å toward subsite +3 (Supplementary Figure S3B).

More strikingly, when comparing the Δ GG binding modes, trisaccharides bind to subsites –1 to +2 and subsites +1 to +3 for Pedsa0632 and AlyF, respectively (Supplementary Figure S3C). AlyF interacts with 17 (G₆), 16 (G₄) and 12 (Δ GG) residues, respectively, (Table III) in addition to the four residues interacting with the catalytic site water (G₄) or binding the catalytic site Na⁺ (Δ GG) (Figure 4E and 4F). As for AlyGC, in addition to these four amino acid residues, a carboxyl group from ligands G₄ or Δ GG interacts with the catalytic site water molecule or the catalytic site Na⁺, respectively (Figure 4E and 4F).

Residue conservation in the alginate-binding site

Analyses of crystal structures of the available PL6 enzymes complexed to a substrate reveal the existence of a common platform made of 11 residues: two catalytic residues, five residues interacting

with the substrate and four residues interacting with the catalytic site water, Na⁺ or Ca²⁺. In AlyF, this platform displays a higher number of interactions when binding G₆, G₄ or Δ GG (Table III). The side chain orientation of the two catalytic residues in a given enzyme superposes well with counterparts in all other enzymes, whether in ligand-free or a ligand-bound form (Figures 3B, 5B and 5E, Supplementary Figure S3).

The same observation is made when considering the four residues that bind to Ca²⁺, Na⁺ or water in the catalytic site of the different enzymes. Indeed, apart from noninteracting residues Asn209 in the structure of Pedsa0632 complexed to a Δ GG, and Asn216 in the structure of Patl3640 complexed to a linear Δ , each of the other residues is similarly positioned (Figure 4). Regarding the substrate-binding residues, Arg215, Ser219, His270, Arg293 and Tyr332 in Pedsa0632 and their counterparts in the other enzymes (Table III), only slight differences can be spotted. Gln377 in AlyF and Ser248 in Pedsa3628 replace the conserved tyrosine and histidine, respectively. However, these two residues are both able to form H-bonds as observed in other enzymes (Supplementary Figure S6A and S6C,

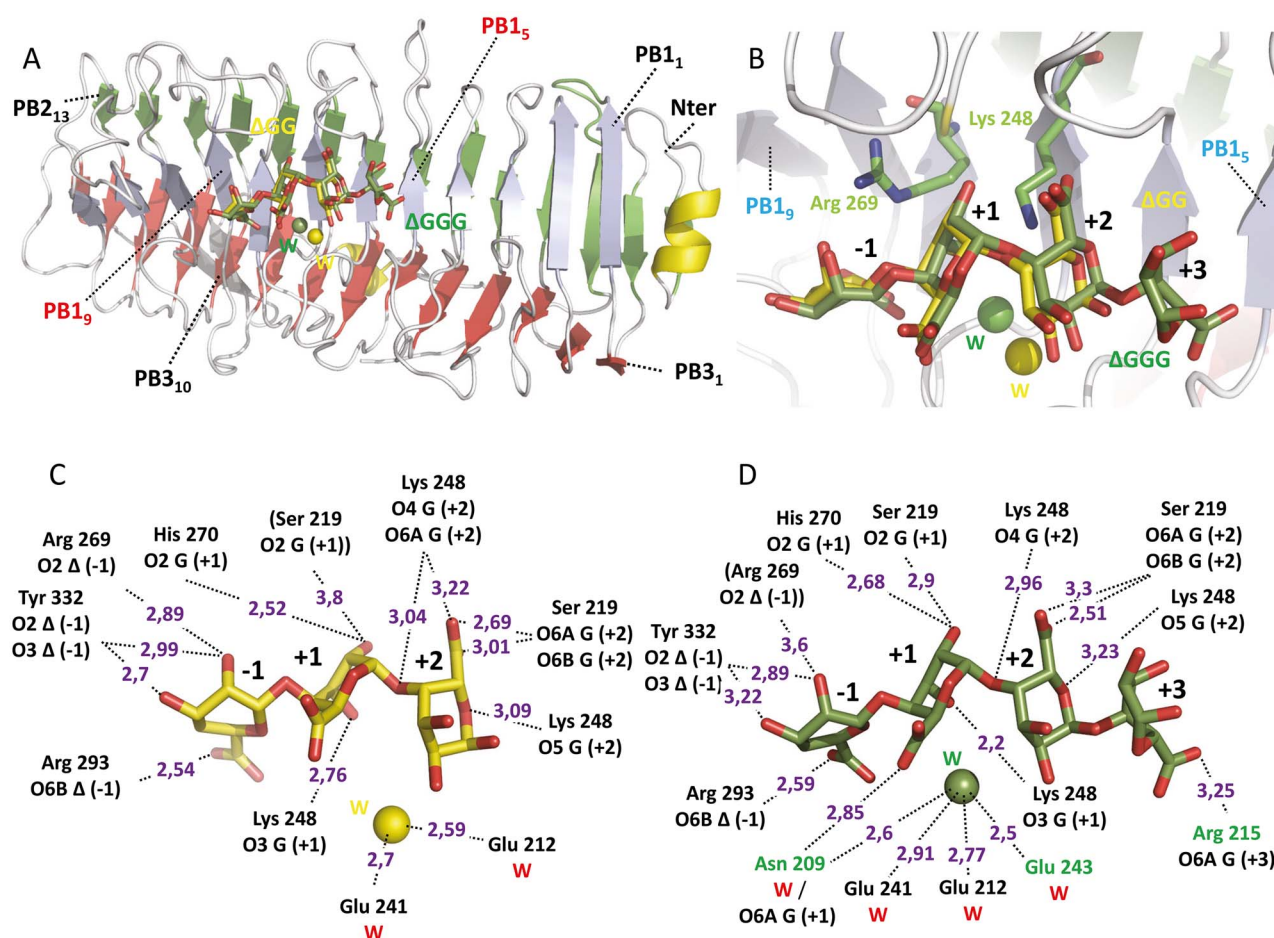


Fig. 3. Crystal structure of Pedsa0632 complexed to Δ GGG and Δ GG: (A) The overall structure of Pedsa0632 presented in a ribbon model. PB1–3 β -sheets are colored respectively in blue, red and green. The two monomers present in the asymmetric unit containing a Δ GGG and a Δ GG, respectively, have been superposed and the respective positions of the two oligosaccharides and the catalytic site water molecules are shown in the cleft formed by the right-handed parallel β -helix. Between β -strands PB1₅ and PB1₉ (blue) resides the Δ GGG- and Δ GG-binding site. (B) Close-up view of Δ GGG (green) and Δ GG (yellow) and the two catalytic site water molecules (same color codes) with subsites numbered from –1 to +3. Catalytic residues Lys248 and Arg269 are shown in green. (C), (D) Residues forming interaction with Δ GGG or Δ GG. Bond lengths (in Å) are given in blue. Residues interacting specifically with Δ GGG are colored in green.

Table III). The main difference is the lack of a serine (Ser219 in Pedsa0632), which is replaced by Tyr191 in Pedsa3628. This residue, positioned in turn T1₆, is differently oriented and can no longer participate in substrate binding (Supplementary Figure S6C). Surprisingly, we observed the same conserved tyrosine in Pedsa3807 and chondroitinase B. Apart from these differences, and the reorientations of Arg239 in AlyF and/or Tyr304 in AlyGC, it appears that the residues are not only conserved but also superposed well (Supplementary Figure S6).

Overall, these comparative studies show that, whatever ligand is bound (or not) to the catalytic site (G₆, G₄, M₄, Δ GG, Δ GGG) and whatever PL6 structure under consideration, the residues are always oriented in the same direction and are not moving to interact with the substrate but rather seem to indicate that the polysaccharides are forced to adapt to the shape of the enzyme.

Hence comparisons do not clearly indicate how each enzyme discriminates between two different substrates nor how an endolytic activity is occasionally preferred to an exolytic activity. One of the most surprising observations is that AlyF, which is the most exotic enzyme in terms of substrate recognition, harbors the very same activity toward polyG (an exolytic cleavage), and the same Δ GG final product as Pedsa0632. In order to search for other specificity

determinants, we performed computer-aided docking experiments with longer substrates in order to explore subsites upstream subsite –1 and downstream subsite +3.

Subsites prediction and comparison

Since no obvious specific features emerged from the comparison of ligand-bound PL6 structures, positions of potential subsites beyond –1 and +3 have been estimated by looking at the surface of the cleft along the right-handed parallel β -helix, which normally accommodates alginate.

In order to determine the position of possible additional subsites, we performed a docking experiment by using Pedsa0632 and a G₇. Herein, G₇ location was predicted in a positively charged groove extending from β -strand PB1₂ to PB1₉ (Supplementary Figures S7 and S8). When superposing onto Δ GGG, it appears that the conformation of the sugar moieties at subsites +1, +2 and +3 is highly similar (Supplementary Figure S8B) and that the same residues are involved in oligosaccharide binding from subsites –1 to +3 (Supplementary Figure S8C, Table III). Despite differences as the lack of interaction with Arg293, or Tyr332 interacting with the sugar moiety in subsite –2 instead of subsite –1, it appears that the binding mode could be well conserved with a longer substrate (Supplementary

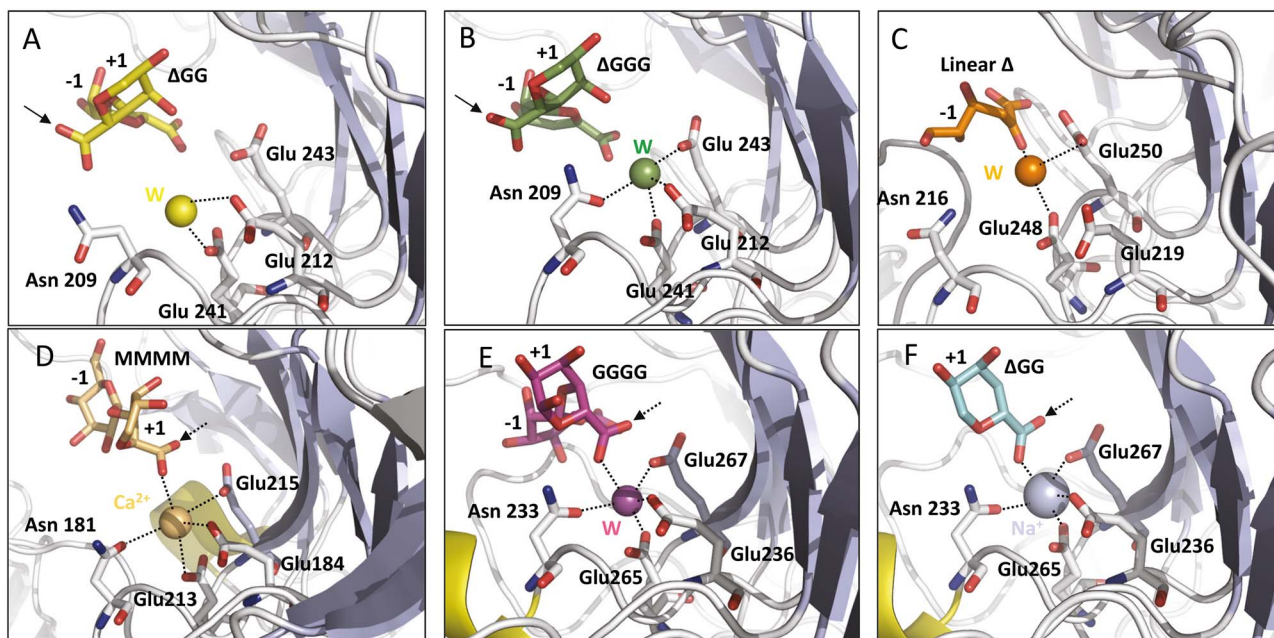


Fig. 4. Close-up view of PL6 catalytic sites with residues involved in the binding of catalytic site water/ Ca^{2+} / Na^{+} : (A) Pedsa0632 complexed to ΔGG . (B) Pedsa0632 complexed to ΔGGG . (C) Patl3640 complexed to a linear Δ . (D) AlyGC complexed to MMMM (pdb entry 5GKQ). (E) AlyF complexed to GGGG (pdb entry 6A40). (F) AlyF complexed to ΔGG (pdb entry 5Z9T). For convenience, only the oligosaccharide residues present in subsites -1 and $+1$ are shown. (A, B) Arrows indicate the position of the oligosaccharide residues carboxylic function turned toward the solvent. (D, E, F) Dotted arrows indicate the position of the oligosaccharide residues carboxylic function interacting with Ca^{2+} , water or Na^{+} .

Figure S8C, Table III). Unexpectedly, only few additional interactions were observed with G₇ in +5 in which Lys123 (T3₃) and Asp125 (PB1₃) interact with the sugar moiety at subsite +5, and no apparent interactions in subsite +4 (Supplementary Figure S8C). Lys123 is poorly conserved and is only present in Pedsa3628 (Lys88), which, surprisingly, does not show any activity toward polyG, thus suggesting that this basic residue could interact with polyMG as well (Supplementary Figure S1). Asp125 is strictly conserved in BCellPL6 and functionally conserved in AlyGC and Patl3640 in which it is substituted by a glutamate. Once again, this residue does not seem to be specific for PolyG activity as BCellPL6 only cleaves PolyM.

At the nonreducing end of the substrate, three new interactions are found. In addition to Tyr332 (T1₁₀) mentioned above, Lys361 (T1₁₁) and Arg362 (T1₁₁) may interact with the sugar moiety at subsite -2 (Supplementary Figure S7C). Tyr332 is well conserved in all six alginate lyases from a functional point of view as it is replaced by a glutamine only in AlyF. Interestingly, a histidine is found in both chondroitinases (Supplementary Figure S1) thereby suggesting that the possibility of making electrostatic interactions in subsite -2 is essential. This residue is found both in exo- and endolyases and can therefore not be considered as a determinant of one type of activity or the other. As opposed to Tyr332, Lys361 is poorly conserved and replaced by Glu332 and Glu367 in AlyGC and Patl3640, respectively, and Gln358 in Pedsa3628. Based on primary and secondary structure alignment performed with PROMALS3D, Lys361 is substituted by Arg363 in BCellPL6 and by Gly416 in AlyF (Supplementary Figure S1). Regarding Arg362, it is conserved in AlyGC (Arg333), Patl3640 (Arg368) and Pedsa3628 (Arg359) and replaced by Gly364 in BCellPL6 and Ile417 in AlyF.

Altogether, this first analysis seems to be not sufficient to apprehend the molecular determinants of substrate recognition. To get a

better view of the roles of residues present in the binding cleft, we thoroughly inspected each enzyme structure and selected all residues for which the side chain is oriented in a manner that it can interact with the substrate. Alignment of the selected residues spanning subsites -2 to $+5$ was performed (Supplementary Figure S9). A WebLogo representation (Crooks et al. 2004) confirmed the existence of a conserved platform of 13 residues hereof 10 being strictly conserved and the lack of conserved residues at subsites $+4$ and $+5$ (Figure 6). No common features observed in this study could be attributed to molecular determinants of substrate specificity with certainty, and it seems that each enzyme has its own recognition mode regardless of the substrate specificity. Interestingly, this observation is confirmed by the recent structure of AlyF in complex with a G₆ extending from subsite $+3$ to -3 (Table III; Zhang et al. 2021). Herein, the residues interacting at subsite -3 are not conserved in any other PL6 structure described.

To further test the validity of this observation, the primary structure of six additional enzymes (all PL6_1 members) was included in the alignment. These two-domain enzymes display exolyase activity on polyG and endolyase activity on polyMG. Once again, a WebLogo representation showed a good conservation of the common platform and very low conservation at subsites -2 , $+4$ and $+5$. If considering the enzymes with polyG exolyase activity, an aromatic residue (subsite $+4$) as well as an asparagine (subsite -1) seem to be conserved. However, residues at these two positions are also present in AlyF, which has endolyase activity (Figure 6). As for enzymes with polyMG endolyase activity, no obvious determinants of specificity could be identified, and only a single specific threonine is shared by the two enzymes having polyG endolyase activity. Finally, BCellPL6, which is the only enzyme with a polyM endolyase activity in this study, possesses three residues that seem to be specific: a glutamine, a threonine and a serine (subsite $+3$ and -1) (Figure 6).

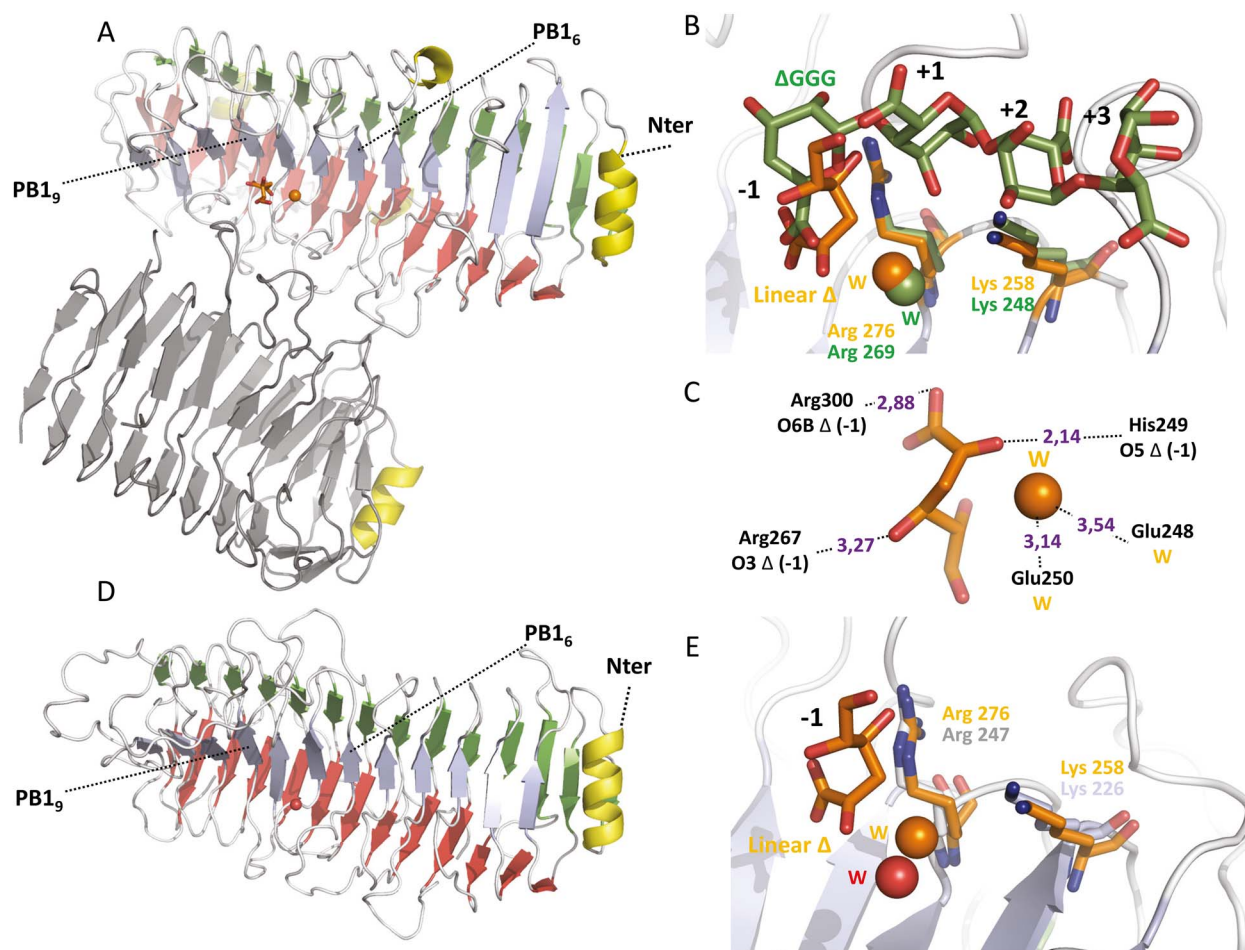


Fig. 5. Crystal structure of ligand-free Pedsa3628 and Patl3640 complexed to a linear Δ : (A, D) Overall crystal structures of Patl3640 and Pedsa3628 in ribbon presentations. The linear Δ and the catalytic site water molecules are shown in the cleft formed by the right-handed parallel β -helix of Patl3640. (B) Close-up view of linear Δ (orange) and the catalytic site water molecule (orange) observed in Patl3640 superposed on Δ GGG (green) and the catalytic site water molecule (green) observed in Pedsa0632. Catalytic residues Lys258 and Arg276 (orange) from Patl3640 and catalytic residues Lys248 and Arg269 (green) from Pedsa0632 show the same side chain orientation. (C) Residues forming interactions with the linear Δ . (E) Close-up view of the linear Δ and the catalytic site water molecule (both orange) observed in Patl3640 superposed on the catalytic site water molecule (red) observed in Pedsa3628. Catalytic residues Lys258 and Arg276 (orange) from Patl3640 and catalytic residues Lys226 and Arg247 (respectively in blue and white) from Pedsa3628 show the same side chain orientation.

Additional features and turns size difference

Upon inspection of primary- and secondary structures, obvious other differences are seen, especially at the amino- and carboxy-terminal ends, along with a number of turns that correspond to loops in certain enzymes (Supplementary Figure S1). The N-terminal parts form the base of the right-handed parallel β -helix, far from the catalytic site and, most probably, do not participate in the binding of the substrate (Figures 3A, 5A and 5D). C-terminal parts, for the single- β -helix enzymes, are opposite to the catalytic site. For the two- β -helix enzymes, the C-terminal part corresponding to the second right-handed parallel β -helix is adjacent to the catalytic site. Even if it does not participate directly in substrate binding in the ligand-bound structures considered in this study, the C-terminal part of the two-helix enzymes forms a constriction that most likely limits the possibility of substrate binding beyond subsite -1 (Supplementary Figure S8A).

Discussion

Alginate lyases have been classified into 14 (May 2021) PL families including PL5, PL6, PL7, PL8, PL14, PL15, PL17, PL18, PL31, PL32,

PL34, PL36 and PL39 and PL41 (www.cazy.org). These families are grouped into four different 3-D structure folds: an $(\alpha/\alpha)_6$ barrel fold associated to family 5 (Yoon et al. 1999; Yoon et al. 2001; Mikami et al. 2012), an $(\alpha/\alpha)_6$ barrel fold and anti-parallel β -sheet associated to families 15 and 17 (Ochiai et al. 2010), a β -jellyroll fold associated to families 7, 14, 18 and 36 (Yamasaki et al. 2004; Yamasaki et al. 2004,2005; Osawa et al. 2005; Ogura et al. 2009; Thomas et al. 2013; Dong et al. 2014; Sim et al. 2017; Qin et al. 2018; Lyu et al. 2019) and a right-handed parallel β -helix associated to families 6 and 31 (Huang et al. 1999; Michel et al. 2004; Park et al. 2014; Xu et al. 2017; Itoh et al. 2019; Lyu et al. 2019; Stender et al. 2019). A phylogenetic analysis of protein sequences has shown that the PL6 family was divided into three subfamilies PL6_1–3 (Mathieu et al. 2016). As opposed to other alginate lyase families and PL6 subfamilies 2 (PL6_2) and 3 (PL6_3), PL6_1 was shown to contain chondroitinase B lyases (Huang et al. 1999; Mathieu et al. 2016).

Eight structures of PL6_1 members have been compared in this study, including six alginate lyases and two chondroitinase B lyases. With the exception of AlyGC and Patl3640, which are both two-domain enzymes, all other PL6_1 members in this study are

Table III. Comparison of interactions involving the residues present in the catalytic site of the different available PL6 structures

Aly F				Pedsa 0632				Aly GC		BcelPL6		Pedsa 3628		Patl 3640	
residues	G ₁ /W	ΔGGG/Na ⁺	G ₁ /Ca ²⁺	residues	ΔGGG/W	ΔGG/W	G ₁	residues	M ₁ /Ca ⁺	residues	Ca ²⁺	residues	W	residues	LinearΔ/W
Arg 127	G (+3)	G (+3)	G (+3)	Asp 125	/	/	G (+5)	Gly 89	/	Asp 115	/	Gly 89	/	Glu 125	/
Lys 165	G (+3) G (+2)	G (+3) G (+2)	G (+3) G (+2)	Asp 154	/	/	/	Gln 123	/	Asp 153	/	Gly 123	/	Ser 158	/
Gln 196	G (+3)	G (+3)	G (+3)	Trp 185	/	/	/	Arg 154	M (+3)	Trp 185	/	Asp 154	/	Arg 189	?
Arg 229	G (-1)	/	G (-2)	/	/	/	/	/	/	/	/	/	/	/	/
Asn 233	G (+1) W	Δ (+1) Na ⁺	G (+1) Ca ²⁺	Asn 209	G (+1) W	G (+1)	G (+1)	Asn 181	Ca ²⁺	Asn 210	Ca ²⁺	Asn 181	W	Asn 216	?
Ser 234	G (+2)	G (+2)	/	Gly 210	/	/	/	Gly 182	/	Gly 211	/	Gly 182	/	Gly 217	/
Glu 236	W	Na ⁺	/	Glu 212	W	W	na	Glu 184	Ca ²⁺	Glu 213	Ca ²⁺	Glu 184	W	Glu 219	?
Arg 239	G (+3)	G (+3)	G (+3)	Arg 215	G (+3)	/	G (+3)	Arg 187	M (+3)	Arg 216	?	Arg 187	?	Arg 222	?
Asp 242	G (+3)	G (+3)	G (+3)	Thr 218	/	/	/	Thr 190	/	His 219	/	Arg 193	/	Thr 225	/
Ser 243	G (+2)	G (+2)	G (+2) G (+1)	Ser 219	G (+2) G (+1)	G (+2) G (+1)	G (+2) G (+1)	Ser 191	M (+2)	Ser 220	/	Tyr 191	/	Ser 226	/
Glu 265	W	Na ⁺	Ca ²⁺	Glu 241	W	W	na	Glu 213	Ca ²⁺	Glu 242	Ca ²⁺	Glu 217	W	Glu 248	W
Arg 266	G (-1) G (+1)	Δ (+1)	G (-3)	Met 242	/	/	/	Val 214	/	Asn 243	/	Met 220	/	His 249	Linear Δ
Glu 267	G (+1) W	Δ (+1) Na ⁺	Ca ²⁺	Glu 243	W	/	na	Glu 215	Ca ²⁺	Glu 244	Ca ²⁺	Glu 221	W	Glu 250	W
Lys 272	(K/A)	Δ (+1) G (+2)	/	Lys 248	G (+2) G (+1)	G (+2) G (+1)	G (+2) G (+1)	Lys 220	M (+2) M (+1)	Lys 249	/	Lys 226	/	Lys 255	?
Arg 293	G (+1) G (-1)	/	G (+1)	Arg 269	(Δ (-1), 3,6 Å)	Δ (-1)	G (-1)	Arg 241	/	Arg 270	/	Arg 247	/	Arg 276	Linear Δ
His 294	G (+1)	Δ (+1)	G (+1)	His 270	G (+1)	G (+1)	G (+1)	His 242	M (+1)	His 271	/	Ser 248	/	His 277	/
Arg 317	G (-1)	/	G (-1)	Arg 293	Δ (-1)	Δ (-1)	/	Arg 265	/	Arg 294	/	Arg 271	/	Arg 300	Linear Δ
Tyr 319	G (-1)	/	G (-1)	Ile 295	/	/	/	Ile 267	/	Ile 296	/	Ala 273	/	Ile 302	/
Arg 336	/	/	G (-3)	/	/	/	/	Tyr 284	/	/	/	/	/	/	/
Asn 343	G (-1)	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Asp 345	/	/	G (-2) G (-3)	/	/	/	/	/	/	/	/	/	/	Leu 320	/
Asn 375	/	/	G (+1)	His 330	/	/	/	Asn 302	/	Thr 337	/	/	/	Asn 336	/
Lys 376	/	/	/	Glu 331	/	/	/	Arg 303	M (-1)	Ser 339	/	Arg 323	/	Arg 338	/
Gln377	/	/	/	Tyr 332	Δ (-1)	Δ (-1)	G (-2)	Tyr 304	M (-1)	Tyr 339	/	Tyr 324	/	Tyr 339	/
Arg 405	/	/	G (-3)	/	/	/	/	/	/	/	/	/	/	/	/

Catalytic residues are highlighted in orange. Residues involved in the binding to water/Ca²⁺/Na⁺ present in the catalytic site are highlighted in blue (water: vertical hatching, Ca²⁺: oblique hatching, Na⁺: horizontal hatching). Conserved residues involved in the binding to an oligosaccharide are highlighted in blue and green. Residues not strictly conserved are hatched in green.

single-domain enzymes sharing the same binding site corresponding to the PB1 β -sheet of the conserved right-handed parallel β -helix. This catalytic site displays four subsites from -1 to +3 with a lysine and an arginine as catalytic residues. Despite these common features, PL6_1 members cleave alginates with an endo- or an exoactivity, with or without Ca²⁺ and, in addition to the respective preferred substrates, are also able to cleave other substrates like e.g. Pedsa0632 that cleaves polyMG but also polyG with an endolyase activity, or AlyGC cleaving polyG but also polyM with an exolyase activity.

Endolyase activity vs. exolyase activity

Endo- or exolyase activities have been suggested to be related to the number of domains (Stender et al. 2019). Indeed, for AlyGC and Patl3640, we observe that the second helix in the C-terminal corresponding to the second domain partly covers the binding site corresponding to the PB1 β -sheet (Supplementary Figures S8A and S8B). This limits the possibility for the substrate to extend beyond subsite -1 and seems to be also true for Pedsa0631, which is a two-domain enzyme with exolyase activity on polyG (Mathieu et al.

2016). However, Mase04135, Patl3659, Nonul2381 and Celly0294, which are all two-domain enzymes as well, cleave both polyG and polyMG with an exo- and an endolyase activity (Mathieu et al. 2016). More surprisingly, FsAlyPL6, also predicted as being a two-domain enzyme, possesses only endolyase activity (Li et al. 2019).

Another characteristic that could potentially be an exolyase activity marker is the dimeric state of these enzymes. Full-length AlyGC has been shown to be a dimer in solution, whereas a Δ CTD (C-terminal domain) variant adopts a monomer in solution. In addition, as compared to full-length AlyGC, the Δ CTD variant lost 94.3% activity toward PG (Xu et al. 2017). We demonstrated that Patl3640 (a two-domain enzyme) is a dimer form in solution (Supplementary Table SII). More interestingly, when looking at the crystal packing of Patl3640, we observe the same dimeric organization as that observed for AlyGC. Unfortunately, for FsAlyPL6, there is no available information concerning its oligomeric state but we speculate that this enzyme may adopt a monomeric organization in solution, explaining its unexpected endolyase activity. Regarding AlyGC (Xu et al. 2017), it was shown that a large loop (Arg627-His638) of the

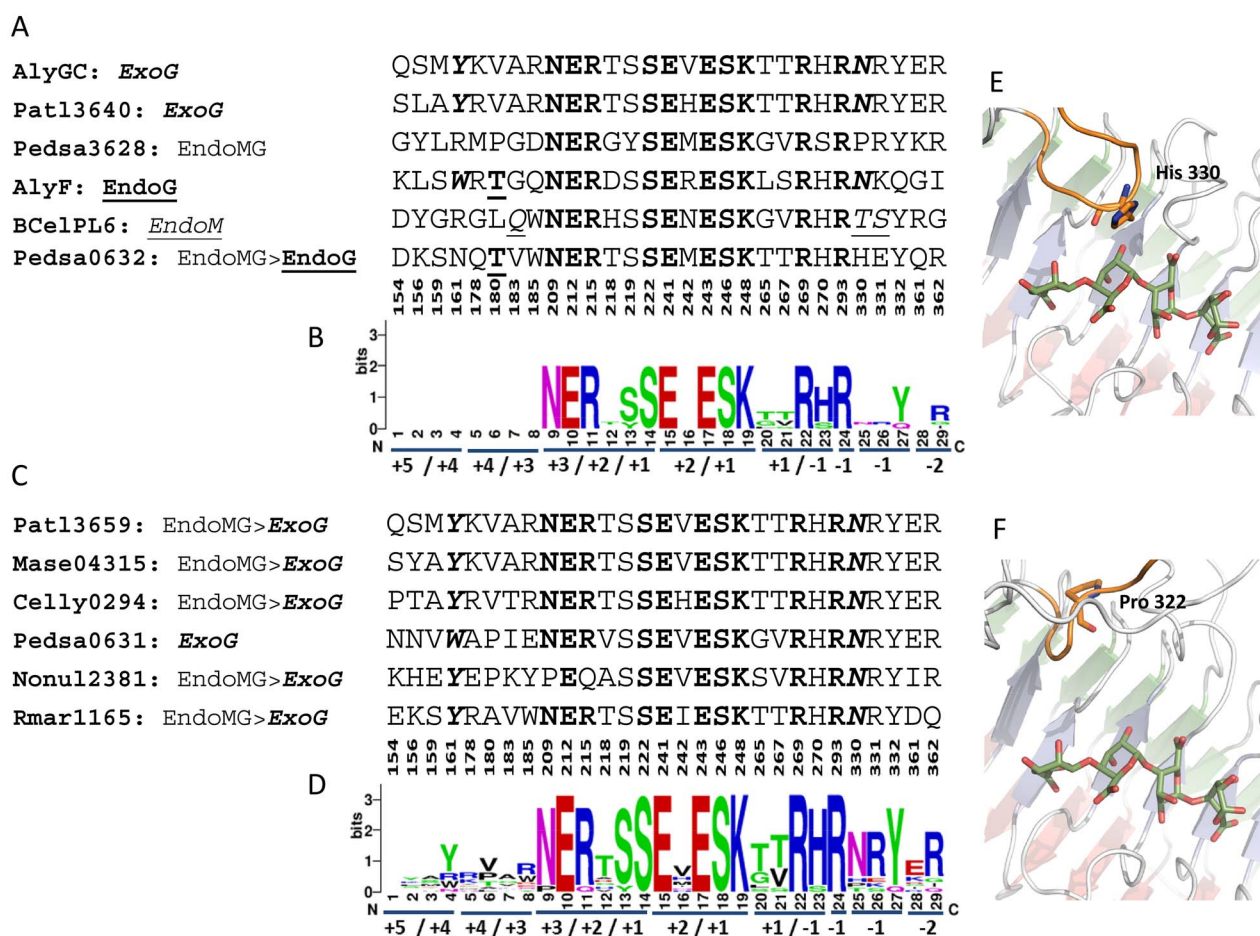


Fig. 6. Sequence alignment of residues involved or putatively involved in substrate binding: (A) The residues of the PL6 family members with a crystal structure and an alginate lyase activity are used for this alignment. (B) WebLogo representation corresponding to the alignment of the first six sequences. (C) Six other PL6_1 family alginate lyases with both endo- and exolyase activity on polyG in addition to endolyase activity on polyMG are used for alignment. (D) WebLogo representation after alignment of the 12 sequences. Numbering above WebLogo representations is according to Pedsa0632. (E) Position of His330 above Δ GGG in Pedsa0632. (F) Position of Pro322 above Δ GGG in Pedsa3628. Δ GGG in Pedsa3628 has been obtained by superimposing Pedsa3628 and Pedsa0632. Loops containing both His330 and Pro322 are in orange.

CTD stretches into the catalytic center and that mutations of Asp631 and Ser633 into alanines led to significant decreases in the enzyme activity. Interestingly, this loop is conserved in these alginate lyases (Supplementary Figure S10).

Substrate specificity

PL6 family members use arginine and lysine as catalytic acid and base, respectively, thereby differing from other PLs in which a tyrosine and a histidine are utilized for these functions (Garron and Cygler 2010; Xu et al. 2017). Single mutation experiments performed on BCellPL6 in which the lysine and the arginine have been replaced by a histidine and a tyrosine, respectively, show a total lack of activity in the mutated enzymes confirming that these two residues are likely an invariant part of the canonic platform of PL6 family members (Stender et al. 2019).

Alignment of the 12 sequences of alginate lyases studied herein show that eight residues are strictly conserved including the two catalytic residues (Figure 6).

It is interesting to note that among these residues, Ser222 and Ser246 (Pedsa0632 numbering) (Figure 6) are not involved in substrate binding. Nevertheless, in Pedsa0632, both serines participate in the overall organization of the binding site: Ser222

by interacting with the main chain carbonyl group of Ser219 present in turn T1₆ and Ser246 by interacting with Glu243 and one of the water molecules forming the water network of the catalytic site. Similar observations are made for residues Ser229 and Ser253 in Patl3640 or Ser194 and Ser224 in Pedsa3628, for which the two serines also participate in the organization of the catalytic site. Apart from Pedsa3628, two other residues are well conserved within the five other studied enzymes, a serine (Ser219 in Pedsa0632) and a histidine (His270 in Pedsa0632) (Table III). As concerns the serine, when mutated into an alanine in AlyF (Ser243), the enzyme shows reduced catalytic activity (Lyu et al. 2019). When mutating the histidine in AlyF (His294Ala) or BCellPL6 (His271Asn), both enzymes show a drastic drop in activity (Lyu et al. 2019; Stender et al. 2019). In Pedsa3628, a serine replaced the conserved histidine, indicating that a strict conservation of this residue is not mandatory for activity, but this difference could explain why Pedsa3628, as opposed to all other members of PL6_1, has no exolyase activity on polyG (Mathieu et al. 2016).

If performing the analysis only on residues conservation, it remains difficult to understand why enzymes of the PL6 family are either endo- or exolyase or why they have preferred substrates. Based on phylogenetic studies, it was suggested (Mathieu et al. 2016) that

PL6 family members could have evolved from strict endo-polyMG lyases to exo-polyG lyase *via* intermediates in which enzymes could have both endo- and exoactivities. Furthermore, they observed that when looking at the end-products of polyMG lyases, a G residue is always present at subsite +1, suggesting that the shift from MG to GG substrate moieties required remodeling of subsite -1 to accommodate a G-residue instead of an M-residue.

The only strict endo-MG in PL6_1 is Pedsa3628. The other enzymes, with the exception of *BCePL6*, all have polyG as substrate with exo- or endoactivities (Figure 6). When focusing on subsite -1, it can be observed that the hydrophobic Pro322 in Pedsa3628 is systematically replaced by a polar residue in the other enzymes (Figure 6). These polar residues do not directly interact with the substrate at subsite -1 but are part of a conserved loop that caps the catalytic site as illustrated by His330 in Pedsa0632 (Figure 6E). In the presence of proline, this loop in Pedsa3628 does not cover the catalytic site anymore (Figure 6F).

Despite a thorough examination of the sequences and 3D structures available, only a few links between a given activity and a given group of residues or structural features have been established. Altogether, we suggest that from a common ancestral structure, several solutions have been found to adapt to the immediate environment. We conclude that, except for the enzymes showing a high degree of similarity like AlyGC and Patl3640, prediction of a substrate specificity based on the sequence comparison alone seems, at this point, precarious. In order to decipher how the PL6 family members specifically cleave their substrates, additional structure-activity studies are needed.

Materials and methods

Sequence alignments

Alignment profiles were generated using ESPrpt (ESPrpt - <http://esprpt.ibcp.fr>; Gouet et al. 2003). Primary- and secondary structure alignments were performed with PROMALS3D (Pei et al. 2008).

Expression and purification

The plasmids carrying the different constructs were transformed into *Escherichia coli* BL21-CodonPlus (DE3)-RIL cells. A single colony was inoculated into 10 mL of Luria Broth containing 25 $\mu\text{g mL}^{-1}$ of kanamycin and grown overnight at 37°C. The overnight culture was added to 2 L of LB with antibiotics and grown at 37°C until the OD₆₀₀ reached ~0.6, hereafter induced with 0.2 mM IPTG (isopropyl- β -D-thiogalactopyranoside) and grown for further 18 h at 20°C. The cells were pelleted by centrifugation at 5000g, 20 min at 20°C and stored at -20°C.

The pellet was resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl - buffer A) and disrupted using a microfluidizer (3 cycles at 15,000 psi). The lysate was centrifuged at 10,000g, 30 min at 4°C and the supernatant was filtered using a 0.45 μm cutoff filter before injection onto a 5 mL HisTrap FF crude using an ÄKTA purifier (GE Healthcare). The column was washed with Buffer A containing 2 M NaCl followed by an elution in buffer A containing 500 mM imidazole. After 12% SDS-PAGE (SDS-polyacrylamide gel electrophoresis) analysis, fractions with relatively pure protein were pooled and concentrated (appropriated cutoff Amicon Ultra-15 centrifugal filter units (EMD Millipore)). Further purification by size-exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare) was done in buffer A. Aliquots of the purified protein were liquid nitrogen flash frozen and stored at -80°C.

Crystallization

Crystallization conditions screening was carried out at 292 K (vapor-diffusion in sitting-drops), using commercially available crystallization kits. For screening, a Mosquito[®] crystallization robot (SPT Labtech Ltd.) was employed using two protein/crystallization agent ratios (200 nL + 200 nL and 300 nL + 100 nL drops equilibrated against 70 μL in MRC Crystallization Plates (Molecular Dimensions)). Proteins were concentrated to 10–40 mg mL⁻¹ in 20 mM Tris-HCl pH 8.0, 150 mM NaCl buffer. Once the crystallization conditions were established, a scale-up was performed in hanging drops mixing 2 μL protein solution with 2 μL reservoir solution (or 3 μL protein and 1 μL reservoir solution) equilibrated against 500 μL reservoir solution in 24-well plates. Crystals of Pedsa0632 grew in 0.2 M K sulfate, 20% (w/v) PEG 3350, while complexes were obtained in 0.2 M ammonium chloride, 20% (w/v) PEG 3350 by co-crystallizing the protein with 10 mM of ΔGGG . Crystals of Pedsa3628 were obtained in 0.2 M K phosphate, 20% (w/v) PEG 3350. Crystals of Patl3640 complexed with Δ were obtained in 0.1 M Na citrate pH 5.6, 2.0 M ammonium sulfate by co-crystallizing the protein with 10 mM of ΔGGG . ΔMGM and ΔGGG were obtained by enzymatic degradation of PolyG and PolyMG by Pedsa0632. They were purified on a semipreparative size-exclusion chromatography system and the structures were confirmed by NMR (Mathieu et al. 2016). The ligand-free form was crystallized in 0.1 M Mg acetate pH 5.6, 0.1 M Na nitrate, 8% (w/v) PEG 10000, 0.15 mM CYMAL-7. As concerns Pedsa3807, the crystallization conditions were 0.2 M tri-Li citrate, 20% (w/v) PEG 3350. Crystals were cryoprotected by adding 15% (v/v) ethylene glycol to initial conditions.

Data collection, structure determination and refinement

X-ray diffraction data were collected at 100 K at the European Synchrotron Radiation Facility (Table I). Data were integrated and scaled with XDS (Kabsch 2010). Data collection statistics are compiled in Table I.

The crystal structures were solved by molecular replacement with the program Phaser (Grayling 2014), using the structure of AlyF (PDB entry 6ITG, Lyu et al. 2019), for Pedsa0632 and Pedsa3628, using the structure of AlyGC (PDB entry 5GKD, Xu et al. 2017) for Patl3640 and using the structure of chondroitinase B (PDB entry 1OFM, Michel et al. 2004) for Pedsa3807 as starting models. Cycles of maximum-likelihood refinement using the program “phenix.refine” and keeping apart 5% of the reflections for cross-validation, were interspersed with manual corrections of the models using COOT (Emsley et al. 2010). Refinement statistics are presented in Table I.

Molecular docking

Docking of substrates into the PL crystal structures were performed using AutoDock Vina embedded in PyRx 0.8 (Trott and Olson 2010; Dallakyan and Olson 2015).

Briefly, the protein structures were prepared for docking by removing unwanted water molecules and bound ligands and by adding polar hydrogens atoms using Discovery Studio Visualizer. The same program was used to build the substrates molecules as PDB files and for energy minimization. PyRx was used for converting all molecules to AutoDock Ligand format (PDBQT). The 3D grid box for molecular docking simulation was obtained using Autodock tools. The Grid box was centered to cover the active site and all essential residues. The docking results were analyzed by comparing

the binding interactions and binding energies between substrate molecules and PL enzymes.

Small angle X-ray scattering

Protein samples were centrifuged prior to measuring their concentration with a Nanodrop spectrophotometer (Thermo). SAXS data were collected in line at the beamline SWING at the synchrotron SOLEIL (David and Perez 2009) after elution on an analytical BioSec-3 (3 mm particle size, 300 Å pore size) column from Agilent, maintained at 288 K, equilibrated in 50 mM Tris–HCl pH 8.0, 150 mM NaCl and operated at a flow rate of 0.3 mL min^{−1}. Individual SAXS frames of 990 ms were collected at a sample-to-detector distance of 2 m, accessing a q range of 0.007 to 0.5 Å^{−1} ($\lambda = 1.03$ Å). All frames were normalized to the intensity of the transmitted beam, radially averaged and background-subtracted using the program Foxtrot. Data were then processed using the HPLC-SAXS module of US-SOMO (Brookes et al. 2016). Then, data were analyzed using the ATSAS 3.0.3 software suite (Manalastas-Cantos et al. 2021). $P(r)$ functions were computed from the scattering curves by an indirect transform method in GNOM (Svergun 1992).

Degradation kinetics

Alginate substrates PolyG (FG = 0.95, DP_n = 20) were prepared from *Laminaria hyperborea* according to Haug et al (Haug et al. 1967). Enzymatic assays were carried out by incubating 150 µL of alginate (0.2% w/v in 50 mM Tris–HCl pH 8, with or without 0.2 mM CaCl₂) with 10 µM of purified enzyme at 25°C. The production of reducing ends was measured using the ferricyanide method at different times of the reaction. Aliquots of 40 µL were transferred to a 200 µL ferricyanide solution (4.55 mM K₃[Fe(CN)₆], 225 mM Na₂CO₃, 5 mM NaOH), which stopped the enzymatic reaction. The solution was heated to 100°C for 10 min and, after cooling, the absorbance of 100 µL of sample was measured at 415 nm with a microplate reader (TECAN M200).

Comparative studies of 3D structures

Crystal structures were compared with existing structures in the Protein Data Bank at Rutgers, RCSB, using the DALI server.

Figure rendering

Figures of 3D structures were drawn with PyMol (Schrödinger, <http://pymol.org>).

Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 7O79 (Pedsa3628), 7O78 (Pedsa3807), 7O7A (Pedsa0632), 7O77 (Patl3640), 7O84 (Pedsa0632/ΔGGG) and 7O7T (Patl3640/Δ), respectively. Other data are available from the corresponding authors upon reasonable request.

Supplementary data

Supplementary data for this article are available at *Glycobiology* online.

Acknowledgments

Technical support from staff on beamlines MX and FIP (both European Synchrotron Radiation Facility), as well as on beamlines PX3 (Swiss Light

Source, Switzerland) and SWING (Soleil, France) is gratefully acknowledged. We are also grateful for assistance from V. Gueguen-Chaignon and F. Delolme of the Protein Science Facility of SFR Biosciences Lyon (UAR3444/US8).

Author contributions

Conceptualization L.B., S.V.; formal analysis, L.B., S.V.; investigation S.V., L.B., F.G., L.Ca., A.T., X.R., V.J., L.Co.; supervision, L.B., N.A., S.V.; validation L.B., S.V., N.A., W.H.; writing L.B., S.V., N.A.

Conflict of interest statement

The authors declare no competing interests.

Abbreviations

CS, chondroitin sulfate; CTD, C-terminal domain; Δ, 4-deoxy-L-erythro-hex-4-eno-pyranosyluronic acid; DS, dermatan sulfate; G, guluronate; HPLC, high-performance liquid chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; M, mannuronate; PEG, polyethylene glycol; PL, polysaccharide lyase; SAXS, small angle X-ray scattering; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)aminomethane.

References

- Aarstad OA, Tøndervik A, Sletta H, Skjåk-Bræk G. 2012. Alginate sequencing: An analysis of block distribution in alginates using specific alginate degrading enzymes. *Biomacromolecules*. 13(1):106–116.
- Brookes E, Vachette P, Rocco M, Pérez J. 2016. US-SOMO HPLC-SAXS module: Dealing with capillary fouling, and extraction of pure component patterns from poorly resolved SEC-SAXS data. *J Appl Cryst*. 49(5):1827–1841.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res*. 14(6):1188–1190.
- Dallakyan S, Olson AJ. 2015. Small-molecule library screening by docking with PyRx. *Methods Mol*. 1263:243–50.
- David G, Perez J. 2009. Combined sampler robot and high-performance liquid chromatography: A fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J Appl Cryst*. 42(5):892–900.
- Dong S, Wei TD, Chen XL, Li CY, Wang P, Xie BB, Qin QL, Zhang XZ, Pang XH, Zhou BC, et al. 2014. Molecular insight into the role of the N-terminal extension in the maturation, substrate recognition, and catalysis of a bacterial alginate lyase from polysaccharide lyase family 18. *J Biol Chem*. 289(43):29558–29569.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 66(4):486–501.
- Garron ML, Cygler M. 2010. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology*. 20(12):1547–1573.
- Gouet P, Robert X, Courcelle E. 2003. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucl. Acids Res*. 31(13):3320–3323.
- Grayling MJ. 2014. phaseR: An R package for phase plane analysis of autonomous ODE systems. *The R Journal*. 6(2):43–51.
- Haug A, Larsen B, Smidsrod O. 1967. Studies on sequence of uronic acid residues in alginic acid. *Acta Chem Scand*. 21:691–704.
- Huang W, Matte A, Li Y, Kim YS, Linhardt RJ, Su H, Cygler M. 1999. Crystal structure of chondroitinase B from *Flavobacterium heparinum* and its complex with a disaccharide product at 1.7 Å resolution. *J Mol Biol*. 294(5):1257–1269.
- Itoh T, Nakagawa E, Yoda M, Nakaichi A, Hibi T, Kimoto H. 2019. Structural and biochemical characterisation of a novel alginate lyase from *Paenibacillus* sp. str. FPU-7. *Sci Rep*. 9(1):14870.

- Kabsch W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr*. 66(2):125–132.
- Li Q, Hu F, Zhu B, Sun Y, Yao Z. 2019. Biochemical characterization and elucidation of action pattern of a novel polysaccharide lyase 6 family alginate lyase from marine bacterium *Flammeovirga* sp. NJ-04. *Mar Drugs*. 17(6):323.
- Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J*. 432(3):437–444.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 42(Database issue):D490–D495.
- Lyu Q, Zhang K, Shi Y, Li W, Diao X, Liu W. 2019. Structural insights into a novel Ca^{2+} -independent PL-6 alginate lyase from *Vibrio* OU02 identify the possible subsites responsible for product distribution. *Biochim Biophys Acta Gen Subj*. 1863(7):1167–1176.
- Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, Panjkovich A, Mertens HDT, Gruzinov A, Borges C, et al. 2021. ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J Appl Cryst*. 54(1):343–355.
- Mathieu S, Henrissat B, Labre F, Skjåk-Bræk G, Helbert W. 2016. Functional exploration of the polysaccharide lyase family PL6. *PLoS One*. 11(7):e0159415.
- Michel G, Pojasek K, Li Y, Sulea T, Linhardt RJ, Raman R, Prabhakar V, Sasisekharan R, Cygler M. 2004. The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery. *J Biol Chem*. 279(31):32882–32896.
- Mikami B, Ban M, Suzuki S, Yoon HJ, Miyake O, Yamasaki M, Ogura K, Maruyama Y, Hashimoto W, Murata K. 2012. Induced-fit motion of a lid loop involved in catalysis in alginate lyase A1-III. *Acta Crystallogr D Biol Crystallogr*. 68(9):1207–1216.
- Ochiai A, Yamasaki M, Mikami B, Hashimoto W, Murata K. 2010. Crystal structure of exotype alginate lyase Atu3025 from *Agrobacterium tumefaciens*. *J Biol Chem*. 285(32):24519–24528.
- Ogura K, Yamasaki M, Yamada T, Mikami B, Hashimoto W, Murata K. 2009. Crystal structure of family 14 polysaccharide lyase with pH-dependent modes of action. *J Biol Chem*. 284(51):35572–35579.
- Osawa T, Matsubara Y, Muramatsu T, Kimura M, Kakuta Y. 2005. Crystal structure of the alginate (poly alpha-L-guluronate) lyase from *Corynebacterium* sp. at 1.2 Å resolution. *J Mol Biol*. 345(5):1111–1118.
- Park D, Jagtap S, Nair SK. 2014. Structure of a PL17 family alginate lyase demonstrates functional similarities among exotype depolymerases. *J Biol Chem*. 289(12):8645–8655.
- Pei J, Kim BH, Grishin NV. 2008. PROMALS3D: A tool for multiple sequence and structure alignment. *Nucleic Acids Res*. 36(7):2295–2300.
- Qin M, Miyakawa T, Inoue A, Nishiyama R, Nakamura A, Asano A, Ojima T, Tanokura M. 2018. Structural basis for controlling the enzymatic properties of polymannuronate preferred alginate lyase FLAlyA from the PL-7 family. *Chem Commun (Camb)*. 54(5):555–558.
- Sim PF, Furusawa G, Teh AH. 2017. Functional and structural studies of a multidomain alginate lyase from *Persicobacter* sp. CCB-QB2. *Sci Rep*. 7(1):13656–13656.
- Stender EGP, Andersen CD, Fredslund F, Holck J, Solberg A, Teze D, Peters GHJ, Christensen BE, Aachmann FL, Welner DH, et al. 2019. Structural and functional aspects of mannuronic acid-specific PL6 alginate lyase from the human gut microbe *Bacteroides cellulosilyticus*. *J Biol Chem*. 294(47):17915–17930.
- Svergun DI. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst*. 25(4):495–503.
- Thomas F, Lundqvist LCE, Jam M, Jeudy A, Barbeyron T, Sandstrom C, Michel G, Czjzek M. 2013. Comparative characterization of two marine alginate lyases from *Zobellia galactanivorans* reveals distinct modes of action and exquisite adaptation to their natural substrate. *J Biol Chem*. 288(32):23021–23037.
- Trott O, Olson, AJ. 2010. AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 31:455–461.
- Xu F, Dong F, Wang P, Cao HY, Li CY, Li PY, Pang XH, Zhang YZ, Chen XL. 2017. Novel molecular insights into the catalytic mechanism of marine bacterial alginate lyase AlyGC from polysaccharide lyase family 6. *J Biol Chem*. 292(11):4457–4468.
- Yamasaki M, Moriwaki S, Miyake O, Hashimoto W, Murata K, Mikami B. 2004. Structure and function of a hypothetical *Pseudomonas aeruginosa* protein PA1167 classified into family PL-7: A novel alginate lyase with a beta-sandwich fold. *J Biol Chem*. 279(30):31863–31872.
- Yamasaki M, Ogura K, Hashimoto W, Mikami B, Murata K. 2005. A structural basis for depolymerization of alginate by polysaccharide lyase family-7. *J Mol Biol*. 352(1):11–21.
- Yoon HJ, Mikami B, Hashimoto W, Murata K. 1999. Crystal structure of alginate lyase A1-III from *Sphingomonas* species A1 at 1.78 Å resolution. *J Mol Biol*. 290(2):505–514.
- Yoon HJ, Hashimoto W, Miyake O, Murata K, Mikami B. 2001. Crystal structure of alginate lyase A1-III complexed with trisaccharide product at 2.0 Å resolution. *J Mol Biol*. 307(1):9–16.
- Zhang K, Liu T, Liu W, Lyu Q. 2021. Structural insights into the substrate-binding cleft of AlyF reveal the first long-chain alginate-binding mode. *Acta Crystallogr D Struct Biol*. 77(3):336–346.
- Zhu B, Yin H. 2015. Alginate lyase: Review of major sources and classification, properties, structure-function analysis and applications. *Bioengineered*. 6(3):125–131.