FEBS Letters 565 (2004) 188–194 FEBS 28298

Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4 ☆

Michael A. McDonough[a,1,2], Renuka Kadirvelraj[a,1,3], Pernille Harris[a,4], Jens-Christian N. Poulsen[a], Sine Larsen[a,b,*]

[a] *Centre for Crystallographic Studies, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, DK, Denmark*
[b] *European Synchrotron Radiation Facility, ESRF-BP 220, F-38043 Grenoble Cedex, France*

Received 30 December 2003; revised 10 March 2004; accepted 18 March 2004

First published online 15 April 2004

Edited by Irmgard Sinning

Abstract Rhamnogalacturonan lyase (RG-lyase) specifically recognizes and cleaves α-1,4 glycosidic bonds between L-rhamnose and D-galacturonic acids in the backbone of rhamnogalacturonan-I, a major component of the plant cell wall polysaccharide, pectin. The three-dimensional structure of RG-lyase from *Aspergillus aculeatus* has been determined to 1.5 Å resolution representing the first known structure from polysaccharide lyase family 4 and of an enzyme with this catalytic specificity. The 508-amino acid polypeptide displays a unique arrangement of three distinct modular domains. Each domain shows structural homology to non-catalytic domains from other carbohydrate active enzymes.

Keywords: Plant cell wall polysaccharide; Pectin degradation; Rhamnogalacturonan; Family 4 polysaccharide lyase; Carbohydrate active enzyme; X-ray crystal structure

1. Introduction

Pectin is of one of the most complex polysaccharides found in the primary cell wall of plants. The pectic network is composed of regions of "smooth" polygalacturonate (homogalacturonan) interspersed within regions of rhamnogalacturonan [1]. The backbone of rhamnogalacturonan I (RG-I) is composed of alternating rhamnose (Rha) and galacturonic acid (GalUA) residues with [,2)-α-L-Rha-(1,4)-α-D-GalUA-(1,] as the repeating unit [2,3], GalUA residues may be acetylated at the O2 or O3 positions. RG-I is often referred to as "hairy" because of multiple branching arabinan, galactan and arabinogalactan side chains attached to C4 of the backbone Rha residues. The complex composition of RG-I explains why an abundance of enzymes present in nature are required to synthesize and degrade pectin [2,4].

The saprophytic filamentous fungus *Aspergillus aculeatus* secretes a variety of enzymes that assist in the degradation of the plant cell wall [4]. Many of the enzymes acting on pectin have been identified and characterized from the commercial enzyme preparation, Pectinex Ultra-SP, which is used for industrial fruit juice clarification [5]. In this product, several enzymes assist in degradation of RG-I. Classification based on sequence homology of these and other non-related carbohydrate active enzymes can be found online in the CAZy database (http://afmb.cnrs-mrs.fr/CAZY/index.html) [6].

The crystal structures are known for three RG-I degrading enzymes, all originating from *A. aculeatus* and being single domain proteins: (1) rhamnogalacturonan acetylesterase (RGAE) from carbohydrate esterase family 12; (2) rhamnogalacturonan hydrolase (RG-hydrolase) from family 28; (3) β-galactanase from glycoside hydrolase family 53 [7–9].

The family 4 polysaccharide lyase, rhamnogalacturonan lyase (RG-lyase), which cleaves the α-1,4 backbone of RG-I through a β-elimination mechanism, has also been isolated from *A. aculeatus* (EC 4.2.2., SwissProt Q00019) [10–12]. Both RG-lyase and RG-hydrolase display an increase in catalytic efficiency towards deacetylated RG-I, thus these enzymes work synergistically with RGAE to degrade RG-I [13]. The mature RG-lyase has 508 residues with a calculated molecular weight of 54202.9 Da, p*I* 5.2 and optimal activity at pH 6.0 and 50 °C [10]. Studies by Mutter et al. concluded that the minimal substrate requirement for RG-lyase was a deacetylated 12-residue oligomer with a preferential cleavage site four residues from Rha at the reducing end. Cleavage via β-elimination introduces a double bond in the fourth residue, GalUA [14]. Removal of the branching galactan side chains

☆ Coordinates have been deposited in the Protein Data Bank (accession code 1NKG).

* Corresponding author. Fax: +45-3532-0299/+33-04-7688-2160.
E-mail addresses: michael.mcdonough@chemistry.oxford.ac.uk (M.A. McDonough), renu_kadirvelraj@hotmail.com (R. Kadirvelraj), ph@kemi.dtu.dk (P. Harris), jcp@ccs.ki.ku.dk (J.-C.N. Poulsen), sine@ccs.ki.ku.dk, slarsen@esrf.fr (S. Larsen).

[1] These persons contributed equally to this work.
[2] Present address: Chemistry Research Laboratory, University of Oxford, Oxford, OX1 3TA, UK. Fax: +44-1865-275-674.
[3] Present address: Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA. Fax: +1-706-542-4412.
[4] Present address: Department of Chemistry, Technical University of Denmark, 2800 Lyngby, DK, Denmark. Fax: +45-45-88-31-36.

Abbreviations: RG-lyase, rhamnogalacturonan lyase; RG-I, rhamnogalacturonan I; Rha, rhamnose; GalUA, galacturonic acid; RGAE, rhamnogalacturonan acetylesterase; RG-hydrolase, rhamnogalacturonan hydrolase; FnIII, fibronectin type III; CBM, carbohydrate binding module

0014-5793/$22.00
doi:10.1016/j.febslet.2004.03.094

attached to Rha decreases the enzymes catalytic efficiency, whereas removal of the branching arabinan side chains results in an increase in enzymatic efficiency [11]. Out of the 14 current polysaccharide lyase families, structures are known for families 1, 3, 5, 6, 7, 8, 9, and 10 [6]. Here, we describe the first crystal structure for a family 4 lyase, which possesses a unique fold and substrate specificity, and employ the structural results in a discussion of its putative substrate binding site and catalytic machinery.

2. Materials and methods

Kadirvelraj et al. [15] have described the crystallization of recombinant *A. aculeatus* RG-lyase expressed in *Aspergillus oryzae* by a stepwise method along with the collection and processing of native and heavy atom data. The structure was solved by multiple isomorphous replacement using phase information from both orange platinum and mercury acetate soaked crystals. The crystals are tetragonal, space group $P4_32_12$ with one molecule in the asymmetric unit. Both heavy atom crystals are isomorphous with the native crystal to a resolution of 3.0 Å. Four mercury and one platinum atomic positions were located and refined using SOLVE resulting in a figure of merit (FOM) of 0.55 [16]. Density modification using RESOLVE improved the phases as was indicated by both an increase in the FOM to 0.73 and a noticeable visual improvement to the electron density maps. Adequate connectivity of electron density enabled tracing of the peptide backbone based on 3.0 Å electron density maps using TURBO-FRODO's TPPR function [17]. Model building was guided by several segments of sequential residues with recognizable side chain electron density, which allowed introduction of the side chains based on sequence and connectivity of the map. The first round of simulated-annealing refinement using CNS with data to 2.5 Å decreased the starting $R = 0.495$ and $R_{free} = 0.490$ to $R = 0.391$ and $R_{free} = 0.463$ (R_{free} based on a random selection of 5% of the data) [18]. The model was adjusted to $F_o - F_c$ and $2F_o - F_c$ electron density maps that were calculated using all data to 1.5 Å. A second round of simulated annealing refinement using data to 2.0 Å converged at $R = 0.382$ and $R_{free} = 0.427$. Cycles of map fitting and refinement, including isotropic *B*-factors and anisotropic scaling with bulk solvent correction using all data to 1.5 Å, continued until R_{free} could no longer be improved, at which point solvent molecules and alternative conformations of side chains were added to the model and refined. Refinement and model statistics can be found in Table 1.

Table 1
Data statistics

Data set	Native	Orange platinum	Mercury acetate
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
$a = b$ (Å)	77.00	77.11	77.35
c (Å)	170.81	170.51	171.50
Resolution range (Å)	19.9–1.5	25.0–3.0	25.0–2.65
Unique reflections	83 086	10 997	15 846
Completeness (%) (highest shell)	100 (100)	100 (100)	99.9 (99.5)
$I/(\sigma)I$ (highest shell)	16.7 (4.5)		
R_{merge}[a] (highest shell)	0.108 (0.292)	0.176 (0.361)	0.097 (0.241)
FOM[b]: SIRAS/solvent flattened	0.55/0.73		
R-factor[c] (highest shell)	0.164 (0.171)		
R_{free} (highest shell)	0.186 (0.204)		
RMS deviation from ideality			
Bonds (Å)	0.011		
Angles (°)	1.7		
Average *B*-factors (Å^2)			
Protein	8.4		
Solvent	21.0		

[a] $R_{merge}(I) = \sum_{hkl} |I_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl} I_{hkl}$, where I_{hkl} is the measured intensity of the reflections with indices *hkl*.
[b] Mean of cosine (phase error).
[c] *R*-factors were calculated using $F > 0\sigma$. *R*-factor $= \sum_{hkl} |F_o| - |F_c| / \sum |F_o|$, $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes for reflection *hkl*, applied to the work (*R*), and test (R_{free}) (5% omitted from refinement) sets, respectively.

3. Results

The X-ray crystal structure was solved by multiple isomorphous replacement (Table 1) and refined using data to 1.5 Å resolution. The final model contains all 508 residues of the mature protein sequence, 723 water molecules, four sulfate ions, and one calcium ion. The structure contains four prolyl *cis*-peptide bonds throughout the structure (residues 122, 223, 376, and 400). The Ramachandran plot has 386 (88.7%) of the residues in most favored regions, 45 (10.3%) in additionally allowed regions, one (0.2%) in generously allowed regions, and three (0.7%) in the disallowed regions. The three residues found in disallowed regions, Gln20, Asn435, and Thr493, have well-defined electron density and *B*-factors corresponding to the average value; all are located in β-turns at the protein surface. The sequence contains one potential N-glycosylation site at Asn331. Contrary to the heavily glycosylated RG-hydrolase and RGAE [7,8], the structure of RG-lyase displays no evidence of glycosylation, a result that is consistent with mass-spectrometry data [15].

The overall structure folds into three distinct modular domains designated I, II, and III and has the shape of a flattened oval disk with approximate dimensions $90 \times 58 \times 40$ Å and a total accessible surface area of 18 900 Å^2 (Fig. 1). The predominantly β-sheet structure is comprised of 38 β-strands, one α-helix, and nine 3_{10} helices. A DALI structural database search for homologues revealed no similar overall structure [19]. However, similar structures for each individual domain could be identified.

Extensive inter-domain interactions link the three domains together. The total buried surface area of the domain interfaces is substantial, 4570 Å^2. The buried area between domain I and II is 1600 Å^2, between domain I and III 1894 Å^2, and between domain II and III 1076 Å^2.

3.1. The domains of RG-lyase

The N-terminal domain I is the largest of the three domains comprising residues 1–257. It folds into a β-supersandwich as defined by the SCOP structural classification database [20]. Its core is composed of two eight-stranded anti-parallel β-sheets. Domain I contains two disulfide bonds, Cys30–Cys73 and Cys164–Cys173, and two associated sulfate ions. Domain II, formed by residues 258–336, displays a common fold that has been described as immunoglobulin-like (Ig) or more specifically fibronectin type III (FnIII) topology. Domain III, formed by residues 337–508, possesses a jelly roll β-sandwich fold structurally homologous to carbohydrate binding modules (CBMs), this domain hosts two sulfate ions and a hexacoordinated calcium ion.

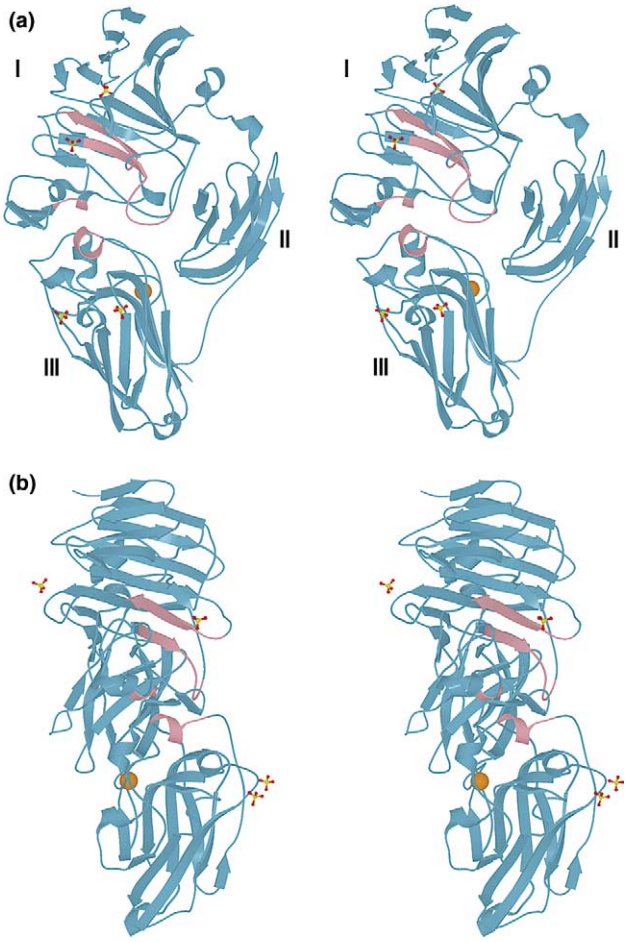

Fig. 1. Overall structure of RG-lyase from *A. aculeatus*. (a) Stereo ribbons representation with the three domains labeled I, II, and III. The four sulfate ions are displayed in ball and stick, and the calcium ion in domain III as an orange CPK sphere. Conserved blocks (outlined in red in the sequence alignment of Fig. 3) are shown with pink secondary structure. (b) Same as (a) rotated 90° around the vertical axis.

4. Discussion

A modular arrangement similar to the one displayed by the RG-lyase has previously not been observed in any known protein structure. However, a similar sequential order of three modules (I, II, and III) determined by sequence analysis was suggested to exist in the *Cellvibrio japonicus* (formerly *Pseudomonas cellulosa*) Rgl11A, a RG-lyase belonging to family 11, for which there is no available structure at present (Fig. 2) [21]. Expression and characterization of the N-terminal domain fragment of Rgl11A shows that this domain possesses RG-lyase activity independent of the FnIII-like and CBM domains and that the activity has a strict requirement for calcium ions. A ClustalW sequence alignment in default mode displays no sequence homology between family 4 RG-lyase domain I and the family 11 Rgl11A N-terminal domain [22]. The family 10 polysaccharide lyase Pel10A also possesses FnIII-like and CBM domains but they are located N-terminal to the catalytic C-terminal domain whose crystal structure [23], solved in the absence of the FnIII-like and CBM domains, displays mostly helical content and is clearly not related to the family 4 RG-lyase domain I.

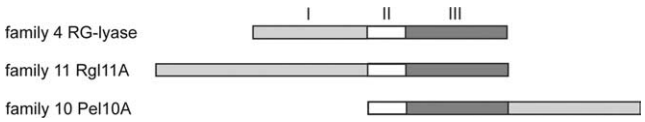

Fig. 2. Comparison of modular domain arrangements among polysaccharide lyase family 4, 10, and 11. Each contains sequential regions similar to domain II (FnIII-like) and domain III (CBM-like). Catalytic domains are light gray, FnIII-like domains are white and CBMs are dark gray.

Although the sequence similarities of structural homologues from a DALI search to the respective domains of RG-lyase are extremely distant, 6–17% identity, the structural similarity is clear [19]. Domain I shows the greatest structural similarity to the C-terminal domain 5 from *Escherichia coli* β-galactosidase, a domain to which no function has yet been assigned. They display identical connectivity and their variation in loop size provides the greatest visible difference (Fig. 3(a)). The C-terminal domain of β-galactosidase contains a substantial buried charged network [24]. It was noted by Juers et al. [24] that residues of this charged network are conserved within β-galactosidase homologues. This feature was explained as a favorable folding over of one surface onto another. Superposition of domain I from RG-lyase with the C-terminal domain 5 from β-galactosidase reveals that although the buried charged network is not conserved in domain I, the corresponding location is in the vicinity of the conserved residues of domain I contributing to the putative active site of RG-lyase (see below). This suggests that the unusual buried charged network in β-galactosidase could be an evolutionary relic of a once functional region. Interestingly, the C-terminal domains of two family 8 polysaccharide lyases, hyaluronate lyase [25] and chondroitin AC lyase [26], also have structural homology to domain I. However, the catalytic activity of both hyaluronate lyase and chondroitin AC lyase is associated with their α-helical N-terminal domain, which does not share any similarity to family 4 RG-lyase.

Domain II shows highest structural similarity to the C-terminal β-sandwich subdomain of the prohormone/propeptide processing enzyme carboxypeptidase gp180 from duck [27] (Fig. 3(b)). Slightly less similar, but more closely related to function, is domain II's similarity to the non-catalytic C-terminal carbohydrate binding domains of cyclodextrin glucosyltransferase (CTGase or α-amylase) [28] and β-amylase [29]. The crystal structures of the C-terminal domains of both CTGase and β-amylase areas are available in complex with carbohydrate molecules. When these complexes are superimposed onto domain II of the RG-lyase, the carbohydrate molecules are located in a deep pocket of the RG-lyase. Part of this deep pocket belongs to domain III, and ends with the conserved residues Arg459 and Gly460.

The highest structural similarity of domain III is with the C-terminal domain of insecticidal protein toxin Cry2Aa [30] (Fig. 3(c)), which is thought to bind to glycosylated receptors. More significant is the similarity of domain III to CBMs from the xylanase 10c fragment (CBM 15-like) [31] and the CBM from endo-1,4-xylanase [32]. Aromatic protein residue–substrate interactions define the polysaccharide binding sites along the concave face of these CBMs. A structural comparison shows, however, that the corresponding region of domain III is relatively planar like those of the CBMs that bind crystalline

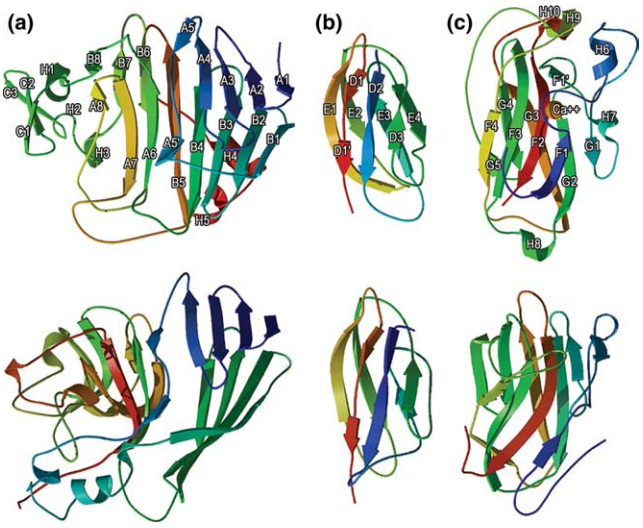

Fig. 3. (a–c) Ribbons representations of the individual domains colored according to sequence progression, N-terminus blue to C-terminus red. Below each domain in the same orientation and coloring scheme, the best structurally homologous domain detected using DALI [19] is shown. (a) The N-terminal domain I (residues 1–257) folds into a β-super-sandwich (SCOP structural classification) with its core having two eight-stranded anti-parallel β-sheets [20]. The homologous structure below is domain 5 (unknown function) from *E. coli* β-galactosidase (PDB ID 1BGL) [38] other structurally homologous domains to the RG-lyase domain I are; hyaluronate lyase (PDB ID 1EGU) [25], chondroitinase AC (PDB ID 1CB8) [26], and α-mannosidase II (PDB ID 1HXK) [39]. (b) Residues 258–336 comprise domain II, the smallest of the three domains. It folds into a Ig-like Greek key β-sandwich made up of one four stranded anti-parallel sheet packed against a three-stranded mixed parallel/anti-parallel β-sheet similar to that of FnIII and pre-albumin. The closest homologue is duck carboxypeptidase GP180 domain II (PDB ID 1QMU) [27] other structurally homologous domains to the RG-lyase domain II are; cyclodextrin-glycosyltransferase (α-amylase family) (PDB ID 1CXL) [28], β-amylase (PDB ID 1B9Z) [29], glucoamylase fragment (PDB ID 1KUM) [40]. (c) Domain III, residues 337–508, folds into a jelly-roll β-sandwich having one sheet of five mixed parallel/anti-parallel β-strands and a second parallel sheet of four anti-parallel β-strands. The only α-helix in the entire structure, H6, is positioned just after the omega shaped calcium ion binding loop between β-strand F1 and G1 makes one turn of helix from residues 357–362. Its closest homologue is *B. thuringiensis* subsp. pesticidal crystal protein Cry2Aa, (PDB ID 1I5P) [30] other structurally homologous domains to the RG-lyase domain III are; xylanase 10c fragment (CBM 15 like) (PDB ID 1GNY) [31], endo-1,4-β-xylanase y fragment X6b CBM 22 (PDB ID 1DYO) [32]. Illustrations are made using MOLSCRIPT [41] and RASTER3D [42].

cellulose. Furthermore, very few strictly conserved residues among the RG-lyases belonging to family 4 are located in this region and there is a noteworthy absence of the surface exposed aromatic residues commonly found in CBMs.

The bound calcium ion represents a unique feature of domain III, as none of the similar structures possesses a calcium ion at the corresponding position. The calcium ion is coordinated by five protein ligands and one water molecule, three of the calcium coordinating residues have their origin in a characteristic omega shaped loop that wraps around the ion. The binding of this calcium ion is very different from the calcium binding in the related pectate lyases. In their active form, pectate lyases contain at least one bound calcium ion, which plays an integral part in substrate binding and activation [33]. Though present evidence suggests that calcium ions are not strictly required for the catalytic activity of polysaccharide lyase family 4 [34,35], the calcium ion may have a unique structural role explaining its positive effects on catalysis.

4.1. Conserved residues in polysaccharide lyase family 4

Fig. 4 shows the multiple sequence alignment of four sequences, assigned to polysaccharide lyase family 4, that show more than 40% identity to *A. aculeatus* RG-lyase. Among conserved residues are those acting as ligands to the calcium ion in domain III, most of the residues interacting with the sulfate ions and the residues engaged in inter-domain interactions. Due to the large complex nature of the substrate, conserved residues involved in substrate binding are expected to cluster over the surface of the protein in an area large enough to accommodate a linear polymer of 12 sugar residues. From an analysis of the sequence alignment and conserved residues, four blocks of conserved residues could be identified, all four are in close proximity in the tertiary structure (Figs. 1 and 5). Three of these conserved blocks are in domain I and one in domain III. The first conserved block [G-E-L-R-F-x-A-R-L] forms β-strand A6, block [S-K-F-Y-S] forms a turn between β-strand C3 and 3_{10} helix H2, and [Y-x-Y-M-x-S-x-H-x-Q-x-E] extends from β-strand A7 into a loop. The last block in domain III [R-G-x-T-R-G] is in a loop located in the same structural region as the three conserved blocks from domain I. Alignment of all sequences from family 4 polysaccharide lyases in CAZy (data not shown), including the distantly related *Arabidopsis thaliana* MYST annotated sequences [36], reveals that Lys150, part of the [S-K-F-Y-S] block in domain I, is strictly conserved. The other strictly conserved residues in this domain are either solvent inaccessible or non-catalytic based on side chain functionality. Low domain I homology (4–9% identity) between fungal and plant sequences may indicate an evolutionary divergent sub-family in which the catalytic machinery has been maintained, while the substrate recognition has undergone major modifications. RhiE, a bacterial RG-lyase from *Erwinia chrysanthemi*, is more related (18–22% identity) to the *A. thaliana* MYST sequences than to *A. aculeatus* RG-lyase (7% identity) and has been shown to cleave to RG-I backbone by a β-elimination mechanism [35]. Such sequence diversity in polysaccharide lyase family 4 may reflect variation in the pattern of arabinose and galactose branching side chains of RG-I from different species.

Pectate lyases have optimal activity at high pH (8–10) and possess a totally conserved arginine that has been identified as the catalytic base in a β-elimination mechanism [23]. Pectate lyases have a strict requirement for calcium but the pectin lyases do not. It should be noted that pectin lyase, which has a similar unique right-hand β-helix fold as the pectate lyase, but no calcium requirement, contains a totally conserved arginine at the same position as the proton accepting arginine of pectate lyase. This could imply that arginine is also the catalytic base in pectin lyases, though this would require an unprecedented pK_a shift of arginine to 6.0–6.5, which could be possible given the appropriate environment. Although pectin lyases contain the equivalent conserved catalytic arginine as the pectate lyases, they have a lower pH optimum 6.0–6.5, which is similar to the family 4 RG-lyases [10,35]. A lysine (Lys150) is the only completely conserved residue found in the sequences in polysaccharide lyase family 4, and the possibility that RG-lyase and pectate/pectin lyases operate by a related mechanism cannot be ruled out.

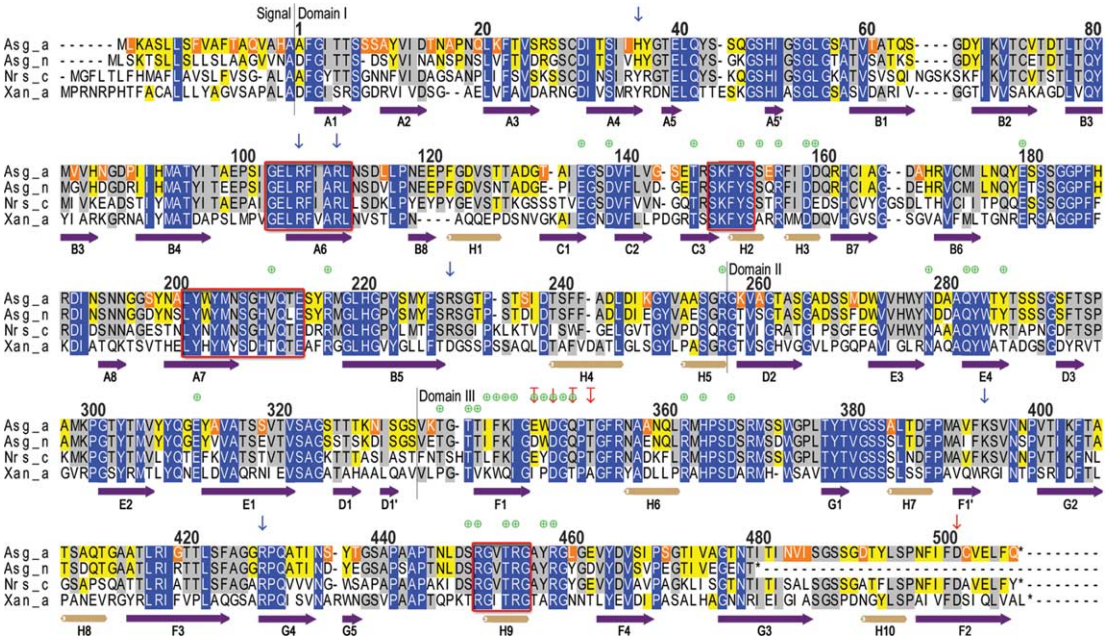

Fig. 4. Multiple sequence alignment performed with CLUSTALW [22] of the family 4 polysaccharide lyases in the CAZy database having identity homology greater than 40%; Asg_a, *A. aculeatus*, Asg_n, *A. niger* (GenBank AJ489944), Nrs_c, *Neurospora crassa* (SwissProt P78710), Xan_a, *Xanthomonas axonopodis* (GenBank AE01999) having sequence identity to Asg_a of 85%, 65%, and 41%, respectively. Residues are colored according to conservation. Blue indicates conservation among the four sequences, gray conservation among Asg_a and at least two other sequences, yellow conservation between Asg_a and at least one other sequence, and orange no conservation between Asg_a and other sequences. Secondary structural elements are labeled as in Figs. 3(a)–(c), purple arrows represent β-strands and light brown cylinders helices. Light gray lines separate the structurally identified domain boundaries. Blue arrows mark residues interacting with sulfate ions in the RG-lyase crystal structure. Red arrows indicate residues having side chain interactions with the calcium ion. Red arrows with a red bar indicate residues having main-chain interactions with the calcium ion. Green circles with crosses indicate residues involved in inter-domain contacts. The four blocks of residues outlined in red create the strictly conserved surface patch encircled in Fig. 5.

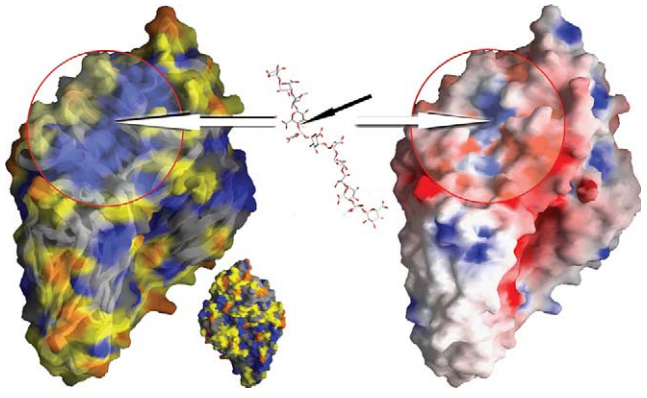

Fig. 5. Left drawing shows the molecular surface colored according to the sequence conservation of the alignment, Fig. 3, with the proposed substrate binding site encircled. It is the only region of strictly conserved residues large enough (approximately 675 Å^2) to accommodate a 12mer of RG-I, which is shown to scale next to the active site in a stick representation (oxygen atoms are colored red and carbon atoms white). The black arrow marks the proposed cleavage site. The structure drawn under to the right is rotated 180° around a vertical axis and displayed in 1:3 scale. White arrows mark the region of positively charged residues. The surface representation images were prepared using GRASP [43].

Arg107 and Arg111 are located in the conserved [G-E-L-R-F-x-A-R-L] block of domain I. Two other conserved arginines (Arg451 and Arg455) are found in the conserved block of domain III. One of the sulfate ions interact with both Arg107 and Arg111. The ability to bind a negative charge at this site suggests that it could be an anchor point for a carboxylate group of the substrate. The close proximity of the two arginine residues could also suggest a lyase mechanism in which one of them acts as the catalytic base analogous to the family 1 and 10 pectate lyases, which have completely different folds, but through convergent evolution have acquired similar enzymatic mechanisms [23,33]. For arginine to act as a base at low pH, the local environment provided by the protein/substrate complex is crucial. To make the picture even more complex, one should also take into account that the RG-lyase pH optimum, around pH 6, could imply that a histidine residue serves as the catalytic base. The proposed active site illustrated in Fig. 6 contains one conserved histidine (His210) located on the conserved [Y-x-Y-M-x-S-x-H-x-Q-x-E] block in a type IV β-turn in the heart of the cluster of conserved residues. Furthermore, the two totally conserved tyrosine residues (Tyr203, Tyr205) that separate the conserved arginines from domain I and III could also be important for catalytic activity or substrate binding.

The assignment of domain I as the one possessing the catalytic activity is supported by the following four observations: 1. The presence of a strictly conserved lysine (Lys150) among distantly related family 4 polysaccharide lyases; 2. The cluster of conserved polar residues among the closely related polysaccharide lyases of family 4 shown in Fig. 5 of which Lys150 is centrally located and solvent accessible; 3. The existence of a shallow water filled groove, matching the size and shape required for substrate binding

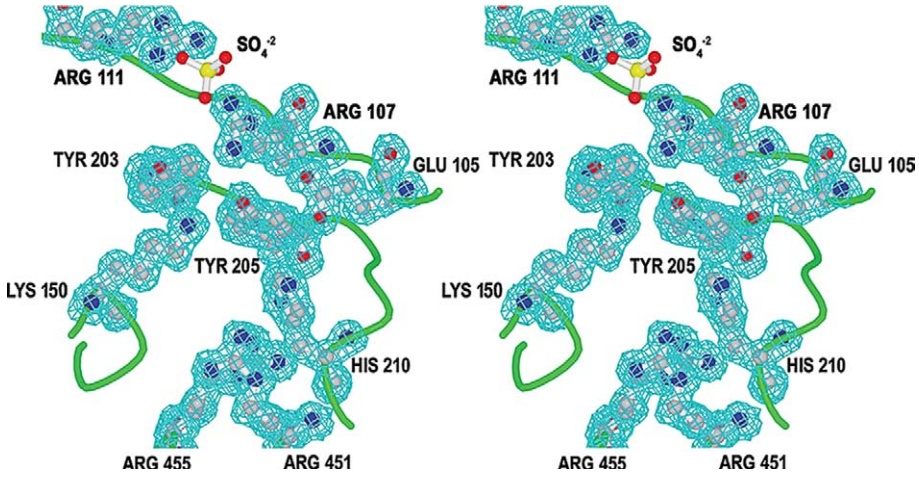

Fig. 6. Details of the proposed conserved active site. The main-chain is represented as a green coil and several conserved residues are shown in ball and stick using CPK coloring. The 1.5 Å $2F_o - F_c$ electron density map in blue-green contoured to 1.5σ. Electron density of the surrounding residues, water molecules, and sulfate ion was removed to improve the visual clarity of the image. The illustration is made using MOLSCRIPT [41] and RASTER3D [42].

(approximately 45×15 Å, with a surface area of 675 Å^2) delineated by the conserved residues; 4. Domains II and III are less likely to carry a catalytic function based on their similarity to well characterized carbohydrate active protein structures, none of which have catalytic activity.

Assuming that domain I is most likely to carry the catalytic function, domains II and III could serve a role as "helper" domains promoting close interactions between the catalytic domain and the substrate, a role exerted by CBMs of other carbohydrate active enzymes [37]. Mutational studies must be conducted to provide more conclusive information regarding the precise details of the RG-lyase active site.

Though a number of points remain to be elucidated before substrate binding and the enzymatic function of the RG-lyase are fully understood, knowledge of its structure has revealed unique structural features of the polysaccharide lyase family 4, and represent a significant addition to our knowledge of the amazing variation of domain arrangements that nature has assembled for interacting with complex carbohydrates. Intriguing evolutionary questions remain on the mechanisms that lead to the assembly of such modular carbohydrate active proteins as this RG-lyase.

Acknowledgements: This research was supported by the Danish National Research Foundation. We thank the people of Novozymes A/S for providing us with the crude extract containing recombinant RG-lyase and one of the referees for the recommendation to include family 4 conserved residues in distantly related (4–9% ID) annotated sequences from plants in our discussion.

References

[1] Vincken, J.P., Schols, H.A., Oomen, R.J., McCann, M.C., Ulvskov, P., Voragen, A.G. and Visser, R.G. (2003) Plant Physiol. 132, 1781–1789.
[2] Ridley, B.L., O'Neill, M.A. and Mohnen, D. (2001) Phytochemistry 57, 929–967.
[3] Willats, W.G., McCartney, L., Mackie, W. and Knox, J.P. (2001) Plant Mol. Biol. 47, 9–27.
[4] de Vries, R.P. and Visser, J. (2001) Microbiol. Mol. Biol. Rev. 65, 497–522.
[5] Schols, H.A., Geraeds, C.C.J.M., Searle-van Leeuwen, M.F., Kormelink, F.J.M. and Voragen, A.G. (1990) Carbohydr. Res. 206, 105–115.
[6] Coutinho, P.M. and Henrissat, B. (1999) in: Recent Advances in Carbohydrate Engineering (Gilbert, H.J., Davies, G., Henrissat, B. and Svensson, B., Eds.), pp. 3–12, The Royal Society of Chemistry, Cambridge.
[7] Molgaard, A., Kauppinen, S. and Larsen, S. (2000) Struct. Fold. Des. 8, 373–383.
[8] Petersen, T.N., Kauppinen, S. and Larsen, S. (1997) Structure 5, 533–544.
[9] Ryttersgaard, C., Lo Leggio, L., Coutinho, P.M., Henrissat, B. and Larsen, S. (2002) Biochemistry 41, 15135–15143.
[10] Kofod, L.V., Kauppinen, S., Christgau, S., Andersen, L.N., Heldt-Hansen, H.P., Dorreich, K. and Dalboge, H. (1994) J. Biol. Chem. 269, 29182–29189.
[11] Mutter, M., Colquhoun, I.J., Beldman, G., Schols, H.A., Bakx, E.J. and Voragen, A.G. (1998) Plant Physiol. 117, 141–152.
[12] Azadi, P., O'Neill, M.A., Bergmann, C., Darvill, A.G. and Albersheim, P. (1995) Glycobiology 5, 783–789.
[13] de Vries, R.P., Kester, H.C., Poulsen, C.H., Benen, J.A. and Visser, J. (2000) Carbohydr. Res. 327, 401–410.
[14] Mutter, M., Renard, C.M., Beldman, G., Schols, H.A. and Voragen, A.G. (1998) Carbohydr. Res. 311, 155–164.
[15] Kadirvelraj, R., Harris, P., Poulsen, J.C., Kauppinen, S. and Larsen, S. (2002) Acta Crystallogr. D 58, 1346–1349.
[16] Terwilliger, T.C. and Berendzen, J. (1999) Acta Crystallogr. D 55, 849–861.
[17] Roussel, A., Cambillau, C. (1989) in: Silicon Graphics Geometry Partners Directory. Silicon Graphics, Mountain View, CA, USA, pp. 77–79.
[18] Brunger, A.T. et al. (1998) Acta Crystallogr. D 54, 905–921.
[19] Holm, L. and Sander, C. (1995) Trends Biochem. Sci. 20, 478–480.
[20] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[21] McKie, V.A., Vincken, J.P., Voragen, A.G., van den Broek, L.A., Stimson, E. and Gilbert, H.J. (2001) Biochem. J. 355, 167–177.
[22] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[23] Charnock, S.J., Brown, I.E., Turkenburg, J.P., Black, G.W. and Davies, G.J. (2002) Proc. Natl. Acad. Sci. USA 99, 12067–12072.
[24] Juers, D.H., Jacobson, R.H., Wigley, D., Zhang, X.J., Huber, R.E., Tronrud, D.E. and Matthews, B.W. (2000) Protein Sci. 9, 1685–1699.
[25] Li, S., Kelly, S.J., Lamani, E., Ferraroni, M. and Jedrzejas, M.J. (2000) EMBO J. 19, 1228–1240.
[26] Fethiere, J., Eggimann, B. and Cygler, M. (1999) J. Mol. Biol. 288, 635–647.

[27] Gomis-Ruth, F.X., Companys, V., Qian, Y., Fricker, L.D., Vendrell, J., Aviles, F.X. and Coll, M. (1999) EMBO J. 18, 5817–5826.
[28] Uitdehaag, J.C., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G. and Dijkstra, B.W. (1999) Nat. Struct. Biol. 6, 432–436.
[29] Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R. and Utsumi, S. (1999) Biochemistry 38, 7050–7061.
[30] Morse, R.J., Yamamoto, T. and Stroud, R.M. (2001) Structure 9, 409–417.
[31] Szabo, L., Jamal, S., Xie, H., Charnock, S.J., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2001) J. Biol. Chem. 276, 49061–49065.
[32] Charnock, S.J., Bolam, D.N., Turkenburg, J.P., Gilbert, H.J., Ferreira, L.M., Davies, G.J. and Fontes, C.M. (2000) Biochemistry 39, 5013–5021.
[33] Scavetta, R.D. et al. (1999) Plant Cell 11, 1081–1092.
[34] Mutter, M., Colquhoun, I.J., Schols, H.A., Beldman, G. and Voragen, A.G. (1996) Plant Physiol. 110, 73–77.
[35] Laatu, M. and Condemine, G. (2003) J. Bacteriol. 185, 1642–1649.
[36] Tavares, R., Aubourg, S., Lecharny, A. and Kreis, M. (2000) Plant Mol. Biol. 42, 703–717.
[37] Bayer, E.A., Shimon, L.J., Shoham, Y. and Lamed, R. (1998) J. Struct. Biol. 124, 221–234.
[38] Jacobson, R.H., Zhang, X.J., DuBose, R.F. and Matthews, B.W. (1994) Nature 369, 761–766.
[39] van den Elsen, J.M.H., Kuntz, D.A. and Rose, D.R. (2001) EMBO J. 20, 3008–3017.
[40] Sorimachi, K., Jacks, A.J., Le Gal-Coeffet, M.-F., Williamson, G., Archer, D.B., Williamson, M.P. (deposited in 1996) to be published.
[41] Kraulis, P. (1991) J. Appl. Crystallogr. 24, 946–950.
[42] Esnouf, R.M. (1997) J Mol Graph Model 15, 132–134, 112–133.
[43] Nicholls, A., Sharp, K.A. and Honig, B. (1991) Proteins 11, 281–296.